Random walks and branching random walks: old and new perspectives

Perla Sousi*

Contents

1 Preliminaries

Let $(X_n)_{n\geq 0}$ be a simple random walk in $\mathbb{Z}^d$ with $d \geq 1$. Note that the time index will always be $\mathbb{N}$. For every $n \in \mathbb{N}$ and $x, y \in \mathbb{Z}^d$ we write $P^n(x,y) = p_n(x,y) = \mathbb{P}_x(X_n = y)$ for the n-step transition probability from x to y. Note that if $x = 0$ we will sometimes omit it from the notation and we will simply write $p_n(y)$ for $p_n(0,y)$. Note that by translation invariance of the walk we have $p_n(x,y) = p_n(0, x-y)$. We say that n and x are of the same parity if $p_n(x) > 0$. For every n and x we write

$$\overline{p}_n(x) = 2 \cdot \left(\frac{d}{2\pi n}\right)^{d/2} \cdot \exp\left(-\frac{d\|x\|^2}{2n}\right).$$

In dimension one, a direct calculation using Stirling's formula immediately yields the following:

Exercise 1.1. Let X be a simple symmetric random walk on $\mathbb{Z}$ starting from 0. Show that for all $n \in \mathbb{N}$ and $m \in \mathbb{Z}$ with $m \leq \sqrt{n}$ we have

$$\mathbb{P}_0(X_{2n} = 2m) = \overline{p}_{2n}(2m)(1 + o(1)) \text{ as } n \to \infty.$$

Hint: Recall Stirling's formula $n! \sim n^n\sqrt{2\pi n} \cdot e^{-n}$ as $n \to \infty$.

*University of Cambridge, Cambridge, UK; p.sousi@statslab.cam.ac.uk

In higher dimensions, one can also obtain an analogous result, but it is more tedious. We state without proof the local CLT that we will use very often in this course.

Theorem 1.2. (Local CLT [6, Proposition 1.2.5]) *Let $d \geq 1$ and let X be a simple random walk in $\mathbb{Z}^d$ started from 0. Suppose that n and x are of the same parity. Let $\alpha < 2/3$. If $\|x\| \leq n^\alpha$, then*

$$p_n(x) = \overline{p}_n(x) \cdot \left(1 + O(n^{3\alpha - 2})\right).$$

Exercise 1.3. Let X be a simple random walk on $\mathbb{Z}^d$ started from 0. Show using Azuma's inequality or otherwise that there exist positive constants c_1 and c_2 so that for all $x \in \mathbb{Z}^d$

$$\mathbb{P}_0(X_n = x) \leq c_1 \exp(-c_2 \|x\|^2/n).$$

Notation: For functions $f, g : \mathbb{N} \to \mathbb{R}_+$ we write $f \lesssim g$ if there exists a positive constant C so that for all $n \in \mathbb{N}$ we have $f(n) \leq Cg(n)$. We write $f \gtrsim g$ if $g \lesssim f$. Finally we write $f \asymp g$ if $f \lesssim g$ and $g \lesssim f$.

By the local CLT we see that

$$p_{2n}(0) \asymp \frac{1}{n^{d/2}}, \tag{1.1}$$

and hence we recover Polya's theorem: when $d \leq 2$, the SRW is recurrent, while when $d \geq 3$ it is transient.

Exercise 1.4. Prove (1.1) using Stirling's formula and the concentration of a Binomial random variable of parameters n and $1/d$ around its mean.

In the transient regime, we now define the Green's function as follows: for $x, y \in \mathbb{Z}^d$ with $d \geq 3$

$$g(x, y) = \mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbf{1}(X_n = y) \right] = \sum_{n=0}^{\infty} p_n(x, y).$$

When $x = 0$ we simply write $g(y)$ for $g(0, y)$. By conditioning on the first step of the random walk it is immediate to deduce the following:

Lemma 1.5. *The Green's function $g : \mathbb{Z}^d \to \mathbb{R}_+$ is harmonic on $\mathbb{Z}^d \setminus \{0\}$.*

Exercise 1.6. Using reversibility and the Cauchy-Schwartz inequality prove that for all $x \in \mathbb{Z}^d$

$$p_{2n}(0, x) \leq \sqrt{p_{2n}(0, 0) p_{2n}(x, x)} = p_{2n}(0, 0). \tag{1.2}$$

Combining (1.1) together with (1.2) we see that the Green's function is well defined when $d \geq 3$, as we get a converging series.

The following asymptotic expression for the Green's function will be used throughout these notes.

Theorem 1.7 (Spitzer)**.** *For all $d \geq 3$ and $\alpha < d$ as $\|x\| \to \infty$ we have*

$$g(x) = \frac{c(d)}{\|x\|^{d-2}} + o(\|x\|^{-\alpha}), \quad \textit{where} \quad c(d) = \frac{d}{2} \Gamma(d/2 - 1) \pi^{-d/2}.$$

Exercise 1.8. Prove Spitzer's result using the local CLT and by approximating the Riemann sum by an integral (see [7, Lemma 12.1.1]).

In the following exercise we obtain an upper bound on $g(x)$ of the correct order but without the sharp constant provided to us by Spitzer's result, which in turn follows by the local CLT.

Exercise 1.9. Let X be a SRW in $\mathbb{Z}^d$ with $d \geq 1$. Without appealing to the local CLT establish the following:

1. For all x of the same parity as 0 and satisfying $\|x\| \leq \sqrt{n}$ prove that
$$p_n(0,x) \asymp \frac{1}{n^{d/2}}.$$
2. Using reversibility prove that
$$\mathbb{P}_0(X_n = x) \leq 2 \cdot \mathbb{P}_0\big(X_n = x, \|X_{\lfloor n/2 \rfloor}\| \geq \|x\|/2\big).$$
3. Using the above and Azuma's inequality, show that there exist positive constants c_1 and c_2 such that for all x
$$\mathbb{P}_0(X_n = x) \leq \frac{c_1}{n^{d/2}} \exp(-c_2\|x\|^2/n).$$
4. Combining all of the above show that
$$g(x) \asymp \|x\|^{2-d}.$$

For a set $A \subseteq \mathbb{Z}^d$ we write
$$H_A = \inf\{n \geq 0 : X_n \in A\} \quad \text{and} \quad \widetilde{H}_A = \inf\{n \geq 1 : X_n \in A\}$$
for the first hitting and first return time to A respectively.

For a finite set A with $x \in A$, we write
$$g_A(x,y) = \mathbb{E}_x\left[\sum_{j=0}^{H_{A^c}} \mathbf{1}(X_j = y)\right]$$
We write $B(0,n) = \{x \in \mathbb{Z}^d : \|x\| < n\}$ for the Euclidean lattice ball of radius n.

Lemma 1.10. *Let $x \in B(0,n/4)$ and $T = \inf\{j \geq 0 : X_j \in \partial B(0,n)\}$. Then for all $y \in \partial B(0,n)$ we have*
$$\mathbb{P}_x(X_T = y) \asymp \frac{1}{n^{d-1}},$$
where the constants appearing in $\asymp$ are universal over all n.

Sketch of proof. Let $\zeta = \inf\{j \geq 1 : X_j \in \{0\} \cup \partial B(0,n)\}$. Then check that
$$\mathbb{P}_x(X_T = y) = g_{B(0,n)}(0,0)\mathbb{P}_y(X_\zeta = 0).$$
Then it suffices to show that $\mathbb{P}_y(X_\zeta = 0) \asymp n^{1-d}$, as $g_{B(0,n)}(0,0) \asymp 1$. To prove this, we define an intermediate scale, i.e. we first wait for the walk to either hit $B(0,n-3)$ or exit $B(0,n)$. We then require the walk to be at $B(0,n-3)$ at this time and estimate the probability that starting from there the walk hits 0 before hitting $\partial B(0,n)$. Finally to achieve this, we use the harmonicity of the Green's function g. $\square$

Theorem 1.11 (Harnack inequality)**.** *Let $f : B(0,n) \to \mathbb{R}_+$ be a harmonic function in $B(0,n-1)$. Then for all $0 < r < 1$, there exists a positive constant $C = C_r$ so that*
$$\sup_{x \in B(0,rn)} f(x) \leq C \inf_{x \in B(0,rn)} f(x).$$

2 Intersections of random walks

In this section we will study the question of intersections of independent simple random walks on $\mathbb{Z}^d$ for all d. We start with the question of collisions to see the analogy.

So let X and Y be two independent simple random walks on $\mathbb{Z}^d$ for $d \geq 1$ starting from 0. Let

$$C_n = \sum_{i=0}^{n} \mathbf{1}(X_i = Y_i).$$

Then taking expectations of both sides we get

$$\mathbb{E}[C_n] = \sum_{i=0}^{n} \mathbb{P}_0(X_{2i} = 0) \asymp \sum_{i=1}^{2n} \frac{1}{i^{d/2}} \asymp \begin{cases} \sqrt{n} \text{ if } d = 1 \\ \log n \text{ if } d = 2 \\ 1 \text{ if } d \geq 3. \end{cases}$$

We thus see from here that dimension 2 is critical for the question of collisions which is of course the well-known theorem of Polya.

Next we move on to intersections. Let I_n be the total number of intersections of X and Y up to time n, i.e.

$$I_n = \sum_{i=0}^{n} \sum_{j=0}^{n} \mathbf{1}(X_i = Y_j).$$

Taking expectations above we get

$$\mathbb{E}[I_n] = \sum_{i=0}^{n} \sum_{j=0}^{n} \mathbb{P}_0(X_{i+j} = 0) \asymp \sum_{i=0}^{2n} i \cdot p_i(0,0).$$

Using the LCLT, we now get that

$$\mathbb{E}[I_n] \asymp \sum_{i=1}^{n} \frac{1}{i^{\frac{d}{2}-1}} \asymp \begin{cases} \sqrt{n} \text{ if } d = 3 \\ \log n \text{ if } d = 4 \\ 1 \text{ if } d \geq 5. \end{cases}$$

We see thus that dimension 4 is the critical dimension when considering intersections analogously to dimension 2 being the critical dimension when considering collisions.

In the next section we are going to calculate the probability that one random walk avoids a two sided random walk in four dimensions. Then we will move to higher dimensions and study large deviations events for the number of intersections, i.e. we will bound the probability that the number of intersections is very large. From the above we see that in high dimensions, the expected number of intersections is of constant order.

2.1 Intersections in four dimensions

As we already discussed, dimension 4 is the critical dimension for the problem of intersections. What is usually expected at the critical dimension is logarithmic corrections to mean field behaviour. The main result of this section is to prove Lawler's result on the non-intersection between a random walk and an independent two-sided random walk in $\mathbb{Z}^4$.

Theorem 2.1 (Lawler (1985))**.** *Let X^1, X^2 and X^3 be three independent simple random walks in $\mathbb{Z}^4$ starting from 0. Then as $n \to \infty$*

$$\mathbb{P}\big(X^1[1,\infty) \cap (X^2[0,n] \cup X^3[0,n]) = \emptyset, 0 \notin X^2[1,n]\big) \sim \frac{\pi^2}{8} \cdot \frac{1}{\log n}$$

The proof that we will present follows Lawler's original argument with some simplifications due to Bai and Wan [3] and Bruno Schapira [9], who generalised it to branching random walks that we will discuss in the final section.

The whole proof is based on the magic equality of Lemma 2.2 which is a consequence of the last exit decomposition formula that we will state shortly. First we need to set up some notation.

Let X be a two-sided simple random walk in $\mathbb{Z}^4$, i.e. $(X_n)_{n\geq 0}$ and $(X_{-n})_{n\geq 0}$ are two independent random walks started from 0. Let $\widetilde{X}$ be an independent simple random walk in $\mathbb{Z}^4$ also started from 0. For every integers $a < b$ we write $\mathcal{R}[a,b] = \{X_a, \ldots, X_b\}$ and for $a, b \in \mathbb{N}$ we set $\widetilde{\mathcal{R}}_n = \{\widetilde{X}_a, \ldots, \widetilde{X}_b\}$ for the ranges of the two walks during the time interval $[a,b]$. Let ξ_n^ℓ and ξ_n^r be two independent geometric random variables of parameter $1/n$ each ($\mathbb{P}\big(\xi_n^\ell = j\big) = 1/n \cdot (1-1/n)^j$ for all j), also independent of the walks. Finally we define

$$\begin{aligned}
\mathcal{A}_n &= \{\widetilde{\mathcal{R}}[1,\infty) \cap \mathcal{R}[-\xi_n^\ell, \xi_n^r] = \emptyset\} \\
e_n &= \mathbf{1}(0 \notin \mathcal{R}[1, \xi_n^r]) \\
G_n &= \sum_{-\xi_n^\ell \leq k \leq \xi_n^r} g(X_k).
\end{aligned}$$

Lemma 2.2. *With the above definitions we have*

$$\mathbb{E}[\mathbf{1}(\mathcal{A}_n) \cdot e_n \cdot G_n] = 1.$$

In Lemma 2.4 below we will show that G_n is concentrated around its mean (which is of order $\log n$). Hence, if we could just pull it out of the expectation above, we would get exactly the statement of the theorem. The proof will then proceed by showing that this is actually correct up to smaller order terms.

Before giving the proof of the magic formula (Lemma 2.2) we state and prove the last exit decomposition formula which is an easy consequence of the Markov property. This result will be used repeatedly throughout these notes.

Lemma 2.3 (Last exit decomposition formula)**.** *Let $d \geq 3$ and let $A \subseteq \mathbb{Z}^d$ be a finite set. Then for all $x \in \mathbb{Z}^d$ we have*

$$\mathbb{P}_x(H_A < \infty) = \sum_{y \in A} g(x,y) \mathbb{P}_y\Big(\widetilde{H}_A = \infty\Big).$$

Proof. Let $L_A = \sup\{t \geq 0 : X_t \in A\}$ be the last time X visits A with the convention that $L_A = -\infty$ if the set is empty. Then by transience of the walk we get $\{H_A < \infty\} = \{0 \leq L_A < \infty\}$, and hence

$$\begin{aligned}
\mathbb{P}_x(H_A < \infty) = \mathbb{P}_x(0 \leq L_A < \infty) = \sum_{n=0}^{\infty} \sum_{y \in A} \mathbb{P}_x(L_A = n, X_n = y) &= \sum_{n=0}^{\infty} \sum_{y \in A} \mathbb{P}_x(X_n = y)\, \mathbb{P}_y\Big(\widetilde{H}_A = \infty\Big) \\
&= \sum_{y \in A} g(x,y) \mathbb{P}_y\Big(\widetilde{H}_A = \infty\Big)
\end{aligned}$$

where for the penultimate equality we used the Markov property. $\square$

Proof of Lemma 2.2. For every nearest neighbour path $(x_1,\dots,x_m)$ we define

$$B(m,x_1,\dots,x_m)=\{\xi_n^\ell+\xi_n^r=m,\ X_{-\xi_n^\ell+k}-X_{-\xi_n^\ell}=x_k,\ \forall\, 1\le k\le m\},$$

and for all $0\le j\le m$ we define

$$B(m,j,x_1,\dots,x_m)=\{\xi_n^\ell=j,\ \xi_n^r=m-j,\ X_{-\xi_n^\ell+k}-X_{-\xi_n^\ell}=x_k,\ \forall\, 1\le k\le m\}.$$

Using the independence of the increments of the walk and the geometric random variables we then obtain

$$\mathbb{P}(B(m,j,x_1,\dots,x_m)\mid B(m,x_1,\dots,x_m))=\frac{1}{m+1}.$$

Setting $x_0=0$, we then have

$$\begin{aligned}&\mathbb{E}[\mathbf{1}(\mathcal{A}_n)\cdot e_n\cdot G_n]\\ =&\sum_{m=0}^{\infty}\sum_{(x_1,\dots,x_m)}\frac{\mathbb{P}(B(m,x_1,\dots,x_m))}{m+1}\cdot\sum_{k=0}^{m}\sum_{j=0}^{m}\mathbf{1}(x_j\notin\{x_{j+1},\dots,x_m\})\\ &\times\mathbb{P}\Big((x_j+\widetilde{\mathcal{R}}[1,\infty))\cap\{x_0,x_1,\dots,x_m\}=\emptyset\Big)\,g(x_j-x_k).\end{aligned}$$

Using the last exit decomposition formula to the set $\{x_0,\dots,x_m\}$ and the starting point x_k we get

$$1=\sum_{j=0}^{m}\mathbf{1}(x_j\notin\{x_{j+1},\dots,x_m\})\times\mathbb{P}\Big((x_j+\widetilde{\mathcal{R}}[1,\infty))\cap\{x_0,x_1,\dots,x_m\}=\emptyset\Big)\,g(x_j-x_k).$$

Substituting this above we obtain

$$\mathbb{E}[\mathbf{1}(\mathcal{A}_n)\cdot e_n\cdot G_n]=\sum_{m=0}^{\infty}\sum_{(x_1,\dots,x_m)}\mathbb{P}(B(m,x_1,\dots,x_m))=1,$$

and this concludes the proof. $\square$

Lemma 2.4. *There exists a positive constant C so that the following holds. Let X be a simple random walk on $\mathbb{Z}^4$ started from 0 and let ξ be an independent geometric random variable of mean n. Then*

$$\mathbb{E}\left[\sum_{i=0}^{\xi}g(X_i)\right]=\frac{4}{\pi^2}\cdot\log n+O(1)\quad\textit{and}\quad\mathrm{Var}\left(\sum_{i=0}^{\xi}g(X_i)\right)\le C\log n.$$

We defer the proof of this lemma to the end of the section and we now give the

Proof of Theorem 2.1. Lemma 2.2 states that

$$\mathbb{E}[\mathbf{1}(\mathcal{A}_n)\cdot e_n\cdot G_n]=1.$$

We now get

$$\mathbb{E}[\mathbf{1}(\mathcal{A}_n)\cdot e_n]=\frac{1}{\mathbb{E}[G_n]}+\frac{1}{\mathbb{E}[G_n]}\cdot\mathbb{E}[\mathbf{1}(\mathcal{A}_n)\cdot e_n\cdot(\mathbb{E}[G_n]-G_n)]\,.\tag{2.1}$$

Let $\varepsilon>0$ and set

$$B=\{|G_n-\mathbb{E}[G_n]\,|\ge\varepsilon\log n\}.$$

Then we have

$$\mathbb{E}[\mathbf{1}(\mathcal{A}_n)\cdot e_n\cdot|\mathbb{E}[G_n]-G_n|]\le\varepsilon\log n\cdot\mathbb{E}[\mathbf{1}(\mathcal{A}_n)\cdot e_n]+\mathbb{E}[\mathbf{1}(\mathcal{A}_n)\cdot|\mathbb{E}[G_n]-G_n|\cdot\mathbf{1}(B)]\,. \tag{2.2}$$

Using Cauchy-Schwartz for the second term together with Lemma 2.4, we obtain

$$\mathbb{E}[\mathbf{1}(\mathcal{A}_n)\cdot|\mathbb{E}[G_n]-G_n|\cdot\mathbf{1}(B)]\le\sqrt{\mathbb{P}(B)\operatorname{Var}(G_n)}\lesssim 1.$$

Substituting this bound into (2.2) and then into (2.1), taking ε sufficiently small, using Lemma 2.4 and rearranging we deduce

$$\mathbb{E}[\mathbf{1}(\mathcal{A}_n)\cdot e_n]\lesssim\frac{1}{\log n}. \tag{2.3}$$

Claim 2.5. *We have*

$$\mathbb{P}(\mathcal{A}_n)\lesssim\frac{1}{\log n}\quad\textit{and}\quad\mathbb{P}\Big(\widetilde{\mathcal{R}}[1,\infty)\cap\mathcal{R}(0,\xi_n^r]=\emptyset\Big)\lesssim\frac{1}{\sqrt{\log n}}. \tag{2.4}$$

We now explain that it suffices to prove that

$$\mathbb{E}[\mathbf{1}(\mathcal{A}_n)\cdot|\mathbb{E}[G_n]-G_n|\cdot\mathbf{1}(B)]\lesssim\frac{1}{(\log n)^{1/4}}. \tag{2.5}$$

Indeed, once this is established, then we get

$$\left|\mathbb{E}[\mathbf{1}(\mathcal{A}_n)\cdot e_n]-\frac{1}{\mathbb{E}[G_n]}\right|\le\varepsilon\cdot\mathbb{E}[\mathbf{1}(\mathcal{A}_n)\cdot e_n]+\mathcal{O}\left(\frac{1}{(\log n)^{5/4}}\right),$$

and, since this holds for any $\varepsilon>0$ and $\mathbb{E}[G_n]\sim 8/\pi^2\log n$ by Lemma 2.4, this concludes the proof in the case where we run the two-sided walk up to two geometric times. To pass to the fixed n case, one needs to use that $\mathbb{P}\big(n/(\log n)^2\le\xi_n^r\le n(\log n)^2\big)=1-(\log n)^{-2}$ and similarly for ξ_n^ℓ. So we now turn to prove (2.5). By the Cauchy-Schwartz inequality we obtain

$$\mathbb{E}[\mathbf{1}(\mathcal{A}_n)\cdot|\mathbb{E}[G_n]-G_n|\cdot\mathbf{1}(B)]\le\sqrt{\mathbb{P}(\mathcal{A}_n\cap B)\cdot\mathbb{E}[(\mathbb{E}[G_n]-G_n)^2]}\le\sqrt{\mathbb{P}(\mathcal{A}_n\cap B)\cdot\log n},$$

where for the last inequality we used Lemma 2.4. It remains to bound the last probability appearing above. To do this we define

$$G_n^1=\sum_{k=-\xi_n^\ell}^{0}g(X_k)\quad\text{and}\quad G_n^2=\sum_{k=0}^{\xi_n^r}g(X_k),$$

and also two events for $i=1,2$

$$B_i=\{|G_n^i-\mathbb{E}\big[G_n^i\big]\,|\ge\varepsilon\log n/2\}.$$

Then it is clear that $B\subseteq B_1\cup B_2$, and hence we deduce

$$\begin{aligned}\mathbb{P}(\mathcal{A}_n\cap B)&\le\mathbb{P}\Big(\widetilde{\mathcal{R}}_\infty\cap\mathcal{R}[-\xi_n^\ell,0]=\emptyset,B_2\Big)+\mathbb{P}\Big(\widetilde{\mathcal{R}}_\infty\cap\mathcal{R}[0,\xi_n^r]=\emptyset,B_1\Big)\\&=2\mathbb{P}\Big(\widetilde{\mathcal{R}}_\infty\cap\mathcal{R}[-\xi_n^\ell,0]=\emptyset\Big)\,\mathbb{P}(B_2)\lesssim\frac{1}{\sqrt{\log n}}\cdot\frac{1}{\log n}.\end{aligned}$$

Note that for the equality we used the independence between the two sides of the walk X and for the last step we used the concentration result, Lemma 2.4, together with (2.4). Altogether this gives

$$\mathbb{E}[\mathbf{1}(\mathcal{A}_n)\cdot|\mathbb{E}[G_n]-G_n|\cdot\mathbf{1}(B)]\lesssim\frac{1}{(\log n)^{1/4}},$$

and this concludes the proof . $\square$

Proof of Claim 2.5. This proof follows closely [9]. Assuming $\mathbb{E}[\mathbf{1}(\mathcal{A}_n)\cdot e_n]\lesssim 1/\log n$, we want to show that

$$\mathbb{P}(\mathcal{A}_n)\lesssim \frac{1}{\log n}.$$

Let σ be the last time that $(X_n)_{n\geq 0}$ is at 0. Then we have

$$\mathbb{P}(\mathcal{A}_n)\leq \mathbb{P}\Big(\widetilde{\mathcal{R}}[1,\infty)\cap(\mathcal{R}[-\xi^{\ell}_{\sqrt{n}},0]\cup\mathcal{R}[\sigma,\sigma+\xi^{r}_{\sqrt{n}}])=\emptyset\Big)+\mathbb{P}\Big(\sigma+\xi^{r}_{\sqrt{n}}\geq\xi^{r}_{n}\Big),$$

where we took $\xi^{r}_{\sqrt{n}}$ to be an independent geometric random variable of parameter $1/\sqrt{n}$. The second probability appearing on the right-hand side above can be bounded as

$$\mathbb{P}\Big(\sigma+\xi^{r}_{\sqrt{n}}\geq\xi^{r}_{n}\Big)\leq\mathbb{P}\Big(\xi^{r}_{n}-\xi^{r}_{\sqrt{n}}<\sqrt{n}\Big)+\mathbb{P}(\sigma\geq\sqrt{n}).$$

Now it is easy to see that both these terms are much smaller than $1/\log n$.

To control the first probability on the right-hand side above we observe that the walk X after time σ has the same law as a walk started from 0 and conditioned on never returning to 0. Hence we get

$$\mathbb{P}\Big(\widetilde{\mathcal{R}}[1,\infty)\cap(\mathcal{R}[-\xi^{\ell}_{\sqrt{n}},0]\cup\mathcal{R}[\sigma,\sigma+\xi^{r}_{\sqrt{n}}])=\emptyset\Big)\leq\frac{1}{\mathbb{P}_0\Big(\widetilde{H}_0=\infty\Big)}\mathbb{E}\Big[\mathbf{1}(\mathcal{A}_{\sqrt{n}})\cdot e_n\Big]\lesssim\frac{1}{\log n},$$

where we used the transience of the walk. This now finishes the proof of the first claim.

We now turn to proving

$$\mathbb{P}\Big(\widetilde{\mathcal{R}}[1,\infty)\cap\mathcal{R}[0,\xi^{r}_{n}]\Big)\gtrsim\frac{1}{\sqrt{\log n}}.$$

By conditioning on $\widetilde{\mathcal{R}}$ we get

$$\begin{aligned}
&\mathbb{P}\Big(\widetilde{\mathcal{R}}[1,\infty)\cap(\mathcal{R}[-\xi^{\ell}_{n},0]\cup\mathcal{R}[0,\xi^{r}_{n}])\Big)=\mathbb{E}\Big[\mathbb{P}\Big(\widetilde{\mathcal{R}}[1,\infty)\cap(\mathcal{R}[-\xi^{\ell}_{n},0]\cup\mathcal{R}[0,\xi^{r}_{n}])\;\Big|\;\widetilde{\mathcal{R}}[1,\infty)\Big)\Big]\\
&=\mathbb{E}\Big[\mathbb{P}\Big(\mathcal{R}[-\xi^{\ell}_{n},0]\cap\widetilde{\mathcal{R}}[1,\infty)=\emptyset\;\Big|\;\widetilde{\mathcal{R}}[1,\infty)\Big)\,\mathbb{P}\Big(\mathcal{R}[0,\xi^{r}_{n}]\cap\widetilde{\mathcal{R}}[1,\infty)=\emptyset\;\Big|\;\widetilde{\mathcal{R}}[1,\infty)\Big)\Big]\\
&=\mathbb{E}\left[\Big(\mathbb{P}\Big(\mathcal{R}[-\xi^{\ell}_{n},0]\cap\widetilde{\mathcal{R}}[1,\infty)=\emptyset\;\Big|\;\widetilde{\mathcal{R}}[1,\infty)\Big)\Big)^2\right]\geq\Big(\mathbb{P}\Big(\mathcal{R}[-\xi^{\ell}_{n},0]\cap\widetilde{\mathcal{R}}[1,\infty)=\emptyset\Big)\Big)^2.
\end{aligned}$$

For the second equality we used the independence of the positive and negative parts of the walk and for the last inequality we used Jensen's inequality. Combining this with the first statement completes the proof. □

Proof of Lemma 2.4. Using the local CLT, it is a direct calculation to check that as $n\to\infty$

$$\mathbb{E}\left[\sum_{i=0}^{n}g(X_i)\right]=\frac{4}{\pi^2}\cdot\log n+O(1).$$

It is straightforward to see that replacing n by a geometric random variable of parameter $1/n$ gives exactly the same asymptotics. It remains to estimate the variance. Note that if instead of the walk we were considering a Brownian motion, then we could divide this sum between the first hitting times of balls of radii 2^i for $i=0,\ldots,\log n/2$ and we would get a sum of independent terms. With the walk one can carry through such an argument too, but there are the lattice effects that have to be taken care of. So as in Lawler's proof we simply estimate the variance using the local CLT. For this we have

$$\mathrm{Var}\left(\sum_{i=0}^{n} g(X_i)\right) = \sum_{i=0}^{n} \mathrm{Var}(g(X_i)) + \sum_{i\neq j} \mathrm{Cov}(g(X_i), g(X_j)).$$

It remains to estimate $\mathbb{E}[g(X_i)g(X_j)]$. This can be done employing the local CLT and for the details we refer the reader to [6]. $\square$

2.2 Capacity

Let $d \geq 3$. Let A be a finite subset of $\mathbb{Z}^d$. The capacity of A is defined as the sum of escape probabilities from A, i.e.

$$\mathrm{Cap}(A) = \sum_{x\in A} \mathbb{P}_x\left(\widetilde{H}_A = \infty\right).$$

We define the equilibrium measure of A to be given by

$$e_A(x) = \mathbb{P}_x\left(\widetilde{H}_A = \infty\right) \cdot \mathbf{1}(x \in A).$$

Exercise 2.6. Let $d \geq 3$ and let $A \subseteq \mathbb{Z}^d$ be a finite set. For all n we let $\mathcal{R}_n = \{X_0, \ldots, X_n\}$ be the range of a simple random walk X in $\mathbb{Z}^d$. Explain why the following limit exists

$$\lim_{n\to\infty} \frac{|\mathcal{R}_n + A|}{n}$$

and identify its value. (Note that $\mathcal{R}_n + A$ denotes the Minkowski sum of $\mathcal{R}_n$ and A.)

Corollary 2.7. *Let $A \subseteq \mathbb{Z}^d$ be a finite subset of $\mathbb{Z}^d$. Then*

$$\mathrm{Cap}(A) = \lim_{\|x\|\to\infty} \frac{\mathbb{P}_x(H_A < \infty)}{g(x)}.$$

Proof. Recall the last exit decomposition formula

$$\mathbb{P}_x(H_A < \infty) = \sum_{y\in A} g(x,y)\mathbb{P}_y\left(\widetilde{H}_A = \infty\right).$$

Dividing both sides of this equality by $g(x)$ we get

$$\frac{\mathbb{P}_x(H_A < \infty)}{g(x)} = \sum_{y\in A} \frac{g(x,y)}{g(x)}\mathbb{P}_x\left(\widetilde{H}_A = \infty\right).$$

Since A is a finite set, using Theorem 1.7 we get

$$\frac{g(x,y)}{g(x)} \to 1 \text{ as } \|x\| \to \infty.$$

Therefore, we conclude

$$\lim_{\|x\|\to\infty} \frac{\mathbb{P}_x(H_A < \infty)}{g(x)} = \sum_{y\in A} \mathbb{P}_x\left(\widetilde{H}_A = \infty\right) = \mathrm{Cap}(A)$$

and this finishes the proof. $\square$

Exercise 2.8. Let $A, B \subseteq \mathbb{Z}^d$ be finite sets. Show that

$$\mathrm{Cap}(A \cup B) \leq \mathrm{Cap}(A) + \mathrm{Cap}(B) - \mathrm{Cap}(A \cap B).$$

Exercise 2.9. Let $r > 0$. Show that

$$\mathrm{Cap}(B(0,r)) \asymp r^{d-2}.$$

Exercise 2.10. Let $x, y \in \mathbb{Z}^d$. Show that

$$\mathrm{Cap}(\{0\}) = \frac{1}{g(0)} \quad \text{and} \quad \mathrm{Cap}(\{x,y\}) = \frac{2}{g(0) + g(x-y)}.$$

Remark 2.11. From the definition of capacity we see that it is intimately related to the question of intersection of a random walk with a set. If we replace the deterministic set A by a random set, then the question of capacity reduces to the question of whether a random walk intersects that independent random set.

Exercise 2.12. Let X be a simple random walk in $\mathbb{Z}^4$ and let $\mathcal{R}_n = \{X_0, \ldots, X_n\}$ be its range up to time n. Using Theorem 2.1 show that

$$\mathbb{E}[\mathrm{Cap}(\mathcal{R}_n)] \sim \frac{\pi^2}{8} \cdot \frac{n}{\log n}.$$

Theorem 2.13. *Let $d \geq 3$ and let $A \subseteq \mathbb{Z}^d$ be a finite subset of $\mathbb{Z}^d$. Then*

$$\frac{1}{\mathrm{Cap}(A)} = \inf \left\{ \sum_{x,y \in A} g(x,y)\mu(x)\mu(y) : \mu \textit{ probability measure on } A \right\}.$$

Proof. First of all using the last exit decomposition formula gives that with

$$\mu(x) = \frac{e_A(x)}{\mathrm{Cap}(A)},$$

we get $\sum_{x,y\in A} g(x,y)\mu(x)\mu(y) = \mathrm{Cap}(A)$. So it suffices to show that for any other probability measure μ supported on A we have

$$\sum_{x,y \in A} g(x,y)\mu(x)\mu(y) \geq \frac{1}{\mathrm{Cap}(A)}. \tag{2.6}$$

To prove this we define an inner product between any two probability measures μ and ν supported on A as follows

$$\langle \mu, \nu \rangle = \sum_{x,y \in A} \mu(x) g(x,y) \nu(y).$$

Then taking $\nu = e_A/\mathrm{Cap}(A)$, the normalised equilibrium measure, and for any μ we get using again the last exit decomposition formula

$$\langle \mu, \nu \rangle = \frac{1}{\mathrm{Cap}(A)}.$$

Now by the Cauchy-Schwartz inequality we obtain

$$\frac{1}{\mathrm{Cap}(A)} = \langle \mu, \nu \rangle \leq \sqrt{\langle \mu, \mu \rangle \langle \nu, \nu \rangle} = \sqrt{\langle \mu, \mu \rangle} \cdot \frac{1}{\sqrt{\mathrm{Cap}(A)}}.$$

Rearranging proves (2.6). □

Exercise 2.14. The goal of this exercise is to show that there exists a universal constant $c > 0$ so that for any finite subset A of $\mathbb{Z}^d$ we have

$$\mathrm{Cap}(A) \geq c \cdot |A|^{1-2/d}. \tag{2.7}$$

1. Show that there exists a positive constant C so that for every $x \in A$

$$\sum_{y \in A} g(x,y) \leq C|A|^{2/d}.$$

2. Taking $\mu = 1/|A|$ in the variational characterisation of capacity and using the above bound prove (2.7).

The following lemma gives yet another equivalent definition of capacity. Its usefulness will be apparent in Lemma 2.16, where the walk is required to spend a certain amount of time at each site of a set A.

Lemma 2.15. *Let $d \geq 3$ and let A be a finite subset of $\mathbb{Z}^d$. Then the capacity of A satisfies*

$$\mathrm{Cap}(A) = \sup\left\{ \sum_{x \in A} \varphi(x) : \varphi : A \to \mathbb{R}_+ \quad \textit{and} \quad \sum_{y \in A} g(x,y)\varphi(y) \leq 1, \ \forall\, x \right\}.$$

Proof. First of all we see that taking $\varphi(x) = e_A(x)$ for $x \in A$ satisfies

$$\sum_{y \in A} g(x,y)\varphi(y) = \mathbb{P}_x(H_A < \infty) \leq 1$$

by the last-exit decomposition formula. Moreover,

$$\sum_{x \in A} \varphi(x) = \mathrm{Cap}(A).$$

Hence, it remains to show that for any function $\varphi : A \to \mathbb{R}_+$ with $\sum_{y \in A} g(x,y)\varphi(y) \leq 1$ for all x, we have that

$$\sum_{x \in A} \varphi(x) \leq \mathrm{Cap}(A).$$

Now observe that using the assumption that $\sum_{y \in A} g(x,y)\varphi(y) \leq 1$ for all x we have

$$\sum_{x \in A} e_A(x) \cdot \sum_{y \in A} g(x,y)\varphi(y) \leq \sum_{x \in A} e_A(x).$$

By the last exit decomposition formula we also obtain

$$\sum_{x \in A} e_A(x) \cdot \sum_{y \in A} g(x,y)\varphi(y) = \sum_{y \in A} \varphi(y) \sum_{x \in A} g(x,y) e_A(x) = \sum_{y \in A} \varphi(y)\mathbb{P}_y(H_A < \infty) = \sum_{y \in A} \varphi(y).$$

Combining this with the above shows that

$$\sum_{y \in A} \varphi(y) \leq \mathrm{Cap}(A)$$

and this completes the proof. $\square$

For a simple random walk X in $\mathbb{Z}^d$ with $d \geq 3$ we write $\ell(x) = \sum_{i=0}^{\infty} \mathbf{1}(X_i = x)$, for $x \in \mathbb{Z}^d$, to denote the local time at x.

Lemma 2.16. *Let A be a finite subset of $\mathbb{Z}^d$ and let $t > 0$. Then*

$$\mathbb{P}(\ell(x) \geq t,\ \forall\ x \in A) \leq 2\exp(-t \cdot \mathrm{Cap}(A)/2).$$

Remark 2.17. We write $f * g$ to denote the convolution of f and g, i.e.

$$f * g(x) = \sum_y f(x-y)g(y).$$

Lemma 2.18. *Let φ be a function satisfying $\|g * \varphi\|_\infty \leq 1$. Then for all $x_0 \in \mathbb{Z}^d$ and all $\theta \in (0,1)$ we have*

$$\mathbb{E}_{x_0}\left[\exp\left(\theta \cdot \sum_x \varphi(x)\ell(x)\right)\right] \leq \frac{1}{1-\theta}.$$

Proof. First of all we notice that we can write this quantity as

$$\sum_x \varphi(x)\ell(x) = \sum_x \varphi(x) \sum_{k=0}^{\infty} \mathbf{1}(X_k = x) = \sum_{k=0}^{\infty} \varphi(X_k).$$

We now upper bounding the n-th moment of $\sum_x \varphi(x)\ell(x)$. For this we have

$$\begin{aligned}
&\mathbb{E}_{x_0}\left[\left(\sum_{k=0}^{\infty}\varphi(X_k)\right)^n\right] = \mathbb{E}_{x_0}\left[\sum_{k_1,\dots,k_n}\varphi(X_{k_1})\cdots\varphi(X_{k_n})\right] \leq n! \sum_{k_1\leq\ldots\leq k_n} \mathbb{E}_{x_0}[\varphi(X_{k_1})\cdots\varphi(X_{k_n})] \\
&= n! \sum_{k_1\leq\ldots\leq k_n}\sum_{x_1,\dots,x_n} \mathbb{P}_{x_0}(X_{k_1} = x_1, \dots, X_{k_n} = x_n)\prod_{i=1}^n \varphi(x_i) \\
&\leq n! \sum_{k_1\leq\ldots\leq k_n}\sum_{x_1,\dots,x_n} P^{k_1}(x_0,x_1)\cdot P^{k_2-k_1}(x_1,x_2)\cdots P^{k_n-k_{n-1}}(x_{n-1},x_n)\prod_{i=1}^n \varphi(x_i) \\
&= n! \sum_{x_1,\dots,x_n} g(x_0,x_1)\cdots g(x_{n-1},x_n)\prod_{i=1}^n \varphi(x_i) \leq n!,
\end{aligned}$$

where in the last step we used the assumption on the function φ. So we now deduce

$$\mathbb{E}_{x_0}\left[\exp\left(\theta\sum_x \varphi(x)\ell(x)\right)\right] = \sum_{n=0}^{\infty}\frac{\mathbb{E}_{x_0}[(\sum_x \varphi(x)\ell(x))^n]}{n!}\cdot\theta^n \leq \frac{1}{1-\theta}$$

and this concludes the proof. □

Proof of Lemma 2.16. Let $\varphi : A \to \mathbb{R}_+$ be a function satisfying $\|g * \varphi\|_\infty \leq 1$. It follows that

$$\{\ell(x) \geq t,\ \forall\ x \in A\} \subseteq \left\{\sum_{x\in A}\ell(x)\varphi(x) \geq t \cdot \sum_{x \in A}\varphi(x)\right\}.$$

By the exponential Chebyshev inequality we now deduce for any $\theta \in (0,1)$

$$\mathbb{P}(\ell(x) \geq t,\ \forall\ x \in A) \leq \mathbb{P}\left(\sum_{x\in A}\varphi(x)\ell(x) \geq t\sum_{x\in A}\varphi(x)\right)$$

$$\leq \exp\left(-\theta \cdot t \cdot \sum_{x\in A} \varphi(x)\right) \cdot \mathbb{E}\left[\exp\left(\theta \cdot \sum_{x} \varphi(x)\ell(x)\right)\right] \leq \frac{1}{1-\theta} \cdot \exp\left(-\theta \cdot t \cdot \sum_{x\in A} \varphi(x)\right).$$

Taking now $\theta = 1/2$, optimising over all functions $\varphi : A \to \mathbb{R}_+$ with $\|g * \varphi\|_\infty \leq 1$ and using Lemma 2.15 shows that

$$\mathbb{P}(\ell(x) \geq t,\ \forall\, x \in A) \leq 2 \cdot \exp(-t \cdot \mathrm{Cap}(A)/2)$$

and this finishes the proof. □

For a set $A \subseteq \mathbb{Z}^d$ we write $\ell(A)$ for the total time spent in A by a simple random walk X, i.e.

$$\ell(A) = \sum_{x\in A} \ell(x).$$

We also write for $x \in \mathbb{Z}^d$

$$g(x, A) := \sum_{y\in A} g(x, y).$$

Lemma 2.19. *Let $d \geq 3$ and X a simple random walk on $\mathbb{Z}^d$. There exist positive constants c and C so that if A is a finite subset of $\mathbb{Z}^d$, then we have*

$$\mathbb{P}(\ell(A) \geq t) \leq C \exp\left(-ct/ \sup_{x\in\mathbb{Z}^d} g(x, A)\right).$$

Proof. Let $\varphi(x) = 1/\sup_{x\in\mathbb{Z}^d} g(x, A)$ for all $x \in A$. Then we have

$$g * \varphi(x) = \sum_{y\in A} g(x, y)\varphi(y) \leq 1.$$

Thus we can apply Lemma 2.18 to obtain for $\theta \in (0, 1)$ that

$$\mathbb{E}\left[\exp\left(\theta \cdot \sum_{x} \varphi(x)\ell(x)\right)\right] \leq \frac{1}{1-\theta}.$$

It is immediate to see that

$$\{\ell(A) \geq t\} \subseteq \left\{\sum_{x\in A} \varphi(x)\ell(x) \geq t/ \sup_{x\in\mathbb{Z}^d} g(x, A)\right\}.$$

Applying exponential Chernoff again we deduce

$$\mathbb{P}\left(\sum_{x\in A} \varphi(x)\ell(x) \geq t/ \sup_{x\in\mathbb{Z}^d} g(x, A)\right) \lesssim \exp\left(-ct/ \sup_{x\in\mathbb{Z}^d} g(x, A)\right)$$

and this completes the proof. □

Remark 2.20. Recall from Exercise 2.14 that there exists a universal constant C so that for all sets A

$$\sup_{x\in\mathbb{Z}^d} g(x, A) \leq C|A|^{2/d}.$$

Plugging this bound into the bound in Lemma 2.19 shows that

$$\mathbb{P}(\ell(A) \geq t) \lesssim \exp(-ct/|A|^{2/d}).$$

2.3 Intersections in higher dimensions

The question on large deviations on intersections of two independent random walks in dimensions $d \geq 5$ was first studied in 1994 by Khanin, Mazel, Shloshman and Sinai [5]. They proved that for all $\varepsilon > 0$ and all t sufficiently large

$$\exp(-t^{1-2/d+\varepsilon}) \leq \mathbb{P}\Big(|\mathcal{R}_\infty \cap \widetilde{\mathcal{R}}_\infty| > t\Big) \leq \exp(-t^{1-2/d-\varepsilon}). \tag{2.8}$$

In 2004, van den Berg, Bolthausen and den Hollander [10] showed that there exists a non-negative rate function $\mathcal{I}$ such that for all $b > 0$

$$\lim_{t\to\infty} \frac{1}{t^{1-2/d}} \log \mathbb{P}\Big(|\mathcal{R}_{\lfloor bt \rfloor} \cap \widetilde{\mathcal{R}}_{\lfloor bt \rfloor}| > t\Big) = -\mathcal{I}(b).$$

(In fact, they established it for Wiener sausages, and it was later adapted to the discrete setup by Phetdrapat [8] in his PhD thesis.)

In 2020, Asselah and Schapira [2] finally managed to settle this open question by proving a large deviations principle for the infinite time horizon.

Theorem 2.21 (Asselah and Schapira [2]). *For $d \geq 5$, the following limit exists and is positive*

$$\mathcal{I}_\infty = \lim_{b\to\infty} \mathcal{I}(b) = \lim_{t\to\infty} -\frac{1}{t^{1-2/d}} \log \mathbb{P}\Big(|\mathcal{R}_\infty \cap \widetilde{\mathcal{R}}_\infty| > t\Big).$$

In these notes we are going to establish the following result of Asselah and Schapira, which removes the power ε from (2.8).

Theorem 2.22 (Asselah and Schapira [2]). *Let $d \geq 5$ and let $\mathcal{R}$ and $\widetilde{\mathcal{R}}$ be two independent ranges. There exist positive constants c_1 and c_2 so that for all $t > 0$*

$$e^{-c_2 t^{1-2/d}} \leq \mathbb{P}\Big(\mathcal{R}_\infty \cap \widetilde{\mathcal{R}}_\infty| > t\Big) \leq e^{-c_1 t^{1-2/d}}.$$

Moreover, Asselah and Schapira are able to identify the strategy for the two walks in order to achieve a large intersection. In particular, they show that given that the size of the intersection is larger than t, a fraction close to t of them happen in a ball of radius $t^{1/d}$.

They first prove a weaker result, namely that there exists a finite number of balls of radius $t^{1/d}$, where most of the intersections happen. To reduce to a single box, they needed to appeal to the large deviations result for the finite time horizon problem.

2.3.1 Lower bound

We start by proving the lower bound of Theorem 2.22. This is the easier direction of this problem as it entails finding a specific strategy for both walks to follow in order to achieve the required event.

The main ingredient of the proof is the following result which gives a lower bound on the probability that a walk visits a fraction of a set.

Proposition 2.23. *Let X be a simple random walk in $\mathbb{Z}^d$ with $d \geq 3$ and let $\mathcal{R}_\infty = X[0,\infty)$ denote its range. There exist positive constants ρ, κ and C so that for all $r > 0$ if $\Lambda \subseteq B(0,r)$ satisfies $|\Lambda| > C$, then*

$$\mathbb{P}(|\mathcal{R}_\infty \cap \Lambda| \geq \rho|\Lambda|) \geq \exp\left(-\kappa \cdot r^{d-2}\right).$$

We start by giving the proof the lower bound and then we proceed with the proof.

Proof of lower bound of Theorem 2.22. Let $\rho < 1$ and $r > 0$ be such that $\rho^2 |B(0,r)| = t$. We then have

$$\{|\mathcal{R}_\infty \cap \widetilde{\mathcal{R}}_\infty| \geq t\} \supseteq \{|\mathcal{R}_\infty \cap B(0,r)| \geq \rho |B(0,r)|\} \cap \{|\widetilde{\mathcal{R}}_\infty \cap (\mathcal{R}_\infty \cap B(0,r))| \geq \rho \cdot |\mathcal{R}_\infty \cap B(0,r)|\}.$$

Using the independence between the two walks and applying Proposition 2.23 we obtain

$$\mathbb{P}\Big(|\mathcal{R}_\infty \cap \widetilde{\mathcal{R}}_\infty| \geq t\Big) \gtrsim \exp(-\kappa \cdot r^{d-2}) = \exp(-\kappa' \cdot t^{1-2/d}),$$

where κ and κ' are positive constants. This completes the proof. $\square$

Let us first give a high level overview of the proof of Proposition 2.23. We consider the balls $B(0,5r)$ and $B(0,10r)$. We are going to count the number of excursions the walk makes across the annulus $B(0,10r) \setminus B(0,5r)$. During each excursion, the walk has a probability $1/r^{d-2}$ of hitting a given vertex of the ball $B(0,r)$. The excursions are approximately independent, as there is enough time for the walk to mix before starting the next one. So during $K_r = K \cdot \rho \cdot r^{d-2}$ excursions, a fraction ρ of the vertices of Λ will be covered. The probability that starting from $\partial B(0,10r)$ the random walk hits $\partial B(0,5r)$ is a positive constant bounded away from 1 and 0, and hence the probability of having at least K_r excursions is of order $\exp(-cK_r)$ which is of the correct order. We now need to make this argument rigorous.

Proof of Proposition 2.23. To this end we first define the successive hitting times of $\partial B(0,5r)$ and $\partial B(0,10r)$. Set $\sigma_0 = 0$ and define recursively for $i \geq 0$

$$\begin{aligned} \tau_i &= \inf\{t \geq \sigma_i : X_t \notin B(0,10r)\} \quad \text{and} \\ \sigma_{i+1} &= \inf\{t \geq \tau_i : X_t \in \partial B(0,5r)\}. \end{aligned}$$

We let $\mathcal{N}$ be the total number of excursions the walk performs, i.e.

$$\mathcal{N} = \sup\{k \geq 0 : \sigma_k < \infty\}.$$

Using the Green's function asymptotics we get that there exists a positive constant c such that

$$\mathbb{P}(\mathcal{N} \geq k) \geq \exp(-c \cdot k). \tag{2.9}$$

Set $K_r = K \cdot \rho \cdot r^{d-2}$. Let $\mathcal{G}$ be the σ-algebra generated by the total number of excursions $\mathcal{N}$ and the entrance and exit time of these excursions, i.e.

$$\mathcal{G} = \sigma(\mathcal{N}, X_{\sigma_i}, X_{\tau_i}, i \leq \mathcal{N}).$$

We now define $\Lambda_1 = \Lambda$ and inductively for $i \geq 1$

$$\mathcal{R}^{(i)} = \{X_{\sigma_i}, \ldots, X_{\tau_i}\} \quad \text{and} \quad \Lambda_{i+1} = \Lambda \setminus (\cup_{j \leq i} \mathcal{R}^{(j)}).$$

Finally set

$$Y_i = |\mathcal{R}^{(i)} \cap \Lambda_i| \mathbf{1}(\sigma_i < \infty).$$

By conditioning on $\mathcal{G}$ we deduce

$$\mathbb{P}\left(\sum_{i=1}^{K_r} Y_i > \rho|\Lambda|, \ \mathcal{N} \geq K_r\right) = \mathbb{E}\left[\mathbf{1}(\mathcal{N} \geq K_r) \cdot \mathbb{P}\left(\sum_{i=1}^{K_r} Y_i > \rho|\Lambda| \;\middle|\; \mathcal{G}\right)\right]. \tag{2.10}$$

Let $\mathcal{H}_i = \sigma(X[0,\sigma_i])$ for every i. We now define

$$M = \sum_{i=1}^{\mathcal{N}\wedge K_r} (Y_i - \mathbb{E}[Y_i \mid \mathcal{H}_i, \mathcal{G}]).$$

Note that by the orthogonality of increments we get

$$\mathbb{E}[M \mid \mathcal{G}] = 0 \quad \text{and} \quad \mathbb{E}\left[M^2 \;\middle|\; \mathcal{G}\right] \leq 2 \sum_{i=1}^{\mathcal{N}\wedge K_r} \mathbb{E}\left[Y_i^2 \;\middle|\; \mathcal{G}\right].$$

Exercise 2.24. Using Harnack's inequality show that for all $i \leq \mathcal{N}$ we have for all $x \in B(0,r)$

$$\mathbb{P}\left(x \in \mathcal{R}^{(i)} \;\middle|\; X_{\sigma_i}, X_{\tau_i}\right) \gtrsim \mathbb{P}\left(x \in \mathcal{R}^{(i)} \;\middle|\; X_{\sigma_i}\right).$$

We also have using the asymptotics of the Green's function

$$\begin{aligned}
\mathbb{P}\left(x \in \mathcal{R}^{(i)} \;\middle|\; X_{\sigma_i} = y\right) &= \mathbb{P}_y(H_x < \infty) - \sup_{z \in \partial B(0,10r)} \mathbb{P}_z(H_x < \infty) \\
&= \frac{g(x,y)}{g(0)} - \sup_{z \in \partial B(0,10r)} \frac{g(x,z)}{g(0)} \\
&= \frac{c_d}{\|x-y\|^{d-2}} - \sup_{z \in \partial B(0,10r)} \frac{c_d}{\|x-z\|^{d-2}} + O(r^{1-d}) \asymp \frac{1}{r^{d-2}}.
\end{aligned}$$

Therefore, putting everything together we deduce that for a positive constant c

$$\mathbb{E}[Y_i \mid \mathcal{G}, \mathcal{H}_i] \geq \frac{c}{r^{d-2}} \cdot |\Lambda_i|,$$

and hence on the event $\{\mathcal{N} \geq K_r\}$ and taking $K = 4/c$ in the definition of K_r this gives

$$\sum_{i=1}^{\mathcal{N}\wedge K_r} \mathbb{E}[Y_i \mid \mathcal{G}, \mathcal{H}_i] \geq K_r \cdot \frac{c}{r^{d-2}} \cdot |\Lambda_{K_r}| = 4\rho|\Lambda_{K_r}|.$$

Since $|\Lambda_{K_r}| = |\Lambda| - \sum_{i=1}^{K_r-1} Y_i$ we get on the event $\{\mathcal{N} \geq K_r\}$ for $\rho \leq 1/2$

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^{K_r} Y_i \leq \rho|\Lambda| \;\middle|\; \mathcal{G}\right) &= \mathbb{P}\left(\sum_{i=1}^{K_r} Y_i \leq \rho|\Lambda|, |\Lambda_{K_r}| \geq |\Lambda|/2 \;\middle|\; \mathcal{G}\right) \leq \mathbb{P}(|M| \geq \rho|\Lambda| \mid \mathcal{G}) \\
&\leq \frac{\mathbb{E}\left[M^2 \;\middle|\; \mathcal{G}\right]}{\rho^2|\Lambda|^2} \leq \frac{2}{\rho^2|\Lambda|^2} \sum_{i=1}^{K_r} \mathbb{E}\left[Y_i^2 \mid \mathcal{G}\right].
\end{aligned} \tag{2.11}$$

Now it remains to bound this last sum of conditional expectations. For this we obtain

$$\begin{aligned}
\mathbb{E}\left[Y_i^2 \;\middle|\; \mathcal{H}_i, \mathcal{G}\right] &= \sum_{(z,z') \in \Lambda_i \times \Lambda_i} \mathbb{P}\left(z \in \mathcal{R}^{(i)}, z' \in \mathcal{R}^{(i)} \;\middle|\; X_{\sigma_i}, X_{\tau_i}\right) \\
&\leq 2 \sum_{(z,z') \in \Lambda_i \times \Lambda_i} \mathbb{P}\left(z \in \mathcal{R}^{(i)}, z' \in \mathcal{R}^{(i)}, H_z < H_{z'} \;\middle|\; X_{\sigma_i}, X_{\tau_i}\right).
\end{aligned}$$

Applying the Harnack inequality again we get that up to a positive constant this last sum is equal to

$$\sum_{(z,z') \in \Lambda_i \times \Lambda_i} \mathbb{P}_{X_{\sigma_i}}(H_z < H_{z'} < \infty) \leq \sum_{(z,z') \in \Lambda \times \Lambda} \frac{1}{r^{d-2}} \cdot \frac{1}{\|z-z'\|^{d-2}+1} \lesssim \frac{1}{r^{d-2}} \cdot |\Lambda|^{1+2/d}.$$

Plugging this bound into (2.11) we see that on the event $\{\mathcal{N} \geq K_r\}$ we have

$$\mathbb{P}\left(\sum_{i=1}^{K_r} Y_i \leq \rho|\Lambda| \,\middle|\, \mathcal{G}\right) \leq \frac{2}{\rho^2|\Lambda|^2} \cdot K_r \cdot \frac{1}{r^{d-2}} \cdot |\Lambda|^{1+2/d} \leq \frac{2}{c\rho|\Lambda|^{1-2/d}} \leq \frac{1}{2},$$

by taking $|\Lambda| > C$ with C a large constant so that $c \cdot \rho \cdot C^{1-2/d} \geq 4$. Plugging this bound back into (2.10) and using also (2.9) with $k = K_r$ completes the proof. □

2.3.2 Upper bound

We devote this section to the proof of the upper bound of Theorem 2.22. Here we are working in dimensions $d \geq 5$.

We first start by reducing the problem to a finite time horizon, as we show that it is very unlikely for intersections to occur at high enough times. More precisely, for every $n \geq 0$ we define

$$A_n = \sum_{i=n}^{\infty} \sum_{j=0}^{\infty} \mathbf{1}(X_i = \widetilde{X}_j).$$

Using the local CLT we then obtain

$$\mathbb{E}[A_n] = \sum_{i=n}^{\infty} \sum_{j=0}^{\infty} \mathbb{P}\Big(X_i = \widetilde{X}_j\Big) = \sum_{k=n}^{\infty} (k+1) p_k(0,0) \asymp n^{(4-d)/2}.$$

By Markov's inequality we get

$$\mathbb{P}\Big(\widetilde{\mathcal{R}}_\infty \cap \mathcal{R}[n,\infty) \neq \emptyset\Big) \leq \mathbb{E}[A_n] \asymp n^{(4-d)/2},$$

and hence taking $n = \exp(t^{1-2/d})$ gives the desired upper bound. So we can focus now on intersections between $\widetilde{\mathcal{R}}_\infty$ and $\mathcal{R}[0,n]$ for this specific value of n.

The following proposition is the main ingredient in the proof of the upper bound.

Proposition 2.25. *There exist positive constants c and C so that if $n = \exp(t^{1-2/d})$, then*

$$\mathbb{P}\left(\sup_{x \in \mathbb{Z}^d} g(x, \mathcal{R}_n) > Ct^{2/d}\right) \leq C \exp(-ct^{1-2/d}).$$

Proof of upper bound of Theorem 2.22. As we explained above it suffices to study the number of intersections between $\widetilde{\mathcal{R}}_\infty$ and $\mathcal{R}_n$. By Proposition 2.25 we obtain

$$\mathbb{P}\Big(|\widetilde{\mathcal{R}}_\infty \cap \mathcal{R}_n| > t\Big) \leq \mathbb{P}\left(|\widetilde{\mathcal{R}}_\infty \cap \mathcal{R}_n| > t, \sup_{x \in \mathbb{Z}^d} g(x, \mathcal{R}_n) \leq Ct^{2/d}\right) + C \exp(-ct^{1-2/d}).$$

Applying Lemma 2.19 to the first probability appearing on the right-hand side above we get

$$\begin{aligned}
&\mathbb{P}\left(|\widetilde{\mathcal{R}}_\infty \cap \mathcal{R}_n| > t, \sup_{x \in \mathbb{Z}^d} g(x, \mathcal{R}_n) \leq Ct^{2/d}\right) \\
&\leq \mathbb{E}\left[\exp(-ct/\sup_{x \in \mathbb{Z}^d} g(x, \mathcal{R}_n)) \mathbf{1}(\sup_{x \in \mathbb{Z}^d} g(x, \mathcal{R}_n) \leq Ct^{2/d})\right] \leq \exp(-ct^{1-2/d})
\end{aligned}$$

and this concludes the proof. □

The main idea behind the proof of the upper bound of Theorem 2.22, and more precisely the proof of Proposition 2.25, is that intersections will happen in the high density regions of the range of each walk, i.e. the regions that are visited a lot by the walk. We are going to perform a multiscale analysis of the high density region and then we will bound the Green's function separately for the high and the low density regions. In particular, we will show that the Green's function for the low density regions can be bounded deterministically, while for the high density we will first prove that with high probability the sizes of these regions are not too large, and then on that we will be able to bound the Green's function, thus proving Proposition 2.25.

For $r > 0$ and $\rho \in (0,1)$ we now let

$$\mathcal{R}_n(\rho, r) = \{x \in \mathcal{R}_n : |\mathcal{R}_n \cap B(x,r)| > \rho |B(x,r)|\},$$

i.e. the set $\mathcal{R}_n(\rho, r)$ contains the points on the range for which a fraction ρ of the ball around them is covered by the range.

The following proposition controls the size of the set of high density regions.

Proposition 2.26. *There exist positive constants C, C_0 and κ such that the following holds. For any $r \geq 1$, $n \in \mathbb{N}$ and any $\rho > 0$ satisfying*

$$\rho \cdot r^{d-2} \geq C_0 \cdot \log n, \tag{2.12}$$

then for any $L \geq 1$ we have

$$\mathbb{P}(|\mathcal{R}_n(\rho, r)| > L) \leq C \exp\left(-\kappa \cdot \rho^{2/d} \cdot L^{1-2/d}\right).$$

Claim 2.27. *There exists a positive constant C so that the following holds. Let A be any finite set and $r \geq 1$ such that*

$$|A \cap B(x,r)| \leq \rho |B(x,r)|, \ \forall \ x \in A.$$

Then for all $R \geq r$ we have for all $x \in \mathbb{Z}^d$

$$|A \cap B(x,R)| \leq C \cdot \rho \cdot |B(x,R)|$$

Proof. To see this, we start by choosing $x_1 \in A \cap B(x,R)$ and then inductively for any $k \geq 0$ choose $x_{k+1} \in A \cap B(x,R) \setminus (\cup_{j \leq k} B(x_j, r))$ until this set becomes empty. Let n be the total number of x_i's picked this way. Then the balls $B(x_i, r/2)$ for $i \leq n$ are disjoint, and hence

$$R^d \asymp |B(x,R)| \geq \sum_{i=1}^{n} |B(x_i, r/2)| = n|B(0,r/2)| \asymp n \cdot r^d,$$

and hence this gives that $n \asymp R^d / r^d$. Therefore, we obtain

$$|A \cap B(x,R)| \leq \sum_{i=1}^{n} |A \cap B(x_i, r)| \leq n \cdot \rho \cdot |B(0,r)| \leq C \cdot \rho \cdot |B(0,R)|,$$

thus establishing the claim. $\square$

Lemma 2.28. *There exists a positive constant C so that the following holds. Let A be any finite set and $r \geq 1$ such that*

$$|A \cap B(x,r)| \leq \rho |B(x,r)|, \ \forall \ x \in A.$$

Then for any $x \in \mathbb{Z}^d$ we have

$$g(x, A \cap B(x,r)^c) \leq C \cdot \rho^{1-2/d} \cdot |A|^{2/d}.$$

Proof. Let $\mathcal{S}_k = B(x, r(k+1)) \setminus B(x, rk)$ for $k \geq 1$. Then we have using integration by parts and Claim 2.27

$$g(x, A \cap B(x,r)^c) = \sum_{k \geq 1} g(x, \mathcal{S}_k \cap A) \leq \sum_{k \geq 1} \frac{|A \cap \mathcal{S}_k|}{(kr)^{d-2}} = \sum_{k \geq 1} \frac{|A \cap B(x, r(k+1))| - |A \cap B(x, rk)|}{(kr)^{d-2}}$$
$$\asymp \frac{1}{r^{d-2}} \cdot \sum_{k \geq 1} \frac{|A \cap B(x, rk)|}{k^{d-1}} \leq \frac{1}{r^{d-2}} \cdot \sum_{k \geq 1} \frac{\min(\rho (rk)^d, |A|)}{k^{d-1}} \asymp \rho^{1-2/d} \cdot |A|^{2/d}$$

and this completes the proof. $\square$

We now give the proof of Proposition 2.25 and then proceed with the proof of Proposition 2.26.

Proof of Proposition 2.25. We define a sequence of densities $\rho_i = 2^{-i}$ and radii r_i for every $i \geq 0$ by setting

$$\rho_i \cdot r_i^{d-2} = C_0 \log n, \tag{2.13}$$

where C_0 is the constant of Proposition 2.26. We now define the sets Λ_i as the regions where the density in the balls of radius r_i is at least ρ_i for the first time at level i. First recall for all $i \geq 0$

$$\mathcal{R}_n(\rho_i, r_i) = \{x \in \mathcal{R}_n : |\mathcal{R}_n \cap B(x, r_i)| > \rho_i \cdot |B(x, r_i)|\}$$

and note that $\mathcal{R}_n(\rho_0, r_0) = \emptyset$. We now set

$$\Lambda_i = \mathcal{R}_n(\rho_i, r_i) \setminus (\cup_{j \leq i-1} \mathcal{R}_n(\rho_j, r_j)) \quad \text{and} \quad \Lambda_i^* = \mathcal{R}_n \setminus (\cup_{0 \leq j \leq i-1} \mathcal{R}_n(\rho_j, r_j)).$$

Let $\mathcal{S}_k = B(x, r_k) \setminus B(x, r_{k-1})$ for every $k \geq 1$. We decompose $g(x, \mathcal{R}_n)$ as follows

$$g(x, \mathcal{R}_n) = g(x, B(x, r_0) \cap \mathcal{R}_n) + \sum_{k=1}^{\infty} g(x, \mathcal{R}_n \cap \mathcal{S}_k).$$

For the first term on the right-hand side above we have

$$g(x, B(x, r_0) \cap \mathcal{R}_n) \leq g(x, B(x, r_0)) \lesssim r_0^2 \lesssim (\log n)^{2/(d-2)} \lesssim t^{2/d}.$$

Now for every $k \geq 1$ we have

$$g(x, \mathcal{R}_n \cap \mathcal{S}_k) = \sum_{i=1}^{k} g(x, \mathcal{S}_k \cap \Lambda_i) + g(x, \mathcal{S}_k \cap \Lambda_{k+1}^*).$$

We control the Green's function of the low density region as follows

$$g(x, \mathcal{S}_k \cap \Lambda_{k+1}^*) \lesssim \frac{|\mathcal{S}_k \cap \Lambda_{k+1}^*|}{r_{k-1}^{d-2}} \lesssim \frac{\rho_k r_k^d}{r_{k-1}^{d-2}} \lesssim \frac{\log n}{r_k^{d-4}},$$

where for the second inequality we used Claim 2.27 and for the final one we used (2.13). Thus taking the sum over all $k \geq 1$ we get

$$\sum_{k \geq 1} g(x, \mathcal{S}_k \cap \Lambda_{k+1}^*) \lesssim \frac{\log n}{r_0^{d-4}} \asymp \frac{\log n}{(\log n)^{(d-4)/(d-2)}} = (\log n)^{2/(d-2)} \lesssim t^{2/d}.$$

Therefore, it remains to treat the high density region. First of all we see that since $\mathcal{R}_n \leq n+1$ we get that $\Lambda_i = \emptyset$ for all i such that $\rho_i r_i^d > n+1$, which using (2.13) gives that $i \gtrsim \log n$.

We have

$$\sum_{k=1}^{\infty}\sum_{i=1}^{k} g(x, \mathcal{S}_k \cap \Lambda_i) = \sum_{i=1}^{\infty}\sum_{k\geq i} g(x, \mathcal{S}_k \cap \Lambda_i) = \sum_{i=1}^{\infty} g(x, \Lambda_i \cap B(x, r_{i-1})^c).$$

We define the good event to be

$$\mathcal{E} = \{|\Lambda_i| \leq \rho_i^{-2/(d-2)} \cdot t, \ \forall \ i \geq 1\}.$$

Applying Proposition 2.26 together with the fact that for $i \gtrsim \log n$ we have $|\Lambda_i| = \emptyset$, it follows that

$$\mathbb{P}(\mathcal{E}^c) \leq \sum_{i=1}^{C\log n} \exp\left(-\kappa \cdot \rho_i^{2/d} \cdot (\rho_i^{-2/(d-2)} t)^{1-2/d}\right) \lesssim \exp\left(-\kappa \cdot t^{1-2/d}\right). \tag{2.14}$$

By definition, the set Λ_i contains all the points of the range that are not of density ρ_j for all $j < i$. Therefore, we see that on the event $\mathcal{E}$ using also Lemma 2.28 we have for all $i \geq 1$

$$g(x, \Lambda_i \cap B(x, r_{i-1})^c) \lesssim \rho_{i-1}^{1-2/d} \cdot |\Lambda_i|^{2/d} \leq C\rho_{i-1}^{1-2/d} \cdot \rho_i^{-4/(d(d-2))} \cdot t^{2/d} = \rho_i^{(d-4)/(d-2)} \cdot t^{2/d}.$$

Taking the sum over all i completes the proof. □

For a set A we write $B(A, r) = \cup_{x\in A} B(x, r)$.

Lemma 2.29. *There exists a positive constant c so that the following holds. Let $\mathcal{C}$ be a set of points in $\mathbb{Z}^d$ at distance at least $2r$ from each other. Then for all $t > 0$ we have*

$$\mathbb{P}(\ell(B(x, r)) \geq t, \ \forall \ x \in \mathcal{C}) \leq \exp\left(-c \cdot t \cdot \mathrm{Cap}(\cup_{x\in\mathcal{C}} B(x, r))/r^d\right).$$

Proof. Let φ be the equilibrium measure of $\cup_{x\in\mathcal{C}} B(x, r)$. Define $\widetilde{\varphi}$ as follows

$$\widetilde{\varphi}(y) = \frac{c_1}{r^d} \sum_{z\in B(x,r)} \varphi(z), \ \forall\, y \in B(x, r),$$

where c_1 is a positive constant to be determined in order to make $g * \widetilde{\varphi} \leq 1$. Let $x_0 \in \mathbb{Z}^d$. We set $A(x_0) = \{x \in \mathcal{C} : \|x - x_0\| \geq 2r\}$. We then have

$$\sum_{x\in\mathcal{C}}\sum_{y\in B(x,r)} g(x_0, y)\widetilde{\varphi}(y) = \frac{c_1}{r^d} \cdot \sum_{x\in\mathcal{C}}\sum_{y\in B(x,r)} g(x_0, y) \sum_{z\in B(x,r)} \varphi(z).$$

We split the sum over $x \in A(x_0)$ and the complement. For $x \notin A(x_0)$, we then get that $g(x_0, y) \lesssim g(x_0, z)$ for any other $z \in \partial B(x, r)$. So we obtain

$$\sum_{x\in A(x_0)}\sum_{y\in B(x,r)} g(x_0, y) \sum_{z\in B(x,r)} \varphi(z) \lesssim r^d \cdot \sum_{x\in A(x_0)}\sum_{z\in B(x,r)} g(x_0, z)\varphi(z) \lesssim r^d$$

by the last exit decomposition formula (recall $g * \varphi \leq 1$). For the sum over $A(x_0)^c$ we get

$$\sum_{x\in A(x_0)^c}\sum_{y\in B(x,r)} g(x_0, y) \sum_{z\in B(x,r)} \varphi(z) \leq \mathrm{Cap}(B(0, r)) \cdot \sum_{z\in B(x_0,3r)} g(x_0, z) \lesssim r^{d-2} \cdot r^2 = r^d.$$

So there exists $c_1 > 0$ so that $g * \widetilde{\varphi} \le 1$. Notice that

$$\sum_{x\in\mathcal{C}} \widetilde{\varphi}(x) = \sum_{x\in\mathcal{C}} \sum_{y\in B(x,r)} \frac{c_1}{r^d} \cdot \varphi(y) = \frac{c_1}{r^d} \cdot \sum_{z\in\cup_{x\in\mathcal{C}} B(x,r)} \varphi(z) = \frac{c_1}{r^d} \cdot \mathrm{Cap}(\cup_{x\in\mathcal{C}} B(x,r)).$$

Applying Lemma 2.18 and exponential Chernoff we finally deduce

$$\mathbb{P}(\ell(B(x,r)) \ge t,\ \forall\ x\in\mathcal{C}) \le \mathbb{P}\left(\sum_{x\in\mathcal{C}} \widetilde{\varphi}(x)\ell(B(x,r)) \ge t\sum_{x\in\mathcal{C}} \widetilde{\varphi}(x)\right) \lesssim \exp(-c\cdot t\cdot \mathrm{Cap}(\cup_{x\in\mathcal{C}} B(x,r))/r^d)$$

and this concludes the proof. □

A final step towards proving Proposition 2.26 is the following bound on the sizes of regions with controlled density from above and below.

For $r > 0$ and $\rho \in (0,1)$ we define

$$\mathcal{R}_n^*(\rho,r) = \{x\in\mathcal{R}_n : \rho|B(x,r)| < |\mathcal{R}_n \cap B(x,r)| \le 2\rho|B(x,r)|\}.$$

These are the points of the range that the balls of radius r around them are visited a lot by the walk. The important step in the proof of the theorem is the following lemma on large deviations of the size of this set.

Lemma 2.30. *There exist positive constants C, C_0 and κ so that for all $r \ge 1$, $n\in\mathbb{N}$ and $\rho > 0$ satisfying*

$$\rho\cdot r^{d-2} \ge C_0\cdot\log n,$$

we have for all $L \ge 1$

$$\mathbb{P}(|\mathcal{R}_n^*(\rho,r)| > L) \le C\exp\left(-\kappa\rho^{2/d}\cdot L^{1-2/d}\right).$$

Proof. Let N be the number of points in $\mathcal{R}_n^*(\rho,r)$ that are at distance at least $2r$ from each other. We start by showing that on the event $\{|\mathcal{R}_n^*(\rho,r)| > L\}$, we must have $N \ge \lfloor L/(2C\rho|B(0,2r)|)\rfloor =: n_0$, where C is a positive constant to be determined. Indeed, first pick $x_1 \in \mathcal{R}_n^*(\rho,r)$. Once we have picked $x_1,\dots,x_n$ we pick x_{n+1} from the set $\mathcal{R}_n^*(\rho,r)\setminus(\cup_{j\le n} B(x_j,2r))$. Then using Claim 2.27 we get

$$|\mathcal{R}_n^*(\rho,r)\cap(\cup_{i=1}^N B(x_i,2r))| \le \sum_{i=1}^N |\mathcal{R}_n\cap B(x_i,2r)| \le C\cdot 2\rho\cdot|B(0,2r)|\cdot N,$$

where C is a positive constant. So we see that taking N as above this upper bound is smaller than $L/2$. This shows that

$$\{|\mathcal{R}_n^*(\rho,r)| > L\} \subseteq \{\exists\ \mathcal{C}\ 2r\text{-separated with } \ |\mathcal{C}| \ge n_0 \ \text{ and } \ |\mathcal{R}_n\cap B(x,r)| \ge \rho|B(x,r)|, \forall\ x\in\mathcal{C}\}$$

The total number of possible sets $\mathcal{C}$ with $|\mathcal{C}| = \ell$ is upper bounded by $(2n)^{d\cdot\ell}$. Using this, Lemma 2.29 and (2.7) we get

$$\mathbb{P}(|\mathcal{R}_n^*(\rho,r)| > L) \le \sum_{\ell\ge n_0} (2n)^{d\cdot\ell}\cdot\exp(-\kappa\cdot\rho\cdot(r^d\ell)^{1-2/d}) = \sum_{\ell\ge n_0}\exp(d\cdot\ell\cdot\log(2n) - \kappa\rho\cdot(r^d\ell)^{1-2/d}).$$

We see that the entropic term that comes from counting all possible subsets dominates in the exponential above. So we would like to reduce the total number of sets $\mathcal{C}$ that we are considering in order to match the two terms appearing in the exponential above. To do this, we will use the following result that shows that every set has a subset of the same capacity up to constants and which is of the same order as its volume.

Theorem 2.31. ([1, Theorem 1.1]) *Suppose $d \geq 3$. There exists a positive constant c so that the following holds. Let A be a finite subset of $\mathbb{Z}^d$ which is $2r$ separated for $r \geq 1$, i.e. any two distinct points of A are at distance at least $2r$ apart. Then there exists a subset U of A with the property that*

$$\mathrm{Cap}(\cup_{x\in U} B(x,r)) \geq c \cdot r^{d-2} \cdot |U| \geq c^2 \cdot \mathrm{Cap}(\cup_{x\in A} B(x,r)).$$

We defer the proof of this to end of the proof of the proposition.

Let $\mathcal{C} = \{x_1, \ldots, x_{n_0}\}$. Applying the above theorem we get that there exists a subset U of $\mathcal{C}$ such that

$$|U| \cdot r^{d-2} \asymp \mathrm{Cap}(B(U,r)) \asymp \mathrm{Cap}(B(\mathcal{C},r)).$$

Using that $\mathrm{Cap}(A) \geq |A|^{1-2/d}$ we get

$$|U| \gtrsim |B(\mathcal{C},r)|^{1-2/d} \cdot r^{2-d} \gtrsim \left(\frac{L}{\rho}\right)^{1-2/d} \cdot r^{2-d}.$$

For every $\ell > 0$ there exist at most $(2n)^{d\cdot\ell}$ possible subsets of $[-n,n]^d$ of size ℓ. So by a union bound we have

$$\begin{aligned}
&\mathbb{P}(|\mathcal{R}_n(\rho,r)| > L) \\
&\leq \sum_{\ell=(L/\rho)^{1-2/d}\cdot r^{2-d}}^{\infty} \mathbb{P}\Big(\exists\, U : |U| = \ell, |U| r^{d-2} \asymp \mathrm{Cap}(B(U,r)), |\mathcal{R}_n \cap B(x,r)| \geq \rho |B(x,r)|,\ \forall\, x \in U\Big) \\
&\leq \sum_{\ell=(L/\rho)^{1-2/d}\cdot r^{2-d}}^{\infty} \exp(c \cdot \ell \cdot \log n) \cdot \exp(-\kappa \cdot \rho \cdot \ell \cdot r^{d-2}),
\end{aligned}$$

where for the final inequality we used Lemma 2.29. By taking the constant C_0 sufficiently large so that $\rho \cdot r^{d-2} \geq C_0 \cdot \log n$, we see that there exists a positive constant κ such that the sum above is upper bounded by

$$\sum_{\ell=(L/\rho)^{1-2/d}\cdot r^{2-d}}^{\infty} \exp(-\kappa \cdot \rho \cdot \ell \cdot r^{d-2}) \lesssim \exp(-\kappa \cdot \rho^{2/d} \cdot L^{1-2/d})$$

and this completes the proof. □

We are now ready to give the

Proof of Proposition 2.26. Clearly we have that

$$\mathcal{R}_n(\rho,r) = \bigcup_{i\geq 0} \mathcal{R}_n^*(2^i\rho, r).$$

Let $\alpha = \sum_{i=0}^{\infty} 2^{-i/(d-2)}$. By a union bound we get

$$\begin{aligned}
\mathbb{P}(|\mathcal{R}_n(\rho,r)| > L) &\leq \sum_{i=0}^{\infty} \mathbb{P}\left(|\mathcal{R}_n^*(2^i\rho,r)| > \alpha \cdot \frac{L}{2^{i/(d-2)}}\right) \\
&\leq \sum_{i\geq 0} \exp\left(-\kappa \cdot (2^i\rho)^{2/d} \cdot \left(\frac{L}{2^{i/(d-2)}}\right)^{1-2/d}\right) \lesssim \exp\left(-\kappa' \cdot \rho^{2/d} \cdot L^{1-2/d}\right)
\end{aligned}$$

and this completes the proof. □

Proof of Theorem 2.31. We first give the proof in the case where $r = 1$. We show that every $\mathcal{C}$ has a subset U satisfying

$$\mathrm{Cap}(U) \geq c \cdot |U| \geq c^2 \mathrm{Cap}(\mathcal{C}).$$

For every $x \in A$, let (X_n^x) be a collection of independent simple random walks in $\mathbb{Z}^d$ with $X_0^x = x$ for every $x \in A$. For every $x \in A$ we write $\widetilde{H}_A^x$ for the first return time to A of the walk X^x. We now define

$$\mathcal{U} = \{x \in A : \widetilde{H}_A^x = \infty\}.$$

We then immediately get that $\mathbb{E}[|\mathcal{U}|] = \mathrm{Cap}(A)$ and $\mathrm{Var}(|\mathcal{U}|) \leq \mathrm{Cap}(A)$ as $|\mathcal{U}|$ is the sum of independent Bernoulli random variables. By Chebyshev's inequality we then obtain

$$\mathbb{P}\left(|\mathcal{U}| \leq \frac{\mathbb{E}[|\mathcal{U}|]}{2}\right) \leq \frac{4}{\mathrm{Cap}(A)} \quad \text{and} \quad \mathbb{P}(|\mathcal{U}| \geq 2\mathbb{E}[|\mathcal{U}|]) \leq \frac{1}{\mathrm{Cap}(A)}.$$

Assuming that $\mathrm{Cap}(A) > 16$, since otherwise the statement holds true, we get that

$$\mathbb{P}\left(2\mathrm{Cap}(A) \geq |\mathcal{U}| \geq \frac{1}{2}\mathrm{Cap}(A)\right) \geq \frac{2}{3}.$$

It remains to show that the capacity of $\mathcal{U}$ is of the same order as the size of $\mathcal{U}$ with high enough probability. To do this we are going to use the variational characterisation of capacity. Let μ be the uniform measure on $\mathcal{U}$. We then deduce

$$\mathrm{Cap}(\mathcal{U}) \geq \frac{|\mathcal{U}|^2}{\sum_{x,y\in\mathcal{U}} g(x,y)}. \tag{2.15}$$

We next upper bound the expectation of the denominator above. By the last exit decomposition formula we have

$$\begin{aligned}\mathbb{E}\left[\sum_{x,y\in\mathcal{U}} g(x,y)\right] &\leq \sum_{x\in A} \mathbb{P}\left(\widetilde{H}_A^x = \infty\right) g(0) + \sum_{x,y\in A} \mathbb{P}\left(\widetilde{H}_A^x = \infty\right) \mathbb{P}\left(\widetilde{H}_A^y = \infty\right) g(x,y) \\ &= g(0)\mathrm{Cap}(A) + \mathrm{Cap}(A) = (g(0)+1)\mathrm{Cap}(A).\end{aligned}$$

Using Markov's inequality we get

$$\mathbb{P}\left(\sum_{x,y\in\mathcal{U}} g(x,y) \leq 4(g(0)+1)\mathrm{Cap}(A)\right) \geq \frac{3}{4}.$$

Therefore, combining all of the above we deduce

$$\mathbb{P}\left(2\mathrm{Cap}(A) \geq |\mathcal{U}| \geq \frac{1}{2}\mathrm{Cap}(A), \sum_{x,y\in\mathcal{U}} g(x,y) \leq 4(g(0)+1)\mathrm{Cap}(A)\right) \geq \frac{5}{12}.$$

By (2.15) we see that on the event appearing in the probability above we get that

$$\mathrm{Cap}(\mathcal{U}) \geq \frac{|\mathcal{U}|^2}{4(g(0)+1)\mathrm{Cap}(A)} \geq c \cdot |\mathcal{U}| \geq c^2 \cdot \mathrm{Cap}(A),$$

where c is a positive constant, and hence, this proves that

$$\mathbb{P}\big(\mathrm{Cap}(\mathcal{U}) \geq c \cdot |\mathcal{U}| \geq c^2 \cdot \mathrm{Cap}(A)\big) \geq \frac{5}{12}.$$

This concludes the proof in the case when $r = 1$.

For $r \geq 1$ we proceed by defining for every $x \in A$ an independent Bernoulli random variable Y_x with parameter

$$\frac{c}{r^{d-2}} \cdot \sum_{y \in \partial B(x,r)} \mathbb{P}_y\Big(\widetilde{H}_{B(A,r)} = \infty\Big),$$

where c is a constant to ensure that the quantity above is smaller than 1 and we write $B(A,r) = \cup_{x \in A} B(x,r)$. We next define the set $\mathcal{U}$ as

$$\mathcal{U} = \{x : Y_x = 1\}.$$

We have for the expectation and the variance

$$\mathbb{E}[|B(\mathcal{U},r)|] = |B(0,r)| \cdot \sum_{x \in A} \frac{c}{r^{d-2}} \sum_{y \in \partial B(A,r)} \mathbb{P}_y\Big(\widetilde{H}_{B(A,r)} = \infty\Big) \asymp r^2 \cdot \mathrm{Cap}(B(A,r)).$$

For the variance as above we get

$$\mathrm{Var}(|B(\mathcal{U},r)|) \leq |B(0,r)| \cdot \mathbb{E}[|B(\mathcal{U},r)|]\,.$$

So with Chebyshev as above we get

$$\mathbb{P}\Big(|B(\mathcal{U},r)| \asymp r^2 \cdot \mathrm{Cap}(B(A,r))\Big) \geq \frac{3}{4}.$$

We finally need to control the sum of the Green's function as before. By taking the uniform measure on $\partial B(\mathcal{U},r)$ we need to control

$$\frac{1}{(|\mathcal{U}| \cdot |\partial B(0,r)|)^2} \sum_{x,x' \in \mathcal{U}} \sum_{y \in \partial B(x,r)} \sum_{y' \in \partial B(x',r)} g(y - y').$$

For $x = x'$ we get

$$\sum_{x \in \mathcal{U}} \sum_{y \in \partial B(x,r)} \sum_{y' \in \partial B(x,r)} g(y - y') \lesssim r^d \cdot |\mathcal{U}|.$$

For $x \neq x'$ we take expectation of the sum involving the Green's function and obtain

$$\begin{aligned}
\mathbb{E}\left[\sum_{x \neq x' \in \mathcal{U}} \sum_{y \in \partial B(x,r)} \sum_{y' \in \partial B(x',r)} g(y - y')\right] &\leq r^{2(d-1)} \cdot \sum_{x \neq x' \in A} \mathbb{P}(Y_x = 1)\, \mathbb{P}(Y_{x'} = 1)\, g(x - x') \\
&= r^d \cdot \sum_{x \neq x' \in A} \sum_{z \in \partial B(x,r)} \mathbb{P}_z\Big(\widetilde{H}_{B(A,r)} = \infty\Big)\, \mathbb{P}(Y_{x'} = 1)\, g(x' - z) \\
&\lesssim r^d \cdot \sum_{x' \in A} \mathbb{P}(Y_{x'} = 1) = r^d \cdot \mathbb{E}[|\mathcal{U}|]\,,
\end{aligned}$$

where in the last inequality we used the last exit decomposition formula. The proof can be completed in the same way as before using Chebyshev's inequality. $\square$

3 Random interlacements

The goal of this section is to define random interlacements on $\mathbb{Z}^d$ with $d \geq 3$. First we will use them in order to prove a very strong coupling result with simple random walk. In the next section we will use them in order to sample uniform spanning forests using the interlacements Aldous Broder algorithm introduced by Tom Hutchcroft.

We will write everything in the case of $\mathbb{Z}^d$ but one can generalise to any transient graph as well. For $n \leq m$ we define $\mathcal{W}(n,m)$ to be the set of graph homomorphisms from $\{n, n+1, \ldots, m\}$ to G that are transient, i.e. they have the property that every vertex is visited a finite number of times. We also define

$$\mathcal{W} = \bigcup (\mathcal{W}(n,m) : -\infty \leq n \leq m \leq \infty).$$

For every path $w \in \mathcal{W}(n,m)$ and a finite set $K \subseteq \mathbb{Z}^d$ we write

$$H_K(w) = \inf\{n \leq i \leq m : w(i) \in K\}$$

for the first time that w hits K and

$$L_K(w) = \sup\{n \leq i \leq m : w(i) \in K\}$$

for the last time w is in K. We write $w_K = w|_{[H_K(w), L_K(w)]}$. We write $\mathcal{W}_K(n,m)$ (resp. $\mathcal{W}_K$) for the paths in $\mathcal{W}(n,m)$ (resp. $\mathcal{W}$) that visit K.

We equip $\mathcal{W}$ with the topology generated by open sets of the form

$$\{w \in \mathcal{W} : w_K = w'_K\},$$

for K a finite subset of $\mathbb{Z}^d$ and $w' \in \mathcal{W}_K$. We also endow $\mathcal{W}$ with the Borel σ-algebra $\mathcal{B}(\mathcal{W})$ generated by this topology.

Finally we define the time shift $\theta_k : \mathcal{W} \to \mathcal{W}$ by assigning to every $w \in \mathcal{W}$ the path $\theta_k(w)(i) = w(i+k)$ for all i and with this we can now also define an equivalence relation $\sim$ by saying that $w_1 \sim w_2$ if there exists k such that $\theta_k(w_1) = w_2$. Lastly, define $\mathcal{W}^* = \mathcal{W}/\sim$ to be the quotient space and $\pi : \mathcal{W} \to \mathcal{W}^*$ for the projection mapping. We define the quotient σ-algebra $\widetilde{\mathcal{W}^*}$ on $\mathcal{W}^*$ by including every set A if and only if $\pi^{-1}(A) \in \mathcal{B}(\mathcal{W})$.

For every finite set $K \subseteq \mathbb{Z}^d$ we define a measure Q_K as follows

$$Q_K(\{w \in \mathcal{W} : w|_{(-\infty,0]} \in A, w(0) = x \text{ and } w|_{[0,\infty)} \in B\}) = \mathbb{P}_x\Big(X \in A, \widetilde{H}_K = \infty\Big)\, \mathbb{P}_x(X \in B)\,,$$

where A and B are Borel subsets of $\cup_{m \geq n \geq 0} \mathcal{W}(n,m)$ and X is a simple random walk. Note that we defined Q_K only on a π-system, but since the σ-algebra is generated by such sets, this uniquely determines Q_K.

From the definition of Q_K we see that $Q_K/\mathrm{Cap}(K)$ is a probability measure on bi-infinite trajectories that hit K at time 0 and $(X_n)_{n \geq 0}$ and $(X_{-n})_{n \geq 0}$ are independent conditionally on X_0 which is distributed according to the normalised equilibrium measure. Moreover, the backward path has the distribution of a simple random walk conditioned on avoiding K and the forward path is an unconditioned simple random walk.

Theorem 3.1 (Sznitman and Teixeira)**.** *There exists a unique σ-finite measure ν on $\mathcal{W}^*$ such that for every set $A \subseteq \mathcal{W}^*$ in the quotient σ-algebra of $\mathcal{W}^*$ and every finite $K \subseteq V$ we have*

$$\nu(A \cap \mathcal{W}^*_K) = Q_K(\pi^{-1}(A)).$$

Sketch of proof. To prove this result, we first show that such a measure is unique if it exists. To prove existence, we show that the measures Q_K are consistent, in the sense that if $K \subseteq K' \subseteq \mathbb{Z}^d$ are both finite subsets, then for any $A \subseteq \mathcal{W}_K^* \subseteq \mathcal{W}_{K'}^*$ we have

$$Q_{K'}(\pi^{-1}(A)) = Q_K(\pi^{-1}(A)).$$

Once this is established, then we can define ν by writing

$$\nu(A) = \sum_{n=1}^{\infty} Q_{E_n}(\pi^{-1}(A \cap (\mathcal{W}_{E_n}^* \setminus \mathcal{W}_{E_{n-1}}^*))),$$

where (E_n) is an increasing sequence of finite subsets of $\mathbb{Z}^d$ with $\cup E_n = \mathbb{Z}^d$.

For a full proof we refer the reader to [4, Theorem 6.2]. $\square$

References

[1] Amine Asselah and Bruno Schapira. Extracting subsets maximizing capacity and folding of random walks. 2020. arXiv:2003.03073.

[2] Amine Asselah and Bruno Schapira. Large deviations for intersections of random walks. *Comm. Pure Appl. Math.*, 76(8):1531–1553, 2023.

[3] Tianyi Bai and Yijun Wan. Capacity of the range of tree-indexed random walk. *Ann. Appl. Probab.*, 32(3):1557–1589, 2022.

[4] Alexander Drewitz, Balázs Ráth, and Artëm Sapozhnikov. *An introduction to random interlacements*. SpringerBriefs in Mathematics. Springer, Cham, 2014.

[5] K. M. Khanin, A. E. Mazel, S. B. Shlosman, and Ya. G. Sinaĭ. Loop condensation effects in the behavior of random walks. In *The Dynkin Festschrift*, volume 34 of *Progr. Probab.*, pages 167–184. Birkhäuser Boston, Boston, MA, 1994.

[6] Gregory F. Lawler. *Intersections of random walks*. Probability and its Applications. Birkhäuser Boston, Inc., Boston, MA, 1991.

[7] Gregory F. Lawler and Vlada Limic. *Random walk: a modern introduction*, volume 123 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

[8] P. Phetdrapat. 2012. PhD thesis, University of Bath.

[9] Bruno Schapira. Branching capacity of a random walk range. arXiv:2303.17830.

[10] M. van den Berg, E. Bolthausen, and F. den Hollander. On the volume of the intersection of two Wiener sausages. *Ann. of Math. (2)*, 159(2):741–782, 2004.