

Performance of the Metropolis algorithm on a disordered tree: the Einstein relation

Pascal Maillard*^{*}; Ofer Zeitouni[†]

The Weizmann Institute of Science

April 1, 2013. Revised July 30, 2013.

Abstract

Consider a d -ary rooted tree ($d \geq 3$) where each edge e is assigned an i.i.d. (bounded) random variable $X(e)$ of negative mean. Assign to each vertex v the sum $S(v)$ of $X(e)$ over all edges connecting v to the root, and assume that the maximum S_n^* of $S(v)$ over all vertices v at distance n from the root tends to infinity (necessarily, linearly) as n tends to infinity. We analyze the Metropolis algorithm on the tree and show that under these assumptions there always exists a temperature $1/\beta$ of the algorithm so that it achieves a linear (positive) growth rate in linear time. This confirms a conjecture of Aldous [Ald98]. The proof is obtained by establishing an Einstein relation for the Metropolis algorithm on the tree.

Keywords. Metropolis algorithm; Einstein relation; branching random walk; random walk in random environment

MSC2010. 60K37, 60J22, 82C41

1 Introduction

Given a d -regular rooted tree, attach to each edge e a random variable $X(e)$, such that the variables are independent and identically distributed. For a vertex v in the tree, denote by $S(v)$ the sum of the variables $X(e)$ over all edges e on the path from the root to v . This defines a branching random walk, a basic model for a disordered tree. It is natural to ask for an efficient algorithm which explores the vertices of this tree in order to find vertices v with a large value of $S(v)$. In fact, Aldous [Ald98] proposed this problem as a benchmark problem for comparing different generic optimization algorithms, since the naïve approach, which would be to simply explore all vertices down to the level n in the tree and taking the one with the maximal value of $S(v)$, is a bad choice for an algorithm because the number of vertices grows exponentially in n .

The Metropolis algorithm is a general recipe for constructing a discrete-time Markov chain on a finite state space for which a given distribution π is stationary and whose transitions

*PM was partially supported by a grant from the Israel Science Foundation.

[†]OZ was partially supported by NSF grant DMS-1203201, the Israel Science Foundation and the Herman P. Taubman chair of Mathematics at the Weizmann Institute.

respect a given graph structure of the state space. In the context of the comparison of algorithms discussed earlier, Aldous [Ald98] suggested to use the Metropolis algorithm to “sample” a certain Gibbs measure on the vertices of a branching random walk tree, namely the one which assigns mass $e^{\beta S(v)}$ to a vertex v , for some parameter $\beta > 0$. In the case where this measure is infinite, for example when there is an infinite number of vertices v with $S(v) \geq 0$, this algorithm should “walk down the tree” and, for an appropriate choice of the parameter β , find vertices v with high values of $S(v)$. Let $|v|$ denote the level of the vertex v in the tree and let V_k be the vertex visited by the Metropolis algorithm at the time k . Aldous raised the following natural question: If the maximum of the branching random walk has positive speed, i.e. if $\lim_{n \rightarrow \infty} \max_{|v|=n} S(v)/n > 0$, does there exist a choice of the parameter β , such that $\liminf_{k \rightarrow \infty} S(V_k)/k > 0$? We will answer this question in the affirmative for a certain class of laws of the variables $X(e)$, including the binomial distribution.

In fact, we show more: Let $v_\beta = \lim_{k \rightarrow \infty} S(V_k)/k$, which exists almost surely [Ald98]. We show that there exists a parameter $\beta_0 > 0$, such that $v_{\beta_0} = 0$ and $(dv_\beta/d\beta)|_{\beta=\beta_0} = \sigma^2/2$, where σ^2 is the asymptotic variance of $S(V_k)$, which we show to be positive and finite. This result was conjectured by Aldous, who gave heuristic arguments and numerical evidence for it (in the case where the variables $X(e)$ only take the values 1 and -1). Results of this type are also known as *Einstein relations* in the domain of random walks in random environments and our methods of proof will indeed rely on many techniques from this field, some of which have been obtained recently.

1.1 Definition of the model and statement of the main result

We are given a d -regular infinite rooted tree, $d \geq 3$. The root is denoted by ρ and the level/depth of a vertex v in the tree by $|v|$. The notation $u \sim v$ denotes that u and v are connected by an edge. The parent of a vertex v is denoted by \vec{v} (with the convention $\vec{\rho} = \rho$). We write $u \leq v$ if u is an ancestor of v and $u < v$ if $u \leq v$ but $u \neq v$. We furthermore use the following handy notation: if $u \leq v$, then $[u, v]$ denotes the set of vertices on the path from u to v , including u and v . The notation $(u, v]$, $[u, v)$ and (u, v) then has obvious meaning. To each edge $e = (\vec{v}, v)$, we then attach a random variable $X(e)$, such that the collection $(X(e))$ is i.i.d. according to the law of a random variable X . Here, orientation of the edges matters, and we will set $X(v, \vec{v}) = -X(\vec{v}, v)$ for all v .

In what follows, we will introduce several assumptions, **which we assume to hold throughout the paper**. We begin with the following assumptions on the law of X .

- (XS) The law of X is of compact support, i.e. $\text{esssup } |X| < \infty$.
- (XR) There exists $\beta_0 > 0$, such that $E[e^{\beta_0 X} f(X)] = E[f(-X)]$ for all bounded measurable functions f .
- (XM) $\inf_{\beta \geq 0} \Lambda(\beta) > 0$, where $\Lambda(\beta) = \log E[e^{\beta X}] + \log(d-1)$.

Note that (XR) is equivalent to the Laplace transform $\beta \mapsto \mathbf{E}[e^{\beta X}]$ being symmetric around $\beta_0/2$. In particular, the constant β_0 is necessarily unique unless $X = 0$ almost surely.

An example for a law satisfying (XS), (XR) and (XM) is the distribution of $2Y - n$, where Y follows a binomial distribution of parameters n and p , with $p \in (p_0, 1/2)$, where $p_0 = (1 - \sqrt{1 - (d-1)^{-2/n}})/2$. In this case, $\beta_0 = \log \frac{1-p}{p}$. In general, in order to construct a law satisfying (XS) and (XR), one can start from a symmetric random variable X taking values

in a compact interval $[-K, K]$ and define a law with Radon–Nikodym derivative proportional to $e^{(-\beta_0/2)X}$ with respect to the law of X . This law will then satisfy (XM) for β_0 small enough.

We remark that assumption (XS) seems not to be crucial and the argument extends to certain distributions with non compact support, at the cost of more complicated technical arguments. To avoid this complication we chose to present the result under this simplifying assumption. On the other hand, assumption (XR) is essential for our treatment, as it ensures, at β_0 , the reversibility of the Markov chain consisting of the environment viewed from the point of view of the particle, see Proposition 2.2; the reversibility will be crucial both in the application of the Kipnis-Varadhan theory, as well as in the proof of validity of the Einstein relation (one may expect a correction term for non reversible chains).

We now define the branching random walk by

$$S(v) = \sum_{u \in (\rho, v]} X(\vec{u}, u), \quad S(\rho) = 0. \quad (1.1)$$

Note that $X(u, v) = S(v) - S(u)$ for every two vertices u and v with $u \sim v$, by the above convention that $X(u, v) = -X(v, u)$. Since $\Lambda(\beta)$ is the log-Laplace transform of this branching random walk, it is known [Big77] that $\lim_{n \rightarrow \infty} \max_{|v|=n} S(v)/n$ exists and is positive under assumption (XM). Note further that assumptions (XM) and (XR) together imply that $\Lambda(\beta) > 0$ for all $\beta \in \mathbb{R}$, such that $\lim_{n \rightarrow \infty} \min_{|v|=n} S(v)/n$ exists and is negative [Big77].

In order to define the Metropolis algorithm, we are given a function $h : \mathbb{R}_+ \rightarrow \mathbb{R}$ satisfying the following conditions (examples are $h(x) = \min(1, x)$ and $h(x) = x/(1+x)$):

- (H1) h takes values in $[0, 1]$, is non-decreasing and satisfies $h(0) = 0$ and $\lim_{x \rightarrow \infty} h(x) = 1$.
- (H2) It is Lipschitz-continuous and continuously differentiable on $(0, 1) \cup (1, \infty)$.
- (H3) It satisfies the functional equation $h(x) = xh(1/x)$ for all $x \geq 0$.

For a given realization of the branching random walk and a parameter $\beta \in \mathbb{R}$, the Metropolis algorithm is then the Markov chain $(V_n)_{n \geq 0}$ on the vertices of the tree with the transition probabilities $P_\beta(v, w)$ given by

$$P_\beta(v, w) = p_\beta(X(v, w)) \text{ for } w \sim v, \quad \text{where } p_\beta(x) = \frac{1}{d} h(e^{(\beta_0 + \beta)x})$$

$$P_\beta(v, v) = 1 - \sum_{w \sim v} P_\beta(v, w).$$

We denote the (annealed, i.e. averaged over the environment) law of the Metropolis algorithm on the branching random walk tree by \mathbb{P}_β and expectation with respect to this law by \mathbb{E}_β . Our main theorem is the following:

Theorem 1.1. *Set $S_n = S(V_n)$.*

1. *The limit $\sigma^2 = \lim_{n \rightarrow \infty} S_n^2/n$ exists \mathbb{P}_0 -almost surely and is a strictly positive and finite constant.*
2. *For each $\beta \in \mathbb{R}$, the deterministic limit $v_\beta = \lim_{n \rightarrow \infty} S_n/n$ exists \mathbb{P}_β -almost surely and satisfies*

$$\lim_{\beta \rightarrow 0} \frac{v_\beta}{\beta} = \frac{\sigma^2}{2}. \quad (1.2)$$

We note that the existence of v_β and the fact that it vanishes at $\beta = 0$ were already shown in [Ald98] (in a slightly more restrictive setup). The main novelty in Theorem 1.1 is the proof of the Einstein relation (1.2), as well as the fact that the right side is strictly positive.

1.2 Related works

Our main inspiration, as noted above, is Aldous' work [Ald98]. In that paper, Aldous makes the crucial observation that a reversible invariant measure for the environment viewed from the point of view of the particle exists at $\beta = 0$, and derives from this that $v_0 = 0$, and the existence of the limit σ^2 under \mathbb{P}_0 ; he also completely analyzes a greedy algorithm and formulates a series of conjectures, some answered here. In the same paper, Aldous also refers speculatively to [LR94] as relevant to the analysis near $\beta = 0$; indeed, the approach of the latter to proofs of the Einstein relation forms the basis of the current paper, as well to recent advances in the analysis of the Einstein relation for disordered systems, as we now discuss.

The Einstein relation (ER) links the asymptotic variance of additive functionals of (reversible) Markov chains in equilibrium to the chains' response to small perturbations. In a weak limit (where the time-scale is related to the strength of the perturbation), Lebowitz and Rost [LR94] provide a general recipe (based on the Kipnis-Varadhan theory, see [KLO12] for a comprehensive account) for the validity of a weak form of the ER in disordered systems. For the tagged particle in the symmetric exclusion process, the ER was proved by Loulakis in $d \geq 3$ [Lou02] by perturbative methods (using transience in an essential way); this approach was adapted to bond diffusion in \mathbb{Z}^d in special environment distributions [KO05a]. For mixing dynamical random environments with spectral gap, a full perturbation expansion was proved in [KO05b].

Significant recent progress was achieved by [GMP12], where the Lebowitz–Rost approach was combined with good *uniform in the environment* estimates on certain regeneration times in the transient regime, that are used to pass from a weak ER to a full ER. These uniform estimates are typically not available for random walks on (random) trees, and a completely different approach, based on explicit recursions, was taken in [BHOZ11], where (biased) random walks on Galton–Watson trees were analyzed. While we still consider walks on trees, the approach we take is closer to that of [GMP12], while replacing their uniform regeneration estimates with probabilistic estimates, in the spirit of [PZ08]. See also [Guo12] for another approach to the proof of the ER in the context of balanced random walks.

1.3 Overview of the proof and outline of the paper

As mentioned above, the starting point is Aldous' observation that under \mathbb{P}_0 , the environment viewed from the point of view of the particle forms a *reversible* Markov chain. We begin by proving this (Proposition 2.2), and then apply the Kipnis–Varadhan theory to deduce an invariance principle for anti-symmetric additive functionals (Lemma 2.4). This allows us to prove the weak ER, Theorem 2.1, following the Lebowitz–Rost recipe.

To handle the perturbation, estimates on regeneration times and distances are crucial. We work with *level* regeneration times, that are introduced in Section 4; these involve the random walk location $\{V_n\}$, not the vertices values $\{S_n\}$; of course, the latter influence the transition probabilities of the random walk. In order to transform the weak ER to a full ER, we need uniform bounds on the moments of the regeneration times. These are obtained in Proposition 4.1, where it is proved that the regeneration times exhibit uniform annealed

stretched-exponential bounds. The proof has two main steps: first, exponential moments are proved for regeneration *distances*, using in a crucial way a structure lemma of Grimmett and Kesten, see Lemma 4.4. Then, the estimates for regeneration times are obtained, using that the walk must visit many well-separated fresh vertices, and between two such visits, the walk has a large enough probability to hit distant levels. In proving the last statement, an argument of Aidékon [Aid08] is used, see Lemma 4.6.

Acknowledgement We thank David Aldous for bringing the problem of the analysis of the Metropolis algorithm on the tree to our attention. We further thank an anonymous referee for his detailed comments improving the exposition of the article.

2 The weak Einstein relation

In this section, we show that the Einstein relation holds for times of the order of β^{-2} . Specifically, we will prove the following result:

Theorem 2.1.

1. $\mathbb{E}_0[S_n] = 0$ for all $n \geq 0$. Furthermore, the limit $\sigma^2 = \lim_{n \rightarrow \infty} \mathbb{E}_0[S_n^2]/n$ exists and is a finite, non-negative constant.
2. Set $S^\beta = \beta S_{\lfloor \beta^{-2} \rfloor}$. Then,

$$\mathbb{E}_\beta[S^\beta] \rightarrow \frac{\sigma^2}{2}, \quad \text{as } \beta \rightarrow 0.$$

The proof of Theorem 2.1 uses a fairly generic and now classical change of measure argument in the spirit of Lebowitz and Rost [LR94]. It uses the crucial concept of the *environment seen from the particle*, which we define as follows.

Let Ω be the space of rooted d -regular unlabelled trees ω with marked edges, i.e. to every two vertices u and v with $u \sim v$ we associate a real number $X_\omega(u, v)$ with $X_\omega(v, u) = -X_\omega(u, v)$. Note that “unlabelled” means that we do not distinguish between the neighbors of a vertex¹, this will be crucial for what follows. The root of every tree ω is denoted by ρ . For every vertex v , we then define the shift operator $\theta_v : \Omega \rightarrow \Omega$, which yields the tree ω “seen from the vertex v ”. A bit more formally², if for a vertex u in ω we denote by $\theta_v^{-1}u$ its corresponding vertex in $\theta_v\omega$, then $X_{\theta_v\omega}(\theta_v^{-1}u, \theta_v^{-1}w) = X_\omega(u, w)$. In particular, if $v \sim \rho$, then the mark of the edge (ρ, v) “changes its sign upon passing from ω to $\theta_v\omega$ ”.

Define $S_\omega(v)$ for each vertex v analogously to (1.1). For $\omega \in \Omega$, we then define the operators \mathcal{L}_ω and L_ω acting on functions $f : \mathbb{R} \rightarrow \mathbb{R}$ by

$$\mathcal{L}_\omega f = \sum_{v \sim \rho} f(S_\omega(v)) = \sum_{v \sim \rho} f(X_\omega(\rho, v)) \quad \text{and} \quad L_\omega f = \mathcal{L}_\omega(pf),$$

where $p(x) = p_0(x) = d^{-1}h(e^{\beta_0 x})$ (see Section 1.1).

¹There are several ways how to render this formal, one of which consists of first defining the space $\tilde{\Omega}$ of labelled rooted d -regular trees with marks, which is homeomorphic to $\mathbb{R}^{\mathbb{N}}$. The space Ω is then defined as the quotient space with respect to the group of graph automorphisms fixing the root. It is endowed with the Borel σ -algebra induced by the quotient topology.

²In order to render this completely formal, one can first define the shift operator on the auxiliary space $\tilde{\Omega}$ (see above) and then show that it induces a well-defined operator on Ω .

Let \mathbf{P} be the law on Ω under which all edges $X_\omega(\vec{u}, u)$ are i.i.d. according to the law of X (see Section 1.1). We denote by \mathbf{E} the expectation with respect to \mathbf{P} . Furthermore, let \mathbb{P}_β be the law of the Metropolis algorithm $(V_n)_{n \geq 0}$ with transition probabilities P_β defined in Section 1.1 and the underlying tree ω_0 distributed according to \mathbf{P} . Expectation w.r.t. \mathbb{P}_β is denoted by \mathbb{E}_β , and we also set $\mathbb{P} = \mathbb{P}_0$ and $\mathbb{E} = \mathbb{E}_0$. Setting $\omega_n = \theta_{V_n} \omega_0$ then defines a Markov chain $(\omega_n)_{n \geq 0}$ on the space Ω , which jumps from ω to $\theta_v \omega$ with probability $p_\beta(X(\rho, v))$ for every $v \sim \rho$. Let $(\mathcal{F}_n)_{n \geq 0}$ be the natural filtration of $(\omega_n)_{n \geq 0}$, augmented by sets of zero measure. The process $S_n = S_{\omega_0}(V_n)$ is then adapted to $(\mathcal{F}_n)_{n \geq 0}$, since it can be almost surely reconstructed from $\omega_0, \dots, \omega_n$.

The following result was already observed by Aldous [Ald98], who had a more complicated proof for it.

Proposition 2.2. *The process $(\omega_n)_{n \geq 0}$ is reversible and ergodic under \mathbb{P} .*

Proof. In order to show reversibility, since $(\omega_n)_{n \geq 0}$ is a Markov process, we only have to show that $\mathbb{E}[F(\omega_0, \omega_1)] = \mathbb{E}[F(\omega_1, \omega_0)]$ for every bounded (Borel) measurable functional $F : \Omega^2 \rightarrow \mathbb{R}$. For this, it is obviously enough to show that $\mathbb{E}[(F(\omega_0, \omega_1) - F(\omega_1, \omega_0)) \mathbb{1}_{\omega_0 \neq \omega_1}] = 0$. Now, we have

$$\begin{aligned} \mathbb{E}[F(\omega_0, \omega_1) \mathbb{1}_{\omega_0 \neq \omega_1}] &= \sum_{v \sim \rho} \mathbb{E}[p(X_{\omega_0}(\rho, v)) F(\omega_0, \theta_v \omega_0)] \\ &= \sum_{v \sim \rho} \mathbf{E}[p(-X_\omega(\rho, v)) e^{\beta_0 X_\omega(\rho, v)} F(\omega, \theta_v \omega)], \end{aligned}$$

where the last equality follows from assumption (H3). Conditioned on $X_\omega(\rho, v)$, the environment $\theta_v \omega$ is distributed as ω but with one edge pointing away from the root bearing the value $-X_\omega(\rho, v)$ (remember that the vertices are unlabelled, such that “it can be any one of them”, which amounts to saying that “we don’t know where we came from”). By assumption (XR), the right-hand side of the last equation is therefore equal to

$$\sum_{v \sim \rho} \mathbf{E}[p(X_\omega(\rho, v)) F(\theta_v \omega, \omega)] = \mathbb{E}[F(\omega_1, \omega_0) \mathbb{1}_{\omega_0 \neq \omega_1}],$$

which finishes the proof of the reversibility.

Ergodicity follows from a classical ellipticity argument which we recall (see also [Zei04, Corollary 2.1.25] for a similar argument): Let Q be a stationary probability measure of the Markov chain $(\omega_n)_{n \geq 0}$ with $Q \ll \mathbf{P}$. We wish to show that $\mathbf{P} \ll Q$, which will imply ergodicity since ergodic measures are the extremal points in the convex set of stationary probabilities. Define the event $E = \{dQ/d\mathbf{P} = 0\}$. By invariance, $E_Q[P \mathbb{1}_E] = E_Q[\mathbb{1}_E] = 0$, where P is the transition kernel of the Markov chain. This further implies $\mathbb{1}_E \geq P \mathbb{1}_E$, \mathbf{P} -almost surely. Since the transition probabilities are strictly positive, we then have $\mathbb{1}_E(\omega) \geq \max_{v \sim \rho} \mathbb{1}_E(\theta_v \omega)$, because $\mathbb{1}_E(\omega)$ takes values in $\{0, 1\}$. Fixing an infinite ray $\rho = v_0, v_1, v_2, \dots$, we then get by iteration of the previous inequality that $\mathbb{1}_E(\omega) \geq \mathbb{1}_E(\theta_{v_i} \omega)$ for every i , whence $\mathbb{1}_E \geq n^{-1} \sum_{i=1}^n \mathbb{1}_E(\theta_{v_i} \omega)$ for every n . But since \mathbf{P} is a product measure and therefore ergodic with respect to the shift along the ray, Birkhoff’s ergodic theorem now gives $\mathbb{1}_E \geq \mathbf{P}(E)$, \mathbf{P} -almost surely, which implies $\mathbf{P}(E) \in \{0, 1\}$. But $Q \ll \mathbf{P}$ by hypothesis, whence $\mathbf{P}(E) = 0$. This finishes the proof. \square

We recall the following basic fact about reversible processes.

Lemma 2.3. For any bounded measurable functionals F and G and every $n \geq 0$, we have

$$\mathbb{E}[F(\omega_0, \dots, \omega_n)G(\omega_n)] = \mathbb{E}[F(\omega_n, \dots, \omega_0)G(\omega_0)].$$

We will need the following result about anti-symmetric additive functionals of reversible ergodic Markov processes. It is implicit in the proof of Theorem 2.1 in [DFGW89] and relies on a celebrated result from [KV86] (see also Chapters 1 and 2 in [KLO12] for a comprehensive account of the theory).

Lemma 2.4. Let $F : \Omega^2 \rightarrow \mathbb{R}$ be an anti-symmetric measurable functional, i.e. $F(\omega, \omega') = -F(\omega', \omega)$ for all $\omega, \omega' \in \Omega$, with $\mathbb{E}[F(\omega_0, \omega_1)^2] < \infty$. Define a sequence of random variables by $\mathfrak{S}_n = \sum_{k=1}^n F(\omega_{k-1}, \omega_k)$ for all $n > 0$. Then

1. The time-variance $\sigma^2 = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[\mathfrak{S}_n^2]$ exists and is finite.
2. There exists a square integrable martingale M_n with stationary ergodic increments, such that $\frac{1}{\sqrt{n}}(\mathfrak{S}_n - M_n)$ converges to 0 in L^2 . In particular, $\frac{1}{\sqrt{n}}\mathfrak{S}_n$ converges in law to a centered Gaussian variable with variance σ^2 .

The following lemma makes precise an expansion of p_β around p for small β . It easily follows from assumptions (H1-3).

Lemma 2.5. There exist measurable functions $q_\beta(x)$ and $q(x)$, such that

1. $p_\beta(x) = p(x) \exp(\beta q_\beta(x))$,
2. $q_\beta(x)$ is uniformly bounded for all $x \in \mathbb{R}$ and small enough β , and $q_\beta(x) \rightarrow q(x)$ as $\beta \rightarrow 0$, for all $x \in \mathbb{R}$,
3. $q(0) = q_\beta(0) = 0$,
4. there exists a constant $c > 0$, such that for small enough β , we have $|\mathbb{L}_\omega(e^{\beta q_\beta} - 1)| \leq \beta c(1 - \mathcal{L}_\omega p)$ and $|\mathbb{L}_\omega q| \leq c(1 - \mathcal{L}_\omega p)$.

We are now ready for the

Proof of Theorem 2.1. First part: We first note that there exists a measurable functional $F : \Omega^2 \rightarrow \mathbb{R}$, such that $S_1 = F(\omega_0, \omega_1)$, \mathbb{P} -almost surely. This follows from the fact that \mathbf{P} -almost surely, the d shifted environments $\theta_v \omega$, $v \sim \rho$ are all different, otherwise there would be at least two identical subtrees of the vertices in the second generation which is an event of probability zero (except if $X = 0$ almost surely, in which case the lemma is trivial). By the definition of S_1 , we can furthermore choose F to be anti-symmetric in the sense of Lemma 2.4. In particular, $\mathbb{E}[S_1] = -\mathbb{E}[S_1] = 0$ by Lemma 2.3, whence $\mathbb{E}[S_n] = 0$ for all $n \geq 0$. The second statement follows from the first part of Lemma 2.4.

Second part: We will use a change of measure argument as in [LR94]. The basic idea is to write the Radon–Nikodym derivative of \mathbb{P}_β with respect to \mathbb{P} as an exponential martingale of the form $\exp(Z_n^\beta - \frac{1}{2}A_n^\beta)$ for some martingale $(Z_n^\beta)_{n \geq 0}$ and to show that the pair $(\beta S_{\lfloor \beta^{-2} \rfloor}, Z_{\lfloor \beta^{-2} \rfloor}^\beta)$ converges in law under \mathbb{P} to a centered Gaussian vector (G_S, G_Z) with covariance $\mathbb{E}[G_S G_Z] = \frac{1}{2} \mathbb{E}[G_S^2] = \frac{1}{2} \sigma^2$. The theorem then follows from a standard change of measure argument for Gaussian variables. Here are the details:

Step 0: Set $\Delta S_n = S_{n+1} - S_n$ for all $n \geq 0$. The Radon–Nikodym derivative of \mathbb{P}_β with respect to \mathbb{P} is given by

$$\log \frac{d\mathbb{P}_\beta}{d\mathbb{P}} \Big|_{\mathcal{F}_n} = \sum_{k=0}^{n-1} \left[\log \frac{p_\beta(\Delta S_k)}{p(\Delta S_k)} + \log \left(\frac{1 - \mathcal{L}_{\omega_k} p_\beta}{1 - \mathcal{L}_{\omega_k} p} \right) \mathbb{1}_{\omega_k = \omega_{k+1}} \right] =: Y_n. \quad (2.1)$$

Note that if $1 - \mathcal{L}_{\omega_k} p = 0$, then $\omega_k \neq \omega_{k+1}$ with probability 1, such that the second summand is well defined. Note also that here and in what follows, the empty sum always has value 0. If q_β and q are the functions from Lemma 2.5, we can write the process Y_n as

$$Y_n = \beta \sum_{k=0}^{n-1} q_\beta(\Delta S_k) + \sum_{k=0}^{n-1} \log \left(1 - \frac{\mathbb{L}_{\omega_k}(e^{\beta q_\beta} - 1)}{1 - \mathcal{L}_{\omega_k} p} \right) \mathbb{1}_{\omega_k = \omega_{k+1}}. \quad (2.2)$$

We then define the process $(A_n)_{n \geq 0}$ by (we suppress the dependence on β from the notation):

$$A_n = -2 \sum_{k=1}^n \mathbb{E}[Y_k - Y_{k-1} | \mathcal{F}_{k-1}] = -2 \sum_{k=0}^{n-1} \left[\beta \mathbb{L}_{\omega_k} q_\beta + (1 - \mathcal{L}_{\omega_k} p) \log \left(1 - \frac{\mathbb{L}_{\omega_k}(e^{\beta q_\beta} - 1)}{1 - \mathcal{L}_{\omega_k} p} \right) \right].$$

(Consistent with the definitions we have that $Y_0 = A_0 = 0$.) Note that by Lemma 2.5, we have for small β ,

$$\begin{aligned} A_{n+1} - A_n &= 2 \left(\mathbb{L}_{\omega_n}(e^{\beta q_\beta} - 1) - \beta \mathbb{L}_{\omega_n} q_\beta \right) - \frac{(\mathbb{L}_{\omega_n}(e^{\beta q_\beta} - 1))^2}{1 - \mathcal{L}_{\omega_n} p} + O(\beta^3) \\ &= \beta^2 (\mathbb{L}_{\omega_n} q^2 + (\mathbb{L}_{\omega_n} q)^2 / (1 - \mathcal{L}_{\omega_n} p) + o(1)). \end{aligned} \quad (2.3)$$

We further define the process $(Z_n)_{n \geq 0}$ (again suppressing the dependence on β) by:

$$Z_n = Y_n - \sum_{k=1}^n \mathbb{E}[Y_k - Y_{k-1} | \mathcal{F}_{k-1}] = Y_n + \frac{1}{2} A_n,$$

such that $(Z_n)_{n \geq 0}$ is a martingale under \mathbb{P} with respect to the filtration $(\mathcal{F}_n)_{n \geq 0}$.

Step 1: We wish to show that the random variable $Z^\beta = Z_{\lfloor \beta^{-2} \rfloor}$ converges in law under \mathbb{P} to a centered Gaussian variable with variance $\sigma_Z^2 < \infty$. Define the \mathbb{P} -martingale $(M_n)_{n \geq 0}$ by

$$M_n = \sum_{k=0}^{n-1} \left(q(\Delta S_k) - \frac{\mathbb{L}_{\omega_k} q}{1 - \mathcal{L}_{\omega_k} p} \mathbb{1}_{\omega_k = \omega_{k+1}} \right),$$

By (2.3) and Lemma 2.5, we then have

$$\mathbb{E}[(Z_n - \beta M_n)^2] = \mathbb{E} \left[\sum_{k=0}^{n-1} [Y_{n+1} - Y_n - \beta(M_{n+1} - M_n) + O(\beta^2)]^2 \right] = o(\beta^2 n), \quad (2.4)$$

whence $\mathbb{E}[(Z^\beta - \beta M_{\lfloor \beta^{-2} \rfloor})^2] \rightarrow 0$ as $\beta \rightarrow 0$.

Note that by the fourth point of Lemma 2.5, M_n is square-integrable and by Proposition 2.2, the sequence of its increments $(M_{n+1} - M_n)_{n \geq 0}$ is stationary and ergodic. By the martingale central limit theorem for stationary ergodic sequences (see e.g. [Dur96, Theorem 7.7.5]), the sequence $(M_n / \sqrt{n})_{n \geq 0}$ then converges in law under \mathbb{P} to a centered Gaussian variable with variance $\sigma_Z^2 = \mathbb{E}[M_1^2] = \mathbb{E}[\mathbb{L}_\omega q^2 + (\mathbb{L}_\omega q)^2 / (1 - \mathcal{L}_\omega p)]$. Together with (2.4), this proves the above-mentioned convergence of Z^β .

Step 2: We wish to show that the random variable $A^\beta = A_{\lfloor \beta^{-2} \rfloor}$ converges in probability to σ_Z^2 under the law \mathbb{P} . Define the process A'_n by $A'_n = \sum_{k=0}^{n-1} [\mathbb{L}_{\omega_k} q^2 + (\mathbb{L}_{\omega_k} q)^2 / (1 - \mathcal{L}_{\omega_k} p)]$. By Proposition 2.2 and the ergodic theorem, the sequence $(A'_n / n)_{n \geq 0}$ converges \mathbb{P} -almost surely

to $\mathbb{E}[A'_1] = \sigma_Z^2$. Together with (2.3), this yields the above-mentioned convergence of A^β as $\beta \rightarrow 0$.

Step 3: Recall the definition $S^\beta = \beta S_{\lfloor \beta^{-2} \rfloor}$. We wish to show that the pair (S^β, Z^β) converges in law under \mathbb{P} to a centered Gaussian vector (G_S, G_Z) with covariance $\mathbb{E}[G_S G_Z] = \frac{1}{2} \mathbb{E}[G_S^2] = \frac{1}{2} \sigma^2$ (with σ^2 from Lemma 2.4). By (2.4), it is enough to show that this convergence holds for the pair $(S_n/\sqrt{n}, M_n/\sqrt{n})$ as $n \rightarrow \infty$. By Lemma 2.4, every linear combination $aS_n + bM_n$ is the sum of a square-integrable martingale with stationary and ergodic increments and a process R_n with $\frac{1}{\sqrt{n}}R_n \rightarrow 0$ in L^2 , as $n \rightarrow \infty$. Again by the martingale CLT for stationary, ergodic sequences ([Dur96, Theorem 7.7.5]), the pair $(S_n/\sqrt{n}, M_n/\sqrt{n})$ then converges in law as $n \rightarrow \infty$ to a centered Gaussian vector (G_S, G_Z) with $\mathbb{E}[G_S^2] = \sigma^2$, $\mathbb{E}[G_Z^2] = \sigma_Z^2$ and $\mathbb{E}[G_S G_Z] = \lim_{n \rightarrow \infty} \mathbb{E}[S_n M_n]/n = \lim_{\beta \rightarrow 0} \mathbb{E}[S^\beta Z^\beta]$.

It remains to show that $\lim_{\beta \rightarrow 0} \mathbb{E}[S^\beta Z^\beta] = \sigma^2/2$. In order to prove this, recall the definition of $\Delta S_n = S_{n+1} - S_n$ and define $\Delta Z_n = Z_{n+1} - Z_n$. We have for every $n \geq 0$,

$$\mathbb{E}[S_{n+1}^2 - S_n^2] = \mathbb{E}[(\Delta S_n)^2 + 2S_n \Delta S_n] = \mathbb{E}[\mathbf{L}_{\omega_n}(x^2) + 2S_n \mathbf{L}_{\omega_n} x], \quad (2.5)$$

where x is the identity function. By Lemma 2.3, we have $\mathbb{E}[S_n \mathbf{L}_{\omega_n} x] = \mathbb{E}[(-S_n) \mathbf{L}_{\omega_0} x]$, whence, summing (2.5) over n , we get,

$$\mathbb{E}[S_n^2] = \sum_{k=1}^n \mathbb{E}[\mathbf{L}_{\omega_0}(x^2) - 2\mathbf{L}_{\omega_0} x \times \sum_{j=0}^{k-2} \Delta S_j].$$

Furthermore, since $(Z_n)_{n \geq 0}$ is a martingale, we have

$$\mathbb{E}[S_{n+1} Z_{n+1} - S_n Z_n] = \mathbb{E}[Z_{n+1} \Delta S_n] = \mathbb{E}[\beta \Delta S_n (q_\beta(\Delta S_n) - \mathbf{L}_{\omega_n} q_\beta) + (\mathbf{L}_{\omega_n} x) Z_n], \quad (2.6)$$

where we made use of the fact that $\Delta S_k = 0$ on the event that $\omega_k = \omega_{k+1}$. Applying Proposition 2.2 to the term $\mathbb{E}[(\mathbf{L}_{\omega_n} x) Z_n]$, we see that the terms corresponding to the second summand in the brackets of (2.1) cancel. Summing (2.6) over n , this yields

$$\mathbb{E}[S_n Z_n] = \beta \sum_{k=1}^n \mathbb{E}[\mathbf{L}_{\omega_0}(x q_\beta) - \mathbf{L}_{\omega_0} x \times \sum_{j=0}^{k-2} (q_\beta(\Delta S_j) - q_\beta(-\Delta S_j))] + \mathbb{E}[S_{n-1} \mathbf{L}_{\omega_0} q_\beta].$$

Now, by assumption (H3) we have $q_\beta(x) - q_\beta(-x) = x$ for every x (This is the critical point!). Moreover, by reversibility, we have $\mathbf{E}[\mathbf{L}_\omega f] = \mathbf{E}[\mathbf{L}_\omega \bar{f}]$ for every function f , where $\bar{f}(x) = f(-x)$. This yields

$$\mathbf{E}[\mathbf{L}_\omega(x q_\beta)] = \frac{1}{2} \times \mathbf{E}[\mathbf{L}_\omega(x q_\beta - x \bar{q}_\beta)] = \frac{1}{2} \times \mathbf{E}[\mathbf{L}_\omega(x^2)].$$

Altogether, the previous equations now yield

$$\mathbb{E}[S^\beta Z^\beta] = \mathbb{E}[(S^\beta)^2]/2 + \beta \mathbb{E}[S_{\lfloor \beta^{-2} \rfloor} \mathbf{L}_{\omega_0} q_\beta].$$

Convergence of the first summand has been established above, and the second tends to 0 by Lemma 2.4 and the Cauchy-Schwarz inequality. Hence, we obtain $\lim_{\beta \rightarrow 0} \mathbb{E}[S^\beta Z^\beta] = \sigma^2/2$ as claimed.

Step 4: We claim that $\mathbb{E}_\beta[S^\beta] \rightarrow \mathbb{E}[G_S \exp(G_Z - \frac{1}{2} \sigma_Z^2)]$, as $\beta \rightarrow 0$. Since (G_S, G_Z) is a centered Gaussian vector with $\mathbb{E}[G_S G_Z] = \frac{1}{2} \mathbb{E}[G_S^2] = \sigma^2/2$ and $\mathbb{E}[G_Z^2] = \sigma_Z^2$, this will finish

the proof of the theorem. By (2.1), $\mathbb{E}_\beta[S^\beta] = \mathbb{E}[S^\beta \exp(Z^\beta - \frac{1}{2}A^\beta)]$ and by the convergences in law established above, it suffices to show that this last expression is uniformly integrable. Now, since Z_n and $\exp(Z_n - \frac{1}{2}A_n)$ are martingales, A_n is a submartingale. A_n being \mathcal{F}_{n-1} -measurable, it is therefore increasing in n . It then remains to show that $S^\beta \exp(Z^\beta)$ is uniformly integrable. By the fourth point of Lemma 2.5, Z_n is a martingale with bounded increments for β small enough. Azuma's inequality [Azu67] then implies that all exponential moments of Z^β are uniformly bounded in β , for small enough β . Furthermore, $\mathbb{E}[(S^\beta)^2]$ is uniformly bounded by the first part of this theorem. Hölder's inequality then yields uniform boundedness of $\mathbb{E}[(S^\beta \exp(Z^\beta))^c]$ for some constant $c > 1$, which finishes the proof. \square

3 Estimates on the branching random walk

In this section, we establish an estimate for the branching random walk (Lemma 3.3 below). We recall that for two vertices u, v with $u \leq v$, we denote by $[u, v]$ the set vertices on the path connecting u and v . Similarly, if $n, m \in \mathbb{N}$, then we define $[n, m]$ to be the set of vertices between levels/depths n and m . More generally, for a vertex u , we let $[n, m]_u$ denote the set of vertices between levels/depths n and m in the subtree rooted at u (which means that these vertices are between levels $n + |u|$ and $m + |u|$ in the original tree), such that $[m, n] = [m, n]_\rho$. Finally, we write $[n]$ for $[n, n]$ and $[n]_u$ for $[n, n]_u$.

Lemma 3.1. *There exist $c \in (0, \infty)$ and $b > 1$, such that for large L ,*

$$\mathbf{P}(\forall v \in [L] : \max_{w \in [\rho, v]} |S(w)| > c \log L) \leq e^{-L^b}.$$

We will first establish the following intermediate bound:

Lemma 3.2. *There exist constants $C_1, C_2 > 0$ such that for all large L ,*

$$\mathbf{P}(\exists v \in [L] : \max_{w \in [\rho, v]} |S(w)| \leq C_1) > C_2.$$

Proof. The proof is a standard first and second moment calculation. Fix $C_1 > 0$ large. Let $(S_n)_{n \geq 0}$ be a random walk starting at 0 with steps distributed according to the law of X . For $n \geq 0$ and $x \in [-C_1, C_1]$, define the event $B_n^{(x)} = \{\forall k \leq n : |S_k + x| \leq C_1\}$ and set $B_n = B_n^{(0)}$. By assumption (XM) and standard large and small deviations estimates, there exists $c_0 < d - 1$, such that for C_1 large enough,

$$\exists L_0 \in \mathbb{N} \forall L > L_0 \forall n \geq 0 : \mathbf{P}(B_n) \geq c_0^{-n}. \quad (3.1)$$

Indeed, this is obtained e.g. by combining the change of measure in the Mogulskii-Varadhan theorem [DZ98, Theorem 5.1.2] with the fact that a centered random walk with bounded i.i.d. increments stays in a tube of width a for time n with probability at least e^{-Cn/a^2} for all n and $a > a_0$ and some constant $C > 0$; for finer estimates see e.g. [Mog74].

In the sequel, we fix such a C_1 once and for all. By an argument similar to the above, there exists a constant C'_1 depending on C_1 only such that

$$(C'_1)^{-1} \sup_{x \in [-C_1, C_1]} \mathbf{P}(B_n^{(x)}) \leq \mathbf{P}(B_n) \leq C'_1 \inf_{x \in [-C_1/2, C_1/2]} \mathbf{P}(B_n^{(x)}). \quad (3.2)$$

To see (3.2), note from the above that $P(B_{n-C})/P(B_n)$ is bounded by a constant depending on C only, uniformly in $n > n_0(C)$, and then couple the walk started at x with the walk started at 0 by time C , with a fixed positive probability.

The second inequality in (3.2) yields the existence of a constant C_1'' (depending on C_1) so that for every $k \leq L$,

$$\mathbf{P}(B_{L-k})\mathbf{P}(B_k) \leq C_1''\mathbf{P}(B_L), \quad (3.3)$$

because conditioned on B_k , the probability that $|S_k| \leq C_1/2$ is bounded from below by a strictly positive constant uniformly in k .

For $v \in [L]$, let

$$A_v = \mathbf{1}_{\{\max_{w \in [\rho, v]} |S(w)| \leq C_1\}}.$$

Further let $A = \sum_{v \in [L]} A_v$. Then,

$$\mathbf{E}[A] = (d-1)^L \mathbf{P}(\forall n \leq L : |S_n| \leq C_1) = (d-1)^L \mathbf{P}(B_L),$$

As for the second moment, denote by $u \wedge v$ the most recent common ancestor of u and v . We then have for large L ,

$$\begin{aligned} \mathbf{E}[A^2] &= \sum_{u, v \in [L]} \mathbf{E}[A_v A_u] \leq \sum_{u, v \in [L]} \mathbf{P}(B_L) \sup_{x \in [-C_1, C_1]} \mathbf{P}(B_{L-|u \wedge v|}^{(x)}) \\ &\leq C_1' C_1'' \mathbf{P}(B_L)^2 \sum_{u, v \in [L]} \mathbf{P}(B_{|u \wedge v|})^{-1} \end{aligned}$$

where the last inequality follows from (3.2) and (3.3). Equation (3.1) now yields,

$$\sum_{u, v \in [L]} \mathbf{P}(B_{|u \wedge v|})^{-1} \leq \sum_{u, v \in [L]} c_0^{|u \wedge v|} \leq C(d-1)^{2L},$$

for some $C > 0$. The lemma now follows from the previous three inequalities together with the Paley–Zygmund bound $\mathbf{P}(A > 0) \geq \mathbf{E}[A]^2 / \mathbf{E}[A^2]$. \square

Proof of Lemma 3.1. Let $c > 0$ and set $H = \lceil c \log L \rceil$. Let C_1 be as in Lemma 3.2. Let $g = \text{esssup } |X|$, which is finite by assumption (XS). The branching random walks spawned by the vertices at level H being independent, we have

$$\mathbf{P}(\forall v \in [L] : \max_{w \in [\rho, v]} |S(w)| > gH + C_1) \leq \mathbf{P}(\forall v \in [L-H] : \max_{w \in [\rho, v]} |S(w)| > C_1)^{(d-1)^H}.$$

The lemma now follows from the last inequality together with Lemma 3.2, by choosing c large enough. \square

Lemma 3.3. *There exist $c \in (0, \infty)$ and $b > 1$, such that for large L ,*

$$\mathbf{P}(\exists u \in [0, L] \forall v \in [L] \text{ with } u \leq v : \max_{w \in [u, v]} |S(w) - S(u)| > c \log L) \leq e^{-L^b}.$$

Proof. Let c be as in the statement of Lemma 3.1. We say that a vertex u is H -bad if for all $v \in [H]_u$ there exists $w \in [u, v]$, such that $|S(w) - S(u)| > c \log L$. Note that if u is H -bad, then it is K -bad for every $K > H$. A simple union bound gives:

$$\mathbf{P}(\exists u \in [0, L] : u \text{ is } (L - |u|)\text{-bad}) \leq \mathbf{P}(\exists u \in [0, L] : u \text{ is } L\text{-bad}) \leq (d-1)^{L+1} \mathbf{P}(\rho \text{ is } L\text{-bad}).$$

The statement then follows from Lemma 3.1. \square

4 Regeneration times

In this section, we establish a regeneration structure for the Metropolis algorithm, which will permit us to prove Theorem 1.1 from the previously established Theorem 2.1. Recall the definition of the Metropolis algorithm (V_n) from Section 1.1, which depends on a parameter $\beta \in \mathbb{R}$. Define the *level regeneration* times $(\tau_n)_{n \geq 0}$ by $\tau_0 = 0$ and τ_{n+1} to be the first time after τ_n where the chain $(V_n)_{n \geq 0}$ hits a level L for the first time, then immediately jumps to level $L + 1$ and never gets back to level L again.

As in Sections 2 and 3, we denote the law of the branching random walk by \mathbf{P} , which is a law on Ω . We further denote the (quenched) law of the Metropolis algorithm $(V_n)_{n \geq 0}$ started from the vertex v and given the branching random walk ω (the *environment*) by $P_{\omega, \beta}^v$. The annealed law is denoted by $\mathbb{P}_\beta^v(d\omega, dV) = \mathbf{P}(d\omega)P_{\omega, \beta}^v(dV)$. We also set $\mathbb{P}_\beta = \mathbb{P}_\beta^\rho$, and note that this agrees with earlier notation. Our goal is to show:

Proposition 4.1. *For each $K > 0$, there exists $a = a(K) > 0$ and $n_a = n_a(K) > 0$ such that for all $n > n_a$ and $|\beta| \leq K$, $\mathbb{P}_\beta(\tau_1 > n) \leq e^{-n^a}$ and $\mathbb{P}_\beta(\tau_2 - \tau_1 > n) \leq e^{-n^a}$.*

The main point in Proposition 4.1 is in *uniformity* (in β) of the tail bounds for the regeneration times. This uniformity is in sharp contrast to other settings discussed in the literature, where the regeneration times usually blow up when the parameter approaches the critical value [BHOZ11, GHS11]. We remark that we actually only need that $\mathbb{E}_\beta[\tau_1^k]$ is uniformly bounded for β in a neighborhood of 0, for some $k > 2$.

In order to prove Proposition 4.1, we will make use of the relation between the Markov chain $(V_n)_{n \geq 0}$ and electrical networks [LP92]: Let $N(v)$ be the set of neighbors of v including v . For $w \in N(v)$, set

$$\begin{aligned} Q(v, w) &= \frac{P_\beta(v, w)}{P_\beta(v, \vec{v})} & Q(v) &= \sum_{w \in N(v)} Q(v, w) \\ C(v, w) &= Q(v, w) \prod_{u \leq v} Q(\vec{u}, u) & C(v) &= \sum_{w \in N(v)} C(v, w). \end{aligned}$$

One checks that for every $w \in N(v)$, $C(v, w) = C(w, v)$ and that

$$C(v, w)/C(v) = Q(v, w)/Q(v) = P_\beta(v, w),$$

whence the Markov chain $(V_n)_{n \geq 0}$ has an interpretation as the random walk on the rooted d -regular tree with loops, induced by the edge conductances $C(v, w)$. By assumption (H3), one has for $u \leq v$,

$$\frac{C(\vec{v}, v)}{C(\vec{u}, u)} = \frac{h(e^{(\beta_0 + \beta)X(v)})}{h(e^{(\beta_0 + \beta)X(u)})} e^{(\beta_0 + \beta)(S(\vec{v}) - S(u))}. \quad (4.1)$$

By assumptions (XS) and (H1-3), this implies the existence of a constant $c > 0$, such that

$$c e^{(\beta_0 + \beta)(S(\vec{v}) - S(u))} < \frac{C(\vec{v}, v)}{C(\vec{u}, u)} < c^{-1} e^{(\beta_0 + \beta)(S(\vec{v}) - S(u))}. \quad (4.2)$$

Define T_L to be the first strictly positive time the chain $(V_n)_{n \geq 0}$ hits the level L . Furthermore, denote by T_u and T_u^* , respectively, the first non-negative and strictly positive times the chain hits a vertex u .

The first lemma gives a uniform bound on the annealed probability that the Metropolis algorithm started from a vertex v escapes to infinity without coming back to its parent \vec{v} . It was essentially already observed by Aldous [Ald98, Lemma 8].

Lemma 4.2. *For each $K > 0$, there exists $c = c(K) > 0$, such that for each vertex $v \neq \rho$ and for all $|\beta| \leq K$, we have*

$$\mathbf{E}[P_{\omega,\beta}^v(T_{\vec{v}} = \infty)] > c.$$

Proof. Fix $v \neq \rho$ and define $f(\beta) := \mathbf{E}[P_{\omega,\beta}^v(T_{\vec{v}} = \infty)]$. Note that f does not depend on v by the definition of the measure \mathbf{P} . As mentioned in Section 1.1, under assumptions (XM) and (XR), there exists almost surely two infinite rays v_0, \dots, v_n, \dots and w_0, \dots, w_n, \dots with $\liminf_{n \rightarrow \infty} S(v_n)/n > 0$ and $\limsup_{n \rightarrow \infty} S(w_n)/n < 0$. By (4.2), the former has finite resistance if $\beta > -\beta_0$, and the latter if $\beta < -\beta_0$, whence $f(\beta) > 0$ for each $\beta \neq -\beta_0$. If $\beta = -\beta_0$, the Metropolis algorithm is just a simple random walk on the d -regular tree and therefore $f(-\beta_0) > 0$ as well. It follows that f is positive for every $\beta \in \mathbb{R}$. Furthermore, $f(\beta)$ is continuous, because it is the decreasing limit as $L \rightarrow \infty$ of $\mathbf{E}[P_{\omega,\beta}^v(T_L < T_{\vec{v}})]$ and each of these quantities depends only on a finite portion of the tree and is therefore continuous in β by assumption (H2). This immediately implies the lemma. \square

The following important lemma controls quenched hitting probabilities and will be used in the evaluation of quenched escape probabilities.

Lemma 4.3. *For each $K > 0$, there exist $c, L_0 > 0$, $b > 1$ depending on K , such that for $|\beta| \leq K$ and $L > L_0$,*

$$\mathbf{P}(\exists v \in [1, L-1] : P_{\omega,\beta}^v(T_L < T_{\vec{v}}) < L^{-c}) < e^{-L^b}.$$

Proof. Let $v \in [1, L-1]$ and let $u \in [L-|v|]_v$, such that $|u| = L$. By (4.2), we have for a fixed environment ω , for some $c > 0$,

$$P_{\omega,\beta}^v(T_u < T_{\vec{v}}) = \left(\sum_{v \leq w \leq u} \frac{C(\vec{v}, v)}{C(\vec{w}, w)} \right)^{-1} > c \left(\sum_{v \leq w < u} e^{(\beta_0 + \beta)(S(w) - S(v))} \right)^{-1}.$$

This gives for $|\beta| \leq K$,

$$P_{\omega,\beta}^v(T_L < T_{\vec{v}}) \geq \max_{u \in [L-|v|]_v} P_{\omega,\beta}^v(T_u < T_{\vec{v}}) > \frac{c}{L} \max_{u \in [L-|v|]_v} \min_{w \in [v, u]} e^{-(K + \beta_0)|S(w) - S(v)|}.$$

The statement now follows from Lemma 3.3. \square

For a vertex v , denote by $\ell(v)$ the depth of the first excursion below v after T_v , i.e.

$$\ell(v) = \sup \{ |V_n| - |v| : n \geq T_v \text{ and } |V_k| > |v| \forall k \in \{T_v + 1, \dots, n\} \} \in \mathbb{N} \cup \{\infty\}. \quad (4.3)$$

Note that since the probability of jumping from v to one of its children does not involve $X(\vec{v}, v)$, the event $\ell(v) > 0$ is independent from $X(\vec{v}, v)$ (conditioned on $T_v < \infty$).

Lemma 4.4. *For each $K > 0$ there exist $\alpha = \alpha(K)$, $L_0 = L_0(K) > 0$, such that for $|\beta| \leq K$ and $L > L_0$,*

$$\mathbf{P}(P_{\omega,\beta}(L \leq \ell(\rho) < \infty) > e^{-\alpha L}) < e^{-\alpha L}.$$

Proof. Fix $\varepsilon \in (0, 1)$. For a vertex v , define the variable $A(v)$ by $A(v) = 1$ if for one of v 's sisters \bar{v} , one has $P_{\omega, \beta}^{\bar{v}}(T_{\bar{v}}^* = \infty, T_{\bar{v}} = \infty) \geq \varepsilon$ and $A(v) = 0$ otherwise. In words, $A(v) = 1$ if the (quenched) probability of a walk, started at an appropriate sister \bar{v} of v , to escape to infinity through the subtree rooted at \bar{v} without visiting again \bar{v} is at least ε . By Lemma 4.2, we can choose ε such that $\mathbf{E}[A(v)] > 1/2$ for all $|\beta| \leq K$. By a result due to Grimmett and Kesten (see [DGPZ02, Lemma 2.2, (2.1)] for this version), there exist then $\alpha, \gamma > 0$, such that $\mathbb{P}_\beta(G_L) \geq 1 - e^{-\alpha L}$ for large L , where

$$G_L = \left\{ \min_{v \in [L]} \sum_{w \in [2, v]} A(w) \geq \gamma L \right\}.$$

Now, let $\omega \in G_L$. We wish to bound $P_{\omega, \beta}(L \leq \ell(\rho) < \infty)$. For this, define T^m for $m = 2, \dots, L - 1$ to be the first time after T_L that the Markov chain $(V_n)_{n \geq 0}$ hits level m . If $T^m < \infty$ and $A(V_{T^m}) = 1$, then by assumptions (XS) and (H1-3), the probability that from V_{T^m} the chain reaches \bar{V}_{T^m} after two steps is bounded from below by δ/ε for some δ sufficiently small. It follows that for $\omega \in G_L$,

$$P_{\omega, \beta}(L \leq \ell(\rho) < \infty) \leq (1 - \delta)^{\sum_{w \in [2, V_T]} A(w)} \leq (1 - \delta)^{\gamma L}.$$

This yields the lemma (reducing the value of α if necessary). \square

Lemma 4.5. *For each $K > 0$ there exist $\alpha = \alpha(K), L_0 = L_0(K) > 0$, such that for $|\beta| \leq K$ and $L > L_0$,*

$$\mathbb{P}_\beta(|V_{\tau_1}| \geq L) < e^{-\alpha L}.$$

Proof. Define a sequence of random numbers L_0, L_1, \dots recursively as follows:

- $L_0 = 1$;
- for $n \in \mathbb{N}$, let $v_n = V_{T_{L_n}}$. If $\ell(v_n) < \infty$, then $L_{n+1} = L_n + \ell(v_n) + 1$;
- otherwise, set $L_m = \infty$ for $m > n$.

Let N be the largest number n , such that $L_n < \infty$. Then by construction, $|V_{\tau_1}| = L_N$. Furthermore, the differences $(L_{n+1} - L_n)_{0 \leq n < N}$ are independent and identically distributed as $\ell + 1$ conditioned on $\ell < \infty$ and N is geometrically distributed with success probability $\mathbf{P}(\ell = \infty) > 0$ (ℓ as in (4.3)). The lemma then follows from Lemma 4.4. \square

Lemma 4.6. *Let \mathcal{G}_n be the σ -field generated by V_0, \dots, V_n and let T be a stopping time with respect to the filtration $(\mathcal{G}_n)_{n \geq 0}$, such that $V_T \neq V_k$ for all $k < T$. Then for each $K > 0$ there exists a constant $c = c(K) > 0$, such that for $|\beta| \leq K$ and all $N \geq 0$, we have*

$$\mathbb{P}_\beta \left(\max_{T \leq j < T+N} |V_j| \geq cN \mid \mathcal{G}_T \right) > c.$$

Proof. We follow the proof of [Aid08, Theorem 1.5]. Throughout the proof, c_0, c_1, \dots will denote positive constants which are uniform in $|\beta| \leq K$.

Step 1. For a vertex $v \neq \rho$, define $\pi_{\omega, \beta}(v) = P_{\omega, \beta}^v(T_{\bar{v}} = \infty)$. Note that the random variables $\pi_{\omega, \beta}(v)$, $v \neq \rho$, are identically distributed under \mathbf{P} (but not independent). Let $\pi_{\omega, \beta}$ denote a random variable with this law. We wish to show that for some constant c_0 ,

$$\mathbf{E}[1/\pi_{\omega, \beta}] \leq c_0. \tag{4.4}$$

Denote by v_1, \dots, v_{d-1} the children of the vertex v . By assumptions (XS) and (H1-3), we have

$$P_{\omega, \beta}^v(V_1 = v_i) \geq c_1, \quad i = 1, \dots, d-1, \quad (4.5)$$

which yields $\pi_{\omega, \beta}(v) \geq c_1 \sum_{i=1}^{d-1} \pi_{\omega, \beta}(v_i)$. Now, note that the variables $\pi_{\omega, \beta}(v_i)$, $i = 1, \dots, d-1$ are independent under \mathbf{P} . The previous inequality then yields for every $x \geq 0$,

$$\mathbf{P}(\pi_{\omega, \beta}(v) \leq x) \leq \mathbf{P}\left(\max_{i=1, \dots, d-1} \pi_{\omega, \beta}(v_i) \leq x/c_1\right) = \mathbf{P}(\pi_{\omega, \beta}(v) \leq x/c_1)^{d-1}. \quad (4.6)$$

Furthermore, by Lemma 4.2, there exists a constant c_2 , such that

$$\mathbf{P}(\pi_{\omega, \beta}(v) \leq 2c_2) \leq 1/2. \quad (4.7)$$

Together with (4.6), this now easily implies (4.4).

Step 2. For a vertex v , let N_v denote the number of times the vertex v has been visited by the Metropolis algorithm $(V_n)_{n \geq 0}$. We wish to show that for each $k \geq 0$:

$$\mathbb{E}_\beta \left[\sum_{|v|=k} N_v \right] \leq c_3. \quad (4.8)$$

Recall that T_v^* denotes the first strictly positive hitting time of the vertex v , such that $E_{\omega, \beta}^v[N_v] = 1/P_{\omega, \beta}^v(T_v^* = \infty)$. By (4.5), we have $P_{\omega, \beta}^v(T_v^* = \infty) \geq c_1 \pi_{\omega, \beta}(v_1)$, such that

$$\mathbb{E}_\beta[N_v] = \mathbf{E} \left[P_{\omega, \beta}^v(T_v < \infty) E_{\omega, \beta}^v[N_v] \right] \leq c_1 \mathbb{P}_\beta(T_v < \infty) \mathbf{E}[1/\pi_{\omega, \beta}(v_1)] \leq c_4 \mathbb{P}_\beta(T_v < \infty), \quad (4.9)$$

by (4.4). Furthermore, we have for every $k \geq 0$,

$$1 \geq \sum_{|v|=k} \mathbb{P}_\beta(T_v < \infty, V_n \geq v_1 \forall n > T_v) \geq c_1 \sum_{|v|=k} \mathbb{P}_\beta(T_v < \infty) \mathbf{E}[\pi_{\omega, \beta}(v_1)], \quad (4.10)$$

and by (4.7), we have $\mathbf{E}[\pi_{\omega, \beta}(v_1)] = \mathbf{E}[\pi_{\omega, \beta}] \geq c_2$. Equations (4.10) and (4.9) now yield (4.8).

Step 3. Recall the notation in the statement of the lemma and let $L \in \mathbb{N}$. Define the event $E_T = \{V_n \geq V_T \forall n > T\}$. A straightforward extension of the proof of the last step permits to prove that for every constant $C > 0$,

$$\mathbb{P}_\beta \left(E_T, \sum_{w \in [V_T, L]} N_w > CL \mid \mathcal{G}_T \right) \leq \frac{1}{CL} \mathbb{E}_\beta \left[\sum_{w \in [V_T, L]} N_w \mathbb{1}_{E_T} \mid \mathcal{G}_T \right] < c_3 C^{-1}.$$

Furthermore, by (4.5), we have $\mathbb{P}_\beta(E_T \mid \mathcal{G}_T) \geq c_1 \mathbf{E}[\pi_{\omega, \beta}] \geq c_1 c_2$. This now yields for every constant $c > 0$ and every $N \in \mathbb{N}$,

$$\mathbb{P}_\beta \left(\max_{T \leq j < T+N} |V_j| \geq cN \mid \mathcal{G}_T \right) \geq \mathbb{P}_\beta \left(E_T, \sum_{w \in [V_T, cN]} N_w \leq N \mid \mathcal{G}_T \right) > c_1 c_2 - c_3 c.$$

Setting $c = c_1 c_2 / 2c_3$ yields the proof. \square

Proof of Proposition 4.1. The proof follows an argument in the spirit of [PZ08, Proposition 3]. Let $K > 0$. Let $L = L(n)$ go to infinity with n (we will later choose $L = n^b$ for some constant b). We have, with T_L denoting the hitting time of level L , and with $\alpha > 0$ as in the statement of Lemma 4.5,

$$\mathbb{P}_\beta(\tau_1 > n) \leq \mathbb{P}_\beta(T_{|V_{\tau_1}|} > n, |V_{\tau_1}| < L) + \mathbb{P}_\beta(|V_{\tau_1}| \geq L) \leq \mathbb{P}_\beta(T_L > n) + e^{-\alpha L}, \quad (4.11)$$

where in the last inequality we used the fact that $T_{L'} \leq T_L$ for $L' \leq L$.

Let c be the constant from Lemma 4.3 and set $\bar{c} = 6(c \vee 1)$. Throughout the proof, all constants will be uniform in β for $|\beta| \leq K$. We write $\bar{L} = L^{\bar{c}}$. Define a vertex v to be *fresh* if it is visited by the random walk for the first time before time \bar{L} . We will upper bound $\mathbb{P}_\beta(T_L > n)$ by showing that on the one hand, there cannot be too few fresh points that are well separated and on the other hand, if there are many such fresh points, it is unlikely that T_L is large.

For a vertex v , let N_v denote the number of visits to v by time \bar{L} and let N_v^u denote the number of times the random walk visits v before time \bar{L} and then, at the next step that it moves, it visits the ancestor of v before time \bar{L} . Clearly, for each v , $N_v^u \leq N_v$, while, using the Markov property, there exist constants $\gamma, \gamma' > 0$ such that

$$\mathbb{P}_\beta(\exists v : N_v \geq \bar{L}^{1/2}, N_v^u \leq \gamma \bar{L}^{1/2}) \leq \bar{L} e^{-\gamma' \bar{L}^{1/2}}. \quad (4.12)$$

On the other hand, the event

$$\{N_v \geq \bar{L}^{1/2}, N_v^u > \gamma \bar{L}^{1/2}, T_L \geq \bar{L}\}$$

implies that the random walk visited v and then hit \bar{v} at least $\gamma \bar{L}^{1/2}$ times. By Lemma 4.3 and the Markov property, the probability that there exists a fresh vertex v satisfying the last event is bounded above by

$$\bar{L}(1 - L^{-c})\gamma \bar{L}^{1/2} \leq \bar{L} e^{-\gamma L^{\bar{c}/2 - c}} \leq \bar{L} e^{-\gamma L^{2(c \vee 1)}}.$$

Combining this with (4.12), we conclude that for all large L ,

$$\mathbb{P}_\beta(\exists v : N_v \geq \bar{L}^{1/2}, T_L \geq \bar{L}) \leq e^{-L}. \quad (4.13)$$

On the event $\{\forall v : N_v < \bar{L}^{1/2}\}$, there are at least $\bar{L}^{1/2}$ fresh points, and therefore there are at least $\bar{L}^{1/4}$ fresh points that are $\bar{L}^{1/4}$ -separated. Let c be the constant from Lemma 4.6. At each arrival to such a fresh point, with (annealed) probability at least c the walk hits level L before time $L/c < \bar{L}^{1/4}$, for large L . It follows that

$$\mathbb{P}_\beta(\forall v : N_v < \bar{L}^{1/2}, T_L \geq \bar{L}) \leq C_0 \bar{L}^{1/4}, \quad (4.14)$$

for some $C_0 \in (0, 1)$. By (4.13) and (4.14), there exists now a constant $b > 0$, such that with $L = n^b$, we have $\mathbb{P}_\beta(T_L > n) \leq 2e^{-L}$. Together with (4.11), this completes the proof of the first statement of Proposition 4.1.

As for the second statement, let $\rho_1, \dots, \rho_{d-1}$ be the children of the root. The law of the subtree of the branching random walk tree rooted at V_{τ_1} is equal in law under \mathbb{P}_β to the subtree rooted at V_1 , conditioned on $V_1 \in \{\rho_1, \dots, \rho_{d-2}\}$ and on $T_\rho^* = \infty$. In particular,

$$\mathbb{P}_\beta(\tau_2 - \tau_1 \geq n) = \mathbb{P}_\beta(\tau_1 \geq n | V_1 \in \{\rho_1, \dots, \rho_{d-2}\}, T_\rho^* = \infty) \leq \mathbb{P}_\beta(\tau_1 \geq n) / c_1 c_2,$$

where c_1 and c_2 are the constants from the proof of Lemma 4.6. This finishes the proof. \square

Proof of Theorem 1.1. Existence and finiteness of the limit $\sigma^2 = \lim_{n \rightarrow \infty} S_n^2/n$ follows from the first part of Theorem 2.1. In order to prove the remaining statements, we will use the regeneration structure established in this section. Note that the random vectors $\{(\tau_{i+1} - \tau_i, S_{\tau_{i+1}} - S_{\tau_i})\}_{i \geq 1}$ are i.i.d. under the law \mathbb{P}_β and independent from (τ_1, S_{τ_1}) . Furthermore, by (XS), $|S_n - S_m| \leq g|n - m|$, where $g = \text{esssup } |X|$. Proposition 4.1 and the \mathbb{P} -almost sure convergence of S_n^2/n to σ^2 now yields

$$\sigma^2 = \lim_{k \rightarrow \infty} \frac{S_{\tau_k}^2}{k} \cdot \frac{k}{\tau_k} = \frac{\mathbb{E}[(S_{\tau_2} - S_{\tau_1})^2]}{\mathbb{E}[\tau_2 - \tau_1]} > 0,$$

which proves the first part of the theorem.

As for the second part, by standard arguments (see e.g. the proof of Theorem 4.1 in [LPP96]), $v_\beta = \lim_{n \rightarrow \infty} S_n/n$ exists almost surely if $\mathbb{E}_\beta[\tau_2 - \tau_1] < \infty$, which is the case for all β by Proposition 4.1. Furthermore, we have $v_\beta = \mathbb{E}_\beta[S_{\tau_2} - S_{\tau_1}]/\mathbb{E}_\beta[\tau_2 - \tau_1]$, which implies in particular that $|v_\beta| \leq g$.

Now let $K_\beta = \inf\{k > 0 : \tau_k > \beta^{-2}\}$. By the optional stopping theorem, we have

$$\mathbb{E}_\beta[S_{\tau_{K_\beta}}] = \mathbb{E}_\beta[S_{\tau_1}] + v_\beta \mathbb{E}_\beta[\tau_{K_\beta} - \tau_1],$$

such that by Proposition 4.1 and assumption (XS), for some constant $C > 0$,

$$|\mathbb{E}_\beta[S_{\lfloor \beta^{-2} \rfloor}] - v_\beta \beta^{-2}| \leq C \mathbb{E}_\beta[\tau_{K_\beta} - \beta^{-2}].$$

Crude moment bounds using Proposition 4.1 yield that the right-hand side of the above equation is $o(\beta^{-1})$, which yields $\lim_{\beta \rightarrow 0} v_\beta/\beta = \lim_{\beta \rightarrow 0} \mathbb{E}_\beta[S_{\lfloor \beta^{-2} \rfloor}] = \sigma^2/2$ by Theorem 2.1. This finishes the proof. \square

References

- [Aid08] E. Aidékon. Transient random walks in random environment on a Galton–Watson tree. *Probability Theory and Related Fields*, 142(3-4):525–559, November 2008.
- [Ald98] D. Aldous. A Metropolis-type optimization algorithm on the infinite tree. *Algorithmica*, 22(4):388–412, December 1998.
- [Azu67] K. Azuma. Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal*, 19(3):357–367, 1967.
- [BHOZ11] G. Ben Arous, Y. Hu, S. Olla, and O. Zeitouni. Einstein relation for biased random walk on Galton–Watson trees. *arXiv:1106.4387*, 2011.
- [Big77] J. D. Biggins. Chernoff’s theorem in the branching random walk. *Journal of Applied Probability*, 14(3):630, September 1977.
- [DFGW89] A. De Masi, P. A. Ferrari, S. Goldstein, and W. D. Wick. An invariance principle for reversible Markov processes. Applications to random motions in random environments. *Journal of Statistical Physics*, 55(3-4):787–855, May 1989.
- [DGPZ02] A. Dembo, N. Gantert, Y. Peres, and O. Zeitouni. Large deviations for random walks on Galton–Watson trees: averaging and uncertainty. *Probability Theory and Related Fields*, 122(2):241–288, February 2002.

- [Dur96] R. Durrett. *Probability: theory and examples*. Duxbury Press, Belmont, California, second edition, 1996.
- [DZ98] A. Dembo and O. Zeitouni. *Large deviation techniques and applications*. Springer, New–York, second edition, 1998.
- [GHS11] N. Gantert, Y. Hu, and Z. Shi. Asymptotics for the survival probability in a killed branching random walk. *Annales de l’Institut Henri Poincaré (B) Probabilités et Statistiques*, 47(1):111–129, February 2011.
- [GMP12] N. Gantert, P. Mathieu, and A. Piatnitski. Einstein relation for reversible diffusions in a random environment. *Communications on Pure and Applied Mathematics*, 65(2):187–228, 2012.
- [Guo12] X. Guo. Einstein relation for random walks in balanced random environment. *arXiv:1212.0255*, 2012.
- [KLO12] T. Komorowski, C. Landim, and S. Olla. *Fluctuations in Markov processes*, volume 345 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Heidelberg, 2012.
- [KO05a] T. Komorowski and S. Olla. Einstein relation for random walks in random environments. *Stochastic Processes and their Applications*, 115(8):1279–1301, August 2005.
- [KO05b] T. Komorowski and S. Olla. On mobility and Einstein relation for tracers in time–mixing random environments. *Journal of Statistical Physics*, 118(3/4):407–435, 2005.
- [KV86] C. Kipnis and S. R. S. Varadhan. Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Communications in Mathematical Physics*, 104(1):1–19, March 1986.
- [Lou02] M. Loulakis. Einstein relation for a tagged particle in simple exclusion processes. *Communications in Mathematical Physics*, 229:347–367, 2002.
- [LP92] R. Lyons and R. Pemantle. Random Walk in a Random Environment and First-Passage Percolation on Trees. *The Annals of Probability*, 20(1):125–136, January 1992.
- [LPP96] R. Lyons, R. Pemantle, and Y. Peres. Biased random walks on Galton–Watson trees. *Probability Theory and Related Fields*, 106(2):249–264, October 1996.
- [LR94] J. L. Lebowitz and H. Rost. The Einstein relation for the displacement of a test particle in a random environment. *Stochastic Processes and their Applications*, 54(2):183–196, December 1994.
- [Mog74] A. A. Mogul’skii. Small deviations in a space of trajectories. *Theory of Probability and its Applications*, 19:726–736, 1974.
- [PZ08] Y. Peres and O. Zeitouni. A central limit theorem for biased random walks on Galton–Watson trees. *Probability Theory and Related Fields*, 140(3-4):595–629, March 2008.

- [Zei04] O. Zeitouni. Random walks in random environment. In *Lectures on probability theory and statistics*, volume 1837 of *Lecture Notes in Mathematics*, pages 189–312. Springer, Berlin, 2004.