Topics in Random Walks in Random Environments

Erwin Bolthausen

University of Zürich

November 18 – December 13, 2013,
February 18 – March 16, 2014,
FMSP Lectures

まえがき

平成25年度に，スイス・チューリッヒ大学数学教室名誉教授の Erwin Bolthausen 氏が２度にわたって東京大学大学院数理科学研究科を訪問されました。最初の滞在期間 2013 年 11 月 18 日〜12 月 13 日の間に，同氏は FMSP レクチャーズとして “Random walks in random environments” と題し，4 回の連続講義を行われました。また，2 度目の滞在期間 2014 年 2 月 18 日〜3 月 16 日の間に，GCOE セミナーとして “The two-dimensional random walk in an isotropic random environment” と題する講演を行われました。

このレクチャーノートは，これらの講義の内容をまとめたものです。作成にあたっては，特任研究員の久保田直樹氏ならびに大学院博士課程２年の角田謙吉氏に大変お世話になりました。また，Bolthausen 氏自身にもノートに加筆するなどしていただきました。

Bolthausen 氏，久保田直樹氏，角田謙吉氏に深く感謝いたします。

2015 年 3 月
舟木直久

Contents

1 Introduction

We consider Markov chains on the lattice $\mathbb{Z}^d$ or on a subset, which have site dependent transition probabilities. We denote by $\mathcal{P}$ a set of probability distributions on $\mathbb{Z}^d$. We assume that $\mathcal{P}$ contains only probabilities on a fixed *finite* subset $\mathcal{E} \subset \mathbb{Z}^d$, for instance $\mathcal{E} := \{e \in \mathbb{Z}^d : |e| = 1\}$. If m is the number points in $\mathcal{E}$, e.g. $m = 2d$ in the example we gave, then $\mathcal{P}$ can be regarded as a subset of $\mathbb{R}^m$, and of course, we will assume that it is Borel measurable. The set of Borel subset of $\mathcal{P}$ is denoted by $\mathcal{B}_{\mathcal{P}}$. A *field* of transition probabilities is described as an element $\omega \in \Omega \stackrel{\text{def}}{=} \mathcal{P}^{\mathbb{Z}^d}$, $\omega = (\omega_x)_{x\in\mathbb{Z}^d}$. Ω is equipped with the product σ-field $\mathcal{F} := \mathcal{B}_{\mathcal{P}}^{\otimes\mathbb{Z}^d}$. Then, the transition probabilities p_ω, which depend on $\omega \in \Omega$, of our "random walk in random environment" (RWRE for short) are 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 $X_0 = x, X_1, X_2, \cdots$. We write Γ for the set of paths in $\mathbb{Z}^d$, equipped with the appropriate σ-field $\mathcal{G}$. The reader will easily check that P defines a Markovian kernel from $\mathbb{Z}^d \times \Omega$ to Γ, i.e., for any $G \in \mathcal{G}$, the mapping $(x, \omega) \mapsto P_{x,\omega}(G)$ is a measurable mapping.

One should remark that to call this a "random walk" is a kind of misnomer, as random walks are usually understood to have transition probabilities which are homogeneous in the space $\mathbb{Z}^d$, which is not necessarily the case here. However, in the context of random environments, which we are just going to introduce, the name is firmly established.

The element $\omega \in \Omega$ is the "environment" for this Markov chain, and we will choose it now 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 $(\mathcal{P}, \mathcal{B}_{\mathcal{P}})$.

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

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

Often, one is only interested in the marginal law on Γ, for which by a slight abuse of nation, we the same symbol $\hat{P}_x$. It has become common to call this the *annealed law*, although this is kind of a misnomer, and one just better call it the *averaged law.* 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 averaged 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.

Remark 1 *Although we 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 reversibility. This is the model of random currents. In that model, the bonds are given random weights, which can be considered as a kind of random current which will be a measure how "easy" this bond can be crossed. To define the model formally, 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 $\omega \in (\mathbb{R}^+)^{\mathbb{B}^d}$, $b \in \mathbb{B}^d$, we write $\xi_b(\omega) \stackrel{\text{def}}{=} \omega_b$. 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 where on of the endpoints equal to x. If y is a nearest neighbor of x, then we define*

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

In other words, for fixed random environment ω, the random walk, being at a time point in x, jumps to the nearest neighbor point y with relative weight $\xi_{\{x,y\}}(\omega)$. In this case, $p_\omega(x,\cdot)$ and $p_\omega(x',\cdot)$ are not independent if $|x-x'|=1$. The main advantage of this model 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 reader can easily check that the transition probabilities given by (1.1) do not satisfy this detailed balance condition, in general.

2 One-dimensional nearest neighbor case

In this section, we consider the strictly one-dimensional nearest neighbor case. We will write an environment ω by $\omega_x = (p_x, q_x)$. p_x is the probability to move from x to $x+1$ and q_x is the probability to move from x to $x-1$. Therefore $q_x = 1 - p_x$ and the environment is described by the sequence $(p_x)_{x \in \mathbb{Z}}$. The following theorem is the main result of this section. It is a well-known result of Solomon [17].

Theorem 2 *Assume that the sequence* $\{\omega_x\} = \{(p_x, q_x)\}_{x\in\mathbb{Z}}$ *is stationary and ergodic under* $\mathbb{P}$ *and* $\mathbb{E}|\log p_x|,\ \mathbb{E}|\log q_x| < \infty$. *Define*

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

Then

a) $\lambda^+ > 0 \Leftrightarrow \lim_{n\to\infty} X_n = -\infty$, $\hat{P}_0$ *almost surely.*

b) $\lambda^+ < 0 \Leftrightarrow \lim_{n\to\infty} X_n = \infty$, $\hat{P}_0$ *almost surely.*

c) $\lambda^+ = 0 \Leftrightarrow \limsup_{n\to\infty} X_n = \infty$, *and* $\liminf_{n\to\infty} X_n = -\infty$, $\hat{P}_0$ *almost surely.*
In particular, $\lambda^+ = 0$ *holds if and only if the RWRE is recurrent.*

Remark 3 *The one-dimensional nearest neighbor RWRE is well studied and much more refined results are known. For instance, the following results are well-known:*

a) Under the assumption $\lambda^+ < 0$, *we have* $\mathbb{E}\frac{q_x}{p_x} < 1 \Leftrightarrow \lim_{n\to\infty}\frac{X_n}{n} > 0$ $\hat{P}_0$ *almost surely.*

b) If $\lambda^+ < 0$ *and* $\mathbb{E}\dfrac{q_x}{p_x} \geq 1$, *then* $\lim_{n\to\infty}\dfrac{X_n}{n} = 0$ *although by Theorem b), one has* $\lim_{n\to\infty} X_n = \infty$

c) In the strictly one-dimensional i.i.d. nearest-neighbor case with $\lambda^+ = \mathbb{E}\log(q_x/p_x) = 0$, *the displacement of the RWRE after time* n *is only of order* $(\log n)^2$. *(not* $\sqrt{n}$*)*

a) (and consequently b)) were proved by Solomon, too. c) is a result by Sinai [16].

We give a proof of Theorem 2 which can easily be adapted to the more complicated situation discussed later. We first introduce the exit distribution from the interval $[a,b] := \{a, a+1, \ldots, b-1, b\}$, $(a, b \in \mathbb{Z}, a < b)$. Define the (random) function $h_{a,b,\omega}(x)$, $x \in [a,b]$,

$$h_{a,b,\omega}(x) := P_{x,\omega}(T_b < T_a), \tag{2.1}$$

where $T_y = \inf\{n \geq 0 : X_n = y\}$ is the first hitting time of y. This quantity will play a crucial role in the proof. By the Markov property of $\{X_n\}$ under $P_{x,\omega}$, we obtain for $a < x < b$

$$h_{a,b,\omega}(x) = p_x h_{a,b,\omega}(x+1) + q_x h_{a,b,\omega}(x-1), \tag{2.2}$$

and the boundary conditions $h(b) = 1,\ h(a) = 0$. In order not to overburden the notation, we often drop the index ω, but the reader should keep in mind that these are *random* functions.

If $a < x$, we define

$$\varphi_a(x) := h_{a,x+1}(x) = P_{x,\omega}(T_{x+1} < T_a). \tag{2.3}$$

For fixed x, this is a non-decreasing random sequence as a function of a, and we therefore have

$$\eta(x) = \lim_{a \to -\infty} \varphi_a(x) = P_{x,\omega}(T_{x+1} < \infty) \in [0,1]\,.$$

For $a < x$, the strong Markov property implies the equation

$$\varphi_a(x) = p_x + q_x \varphi_a\,(x-1)\,\varphi_a\,(x)\,, \tag{2.4}$$

and letting $a \to -\infty$

$$\eta\,(x) = p_x + q_x \eta\,(x-1)\,\eta\,(x)\,. \tag{2.5}$$

For each $x \in \mathbb{Z}$, define $A_x := q_x/p_x$. Since we assumed that $\mathbb{E}|\log p_x|$, $\mathbb{E}|\log q_x| < \infty$, A_x is well-defined for $\mathbb{P}$-a.s. ω. By the ergodic theorem and the assumed ergodicity of the environment

$$\lim_{n\to\infty} \frac{1}{n} \log A_n \cdots A_1 = \frac{1}{n} \sum_{i=1}^{n} \log A_i = \mathbb{E} \log \frac{q_1}{p_1} = \lambda^+.$$

almost surely. Now we will prove a dichotomy. The following lemma follows easily from ergodicity and stationarity

Lemma 4 *Either* $\mathbb{P}\,(\eta(x) < 1) = 1$ *holds for all* $x \in \mathbb{Z}$ *or* $\mathbb{P}\,(\eta(x) = 1) = 1$ *holds for all* $x \in \mathbb{Z}$.

Proof. Let $B_x = \{\eta(x) < 1\}$. From (2.5) and $\eta\,(x-1) \le 1$, we obtain

$$\eta(x) \le p_x + q_x \eta(x-1).$$

Therefore, $\eta\,(x-1) < 1$ implies $\eta\,(x) < 1$, that is $B_{x-1} \subset B_x$ holds for all $x \in \mathbb{Z}$. On the other hand, by stationarity, $\mathbb{P}(B_x) = \mathbb{P}(B_{x-1})$. Putting these together, we see that $B_x \setminus B_{x-1}$ has $\mathbb{P}$-measure 0, that is, up to a $\mathbb{P}$-nullset, the event B_x is invariant under shifts. Therefore, by ergodicity, $\mathbb{P}(B_x) = 1$ is true for all x or $\mathbb{P}(B_x) = 0$ holds for all x. This finishes the proof of the lemma. ■

Let $\lambda_\eta = \mathbb{E} \log \eta(0)$. Remark first that $p_0 \le \eta\,(0) \le 1$, and as we assume $\mathbb{E}\,|\log p_0| < \infty$, we have that $\log \eta\,(0)$ is integrable and λ_η is well-defined in $(-\infty, 0]$. $\lambda_\eta = 0$ is equivalent with $\eta\,(0) = 1$ $\mathbb{P}$-a.s., which by Lemma 4 is equivalent with $\mathbb{P}\,(\eta(x) = 1,\ \forall x) = 1$. Applying again the ergodic theorem, we also have the relation

$$\lambda_\eta = \lim_{k\to\infty} \frac{1}{k} \log \Pi_{j=0}^{k-1} \eta(j),\ \mathbb{P}\text{-a.s.} \tag{2.6}$$

Lemma 5 *a)* $\lambda_\eta = 0$ *holds if and only if for any starting point* $x \in \mathbb{Z}$, $\limsup_{n\to\infty} X_n = \infty$ $\hat{P}_x$*-a.s., i.e.* $P_{x,\omega}\left(\limsup_{n\to\infty} X_n = \infty\right) = 1$ *for* $\mathbb{P}$ *almost all* ω.

b) $\lambda_\eta < 0$ *holds if and only if for any starting point* $x \in \mathbb{Z}$, $\lim_{n\to\infty} X_n = -\infty$ $\hat{P}_x$*-a.s.*

Proof. Assume that $\lambda_\eta = 0$. From Lemma 4, we have $\mathbb{P}\left(\eta\left(y\right) = 1,\ \forall y\right) = 1$. If $\eta\left(x\right), \eta\left(x+1\right), \ldots$ are all 1, then from the Markov property, one sees that for any $z > x,\ z \in \mathbb{Z}$, one has

$$P_{x,\omega}\left(T_z < \infty\right) = 1.$$

Therefore

$$P_{x,\omega}\left(T_z < \infty,\ \forall z > x\right) = 1,$$

that is

$$P_{x,\omega}\left(\limsup_{n\to\infty} X_n = \infty\right) = 1.$$

So, we have proved that this holds true for $\mathbb{P}$-almost all ω, i.e. we have proved a)

Assume that $\lambda_\eta < 0$. For the sake of notational simplicity, we assume that the starting point x is 0. We use (2.6). As $\lambda_\eta < 0$, we have $\lambda_\eta < \lambda_\eta/2 < 0$. Therefore, for $\mathbb{P}$-almost all ω, there exists $N(\omega) \in \mathbb{N}$ such that

$$\frac{1}{k}\log \Pi_{j=0}^{k-1}\eta_{_\omega}(j) \leq \frac{\lambda_\eta}{2}\ ;\ k \geq N(\omega).$$

We emphasize here the fact that the η are *random*, i.e. depend on ω. The product $\Pi_{j=0}^{k-1}\eta_{_\omega}(j)$ is by the Markov property simply $P_{0,\omega}(T_k < \infty)$ and we therefore conclude that

$$P_{0,\omega}(T_k < \infty) \leq \exp\left[k\lambda_\eta/2\right]$$

for $k \geq N(\omega)$. Since we assumed $\lambda_\eta < 0$ we have $\sum_k P_{0,\omega}(T_k < \infty) < \infty$, and using the Borel-Cantelli Lemma, we conclude

$$P_{0,\omega}(T_k < \infty,\ \text{i.o.}) = 0.$$

This implies that

$$P_{0,\omega}\left(\sup_n X_n < \infty\right) = 1$$

and we have proved that this holds for $\mathbb{P}$ almost all ω.

We now argue that this implies that $\lim_{n\to\infty} X_n = -\infty$. For $\mathbb{P}$ almost surely, all p_x, q_x are positive. The Markov chain is therefore irreducible. If $\sup_n X_n < \infty$, the chain is therefore transient and visits every point only finitely often. This however implies, together with $\sup_n X_n < \infty$, that $\lim_{n\to\infty} X_n = -\infty$. ■

Remark 6 *a) The above lemma sets up a dichotomy: Either one has*

$$\limsup_{n\to\infty} X_n = \infty,\ \hat{P}\text{-a.s.}$$

(irrelevant of the starting point) or

$$\lim_{n\to\infty} X_n = -\infty,\ \hat{P}\text{-a.s.}$$

This also means that if $\hat{P}\left(\lim_{n\to\infty} X_n = -\infty\right) > 0$*, then this probability is* 1.

b) As the whole concept of a RWRE is invariant under a reflection of $\mathbb{Z}$*, just by exchanging* p_x *with* q_{-x} *and* q_x *with* p_{-x}*, there is also the similar statement: Either one has* $\limsup_{n\to\infty} X_n = -\infty$ $\hat{P}$*-a.s. (irrelevant of the starting point) or* $\lim_{n\to\infty} X_n = \infty$ $\hat{P}$*-a.s.*

Proof of Theorem 2. Let $\Delta_a(x) := 1 - \varphi_a(x)$. Recall the equation (2.4) and perform a few elementary computations:

$$\begin{aligned}
\varphi_a(x) &= p_x + q_x\varphi_a\left(x-1\right)\varphi_a\left(x\right)\\
p_x &= \varphi_a(x) - q_x\varphi_a\left(x-1\right)\varphi_a\left(x\right)\\
p_x - p_x\varphi_a(x) &= q_x\varphi_a(x) - q_x\varphi_a\left(x-1\right)\varphi_a\left(x\right)\\
1 - \varphi_a(x) &= \frac{q_x}{p_x}\left[1-\varphi_a\left(x-1\right)\right]\varphi_a\left(x\right).
\end{aligned}$$

Which with the abbreviation $A_x := q_x/p_x$ reads as

$$\Delta_a(x) = A_x\Delta_a(x-1)\varphi_a(x).$$

Iterating this equality for $a \le y < x$ gives us

$$\Delta_a(x) = A_xA_{x-1}\cdots A_{y+1}\Delta_a(y)\varphi_a(y+1)\varphi_a(y+2)\cdots\varphi_a(x). \tag{2.7}$$

We can also let $a \to -\infty$ in this relation (with $y < x$ fixed), and get

$$\left(1-\eta\left(x\right)\right) = A_xA_{x-1}\cdots A_{y+1}\left(1-\eta\left(y\right)\right)\eta(y+1)\eta(y+2)\cdots\eta(x). \tag{2.8}$$

We first prove the direction $\Rightarrow$ in a) and b) of Theorem 2 and in fact first in case b). Assume that $\lambda^+ < 0$. Applying (2.7) to $y = a$, we obtain

$$\Delta_a(x) \le A_xA_{x-1}\cdots A_{a+1}.$$

Using the assumption $\lambda^+ = \mathbb{E}A_x < 0$, we obtain from the ergodic theorem in the exactly the same way as in the proof of Lemma 5 that for any fixed a, $\Delta_a(x)$ is $\mathbb{P}$-a.s. exponentially decaying in x for $x \to \infty$. This means that for $\mathbb{P}$-almost all ω, there exist $N\left(\omega, a\right) \in \mathbb{N}$ and $\varepsilon > 0$ such that

$$P_{x,\omega}(T_{x+1} > T_a) \le \exp\left[-\varepsilon\left(x-a\right)\right]$$

for $x \geq a + N(\omega, a)$. In particular, by enlarging $N(\omega, a)$ if necessary, we get

$$\sum_{x=a+N(\omega,a)}^{\infty} P_{x,\omega}(T_{x+1} > T_a) \leq \frac{1}{2}.$$

Using the (strong) Markov property, we conclude that if the Markov chain starts at $b := a + N(\omega, a)$, the probability it never reaches a is at least $1/2$. Therefore

$$P_{b,\omega}\left(\lim_{n\to\infty} X_n = \infty\right) \geq 1/2.$$

Hence, from Lemma 5, in the form of Remark 6 b), we can conclude that $P(\lim_{n\to\infty} X_n = \infty) = 1$.

To prove $\Rightarrow$ in a), we use the reflection to which we alluded in Remark 6 b): Define $\widetilde{p}_x := q_{-x}$, $\widetilde{q}_x := p_{-x}$. Then $\widetilde{\lambda}^+ = -\lambda^+$, where of course

$$\widetilde{\lambda}^+ := \mathbb{E} \log \frac{\widetilde{q}_x}{\widetilde{p}_x}.$$

So $\lambda^+ > 0$ is equivalent with $\widetilde{\lambda}^+ < 0$, and so we can apply what we have already proved to the reflected situation which proves $\lambda^+ > 0 \Rightarrow \lim_{n\to\infty} X_n = -\infty$ almost surely.

We next prove the directions $\Leftarrow$ in a) and b). By the same reflection argument as just explained, it suffices to cope with one case, and we take a). Therefore, we assume that $\lim_{n\to\infty} X_n = -\infty$, a.s.. Then, from Lemma 4 and Lemma 5, we have $\mathbb{P}$-a.s. $\forall x$, $0 < \eta(x) < 1$, and $\lambda_\eta < 0$. From the equality (2.8) with $y = 0 < x$, we have

$$1 - \eta(x) = A_x A_{x-1} \cdots A_1 (1 - \eta(0)) \eta(1) \cdots \eta(x).$$

Taking a logarithm in this equality,

$$\frac{1}{x} \log(1 - \eta(x)) = \frac{1}{x} \sum_{j=1}^{x} \log A_j + \frac{1}{x} \sum_{j=1}^{x} \log \eta(j) + \frac{1}{x} \log(1 - \eta(0)).$$

Letting now $x \to \infty$, the left hand side and the third summand on the right hand side converge to 0, and we obtain

$$\lambda^+ + \lambda_\eta = 0,$$

which, because of $\lambda_\eta < 0$ implies $\lambda^+ > 0$.

It remains c) which however is a consequence of a) and b) and Remark 6: According to a) and b), $\lambda^+ = 0$ is equivalent with that neither $\lim_{n\to\infty} X_n = \infty$ $\hat{P}$-a.s., nor $\lim_{n\to\infty} X_n = -\infty$ $\hat{P}$-a.s. By Remark 6 this is equivalent with that $\limsup_{n\to\infty} X_n = \infty$, and $\liminf_{n\to\infty} X_n = -\infty$, $\hat{P}$ almost surely. ■

Remark 7 *The standard proofs of Solomon's theorem use the fact that the difference equation (2.2) has an explicit solution which can be analyzed quite easily. The proof above has advantage that it can be generalized to situation where such an explicit solution is no longer possible.*

3 Quasi-one-dimensional RWRE

3.1 Statement of the result

A precise discussion of the a RWRE on $\mathbb{Z}$ with finite range jumps, not just nearest neighbor ones, is considerably more delicate. In that case $\omega_x,\ x \in \mathbb{Z}$, are random variables taking values in the set of probability measures on $\{-R, \cdots, R\}$, where R is some fixed natural number. $\omega_x(y)$ is then the probability with which the RWRE (under the quenched law) jumps from x to $x+y$, i.e. one has for fixed $\omega = \{\omega_x\}$ the law $P_{x,\omega}$ of a Markov chain $\{X_n\}_{n\geq 0}$ satisfying

$$\begin{aligned} P_{x,\omega}\left(X_0 = x\right) &= 1 \\ P_{x,\omega}\left(X_{n+1} = x + y | X_n = x\right) &= \omega_x\left(y\right). \end{aligned}$$

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

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

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

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

and boundary conditions $h = 1$ on $[b, \infty)$, $h = 0$ on $(-\infty, a]$. Evidently, if $\omega_x\left(R\right) > 0$, one can express $h\left(x+R\right)$ through $h(x+R-1), \ldots, h(x-R)$ which leads to following matrix expression for the vectors

$$\mathbf{h}_R\left(x\right) := \begin{pmatrix} h(x+R) \\ \vdots \\ \vdots \\ \vdots \\ h(x-R+1) \end{pmatrix}.$$

$$\mathbf{h}_R\left(x\right) = \begin{pmatrix} -\frac{\omega_x(R-1)}{\omega_x(R)} & \cdots & \frac{1-\omega_x(0)}{\omega_x(R)} & \cdots & -\frac{\omega_x(-R)}{\omega_x(R)} \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & & & \vdots \\ \vdots & & \ddots & & \vdots \\ 0 & & & 1 & 0 \end{pmatrix} \mathbf{h}_R\left(x-1\right).$$

One is then naturally led to the investigation of the products of the above random matrices. A difficulty is to match the boundary conditions. This is the approach of Key [12] who generalized the Solomon's 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 Bŕemont [5], [6].

We introduce a somewhat different approach in a slightly more general setup which has the advantage that the arguments of the previous section can be modified. This approach was developed in [2].

We now consider a RWRE in a *strip* of which $m : \mathrm{ST}_m := \mathbb{Z}\times\{1,\cdots,m\}$. We call the subset $\{k\}\times\{1,\cdots,m\}$ the k-th layer, and write sometimes as $\mathrm{LAY}(k)$ for it. Transitions in one step are possible from $\mathrm{LAY}(k)$ to $\mathrm{LAY}(k+1)$ and to $\mathrm{LAY}(k-1)$, but also inside $\mathrm{LAY}(k)$. The transitions probabilities are therefore described by a sequence of triplets $\omega = ((P_k, Q_k, R_k))_{k\in\mathbb{Z}}$ of positive $m\times m$-matrices, where P_k describes the transition probabilities from $\mathrm{LAY}(k)$ to $\mathrm{LAY}(k+1)$, Q_k from $\mathrm{LAY}(k)$ to $\mathrm{LAY}(k-1)$, and R_k the transitions inside $\mathrm{LAY}(k)$. Given an environment $\omega = ((P_k, Q_k, R_k))_{k\in\mathbb{Z}}$, we consider a Markov chain $X_n = (Y_n, Z_n) \in \mathrm{ST}_m$, $n \geq 0$, where Y_n is the component in $\mathbb{Z}$ and Z_n is the component in $\{1,\ldots,m\}$. This Markov chain has the following transition probabilities:

$$\begin{aligned} P_{(x,i),\omega}(X_0 = (x,i)) &= 1,\\ P.(X_{n+1} = (x+1,j)|X_n = (x,i)) &= P_x(i,j),\\ P.(X_{n+1} = (x-1,j)|X_n = (x,i)) &= Q_x(i,j),\\ P.(X_{n+1} = (x,j)|X_n = (x,i)) &= R_x(i,j). \end{aligned}$$

Of course, we have the assumption that $P_x + Q_x + R_x$ is a stochastic matrix. In this way, the law of the Markov chain on ST_m is uniquely defined after fixing the starting point $(x,i) \in \mathrm{ST}_m$.

Remark 8 *It is clear that a RWRE on $\mathbb{Z}$ with 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 pieces to be the layers. Due to the restrictions of the size of the jumps, in the layered situation only jumps to the nearest neighbor layers are possible.

We use the matrix norm

$$\|A\| := \max_i \sum_j |A(i,j)|,$$

and on $\mathbb{R}^m$, we will use the supremums norm $\|x\| := \max_i |x_i|$.

Condition 9 *Our basic assumptions for the main theorem are*

(C1) The sequence $(P_x, Q_x, R_x)_{x\in\mathbb{Z}}$ is stationary and ergodic.

(C2) For all $j \in \{1, \cdots, m\}$, and all $x \in \mathbb{Z}$, $\sum_i P_x(i,j)$ and $\sum_i Q_x(i,j)$ are strictly positive $\mathbb{P}$-almost surely.

(C3)

$$\mathbb{E} \log (1 - \|R_0 + P_0\|)^{-1} < \infty,$$
$$\mathbb{E} \log (1 - \|R_0 + Q_0\|)^{-1} < \infty.$$

In order to simplify somewhat the presentation, we use below a stronger assumption than (C2), namely

(C2') $P_x(i,j), Q_x(i,j) > 0$ $\mathbb{P}$-almost surely, for all $x \in \mathbb{Z}$, $i, j \in \{1, \ldots, m\}$.

This is not necessary but simplifies irreducibility considerations.

Three sequences of matrices: Crucial in our approach is the construction of three sequences of (random) $m \times m$-matrices φ_x, ψ_x, and A_x, $x \in \mathbb{Z}$, which will now be introduced. $\varphi_{a,x}$ will be similarly defined as in (2.3) with the additional information about the transition probabilities for the elements within the layers: If $a < x$ and $i, j \in \{1, \ldots, m\}$ we define

We define T_x to be the first entrance time of the Markov chain into $\mathrm{LAY}(x)$, $x \in \mathbb{Z}$. Then

$$\varphi_{a,x}(i,j) := P_{(x,i),\omega}\left(T_{x+1} < T_a,\ Z_{T_{x+1}} = j\right).$$

This is not a stochastic matrix, but it is a positive substochastic one: $\sum_j \varphi_{a,x}(i,j) < 1$. $\varphi_x(i,j)$ is the probability that the Markov chain, when starting in (x,i) enters the $\mathrm{LAY}(x+1)$ at j before reaching $\mathrm{LAY}(a)$. Evidently

$$\sum_j \varphi_{a,x}(i,j) = P_{(x,i),\omega}(T_{x+1} < T_a).$$

Due to (C2'), this is strictly smaller than 1. By the Markov property, one has the matrix identities

$$\varphi_{a,x} = P_x + R_x \varphi_{a,x} + Q_x \varphi_{a,x-1} \varphi_{a,x} \tag{3.1}$$

from which we can compute $\varphi_{a,x}$ in terms of $\varphi_{a,x-1}$:

$$\varphi_{a,x} = (I - R_x - Q_x \varphi_{a,x-1})^{-1} P_x.$$

Here I is the $m \times m$ identity matrix. Remark that the right hand side is well defined as we had assumed that

$$\|R_x + Q_x \varphi_{a,x-1}\| \leq \|R_x + Q_x\| < 1$$

almost surely. The boundary condition for $\varphi_{a,x}$ is $\varphi_{a,a} = 0$.

We next introduce the matrices ψ_x which satisfy the same recursion relation (3.1), but $\{\psi_x\}$ and $\{\varphi_x\}$ have a different boundary condition at LAY (a). ψ_x is always a *stochastic* matrix.

Given $a \in \mathbb{Z}$, $x > a$ and a stochastic matrix $\rho = (\rho(i,j))_{1\le i,j\le m}$, we define $\psi_x = \psi_{a,\rho,x}$ by

$$\psi_{a,\rho,x}(i,j) = \widetilde{P}_{(x,i),\omega}(Z_{T_{x+1}} = j),$$

where $\widetilde{P}_{(x,i),\omega}$ is the distribution of the Markov chain on $[a,\infty)\times\{1,\dots,m\}$ which has the above defined transition probabilities, except on LAY (a) where we have a reflection to the right with the matrix ρ, i.e. on this layer, we replace (P_a, Q_a, R_a) by $(\rho, 0, 0)$. It is evident that $\psi_{a,\rho,x}$ is a stochastic matrix, as the chain, when starting on LAY (x) has eventually to leave the finite set $[a,x]\times\{1,\dots,m\}$, and due to the reflection at LAY (a) it cannot do so on the left side. By the Markov property, we have the equations

$$\begin{aligned} \psi_{a,\rho,x} &= P_x + R_x\psi_{a,\rho,x} + Q_x\psi_{a,\rho,x-1}\psi_{a,\rho,x}, \\ \psi_{a,\rho,x} &= (I - R_x - Q_x\psi_{a,\rho,x-1})^{-1}P_x, \end{aligned} \tag{3.2}$$

for $x > a$, which are the same as for the $\varphi_{a,x}$, but now we have the different boundary condition $\psi_{a,\rho,a} = \rho$.

Proposition 10 *a) $\mathbb{P}$-a.s., there exists a sequence $\{y_x\}_{x\in\mathbb{Z}}$ of (random) stochastic matrices such that $\lim_{a\to-\infty}\sup_\rho\|\psi_{a,\rho,x} - y_x\| = 0$ $\mathbb{P}$-a.s.*

b) The sequence $\{y_x\}_{x\in\mathbb{Z}}$ is the unique sequence of stochastic matrices satisfying the equation $y_x = (I - Q_x y_{x-1} - R_x)^{-1}P_x$.

c) The sequence $\{(P_x, Q_x, R_x, y_x)\}_{x\in\mathbb{Z}}$ is stationary and ergodic.

Sketch of proof. We don't give a full proof which can be found in [2], Theorem 1, but we give an explanation what probabilistically is behind this result.

One has to distinguish two cases. The simple one is the case where $P_{(x,i),\omega}(T_{x+1} < \infty) = 1$ $\mathbb{P}$-a.s. It is not difficult to see from our assumptions that if this is true for one (x,i), then it is true for all others. In this case, the φ_x are stochastic matrices, and it is evident that

$$\psi_{a,\rho,x} \to \varphi_x$$

as $a \to -\infty$, uniformly in ρ. This simply comes from the fact that the chain, starting in (x,i) visits LAY (a) only with very small probability when $a \ll x$.

The more delicate case is when $P_{(x,i),\omega}(T_{x+1} < \infty) < 1$ with positive $\mathbb{P}$-probability, which actually implies by ergodicity, that it is true with $\mathbb{P}$-probability 1. In that case, there is a non-vanishing $P_{(x,i),\omega}$-probability that

the chain reaches LAY (a) with $a \ll x$, even in $a \to -\infty$ limit. This probability of course depends on ω. Under $\widetilde{P}_{(x,i),\omega}$ the chain is however now reflected to the right at LAY (a) where the law of the reflection is given by ρ. Due to this reflection, eventually, the chain reaches LAY $(x+1)$, but it is at first sight not completely clear why the first entrance distribution should not depend on ρ. Actually, it does, but with less and less dependence the smaller a is. The chain, when starting in LAY (a) has a high chance to return to LAY (a) before reaching LAY $(x+1)$ when $a \ll x$. Typically, the probability is exponentially close to 1, exponentially in $x-a$. Therefore, it needs an exponential number of "trials" to escape from LAY (a) and reach LAY $(x+1)$, but for fixed a, the probability is 1 that this finally (after a long time) happens. This successful "escape" from LAY (a) has to go through the strip on a the distance $x+1-a$, and due to the mixing properties of the transition probabilities, the influence of ρ gets lost for $a \ll x$.

It is possible to describe the excursion from LAY (a) to LAY $(x+1)$ probabilistically rather precisely. A complication is coming from the fact that in the RWRE case, the transitions, described by (P_y, Q_y, R_y) vary from layer to layer $a < y < x+1$.

For the formal proof, which is a bit tricky, see [2]. ■

Using the matrices ψ and y, we define the matrices $A_{\rho,a,x}$, $a < x$, and B_x in the following way:

$$A_{\rho,a,x} := (I - R_x - Q_x \psi_{\rho,a,x-1})^{-1} Q_x,$$
$$B_x := (I - R_x - Q_x y_{x-1})^{-1} Q_x.$$

It should be remarked that in the classical $m = 1$ case discussed in the previous chapter, one evidently has $\psi_{\rho,a,x} = y_x = 1$ for all x, a, ρ and therefore $A_x = B_x = \dfrac{q_x}{p_x}$ in agreement with the setting there.

By Proposition 10 c), The sequence $\{B_x\}$ is a stationary and ergodic sequence of non-negative random matrices. By Kingman's subadditive ergodic theorem (see [13], [9]) the following Lyapunov number exists, and is non-random

$$\lambda^+ = \lim_{N\to\infty} \frac{1}{N} \log \|B_N B_{N-1} \cdots B_1\|, \tag{3.3}$$

One should remark that in the $m = 1$ case, this is exactly what we had in the last chapter: $\lambda^+ = \mathbb{E} \log (q_x/p_x)$.

The following theorem describes the recurrence-transience behavior of the RWRE in terms of λ^+. It is an extension of the Solomon's Theorem.

Theorem 11 *Under the assumptions 9, the following hold.*

a) $\lambda^+ > 0 \Leftrightarrow \lim_{n\to\infty} Y_n = -\infty$ *almost surely.*

b) $\lambda^+ < 0 \Leftrightarrow \lim_{n\to\infty} Y_n = \infty$ *almost surely.*

c) $\lambda^+ = 0 \Leftrightarrow \limsup_{n\to\infty} Y_n = \infty$ *and* $\liminf_{n\to\infty} Y_n = -\infty \Leftrightarrow$*the RWRE is recurrent.*

Remark 12 *The theory of RWRE's in the quasi one-dimensional case has been developed considerably in the past years. First, the Sinai type behavior of Remark 3 has been proved in [4]. Then, also results similar to 3 a) and b) have been proved in [8].*

3.2 Proof of Theorem 11

The proof of the theorem runs parallel to the one for $m = 1$, given Proposition 10. There is however one part of the argument which is considerably more complicated for general m, namely the symmetry under reflection with regards to the $\mathbb{Z}$-component of the strip. We explain this first.

For $x \leq b$ set

$$\varphi^-_{b,x}(i,j) := P_{(x,i)}\left(T_{x-1} < T_b,\ Z_{T_{x-1}} = j\right).$$

The matrices with a reflection with a stochastic matrix ρ at b are defined by

$$\psi^-_{b,\rho,x}(i,j) := \widetilde{P}_{(x,i)}\left(Z_{T_{x-1}} = j\right), \tag{3.4}$$

where $\widetilde{P}_{(x,i)}$ has the same transition probabilities as before except at LAY (b) where (P_b, Q_b, R_b) is replace by $(0, \rho, 0)$. Furthermore

$$y^-_x := \lim_{b\to\infty} \psi^-_{b,x}(i,j).$$

These quantities satisfy similar equations as the original ones, for instance

$$\varphi^-_{b,x} = Q_x + R_x\varphi^-_{b,x} + P_x\varphi^-_{b,x+1}\varphi^-_{b,x}.$$

for $x < b$. We again define

$$A^-_{\rho,b,x} := (I - R_x - P_x\psi^-_{\rho,a,x+1})^{-1}P_x,$$
$$B^-_x := (I - R_x - P_x y^-_{x+1})^{-1}P_x,$$

and then

$$\lambda^- := \lim_{N\to\infty} \frac{1}{N} \log \|B^-_{-N}B^-_{-N+1}\cdots B^-_{-1}\|.$$

Lemma 13

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

The proof of this lemma will be given in Subsection 3.3.

Similarly to the η_x of Section 2, we define the η_x here as matrices:

$$\eta_x(i,j) = \lim_{a\downarrow-\infty} \varphi_{a,x}(i,j) = P_{(x,i)}(T_{x+1} < \infty, Z_{T_{x+1}} = j).$$

Here, the convergence is trivial as the $\varphi_{a,x}(i,j)$ are monotone increasing if a decreases. The η_x are also sub stochastic. By the Markov property, they satisfy the equation

$$\eta_x = P_x + R_x\eta_x + Q_x\eta_{x-1}\eta_x.$$

The following two lemmas are similar to the case $m = 1$. First the analogue to Lemma 4:

Lemma 14 *Under the Condition 9, one has the following alternative*

(i) Either for all $x \in \mathbb{Z}$, one has $\eta_x \cdot \mathbf{1} = \mathbf{1}$ $\mathbb{P}$-a.s. (meaning that the η_x are all stochastic matrices), where $\mathbf{1}$ denotes the vector with all components equal to 1.

(ii) or $\sum_j \eta_x(i,j) < 1$ holds for all (x,i) $\mathbb{P}$-a.s.

The proof of the Lemma is very similar to the proof of Lemma 4, and we leave it to the reader to fill in the details, or go to [2] Lemma 5 and Corollary 2. Kingman's subadditivity theorem and ergodicity gives the existence (and finiteness) of the Lyapunov exponent

$$\lambda_\eta = \lim_{n\to\infty} \frac{1}{n} \log \|\eta_1 \cdots \eta_n\| \leq 0.$$

Lemma 15 *Case (i) in Lemma 14 is equivalent to $\lambda_\eta = 0$ in which case one has $\limsup_{n\to\infty} Y_n = \infty$, $\hat{P}_{(x,i)}$-a.s. for all $(x,i) \in \mathrm{ST}_m$. (ii) is equivalent to $\lambda_\eta < 0$ which implies $\lim_{n\to\infty} Y_n = -\infty$, $\hat{P}_{(x,i)} - a.s.$ for all $(x,i) \in \mathrm{ST}_m$.*

Remember that $\hat{P}_{(x,i)}$ was the annealed law. $\hat{P}_{(x,i)}$-a.s. just means $P_{(x,i),\omega}$-a.s. for $\mathbb{P}$ almost all ω.

The proof of the above lemma is again very similar to the proof of Lemma 5, and so we leave the small necessary modifications to the reader.

Proof of Theorem 11. By Lemma 13, a) is equivalent to b). As in the $m = 1$ case, it is convenient to prove the implication $\Longrightarrow$ in case b) and the reverse implication in case a). Given Lemma 13, this proves a) and b) fully.

We start by introducing notations. For each a, x, $a < x$, define $\Delta_{a,x} := y_x - \varphi_{a,x}$. It is easy to see that the $\Delta_{a,x}$ are non-negative matrices: y_x and $\varphi_{a,x}$ both satisfy the same recursion relations (3.1), but the boundary condition at LAY (a) is 0 for φ and the non-negative matrix y_a for y. From that, it follows by induction on x that $y_x(i,j) \geq \varphi_{a,x}(i,j)$ for all $x > a$, $i,j \in \{1,\ldots,m\}$. (Actually, under Condition (C2'), $y_x(i,j) > \varphi_{a,x}(i,j)$). From the definitions of matrices, one sees, using the matrix identity $A^{-1} - B^{-1} = A^{-1}(B-A)B^{-1}$, that

$$\begin{aligned}
\Delta_{a,x} &= [(I - R_x - Q_x y_{x-1})^{-1} - (I - R_x - Q_x\varphi_{a,x-1})^{-1}]P_x \\
&= (I - R_x - Q_x y_{x-1})^{-1} Q_x (y_{x-1} - \varphi_{a,x-1})(I - R_x - Q_x\varphi_{a,x-1})^{-1} P_x \\
&= B_x \Delta_{a,x-1}\varphi_{a,x}.
\end{aligned}$$

Inductively, we obtain, for $a \leq u < x$,

$$\Delta_{a,x} = B_x \cdots B_{u+1} \Delta_{a,u} \varphi_{a,u+1} \cdots \varphi_{a,x}.$$

We also consider the $a \to -\infty$ limit of the above relations. Remember that $\eta_x := \lim_{a \to -\infty} \varphi_{a,x}$. We define $\hat{\Delta}_x := y_x - \eta_x$ which satisfies

$$\hat{\Delta}_x = B_x \cdots B_{u+1} \hat{\Delta}_u \eta_{u+1} \cdots \eta_x.$$

Now assume that $\lambda^+ < 0$. Fixing $a \in \mathbb{Z}$ we then have that $\|B_x \cdots B_{a+1}\|$ decays $\mathbb{P}$-almost surely exponentially fast if $x \to \infty$, and so, with $\varepsilon := -\lambda^+/2 > 0$, we find for almost all ω, a natural number $N(\omega, a) > x$ such that

$$\|\Delta_{a,x}\| \leq \mathrm{e}^{-\varepsilon(x-a)}, \ x \geq N(\omega, a).$$

For $a < x < b$

$$\begin{aligned} P_{(x,i)}(T_{x+1} < T_a) &= (\varphi_{a,x}\mathbf{1})(i) = ((y_x - \Delta_{a,x})\mathbf{1})(i) \geq 1 - \|\Delta_{a,x}\| \\ &\geq 1 - \mathrm{e}^{-\varepsilon(x-a)} \end{aligned}$$

the last inequality for $x \geq N(\omega, a)$. Arguing now in exactly the same way as for $m = 1$, one concludes first that $\hat{P}_{(a,i)}(\lim_{n\to\infty} Y_n = \infty) > 0$, and then it has to be 1.

We now deal the reverse implication in a). Assume that $\lim_{n\to\infty} Y_n = -\infty$. Again from Lemma 15, we conclude that for all $x \in \mathbb{Z}$, $\eta_x \mathbf{1} < \mathbf{1}$ $\mathbb{P}$-a.s.

$$\hat{\Delta}_x = B_x \cdots B_1 \hat{\Delta}_0 \eta_1 \cdots \eta_x. \tag{3.5}$$

Because of $\eta_x \mathbf{1} < \mathbf{1}$, $\left\{\hat{\Delta}_x\right\}$ is a stationary ergodic sequence of strictly sub-stochastic matrices. Therefore, using (C3), one has

$$\lim_{x\to\infty} \frac{1}{x} \log \left\|\hat{\Delta}_x\right\| = \lim_{x\to\infty} \frac{1}{x} \log \left\|\hat{\Delta}_0\right\| = 0$$

almost surely. Therefore,

$$\lim_{x\to\infty} \frac{1}{x} \log \|B_x \cdots B_1\| + \lim_{x\to\infty} \frac{1}{x} \log \|\eta_1 \cdots \eta_x\| = 0.$$

Hence $\lambda^+ = -\lambda_\eta > 0$, which finishes the proof of a), b) of Theorem 11.

The first implication in c) follows from a), b) and Lemma 15 in the same way as in the $m = 1$ case. The statement about recurrence and transience needs a slight additional argument which we only sketch: If $\limsup_{n\to\infty} Y_n = \infty$ and $\liminf_{n\to\infty} Y_n = -\infty$ then the Markov chain has to pass through $\mathrm{LAY}(x)$ infinitely often, for every x. It then follows from elementary Markov chain theory (and (C2')) that every point $(x, i) \in \mathrm{ST}_m$ is visited infinitely often, with probability 1. On the other hand, if $\lim_{n\to\infty} Y_n = \infty$ or $\lim_{n\to\infty} Y_n = -\infty$, then the Markov chain can visit points only finitely often. ■

3.3 Proof of Lemma 13

The proof of $\lambda^+ + \lambda^- = 0$ is not completely trivial (for $m > 1$). It depends on the existence of a unique invariant measure $(\pi_x)_{x\in\mathbb{Z}}$, $\pi_x = (\pi_x(i))_{1\le i\le m}$, unique up to normalization which satisfies

$$\pi_x = \pi_{x+1}Q_{x+1} + \pi_x R_x + \pi_{x-1}P_{x-1}.$$

We start by fixing $a < b$, $a, b \in \mathbb{Z}$ and impose reflecting boundary conditions on LAY (a) and LAY (b) by matrices ρ_a and ρ_b, i.e. we replace (P_a, Q_a, R_a) by $(\rho_a, 0, 0)$ and (P_b, Q_b, R_b) by $(0, \rho_b, 0)$, where ρ_a and ρ_b stochastic matrices. Since this modified Markov chain is defined on a finite set and is irreducible, there exists a unique stationary measure (unique modulo multiplying by a constant). We denote this stationary measure by $\{\pi_{a,b,x}\}_{x\in\mathbb{Z}}$, where $\pi_{a,b,x} = \pi_{a,b,\rho_a,\rho_b,x}$ are strictly positive vectors in $\mathbb{R}^m$. By the irreducibility assumptions (C2) and (C3) it is easily checked that the stationary distribution is positive everywhere. For the moment, keep a, b, ρ_a, ρ_b fixed. For notational simplicity, we just write $\pi_{a,b,x}$, but the dependence on ρ_a, ρ_b should be kept in mind.

Usually, the stationary measure (on a finite set) is normalized to be a probability measure, but for us, it is more convenient to normalize it differently. We assume that $a < 0 < b$, an then we assume that

$$\|\pi_{a,b,0}\| = \max_i \pi_{a,b,0}\,(i) = 1. \tag{3.6}$$

This defines $\{\pi_x\}_{a\le x\le b}$ uniquely, given the above boundary conditions ρ_a, ρ_b and the sequence $\{(P_x, Q_x, R_x)\}_{a<x<b}$. Uniqueness follows from the irreducibility condition (C2'). It is important to remember, that π_x of course depends on ω.

Lemma 16 *For $a \le x < b$, one has*

$$\pi_{a,b,x} = \pi_{a,b,x+1}\alpha_{a,x}$$

with

$$\begin{aligned}\alpha_{a,a} &= Q_{a+1}\\ \alpha_{a,x} &= Q_{x+1}(I - R_x - Q_x\psi_{a,x-1})^{-1},\ x > a.\end{aligned}$$

(Remark that α's depend also on ρ_a but not on b and ρ_b).

Proof. We write $\psi_{a,x}$ for $\psi_{a,\rho_a,x}$ for notational simplicity which satisfies the boundary condition $\psi_{a,a} = \rho_a$.

The relation $\pi_{a,b,a} = \pi_{a,b,a+1}Q_{a+1}$ is trivial. The other is a consequence of simple computations. Notice that

$$\begin{aligned}\pi_{a,b,a+1} &= \pi_{a,b,a}\rho_a + \pi_{a,b,a+1}R_{a+1} + \pi_{a,b,a+2}Q_{a+2}\\ &= \pi_{a,b,a+1}(Q_{a+1}\rho_a + R_{a+1}) + \pi_{a,b,a+2}Q_{a+2}.\end{aligned}$$

Since $\rho_a = \psi_{a,a}$, the above equality gives the result for $x = a+1$.

Inductively, we have

$$\begin{aligned}\pi_{a,b,x} &= \pi_{a,b,x-1}P_{x-1} + \pi_{a,b,x}R_x + \pi_{a,b,x+1}Q_{x+1} \\ &= \pi_{a,b,x}(\alpha_{a,x-1}P_{x-1} + R_x) + \pi_{a,b,x+1}Q_{x+1}.\end{aligned}$$

Therefore

$$\pi_{a,b,x}(I - \alpha_{a,x-1}P_{x-1} - R_x) = \pi_{a,b,x+1}Q_{x+1}.$$

From induction assumption, we have

$$\alpha_{a,x-1}P_{x-1} = Q_x(I - R_{x-1} - Q_{x-1}\psi_{a,x-2})^{-1}P_{x-1} = Q_x\psi_{a,x-1},$$

where we have used the recursion relation (3.2). Hence,

$$\pi_{a,b,x} = \pi_{a,b,x+1}Q_{x+1}(I - Q_x\psi_{a,x-1} - R_x)^{-1}.$$

This finishes the proof. ■

Remark 17 *Completely similarly, we obtain a representation in the other direction: For $a < x \leq b$*

$$\pi_{a,b,x} = \pi_{a,b,x-1}\beta_{b,x}$$

with

$$\begin{aligned}\beta_b &= P_{b-1} \\ \beta_{b,x} &= P_{x-1}\left(I - R_x - P_x\psi^-_{b,x+1}\right)^{-1},\ x < b,\end{aligned}$$

where the ψ^- were introduced in (3.4).

Lemma 18 *The limits*

$$\begin{aligned}\bar{\alpha}_x &:= \lim_{a\to-\infty}\alpha_{x,a} = Q_{x+1}(I - R_x - Q_x y_{x-1})^{-1}, \\ \bar{\beta}_x &:= \lim_{b\to\infty}\beta_{x,b} = P_{x-1}\left(I - R_x - P_x y^-_{x+1}\right)^{-1}\end{aligned}$$

exist and do not depend on the sequences $\{\rho_a\}, \{\rho_b\}$ for $a \to -\infty$ and $b \to \infty$.

Proof. This follows from the convergence property of $\psi_{a,x-1}$ for $a \to -\infty$, and $\psi^-_{b,x+1,b}$ for $b \to \infty$. ■

We can now finish the proof of Lemma 13.

We argue that $\lim_{a\to-\infty,b\to\infty}\pi_{a,b,0}$ exists. We first remark that

$$\begin{aligned}\pi_{a,b,0} &= \pi_{a,b,1}\alpha_{a,0} = \pi_{a,b,0}\beta_{b,1}\alpha_{a,0} \\ &= \pi_{a,b,n}\alpha_{a,n-1}\cdot\cdots\cdot\alpha_{a,0} \\ &= \pi_{a,b,0}\beta_{b,1}\cdot\cdots\cdot\beta_{b,n-1}\cdot\alpha_{a,n-1}\cdot\cdots\cdot\alpha_{a,0}\end{aligned}$$

for $a < 0 < n < b$. So, from the convergence of $\alpha_{x,a}, \beta_{x,b}$ we obtain that any possible limit $\bar{\pi}_0$ along subsequences of a and b satisfies

$$\bar{\pi}_0 = \bar{\pi}_0 \bar{\beta}_1 \cdot \cdots \cdot \bar{\beta}_{n-1} \cdot \bar{\alpha}_{n-1} \cdot \cdots \cdot \bar{\alpha}_0. \tag{3.7}$$

Our assumptions 9 however imply that the matrices $\bar{\beta}_1 \cdot \cdots \cdot \bar{\beta}_{n-1} \cdot \bar{\alpha}_{n-1} \cdot \cdots \cdot \bar{\alpha}_0$ are $\mathbb{P}$-almost surely irreducible non-negative matrices, and therefore $\bar{\pi}_0$ is uniquely define and we have proved that

$$\bar{\pi}_0 = \lim_{a \to -\infty, b \to \infty} \pi_{a,b,0}.$$

Similarly, it follows that

$$\bar{\pi}_x = \lim_{a \to -\infty, b \to \infty} \pi_{a,b,x}$$

exists (and does not depend on the boundary conditions ρ_a, ρ_b chosen), and $(\bar{\pi}_x)_{x \in \mathbb{Z}}$ is a stationary distribution for our Markov chain.

We can now prove $\lambda^+ + \lambda^- = 0$ by relating these Lyapunov exponents to the $\{\bar{\alpha}_x\}$, $\{\bar{\beta}_x\}$ sequences.

From (3.7), and $\pi_n = \bar{\pi}_0 \bar{\beta}_1 \cdot \cdots \cdot \bar{\beta}_{n-1}$ we obtain

$$1 = \|\pi_0\| = \|\pi_0 \bar{\beta}_1 \cdots \bar{\beta}_N\| \| \frac{\pi_N}{\|\pi_N\|} \bar{\alpha}_{N-1} \cdots \bar{\alpha}_0 \|. \tag{3.8}$$

Define two exponents $\lambda_{\bar{\alpha}}$ and $\lambda_{\bar{\beta}}$ by

$$\lambda_{\bar{\alpha}} := \lim_{n \to \infty} \frac{1}{n} \log \|\bar{\alpha}_{n-1} \cdots \bar{\alpha}_0\|, \ \lambda_{\bar{\beta}} := \lim_{n \to \infty} \frac{1}{n} \log \|\bar{\beta}_1 \cdots \bar{\beta}_n\|.$$

Then (3.8) implies that

$$\lambda_{\bar{\alpha}} + \lambda_{\bar{\beta}} = 0.$$

On the other hand,

$$\begin{aligned} \bar{\alpha}_{n-1} \cdots \bar{\alpha}_0 &= Q_n (I - R_{n-1} - Q_{n-1} y_{n-1})^{-1} Q_{n-1} \cdots (I - R_0 - Q_0 y_{-1})^{-1} \\ &= Q_n B_{n-1} \cdots B_1 (I - R_0 - Q_0 y_{-1})^{-1}. \end{aligned}$$

Therefore, taking the limit as $n \to \infty$ in the above equation gives us $\lambda_{\bar{\alpha}} = \lambda^+$. Analogously, we can conclude that $\lambda_{\bar{\beta}} = \lambda^-$. Hence $\lambda^+ + \lambda^- = 0$, which completes the proof of Lemma 13.

4 Exit distributions of RWRE on $\mathbb{Z}^d$ for $d \geq 3$

Throughout this section, we always assume the following.

Condition 19 *The law μ of a random environment ω_0 (see Section 1) satisfies:*

1. *$\mu(\mathcal{P}_\epsilon) = 1$ holds for some $\epsilon \in (0, 1/(2d))$, where*
$$\mathcal{P}_\epsilon := \{q \in \mathcal{P} : |q(e) - 1/(2d)| \leq \epsilon \text{ for all } e \in \mathcal{E}\}.$$
2. *For any orthogonal mapping O, leaving the lattice $\mathbb{Z}^d$ invariant, the laws of $(\omega_0(O(e)))_{e\in\mathcal{E}}$ and $(\omega_0(e))_{e\in\mathcal{E}}$ coincide.*

The aim of this section is to review the method developed in [1] and [3] to prove that, in dimension $d \geq 3$, the exit distributions of the RWRE on large balls (or more general sets) is close to the one of the ordinary random walk (ORW for short) on $\mathbb{Z}^d$, if the above condition is satisfied, for small enough ε.

To begin with, let us introduce some notations. For $x, y \in \mathbb{Z}^d$ set $I(x, y) = 1$ if $x = y$, and $I(x, y) = 0$ otherwise. A function $F : \mathbb{Z}^d \times \mathbb{Z}^d \to \mathbb{R}$, is also called a *kernel*. We will always assume that for any $x \in \mathbb{Z}^d$ the set $\left\{y \in \mathbb{Z}^d : F(x, y) \neq 0\right\}$ is finite. For two kernels F, G, we write FG for the (matrix) product
$$FG(x, y) := \sum_{z\in\mathbb{Z}^d} F(x, z)G(z, y),$$
which evidently is a kernel with the above finiteness property. For a given F, we define the powers F^k in the usual way, also be setting $F^0(x, y) := I(x, y)$.

We write $\|\cdot\|_{\text{tv}}$ for the total variation norm. For a finite subset V of $\mathbb{Z}^d$, denote the exit distributions from V for the RWRE by $\Pi_{V,\omega}(x, y)$, i.e., $\Pi_{V,\omega}(x, y) := P^x_\omega(X_{\tau_V} = y)$ where τ_V is the exit time for the RWRE from V. Similarly, $\pi_V(x, y)$ is the exit distribution from V for the ORW. Moreover, for $L \in \mathbb{R}^+$ and $x \in \mathbb{Z}^d$, let $V_L := \{v \in \mathbb{Z}^d; |v| \leq L\}$ and $V_L(x) := x + V_L$.

The main objective of this section is to review the following theorem.

Theorem 20 *Assume $d \geq 3$ and Condition 19. There exists $\epsilon_0 > 0$, depending only on the dimension such that for $\epsilon \in (0, \epsilon_0]$, the following statement is true: Consider an arbitrary smooth probability density $\varphi : \mathbb{R}^d \to \mathbb{R}^+$ of support in the unit ball. Then, for any $t \geq 1$ define a discrete smoothing kernel by*
$$\varphi_t(x, y) = \frac{\varphi((y - x)/t)}{\sum_{z\in\mathbb{Z}^d} \varphi(z/t)},\ x \in \mathbb{Z}^d,$$
which is well defined, at least for large enough t. For any function λ of L with $\lambda(L) \nearrow \infty$ as $L \to \infty$, we have
$$\mathbb{P}\left(\lim_{L\to\infty} \left\|(\Pi_{V_L} - \pi_{V_L})\varphi_{\lambda(L)}(0, \cdot)\right\|_{\text{tv}} = 0\right) = 1. \tag{4.1}$$

Remark 21 *In the proof, we work mainly with the following very special smoothing kernel $\hat{\pi}_t$ instead of φ_t which is defined by*

$$\hat{\pi}_t(x,y) := \int_1^2 \pi_{V_{\ell t}(x)}(x,y)\phi(\ell)\, d\ell,$$

where $\phi : \mathbb{R} \to \mathbb{R}^+$ is a smooth probability density with support on $[1,2]$. This choice is important for performing the induction explained below. In the end, with a simple additional argument, one can show the main result (4.1) for an arbitrary smoothing kernel. We will leave out the technical details of the proof of this claim.

Remark 22 *It is fairly obvious that $\|(\Pi_L - \pi_L)(0,\cdot)\|_{\mathrm{tv}}$ cannot go to 0. This is coming from the disorder close to the boundary which certainly has a non-vanishing effect on the total variation norm. However, it turns out that this effect of the disorder close to the boundary is very local, and is smeared out by a smoothing whose scale is increasing at an arbitrary small rate $\lambda(L)$. We give some more discussion about that later.*

There are only a few results on RWREs which satisfy conditions of the type above. The first one is a celebrated paper by Brimont and Kupiainen [7] which proved, under similar conditions, that the RWRE is diffusive for $d \geq 3$. For similar processes in continuous space and time, diffusivity has been proved by Sznitman and Zeitouni [18]. Theorem 20 was first proved in [3]. In [1] the statement was proved under a weaker condition than the isotropy condition. There, we assume only that μ is invariant under all d reflections $O_i : \mathbb{R}^d \to \mathbb{R}^d$ mapping the unit vector e_i to its inverse for each $i = 1, \ldots, d$.

Before starting with a discussion of the main technical steps, we give a quick heuristic argument which shows why $d \geq 3$ is important for the result. This argument indicates that the disorder is disappearing in the $L \to \infty$ limit for $d \geq 3$, but not for $d = 1$, with $d = 2$ being the difficult border line case.

Let $\xi_x(\omega)$ be the "quenched" expectation of the RWRE after one step. "Quenched" refers to keeping ω fixed, i.e.

$$\xi_x(\omega) := \sum_e \omega_x(e)\, e.$$

Evidently, from the basic assumptions, and the isotropy property, ξ_x has the same distribution for all x, and $\mathbb{E}\xi = 0$ and the covariance matrix is a multiple of the identity:

$$\mathrm{cov}(\xi) = \delta I_d,$$

I_d being the $d \times d$ identity matrix. δ depends on the distribution of μ, but as $(2d)^{-1}\sum_e e = 0$ and $|\omega_x(e) - (2d)^{-1}| \leq \varepsilon$, one has $\delta \leq \varepsilon^2$.

We now define

$$\xi_L(\omega) \stackrel{\text{def}}{=} \frac{E_{0,\omega}(X_{\tau_L})}{L},$$

and are interested in the quenched expectation of the exit position, scaled down by L in order to have an object which can be compared with ξ. It is plausible that for small ε, the leading contribution comes from a "kick" given by the disorder just at one location of V_L, and therefore

$$\xi_L(\omega) \approx \frac{1}{L} \sum_{y \in V_L} g_L(0,y)\, \xi_y, \tag{4.2}$$

where $g_L(0,y)$ is the expected number of visits of $y \in V_L$ by an ordinary random, before exiting V_L. This is a special case of the perturbation expansion (4.7) which will be discussed in details below. By the isotropy: $\mathbb{E}\xi_L = 0,\ \mathrm{cov}(\xi_L) = \delta_L I_d$.

How does δ_L behave for small ε? Using the above approximation, we get

$$\mathrm{cov}_{\mathbb{P}}(\xi_L) \approx \left[L^{-2} \sum\nolimits_{x \in V_L} g_L(0,x)^2\right] \delta I_d. \tag{4.3}$$

The right hand side is easy to evaluate in all dimensions:
$\underline{d=1:}$ In that case, for x not being close to the boundary, $g_L(0,x)$ is of order L, and therefore

$$\mathrm{cov}_{\mathbb{P}}(\xi_L) \approx \mathrm{const} \times L \delta I_d.$$

It follows that the disorder parameter δ is multiplied by a (big) factor of order L.
$\underline{d=2:}$ In this case, $g_L(0,x) \approx \log L$ for $x = 0$, and $g_L(0,x) \approx \frac{2}{\pi} \log \frac{L}{|x|}$, $x \neq 0$, close to 0. At the boundary of V_L, $g_L(0,x)$ behaves differently. One can use an approximation of g_L by the corresponding Green's function for the Brownian motion in the unit ball which is explicitly known, and some computation give for large L (and small δ)

$$\mathrm{cov}_{\mathbb{P}}(\xi_L) \approx \frac{2}{\pi} \delta I_d.$$

Therefore, it looks that the covariances stay always of order δ. However, one has to take into consideration that we have been very imprecise about the limits $L \to \infty,\ \delta \approx 0$, and the meaning of $\approx$ in (4.2). We will discuss this issue more careful below. What is quite easy to prove, however, is that for any fixed (large) L, one has

$$\lim_{\delta \to 0} \frac{1}{\delta} \mathrm{cov}_{\mathbb{P}}(\xi_L) = c(L)\, I_d,$$

where

$$\lim_{L \to \infty} c(L) = \frac{2}{\pi}.$$

This is just a simple exercise I leave to the reader to check. We however believe that for any fixed small enough δ, one has

$$\lim_{L\to\infty} \mathrm{cov}_{\mathbb{P}}\left(\xi_L\right) = 0,$$

but this is an extremely challenging open problem.
$\underline{d \geq 3:}$ In that case, $g_L\left(0,0\right)$ stays of order 1 for L large, and $g_L\left(0,x\right)$ decays like $|x|^{-d+2}$ for x not close to the boundary of V_L where it is smaller. Well known approximations for g_L (see for instance in [14]) show that

$$\sum_{x\in V_L} g_L\left(0,x\right)^2 \approx \mathrm{const} \times \begin{cases} L & \text{for } d=3 \\ \log L & \text{for } d=4 \\ 1 & \text{for } d\geq 5 \end{cases},$$

so that

$$\delta_L = \mathrm{const} \times \begin{cases} L^{-1}\delta & \text{for } d=3 \\ \left(\log L\right) L^{-2}\delta & \text{for } d=4 \\ L^{-2}\delta & \text{for } d\geq 5 \end{cases}.$$

Then, it appears, that the disorder is indeed contracting strongly for $d \geq 3$. A moments reflection however reveals that the above rough computations don't tell us much about the problem we are really after. The above approximations give indeed the correct behavior for arbitrary (large) L *when* $\delta \to 0$. This again is just a simple exercise. That's however not what we are after. We want $\delta > 0$ fixed (small), and $L \to \infty$. For that, the approximation (4.3) is totally useless, as the covariances of ξ_L for large L are certainly not determined by "kicks" at one place.

The way out of this difficulty is to do a multiscale analysis which will be explained in some details below.

4.1 Preliminaries

For a sub-Markov kernel $p = \left(p(x,y)\right)_{x,y\in\mathbb{Z}^d}$ and a finite subset V of $\mathbb{Z}^d$, we write

$$\left(\mathbf{1}_V p\right)(x,y) \stackrel{\mathrm{def}}{=} \begin{cases} p(x,y), & x\in V, \\ 0, & x\notin V, \end{cases}, \ x,\, y\in\mathbb{Z}^d.$$

The Green's function of p is defined by

$$g(p)(x,y) \stackrel{\mathrm{def}}{=} \sum_{n=0}^{\infty} p^n(x,y),$$

assuming that the sum converges. We will always be in situations where the convergence will be evident through cutoffs outside a finite region. We write $g_V(p) \stackrel{\mathrm{def}}{=} g(\mathbf{1}_V p)$.

Note that if $((P_x)_{x\in\mathbb{Z}^d}, (X_n)_{n=0}^{\infty})$ is a Markov chain with a transition probabilities $(p(x,y))_{x,y\in\mathbb{Z}^d}$, then for $x \in \mathbb{Z}^d$, $y \notin V$,

$$g_V(p)(x,y) = P_x(X_{\tau_V} = y),$$

where τ_V is the first exit time for the Markov chain from V. In particular, if $x \notin V$, then $g_V(p)(x,y) = \delta_{x,y}$.

Since $g(q) = I + qg(q) = I + g(q)q$, we have for sub-Markov kernels q_1 and q_2,

$$\begin{aligned} g(q_2)(q_1 - q_2)g(q_1) &= g(q_2)q_1g(q_1) - g(q_2)q_2g(q_1) \\ &= g(q_2)(g(q_1) - I) - (g(q_2) - I)g(q_1) \\ &= g(q_1) - g(q_2), \end{aligned}$$

and therefore

$$g(q_1) = g(q_2) + g(q_2)(q_1 - q_2)g(q_1). \tag{4.4}$$

One can iterate this equation, by replacing $g\,(q_1)$ on the right hand side again:

$$\begin{aligned} g(q_1) &= g(q_2) + g(q_2)(q_1 - q_2)\,[g(q_2) + g(q_2)(q_1 - q_2)g(q_1)] \\ &= g(q_2) + g(q_2)(q_1 - q_2)g(q_2) + g(q_2)(q_1 - q_2)g(q_2)(q_1 - q_2)g(q_1). \end{aligned}$$

Repeating this procedure inductively, we arrive for any $n \in \mathbb{N}$ at

$$g(q_1) = g\,(q_2) + \sum_{k=1}^{n} [g(q_2)(q_1 - q_2)]^k\, g\,(q_2) + [g(q_2)(q_1 - q_2)]^{n+1}\, g\,(q_1)\,. \tag{4.5}$$

In case that the last summand on the right hand side converges to 0 as $n \to \infty$, we get

$$g(q_1) = g\,(q_2) + \sum_{k=1}^{\infty} [g(q_2)(q_1 - q_2)]^k\, g\,(q_2)\,.$$

We then have expressed $g\,(q_1)$ through $g\,(q_2)$ and the differences $q_1 - q_2$.

Specializing to $q_1 = \mathbf{1}_V p_1,\ q_2 = \mathbf{1}_V p_2$, we get for the exit distributions $\pi_V^{(1)}, \pi_V^{(2)}$ of p_1, p_2 from V:

$$\pi_V^{(1)} = \pi_V^{(2)} + \sum_{k=1}^{\infty} [g(\mathbf{1}_V p_2)(\mathbf{1}_V p_1 - \mathbf{1}_V p_2)]^k\, \pi_V^{(2)}$$

provided that

$$\lim_{n\to\infty} [g(\mathbf{1}_V p_2)(\mathbf{1}_V p_1 - \mathbf{1}_V p_2)]^n = 0.$$

We will use these equations for various versions of the RWRE transitions p_1, and the ordinary random p_2. However, the transition kernels will enter in a

modified form which is suitable for the multiscale approach. Details will be given in the next section.

Let us first look at the first iteration where $p_1(x, x+e) := \omega_x(e)$ and $p_2(x, x+e) = (2d)^{-1}$. If $V \subset \mathbb{Z}^d$ is a finite set, Π_V is the exit distribution of the RWRE and π_V the one of the ordinary random walk, we get with $\Delta(x, x+e) = \omega_x(e) - (2d)^{-1}$ the perturbation expansion

$$\Pi_V = \pi_V + \sum_{k=1}^{\infty} [g_V \mathbf{1}_V \Delta]^k \pi_V. \tag{4.6}$$

On the right hand side, the disorder sits only in Δ. In case that the right hand side is dominated by the $k = 1$ summand, we get

$$\Pi_V \approx \pi_V + g_V \mathbf{1}_V \Delta \pi_V. \tag{4.7}$$

This immediately leads to (4.2). However, as remarked before, it is clear that for V_L, $L \to \infty$, there is absolutely no reason to believe that this is a good approximation, and in fact, the $k \geq 2$ terms dominate the series in (4.6). However, for fixed L and ε small, the $k = 1$ summand dominates. The key idea is to choose $L = L_1$ and ε such that (4.7) is a good approximation. Then, one concludes that $\Pi_{V_{L_1}} - \pi_{V_{L_1}}$ is small, with high probability, measured in an appropriate norm. Define now

$$\pi_L(x, y) \stackrel{\text{def}}{=} \pi_{V_L(x)}(x, y),$$
$$\Pi_L(x, y) \stackrel{\text{def}}{=} \Pi_{V_L(x)}(x, y),$$

which are our transition kernels on scale L, the first one coming from the ordinary random walk, and the second one from RWRE[1]. As explained, the difference

$$\Delta_L(x, \cdot) \stackrel{\text{def}}{=} \Pi_L(x, \cdot) - \pi_L(x, \cdot)$$

for $L = L_1$ should be smaller than the one on level $L = 1$, but one should keep in mind, that this has to be a probabilistic statement: There is a small probability that Δ_L is quite big. As a simple example, take a RWRE which has one randomly distributed preferred direction:

$$\omega_x(e) = \begin{cases} \frac{1}{2d} + \varepsilon & \text{if } e = \mathcal{E}_x(\omega) \\ \frac{1}{2d} - \frac{\varepsilon}{2d-1} & \text{if } e \neq \mathcal{E}_x(\omega) \end{cases},$$

where $\{\mathcal{E}_x\}_{x \in \mathbb{Z}^d}$ are independent uniformly chosen vectors of length 1 from the lattice. Then, in the box V_{L_1}, with probability $(2d)^{-|V_{L_1}|+1}$, all $\mathcal{E}_x$ are

[1] I hope, the notations will not confuse the reader. $\pi_{V_L}(x, \cdot)$ is the exit distribution for a random walk, starting in x, with exits from the ball V_L *centered* at 0. $\pi_L(x, \cdot)$ however is the exit distribution from a ball $V_L(x)$, centered at x, and for a walk starting in x. Evidently, $\pi_L(x, y) = \pi_{V_L}(0, y - x)$. There will however be modifications on later levels where the distinction is useful.

the same, and in that case, $\Delta_{L_1}(0,\cdot)$ is large in total variation, unless ε is really tiny. This event has however a probability which is exponentially small in $|V_{L_1}|$. By ergodicity, one has that almost surely there exist infinitely many $x \in \mathbb{Z}^d$, where $\Delta_{L_1}(x,\cdot)$ is large. The hope evidently is that these "bad" points x are thinly placed.

Neglecting this important point for the moment, one chooses $L_2 > L_1$ and one expresses $\Pi_{V_{L_2}} - \pi_{V_{L_2}}$ through the perturbation expansion in terms of $\Pi_{L_1}(x,\cdot) - \pi_{L_1}(x,\cdot)$, which are, appropriately measured, smaller than the original differences, and one chooses L_2 in such a way that again the $k=1$ term in the expansion dominates, and goes in this way, along a sequence of scales $L_1 < L_2 < L_3 < \cdots$.

As remarked, we expect that the Δ_{L_k} are getting smaller with k increasing, so that one can also choose the sequence $\{L_k\}$ in such a way that L_{k+1}/L_k is increasing. That will turn out to be an important technical point. Writing $\Delta_k(x.) \stackrel{\text{def}}{=} \Pi_{L_k}(x,\cdot) - \pi_{L_k}(x,\cdot)$, one gets in a somewhat schematic way:

$$\Delta_{k+1} = g_{k,k+k}\Delta_k\pi_{k,k+1} + g_{k,k+1}\Delta_k g_{k,k+1}\Delta_k\pi_{k,k+1} + \cdots, \tag{4.8}$$

where $g_{k,k+1}$ is the Green's function on $V_{L_{k+1}}$ based on random walk steps with transitions given by π_{L_k}, and $\pi_{k,k+1}$ is the exit distribution from $V_{L_{k+1}}$ coming from the same transitions on level k.

There are essentially two problems: The first one is that the above expression is not quite correct, as we can of course not express the exit distribution from $V_{L_{k+1}}$ through the exit distributions from $V_{L_k}(x)$, $x \in V_{L_{k+1}}$, because of problems near the boundary, where the exits from the smaller balls would overshoot the boundary of the bigger box. This is an annoying technical but essentially minor problem which is solved by adapting the transitions close to the boundary of $V_{L_{k+1}}$.

A more serious problem is coming from the fact that, as explained above, we cannot expect that differences of the exit distributions on large balls are close to 0, when measured for instance in total variation. They can only decrease if some smoothing is applied. In the induction step, one however wants to apply inductively smallness properties of Δ_k to prove that Δ_{k+1} is even smaller. In the perturbation expansion (4.8), Δ_k enters through $\Delta_k g_{k,k+1}$. $g_{k,k+1}$ is however not really a good smoothing operator.

The next serious problem is coming from the fact that there are always some x where $\Delta_k(x,\cdot)$ is large.

We cannot give full details of the approach how to solve all these issues, but let us give some indications of the key issues in the next sections.

4.2 The precise recursion scheme

The main scheme is an induction to transfer information about the exit distributions on one scale to information on a bigger scale, as shortly explained

in the last section. We first fix the sequence of scales: Start with $L_0 > 1$, and for $k \geq 0$ define $L_{k+1} \stackrel{\text{def}}{=} L_k(\log L_k)^3$ inductively.

One of the key difficulties to follow the ideas explained in the last section is that, as explained, the disorder can on big scales only decay if the exit distributions are smoothed. This implies that one has to work in the induction with smoothed transitions. We do a smoothing with averaging the exits over the radius, and, in addition, convolute the exit distribution with a smoothing kernel. Fix once for all a smooth probability density φ on $\mathbb{R}$ with support in $[1,2]$. Then we define

$$\hat{\pi}_L(x,y) \stackrel{\text{def}}{=} \int_1^2 \pi_{\ell L}(x,y)\,\varphi(\ell)\,d\ell,$$
$$\hat{\Pi}_L(x,y) \stackrel{\text{def}}{=} \int_1^2 \Pi_{\ell L}(x,y)\,\varphi(\ell)\,d\ell.$$

In order not to overburden the notations, we use some abbreviations, and write $\hat{\pi}_k$ for $\hat{\pi}_{L_k}$, and similarly $\hat{\Pi}_k$. Also, we write V_k instead of V_{L_k}. $\hat{\pi}_k, \hat{\Pi}_k$ will be our basic transition kernels with which the path moves inside a bigger box. The exit distributions without the averaging over the radius will be written without the hat: π_k, Π_k. As remarked, there is the problem to represent the exits from a bigger box $V_{L'}$, $L' > L$ through these transitions. In order to do this, we need a shrinking of the radius of the jumps inside the bigger box when the Markov chain approaches the boundary. We divide V_{k+1} into the following three regions:

$$A_{k+1}^{(1)} \stackrel{\text{def}}{=} \left\{x \in V_{k+1} : d(x, \partial V_{k+1}) \geq 4L_k = L_{k+1}/(\log L_k)^3\right\},$$

where $d(x, \partial V_{k+1})$ is the Euclidean distance of x from the boundary.

$$A_{k+1}^{(2)} \stackrel{\text{def}}{=} \left\{x \in V_{k+1} : L_{k+1}/(\log L_k)^{10} \leq d(x, \partial V_{k+1}) < 4L_{k+1}/(\log L_k)^3\right\},$$
$$A_{k+1}^{(3)} \stackrel{\text{def}}{=} \left\{x \in V_{k+1} : d(x, \partial V_{k+1}) < L_{k+1}/(\log L_k)^{10}\right\}.$$

Then, we define for $x \in V_{k+1}$, a transition kernel

$$\hat{p}_{k,k+1}(x,\cdot) \stackrel{\text{def}}{=} \begin{cases} \hat{\pi}_k(x,\cdot) & \text{for } x \in A_{k+1}^{(1)} \\ \hat{\pi}_{d(x,\partial V_{k+1})/4}(x,\cdot) & \text{for } x \in A_{k+1}^{(2)} \\ \pi_{V_{k+1} \cap V_{10L_{k+1}/(\log L_k)^{10}}(x)}(x,\cdot) & \text{for } x \in A_{k+1}^{(3)} \end{cases}$$

Similarly, define $\hat{P}_{k,k+1}$ for the RWRE. In addition, set

$$\Delta_{k,k+1} \stackrel{\text{def}}{=} \hat{P}_{k,k+1} - \hat{p}_{k,k+1}$$

and $g_{k,k+1}$ is the Green's function of $\hat{p}_{k,k+1}$, with killing at the boundary of V_{k+1}. In words: We start to shrink the radius for the transition as soon as

the chain is closer than $4L_{k+1}/\left(\log L_k\right)^3 = 4L_k$ to the boundary of V_{k+1}. The jump radius in this region is then always proportional to the distance to the boundary. We however stop this shrinking if the chain is closer than $L_{k+1}/(\log L_k)^{10}$ to the boundary, and when reaching such points, we just jump with the non-smoothed transition kernel, also cut at the boundary ∂V_{k+1}. The motivation for stopping the shrinking is probably difficult to see at the moment. Essentially, the point is that we don't want to cope with too many "bad" boxes, i.e. regions where $\Delta_{k,k+1}$ is larger than a certain value which will be specified below. One should however remark that in the last layer $A_{k+1}^{(3)}$, we can essentially not use any induction hypothesis for $\Delta_{k,k+1}$, as in this region, the exits are no longer described through exits from centered balls. However, $L_{k+1}/\left(\log L_k\right)^{10}$ is chosen because what happens on this scale is essentially irrelevant for the induction hypothesis we are now going to formulate.

Define $\varepsilon_k \stackrel{\text{def}}{=} \left(\log L_k\right)^{-9}$, and formulate the event

$$\mathrm{GOOD}^{(1)}\left(k\right) \stackrel{\text{def}}{=} \left\{ \left\| \left(\hat{\Pi}_k - \hat{\pi}_k \right) \hat{\pi}_k \left(0, \cdot\right) \right\|_{\mathrm{tv}} < \varepsilon_k \right\}.$$

Essentially, what we want to prove is the following implication: Assume that L_1 is large enough. Then for all $k \in \mathbb{N}$

$$\begin{aligned} &\mathbb{P}\left(\mathrm{GOOD}^{(1)}\left(j\right)\right) \geq 1 - \exp\left[-\left(\log L_j\right)^2\right], \ \forall j \leq k \\ \Longrightarrow\ &\mathbb{P}\left(\mathrm{GOOD}^{(1)}\left(k+1\right)\right) \geq 1 - \exp\left[-\left(\log L_{k+1}\right)^2\right]. \end{aligned} \tag{4.9}$$

This would be sufficient to prove our main theorem. We would start with taking L_0 appropriately large, and then choose ε small enough such that the claim is correct for $k = 0$. This can trivially be achieved. From that we would conclude that the statement holds for all k. This is not quite what we need in the theorem, as we use here a special smoothing kernel $\hat{\pi}_{L_k}$, whereas in the theorem, we allowed an arbitrary one. Also, in the expression $\hat{\Pi}_k - \hat{\pi}_k$, there is an averaging over the radii L with $L_k \leq L \leq 2L_k$ which is not present in the statement of the theorem. These are a minor technical points which can be taken care of at the end, and I am not going into details about it.[2]

There is a more serious problem, and in fact we have not been able to prove the implication (4.9). The difficulty comes from the fact that the Green's function is not a good smoothing kernel, and so, it is difficult to implement properties of $(\hat{\Pi}_j - \hat{\pi}_j)\hat{\pi}_j$ into the perturbation expansion. It however turns out, that only a relatively modest information on the *non-smoothed* exits is needed to remedy the situation. The point is, as had been

[2]The averaging over the radius is done only to be able to prove good estimates for the Green's function coming from the transitions on scale L_k. These estimates require "smooth" transition kernels. As we are not going to discuss these estimates at all, the reader can as well "forget" this averaging over the radius.

remarked before, the reason that $\|\Pi_k - \pi_k\|_{\text{tv}}$ cannot go to 0 as $k \to \infty$, comes only from effects of the disorder near the boundary. Although this effect does not disappear on large scales, it however is produced by essentially independent small regions near the boundary, and therefore, by a kind of law of large numbers, there should be good tail estimates. To formulate it, fix a $\delta \in (0,1)$ and define

$$\text{GOOD}^{(2)}(\delta, k) \stackrel{\text{def}}{=} \left\{ \left\| \hat{\Pi}_k(0,\cdot) - \hat{\pi}_k(0,\cdot) \right\|_{\text{tv}} < \delta \right\},$$

and put

$$\text{GOOD}(\delta, k) \stackrel{\text{def}}{=} \text{GOOD}^{(1)}(k) \cap \text{GOOD}^{(2)}(\delta, k)$$

Then the proper induction scheme is to prove that for properly chosen L_1, ε, δ, one has

$$\begin{aligned} &\mathbb{P}(\text{GOOD}(\delta, j)) \geq 1 - \exp\left[-(\log L_j)^2\right], \ \forall j \leq k \qquad (4.10) \\ \Longrightarrow\ &\mathbb{P}(\text{GOOD}(\delta, k+1)) \geq 1 - \exp\left[-(\log L_{k+1})^2\right]. \end{aligned}$$

There will still be a small technical modification needed, which we mention a bit later.

4.3 The induction: Outline of the proof

We will give some details about the implication from the left hand side of (4.10) to derive

$$\mathbb{P}\left(\text{GOOD}^{(1)}(k+1)\right) \geq 1 - \frac{1}{2}\exp\left[-(\log L_{k+1})^2\right].$$

In particular, we will explain why in the induction hypothesis, one needs the event $\text{GOOD}^{(2)}(\delta, j)$, $j \leq k$. Of course, one then still has to prove

$$\mathbb{P}\left(\text{GOOD}^{(2)}(\delta, k+1)\right) \geq 1 - \frac{1}{2}\exp\left[-(\log L_{k+1})^2\right],$$

but that, we leave out, and give just some rough indications.

We apply the perturbation expansion with $V = V_{k+1}$, $q_1 := \mathbf{1}_{V_{k+1}} \hat{P}_{k,k+1}$ and $q_2 := \mathbf{1}_{V_{k+1}} \hat{p}_{k,k+1}$ leading to

$$\Pi_{k+1}(0,\cdot) - \pi_{k+1}(0,\cdot) = \sum_{i=1}^{\infty} X_i(0,\cdot)$$

where

$$X_i(0,\cdot) := \left(g_{k,k+1}\Delta_{k,k+1}\right)^i \pi_{V_{k+1}}(0,\cdot), \ i \in \mathbb{N}.$$

Here the kernel $\pi_{V_{k+1}}(x,y)$ is appearing, which, as the reader may remember, is the exit distribution from V_{k+1} (which is centered at 0), by a random walk, starting in x and *not* 0.

We may also add a convolution with $\hat{\pi}_{k+1}$ to get information about Δ_{k+1} :

$$\Delta_{k+1}^{0} \stackrel{\text{def}}{=} \Pi_{k+1}\hat{\pi}_{k+1} - \pi_{k+1}\hat{\pi}_{k+1} = \sum_{i=1}^{\infty} X_1\hat{\pi}_{k+1}.$$

In the end, we also have to perform the averaging over a radius L between L_{k+1} and $2L_{k+1}$ in order to get Δ_{k+1}, but this is trivial step after having estimated Δ_{k+1}^{0}.

As remarked before, the scheme is chosen in such a way that the X_1-term, i.e. the first term in the perturbation expansion dominates, and the contribution of the other terms is in the end negligible. We in fact use somewhat sophisticated estimates only in the summand with $i = 1$, and estimate the others rather crudely.

We define $\mathrm{GOOD}^{(1)}(k,x)(\omega) \stackrel{\text{def}}{=} \mathrm{GOOD}^{(1)}(k)(\theta_x\omega)$, where $(\theta_x\omega)_y \stackrel{\text{def}}{=} \omega_{y+x}$, and similarly the shifted events, $\mathrm{GOOD}^{(2)}(\delta,k,x)$, $\mathrm{GOOD}(\delta,k,x)$.

The tasks one has to perform can be summarized roughly as follows. In order to get estimates of $\mathbb{P}\left(\mathrm{GOOD}^{(1)}(k+1)\right)$ one has to do:

1. Estimate $\|X_1\hat{\pi}_{k+1}\|_{\mathrm{tv}}$ on the event

$$\mathrm{ALLGOOD} := \bigcap\nolimits_{x\in V_{L_{k+1}}} \mathrm{GOOD}(\delta,k,x).$$

   Actually one has also to consider $\mathrm{GOOD}(\delta,j,x)$ for $j \le k$, for x near the boundary of $V_{L_{k+1}}$, but we leave out this technical nuisance.

2. Estimate $\|X_1\hat{\pi}_{k+1}\|_{\mathrm{tv}}$ on the complement of this event, but some $x \in V_{L_{k+1}}$ where $\mathrm{GOOD}(\delta,k,x)$ fails are concentrated in a subcube of $V_{L_{k+1}}$ of side length of order L_k. We call this event ONEBAD.

3. Estimate the probability of the complement of $\mathrm{ALLGOOD} \cup \mathrm{ONEBAD}$.

4. Estimate $\|X_i\hat{\pi}_{k+1}\|_{\mathrm{tv}}$ for $i \ge 2$ under the above events ALLGOOD and ONEBAD. As we will have an overall estimate of Task 3, we don't have to care for any details on $(\mathrm{ALLGOOD} \cup \mathrm{ONEBAD})^c$.

The reader should however keep in mind, that this solves only "half" of the problem, as we still remain to have to obtain a similar estimate for $\mathbb{P}\left(\mathrm{GOOD}^{(2)}(\delta,k+1)\right)$.

We start writing

$$\begin{aligned}(X_1\hat{\pi}_{k+1})(0,z) = &\sum_{x\in V_{k+1},y,y'\in V_{k+1}\cup\partial V_{k+1}} g_{k,k+1}(0,x)\,\Delta_{k,k+1}(x,y)\\ &\times\left(\pi_{V_{k+1}}\right)(y,y')\,\hat{\pi}_{k+1}(y',z).\end{aligned}$$

In principle, we would only be interested in probabilistic properties of $X_1(0,\cdot)$, but as explained before, we apply the smoothing kernel $\hat{\pi}_{k+1}$ in order to obtain a quantity which in total variation is smaller than the one on the previous level k. It should also be remarked that randomness in the above expression is only in $\Delta_{k,k+1}$. In order to be able to apply the inductive hypothesis, we observe that the exit distribution $\pi_{V_{k+1}}$ from V_{k+1} can be written as

$$\pi_{V_{k+1}} = \hat{p}_{k,k+1}\pi_{V_{k+1}},$$

simply because, by the strong Markov property, we can perform one step with the coarse grained jump probabilities given by $\hat{p}_{k,k+1}$, and then exit again according to the standard random walk. In order to apply the strong Markov property, one only has to observe that $\hat{p}_{k,k+1}(x,\cdot)$ is the probability distribution of the ordinary random walk, stopped at a (randomized) stopping time.

Observe furthermore that

$$\sum_y \Delta_{k,k+1}(x,y) = 0,$$

and therefore also

$$\sum_y \Delta_{k,k+1}\hat{p}_{k,k+1}(x,y) = 0.$$

Thus, we can write

$$\begin{aligned} &\sum_{x,y,y'} g_{k,k+1}(0,x)\,\Delta_{k,k+1}(x,y)\left(\pi_{V_{k+1}}\right)\left(y,y'\right)\hat{\pi}_{k+1}\left(y',z\right) \qquad (4.11)\\ &= \sum_{x,y} g_{k,k+1}(0,x)\,\xi(x,y)\left[\sigma(y,z) - \sigma(x,z)\right], \end{aligned}$$

with

$$\xi := \Delta_{k,k+1}\hat{p}_{k,k+1},\ \sigma := \pi_{V_{k+1}}\hat{\pi}_{k+1}.$$

On $\Delta_{k,k+1}\hat{p}_{k,k+1}(x,\cdot)$ we can apply the induction hypothesis.

There are some technical complications near the boundary which we don't want to discuss in details. Therefore we just look the case where $x \in A^{(1)}_{k+1}$. We use two facts about $\Delta_{k,k+1}\hat{p}_{k,k+1}$ in this region:

- Up to a probability $\exp\left[-(\log L_k)^2\right]$, $\left\|\Delta_{k,k+1}\hat{p}_{k,k+1}(x,\cdot)\right\|_{\text{var}}$ is $\leq \varepsilon_k$.
- If $d(x,x') \geq 4L_k$ then $\Delta_{k,k+1}\hat{p}_{k,k+1}(x,\cdot)$ and $\Delta_{k,k+1}\hat{p}_{k,k+1}(x',\cdot)$ are independent.

Essentially $g_{k,k+1}(x,y)$ should be like the Green's function of the ORW with a scaling, and the ORW Green's function should be like the Brownian motion Green's function.

We first have to discuss how $g_{k,k+1}(0,y)$ behaves. One should remember that it is the Green's function of a random walk in V_{k+1} jumping with smeared out exit probabilities from balls, centered at x, of radius between L_k and $2L_k$. It is therefore clear, that without the killing of the walk outside $V_{L_{k+1}}$, the Green's function on all of $\mathbb{Z}^d$ would be of order

$$L_k^{-d}\left(\frac{1}{1+|y|\,L_k^{-1}}\right)^{d-2}.$$

With the killing, the Green's function is only smaller, but for y not close to the boundary, the above form is essentially the right one. Most of the y-summation comes from y's which are at distance L_{k+1} from the origin, so we simplify things by replacing $g_{k,k+1}$ by $L_k^{-d}(L_{k+1}/L_k)^{-d+2}$. In particular, it is not difficult to prove that

$$\begin{aligned}\sum_{x\in V_{L_{k+1}}} g_{k,k+1}(0,x) &= O\left(L_{k+1}^d L_k^{-d}(L_{k+1}/L_k)^{-d+2}\right) = O\left((L_{k+1}/L_k)^2\right)\\ &= O\left((\log L_k)^6\right),\end{aligned}$$

as we had chosen $L_{k+1} = L_k(\log L_k)^3$. Of course, for a correct argument, one has to carefully check that the summation of y close to 0 does not spoil things, but this is an easy technical point. So we don't discuss it here.

Next, there is a very important observation which is the only one point where the isoptropy assumptions really enters: In order to apply exponential inequalities, we have to center the expression on the right hand side of (4.11) and write it as

$$\begin{aligned}&\sum_{x,y} g_{k,k+1}(0,x)\left[\xi(x,y)-\mathbb{E}\xi(x,y)\right]\left[\sigma(y,z)-\sigma(x,z)\right] \qquad (4.12)\\ &+\sum_{x,y} g_{k,k+1}(0,x)\,\mathbb{E}\xi(x,y)\left[\sigma(y,z)-\sigma(x,z)\right].\end{aligned}$$

Of course, $\mathbb{E}\xi(x,y)=\mathbb{E}\xi(0,y-x)$, and this inherits the invariance properties from the original random environment: It is invariant under lattice isometries. Together with the fact that the function $y\mapsto\sigma(y,z)$ is harmonic with respect to the transition kernel $\hat{\pi}_k$, this leads to two cancellation which are crucial, as we roughly explain:

Remember that we are finally interested in the total variation norm. So, it is natural to investigate $\|\sigma(y,\cdot)-\sigma(x,\cdot)\|_{\mathrm{var}}$ for x,y which are at a distance of order L_k, which is the relevant distance in the above expression. As σ is produced via the exit distributions from V_{k+1} convoluted with a smoothing kernel at scale L_{k+1}, it is plausible, and easy to prove, that

$$\|\sigma(y,\cdot)-\sigma(x,\cdot)\|_{\mathrm{var}} \le C\frac{L_k}{L_{k+1}} = C\frac{1}{(\log L_k)^3}.$$

A moment's reflection reveals that this is far from sufficient. We cannot expect that $\|\mathbb{E}\xi(x,\cdot)\|_{\text{var}}$ is better than $(\log L_k)^{-9}$. This follows from the induction hypothesis on level k. Using this estimate, we therefore would get for the second summand in (4.12) an estimate of order $(\log L_k)^{-6}$ which is far from the desired $(\log L_{k+1})^{-9}$. A finer argument uses the fact that $\sigma(x,\cdot)$ is essentially differentiable in x. Of course, one has to be careful here as we are on a lattice. Anyway, one can easily prove that there is a function $D\sigma(x,z)$ such that

$$\|\sigma(y,\cdot)-\sigma(x,\cdot)-\langle y-x, D\sigma(x,\cdot)\rangle\|_{\text{var}} \le C\frac{1}{(\log L_k)^{-6}},$$

for $|x-y|$ of order L_k. As evidently, $\sum_y (y-x)\,\mathbb{E}\xi(x,y)=0$ from the symmetry properties, one would get with this better approximation an estimate of order $(\log L_k)^{-9}$ for the second summand in (4.12). This is not quite enough as one needs something like $\frac{1}{2}(\log L_{k+1})^{-9}$, hoping of course that one gets the same for the first summand. In fact, one needs to go to a second derivative of $\sigma(x,\cdot)$ in x, and use a second cancellation coming again from the invariance property of $\mathbb{E}\xi(x,y)$ under discrete rotations and harmonicity of σ in x. In fact, we proved in [3] (see also Proposition 3.1 of [1]) that

$$\left\|\sum\nolimits_{x,y} g_{k,k+1}(0,x)\,\mathbb{E}\xi(x,y)\,[\sigma(y,\cdot)-\sigma(x,\cdot)]\right\|_{\text{var}} \le C(\log L_{k+1})^{-9-\eta},$$

for some $C,\eta>0$. Probably, with some efforts, one could take $\eta=3$.

From this one sees that the second summand in (4.12) is fine for the desired bound, and therefore, the first remains.

For that, one has to use probabilistic arguments. In fact, we definitely cannot assume that $\|\xi(x,\cdot)-\mathbb{E}\xi(x,\cdot)\|_{\text{var}}$ is of order $(\log L_k)^{-9}$ for all x. There is the additional technical problem that $\xi(x,\cdot)$ and $\xi(x',\cdot)$ are dependent if $|x-x'|$ are of order L_k. We divide $\mathbb{Z}^d$ into disjoint hypercubes $C_{\mathbf{j}}$ of side length $4L_k$, $\mathbf{j}\in\mathbb{Z}^d$. $\mathcal{C}_k$ is the set of such cubes which intersect V_{k+1}. We call a cube $C_{\mathbf{j}}$ in $\mathcal{C}_k$ *bad*, if there exists $x\in C_{\mathbf{j}}\cap V_{L_{k+1}}$ such that

$$\|(\Delta_k\hat{\pi}_{L_k})(x,\cdot)\|_{\text{var}} > (\log L_k)^{-9}. \tag{4.13}$$

As indicated before, we define ALLGOOD to be the event that there is no bad cube in $\mathcal{C}_k$. The event ONEBAD is the event that (4.13) occurs for some $x\in V_{k+1}$, but the set of such x is confined to at most two adjacent cubes. The event TWOBAD then defined as $(\text{ALLGOOD}\cup\text{ONEBAD})^c$. All these notions depend of course on k.

With these notions, our tasks 1-4 can now be attacked. We start with the simplest one:

Task 3:

$$\text{TWOBAD} \subset \bigcup\nolimits_{x,x' \in V_{k+1}:|x-x'|>4L_k} \Big[\left\{ \left\| \left(\Delta_k \hat{\pi}_k\right)(x,\cdot) \right\|_{\text{var}} > \left(\log L_k\right)^{-9} \right\} \\ \cap \left\{ \left\| \left(\Delta_k \hat{\pi}_k\right)\left(x',\cdot\right) \right\|_{\text{var}} > \left(\log L_k\right)^{-9} \right\} \Big].$$

If $|x-x'| > 4L_k$, then the events which are intersected are independent. Moreover, each event has, by the induction hypothesis, probability at most $\exp[-\left(\log L_k\right)^2]$. Therefore

$$\mathbb{P}\left(\text{TWOBAD}\right) \leq C\left(L_{k+1}\right)^2 \exp\left[-2\left(\log L_k\right)^2\right] \ll \exp\left[-\left(\log L_{k+1}\right)^2\right].$$

So this is done.

Task 1:

We have to estimate

$$\mathbb{P}\left(\left\|\left(\Pi_{k+1}-\pi_{k+1}\right)\hat{\pi}_{k+1}\right\|_{\text{tv}} \geq \varepsilon_{k+1},\ \text{GOOD}\right), \tag{4.14}$$

where $d \geq 3$ is crucial.

We again consider the cubes $C_{\mathbf{j}}$ as above. Put

$$y(x,z) \stackrel{\text{def}}{=} \sum_{y \in V_{k+1}} g_{k,k+1}(0,x)\{\xi(x,y) - \mathbb{E}\xi(x,y)\}\{\sigma(y,z) - \sigma(x,z)\},$$

so that

$$\left\|\left(\Pi_{k+1}-\pi_{k+1}\right)\hat{\pi}_{k+1}\left(0,\cdot\right)\right\|_{\text{tv}} = \sum_z \left| \sum\nolimits_{x \in V_{k+1}} y(x,z) \right|.$$

and we split the x-summation according to the cubes $C_{\mathbf{j}}$: For any $u \in C_{\mathbf{0}}$ we consider the translates by multiples of $4L_k$: $u + 4L_k\mathbf{n}$, $\mathbf{n} \in \mathbb{Z}^d$

$$\sum_z \left| \sum\nolimits_{x \in A_{k+1}^{(1)}} y\left(x,z\right) \right| = \sum_z \left| \sum\nolimits_{x \in A_{k+1}^{(1)}} y\left(x,z\right) \right| \\ \leq \sum_z \sum_{u \in C_{\mathbf{0}}} \left| \sum\nolimits_{\mathbf{n}: u+4L_k\mathbf{n} \in A_{k+1}^{(1)}} y\left(u+4L_k\mathbf{n},z\right) \right|.$$

The summation over z is over at maximum const $\times L_{k+1}^d$ points and the summation over $C_{\mathbf{0}}$ is over const $\times L_k^d$ points. The variables inside the absolute value on the right hand side are independent and have mean 0 and the number of them is const $\times \left(L_{k+1}/L_k\right)^d$. We have to know how big $y\left(x,z\right)$ can be, but it has to be remembered that we restrict for the moment to be inside GOOD where $\left\|\xi(x,\cdot) - \mathbb{E}\xi(x,\cdot)\right\|_{\text{var}}$ is at most $\left(\log L_k\right)^{-9}$. Then, for individual z, $\sigma\left(x,z\right)$ is of order L_{k+1}^{-d} but as we consider the differences $\sigma(y,z) - \sigma(x,z)$ with $|y-x| \approx L_k$, we gain an additional factor L_k/L_{k+1}

through a discrete differentiation. Therefore $|\sigma(y,z)-\sigma(x,z)|$ is of order $L_k L_{k+1}^{-d-1}$. Finally, $g_{k,k+1}(0,x)$ is of order $L_k^{-d}\left(L_k/L_{k+1}\right)^{d-2}$. As remarked before, this is not correct for x close to 0, but a more exact computation reveals that this does not change anything. Therefore, we get

$$\begin{aligned}|y(x,z)| &\approx L_k L_{k+1}^{-d-1} L_k^{-d}\left(L_k/L_{k+1}\right)^{d-2}\left(\log L_k\right)^{-9}\\ &= L_k^{-1} L_{k+1}^{-2d+1}\left(\log L_k\right)^{-9}.\end{aligned}$$

The Hoeffding inequality states that for independent centered random variables X_i with $|X_i|\le b_i$, one has

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i\right|\ge t\right)\le 2\exp\left[-\frac{2t^2}{\sum_i b_i^2}\right].$$

Applying this we get

$$\begin{aligned}\mathbb{P}\left(\sum_z\left|\sum\nolimits_{x\in A_{k+1}^{(1)}} y\left(x,z\right)\right|\ge t\right) &\le \exp\left[-t^2(\log L_k)^{18}\left(\frac{L_{k+1}}{L_k}\right)^{d-2}\right]\\ &= \exp\left[-t^2\left(\log L_k\right)^{3d+12}\right].\end{aligned}$$

For $t=\left(\log L_{k+1}\right)^{-9}$, we get a bound

$$\exp\left[-\operatorname{const}\times\left(\log L_k\right)^{3d+12}\left(\log L_k\right)^{-18}\right]\le\frac{1}{10}\exp\left[-\left(\log L_{k+1}\right)^2\right],$$

provided L_k is large enough. This can always be achieved by choosing already L_0 large enough, and adapt the ε for the starting condition small enough.

This settles Task 1, except that we have restricted the summation above to $x\in A_{k+1}^{(1)}$, and there remain the other two regions. $x\in A_{k+1}^{(2)}$ can be handled in essentially the same way, with some slight additional technical considerations. Here, we still use exits from smaller balls, but in a scale between $\approx L_k$ and $\approx L_{k+1}/\left(\log L_k\right)^{10}$. In order to do it, we need the induction hypothesis for certain $j<k$. The summation over $x\in A_{k+1}^{(3)}$ has however be handled differently, as we no longer can use fully the induction hypothesis. The exits from $x\in A_{k+1}^{(3)}$ are performed via truncated smaller balls. What helps is that the scale with which the random walk jumps when being in $A_{k+1}^{(3)}$ is now of order $L_k/\left(\log L_k\right)^{10}$ and some crude technical estimates give then that the contribution of the summation over $x\in A_{k+1}^{(3)}$ to the estimate of the expression in (4.12) is at most $\left(\log L_k\right)^{-10}\ll\left(\log L_{k+1}\right)^{-9}=\varepsilon_{k+1}$. This is technically slightly annoying, but not difficult to prove. We skip the argument for that.

Task 2:

Unfortunately, here, there is an additional difficulty here which we have not addressed till now: If $\|(\hat{\Pi}_k - \hat{\pi}_k)\hat{\pi}_k(x,\cdot)\|_{\text{tv}}$ is really bad in one or several points x in one of the cubes in $\mathcal{C}_k$, then $\|X_1\hat{\pi}_{k+1}(0,\cdot)\|_{\text{var}}$ will typically not be of order $(\log L_{k+1})^{-9}$, how good the situation in the other cubes may be. The reader can easily construct an example which shows that this is the case, even when the disorder outside the bad cube is as good as it can, i.e. the disorder that has disappeared completely.

However, using the same type of arguments as above, one can prove that if there is just one (or two adjacent) bad cubes, then $\|X_1\hat{\pi}_{k+1}(0,\cdot)\|_{\text{var}} \leq (\log L_k)^{-3}$ with high probability. More precisely, if $\|(\hat{\Pi}_k - \hat{\pi}_k)\hat{\pi}_k(x,\cdot)\|_{\text{tv}}$ is bigger than ε_k for some x in one (or two adjacent) cubes, but at maximum still $\leq (\log)^{-\alpha}$, with $0 \leq \alpha$, then $\|X_1\hat{\pi}_{k+1}(0,\cdot)\|_{\text{var}} \geq (\log L_k)^{-\min(9,\alpha+3)}$ has again probability $\ll \exp[-(\log L_{k+1})^2]$. I don't want to go into the technical details as they quite cumbersome, but essentially, the "badness" of the one (or two adjacent) cubes is washed out by the increase of the scale with the factor $(\log L_k)^3$, where the contributions of the good cubes is estimated in the same way as in the good situation.

Unfortunately, however, this requires a modification of the basic induction scheme (4.10) and one has to work with 3 levels of badness. Actually we did it with 4, in order to keep more flexibility: ONEBAD is split up according to $\sup_x \|(\Pi_k - \pi_k)\hat{\pi}_k(x,\cdot)\|_{\text{tv}}$, and (4.10) has to be adapted accordingly. There is no point to discuss that in technical details here, as it would take a couple of pages.

The reader should remember that up to now, we have only investigated the level $k+1$ based on the first term in the perturbation expansion, and we clearly have to check that the others $X_2, X_3, \ldots$ are not spoilsports. (Actually they are for $d = 2$). This is Task 4.

Task 4:

To see the problems which arise, we start by looking formally at the issue: With $g \stackrel{\text{def}}{=} g_{k,k+1}$, $\hat{\pi} \stackrel{\text{def}}{=} \pi_{V_{k+1}}\hat{\pi}_{k+1}$, for abbreviation, we get for $m \geq 2$,

$$X_m\hat{\pi}_{k+1} = g\,(\Delta g)^{m-1}\,\Delta\hat{\pi}.$$

First, we should observe that

$$\begin{aligned}\Delta g(x,y) &= \sum_u \Delta(x,u)\,g(u,y) \\ &= \sum_u \Delta(x,u)\,[g(u,y) - g(x,y)]\end{aligned}$$

(see the notation above (4.6)). This means that we "gain" a discrete derivative in the Green's function. In fact, it is not difficult to see that for x in the bulk and u at distance of order L_k from x, one gets

$$\sum_{y \in V_k} |g(u,y) - g(x,y)| \approx (\log L_k)^3,$$

where, as remarked before, $\sum_y g(x,y) \approx (\log L_k)^6$. Forgetting for the moment the issue of the necessity of smoothing of the Δ, we expect a behavior $\|\Delta(x,\cdot)\|_{\text{var}} \leq (\log L_k)^{-9}$ from the induction hypothesis. Therefore, in the above expression for $X_m\hat{\pi}_{k+1}$, we have the contribution of the first g, which, as there is no gain from a derivative, is of $(\log L_k)^6$, then m times the Δ, giving $(\log L_k)^{-9m}$, and $m-1$ times the g with the discrete gradient, giving $(\log L_k)^{(m-1)3}$. So, in the end we have, without doing any type of probabilistic estimates, however assuming that all cubes are good, that $X_n\hat{\pi}_{k+1}$ should in total variation behave like $(\log L_k)^3 (\log L_k)^{-6m}$. At least for $m \geq 3$, this would be totally fine as $(\log L_k)^{-n} \ll (\log L_{k+1})^{-9}$, for $n > 9$.

There is however the problem that the induction hypothesis does not give an estimate, which decays to 0, on $\|\Delta(x,\cdot)\|_{\text{var}}$ but only after smoothing. g is however not quite a good smoothing kernel, but we now observe, neglecting for simplicity the killing at the boundary,

$$g_{k,k+1} = I + \hat{\pi}_k g_{k,k+1}.$$

If the I would not be there, we would be fine, as we the kernel $\hat{\pi}_k$ appears which is smoothing Δ to $(\log L_k)^{-9}$, in good regions. Actually, this point is the very reason that we are working with the $\hat{\pi}$ as smoothing kernels. But we have to take care of the I which does not smooth Δ at all. For instance with $m = 2$, one has

$$\begin{aligned} X_2\hat{\pi}_{k+1} = {} & g_{k,k+1}\Delta^2_{k,k+1}\hat{\pi}_k\pi_{V_{k+1}}\hat{\pi}_{k+1} \\ & + g_{k,k+1}(\Delta_{k,k+1}\hat{\pi}_k)g_{k,k+1}\Delta_{k,k+1}\hat{\pi}_k\pi_{V_{k+1}}\hat{\pi}_{k+1}. \end{aligned}$$

In the second summand, one has twice smoothed Δ: $\Delta_{k,k+1}\hat{\pi}_{L_k}$ which are, by the induction hypothesis, of order at most $(\log L_k)^{-9}$, each, with high probability. There is still some work to be done along the line as in the X_1-case, but essentially, this part behaves much better than the $X_2\hat{\pi}_{k+1}$.

In the first summand, things are less pleasant. One should also remark that, in the end, we have to estimate $X_m\hat{\pi}_{k+1}$ and we get there a summand of the form $g_{k,k+1}\Delta^m_{k,k+1}\hat{\pi}_k\pi_{V_{k+1}}\hat{\pi}_{k+1}$, and we finally have to sum over m. It is here that we need the second part of the induction assumptions, namely that for some $\delta < 1$, one has $\|\Delta_{k,k+1}\|_{\text{var}} \leq \delta$, with high probability. Therefore, up to small probability, one gets

$$\left\|\Delta^m_{k,k+1}\hat{\pi}_{L_k}(x,\cdot)\right\|_{\text{var}} \leq \delta^{m-1}(\log L_k)^{-9},$$

and this can be summed over m.

In this way, one can finally prove that the left hand side of (4.10) implies $\mathbb{P}(\text{GOOD}^{(1)}(k+1)) \geq 1 - \frac{1}{2}\exp[-(\log L_{k+1})^2]$, where, however, we have neglected the additional complication coming from the different levels of "badness", as we remarked above.

There now remains the task to prove that the left hand side of (4.10) also implies

$$\mathbb{P}\left(\mathrm{GOOD}^{(2)}\left(\delta, k+1\right)\right) \geq 1 - \frac{1}{2}\exp\left[-\left(\log L_k\right)^2\right].$$

This however cannot be achieved through the shrinking scheme where we stop shrinking the balls on the lower scale at $L_{k+1}/\left(\log L_k\right)^{10}$. We need here a scheme where the balls from which we exit shrink proportional to the distance from the boundary of V_{k+1}. Essentially, that does not change the above procedure, except that on the small balls close to the boundary of V_{k+1}, say of size L_j, we of course only have estimates by ε_j, so there is no hope that we get an estimate for the total variation $\|\hat{\Pi}_{k+1} - \hat{\pi}_{k+1}\|_{\mathrm{tv}}$ which shrinks to 0. But that is no problem, as we know that there *cannot* be such an estimate. Therefore, in the case, that for all these exits in the smaller scale, we could use the bounds ε_j, then some easy modification of the scheme discussed would to the job. In fact, the corresponding Tasks 1, 2, and 4 are a straightforward rerun of the ones before.

However, there is the problem there can be no good estimate of

$$\mathbb{P}\left(\left(\mathrm{ALLGOOD} \cup \mathrm{ONEBAD}\right)^c\right)$$

in the modified setting, as this probability simply goes to 1 as $k \to \infty$. In fact, the reason we chose $L_{k+1}/\left(\log L_k\right)^{10}$ as the smallest radius for the exits was done exactly to avoid that.

That

$$\mathbb{P}\left(\left(\mathrm{ALLGOOD} \cup \mathrm{ONEBAD}\right)^c\right) \to 1$$

as $k \to \infty$, when we shrink the exits to scale of order 1 near the boundary of V_{k+1} can easily be seen. Whatever the size of V_L is, there is a nonvanishing probability that Δ, on this scale, is bad in total variation, say bigger than a constant. In our new scheme, we therefore have, with probability close to 1, many such non-intersecting cubes of a fixed size L to consider, and therefore, there are always many, for large k, which are bad.

In order to cope with this problem, one needs an additional probabilistic argument which shows that with large probability these bad regions, although being many, are thinly spread, and so they don't influence the total variation of the exit distribution from V_{k+1} too much, at least of one is only interested to get a total variation estimate by some (small) δ.

References

[1] Baur, E. and Bolthausen, E.: *Exit laws from large balls of (an) isotropic random walks in random environment.* arXiv:1309.3169, to appear in Ann. Prob.

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

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

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

[5] Brémont, J.: *On some random walks on Z in random medium,* Ann. Probab. **30**, 1266-1312 (2002).

[6] Brémont, J.: *Behavior of random walks on Z in Gibbsian medium*, C. R. Acad. Sci. Série 1 Math. 338, **11**, 895-898 (2004).

[7] Bricmont, J., and Kupianinen, A.: *Random walks in asymmetric random environments.* Comm. Math. Phys. **142** , 345–420 (1991).

[8] Dolgopyat, D.; Goldsheid, I.: *Limit theorems for random walks on a strip in subdiffusive regimes.* Nonlinearity **26**, 1782 (2013).

[9] Furstenberg, H., and Kesten, H.: *Products of random matrices.* Ann. Math. Statist. **31**, 457–469 (1960).

[10] Golosov, A.: *Localization of random walks in one-dimensional random environments.* Comm. Math. Phys. **92**, 491-506 (1984).

[11] Golosov, A.: *On the limit distributions for a random walk in a critical one-dimensional random environment.* Uspekhi Mat. Nauk **41**, no 2, 189-190 (1986).

[12] Key, E.: *Recurrence and transience criteria for random walk in a random environment.* Ann. Prob. **12**, 529–560 (1984).

[13] Kingman, J.: *Subadditive ergodic theory,* Ann. Probab. **1,** 883-909 (1973).

[14] Lawler, G., and Limic, V.: *Random Walk: A Modern Introduction.* Cambridge Studies in Advanced Mathematics 2010.

[15] Letchikov, A.V.: *Localization of one-dimensional random walks in random environment.* Soviet Scientific Reviews Section C: Mathematical Physics Reviews. Harwood Academic Publishers, 173-220, 1989.

[16] Sinai, Ya. G.: *The limiting behavior of a one-dimensional random walk in a random medium.* Theory Prob. Appl. **27**, 256-268 (1982).

[17] Solomon, F.: *Random walks in a random environment.* Ann. Prob. **3**, 1–31 (1975).

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

[19] Zeitouni, O.: *Random walks in random environment,* XXXI Summer school in Probability, St. Flour (2001). Lecture Notes in Math. **1837**, 193-312, Springer, Berlin, 2004.