On the First Passage Times of Branching Random Walks in $\mathbb{R}^d$

Jose Blanchet[*], Wei Cai[†], Shaswat Mohanty[‡], and Zhenyuan Zhang[§]

16th April 2024

Abstract

We study the first passage times of discrete-time branching random walks in $\mathbb{R}^d$ where $d \geqslant 1$. Here, the genealogy of the particles follows a supercritical Galton-Watson process. We provide asymptotics of the first passage times to a ball of radius one with a distance x from the origin, conditioned upon survival. We provide explicitly the linear dominating term and the logarithmic correction term as a function of x. The asymptotics are precise up to an order of $o_{\mathbb{P}}(\log x)$ for general jump distributions and up to $O_{\mathbb{P}}(\log \log x)$ for spherically symmetric jumps. A crucial ingredient of both results is the tightness of first passage times. We also discuss an extension of the first passage time analysis to a modified branching random walk model that has been proven to successfully capture shortest path statistics in polymer networks.

1 Introduction

We consider a discrete-time branching random walk (BRW) on $\mathbb{R}^d$, where particles in $\mathbb{R}^d$ perform independent random walks until they randomly branch into more particles or die. Spatial branching processes, including BRW and its continuous-time sibling branching Brownian motion (BBM), have a long history and a wide range of applications in ecology, population biology, and modeling epidemics ([25, 28, 34, 35, 36, 49]). The study of the extremal behavior of spatial branching processes has gained increasing attention over recent years. The mathematical techniques applied in those studies are related to developments in numerous other fields, e.g., discrete Gaussian free field [14, 24], random unitary matrices [3], Riemann zeta function [4], random multiplicative functions [31], spin glasses [12], among many others that exhibit a log-correlated structure. We refer to [46, 52] for lecture notes on BRW and related topics.

The study of the extrema for one-dimensional BBM traces back to [17], which established a precise asymptotic for the maximum as a function of time. For general branching random walks, [1] provided a precise asymptotic in terms of the jump distribution of the process. Finer behavior near the frontier is also well understood since the works of [2, 13, 38, 39]. Nevertheless, less is known in dimensions greater than one. To name a few examples, there are studies of the maximum norm for BBM ([7, 32, 33, 40, 47]) and very recently of BRW ([8]), as well as the range of spatial branching processes ([37, 53]). These works in higher dimensions typically assumed that the increment distribution is spherically symmetric. A notable exception is [48], which provided estimates on the number of particles in a linearly shifted ball in time for a continuous-time version of the BRW.

[*]Department of Management Science and Engineering, Stanford University. Email: jose.blanchet@stanford.edu

[†]Department of Mechanical Engineering, Stanford University. Email: caiwei@stanford.edu

[‡]Department of Mechanical Engineering, Stanford University. Email: shaswatm@stanford.edu

[§]Department of Mathematics, Stanford University. Email: zzy@stanford.edu

Our primary focus is on the first passage times (FPT) for branching random walks in $\mathbb{R}^d$. By definition, the FPT is the first time for particles to reach a prescribed region. For BBM, [54] established a precise asymptotic formula (up to $O_{\mathbb{P}}(1)$) for the FPT; see (62) below. Our interest in estimating the FPT of spatial branching processes stems from their relevance to polymer physics. The mechanical behavior of polymer networks appears to be controlled by the shortest paths between distant nodes in the polymer network, as shown by recent works [50, 51] using coarse-grained molecular dynamics (CGMD) simulations. The shortest path length (SP) serves as an indicator of how stretched the average load-bearing chains are in the polymer network. Theories based on SP go beyond traditional polymer models focused on elastic behaviors and aim to explain the inelastic behaviors exhibited by such materials.

The SP between (fixed) distant nodes is equivalent to the FPT of the BRW beginning from the origin (source node) and terminating at a desired point in space (destination node). Moreover, the occurrence of the cross-links along the polymer backbone mimics a Poisson process. As a result, a cross-link between two chains is equivalent to a branching event observed in the BRW genealogy tree. The agreement of the mean FPT from the BRW theory with the mean SP from the CGMD simulation results has been shown in our recent work [54]. Furthermore, the numerical implementation of the BRW model shows an excellent agreement between its FPT distribution and the SP distribution from the CGMD simulation. Our earlier work initially focused on constant length jumps with the BRW model where there existed a correlation between the components of a jump step in the d dimensions by virtue of being sampled from the unit sphere $\mathbb{S}^{d-1} = \{ \mathbf{x} \in \mathbb{R}^d \mid \|\mathbf{x}\| = 1 \}$. This was later shown to be represented equally well by the Gaussian BRW and BBM in which the components of each jump step are independent in the d dimensions. However, the analysis was restricted to spherically symmetric jumps to accurately model the as-prepared polymer network prior to deformation.

In this paper, we provide FPT asymptotics for BRW that may not have a spherically symmetric jump distribution (while the case of spherically symmetric jumps is of special interest for our applications). The main results are explained in Section 1.2 and their proofs are presented in Section 2. Applications to polymer physics will be detailed in Section 3.1.

1.1 Notation and assumptions

Let us start with a formal definition of the branching random walk (BRW). We consider a discrete-time BRW model with offspring distribution $\{p_i\}_{i \geq 0}$. This means that at each time step, each particle independently reproduces i particles at the same location with probability p_i. The case of $i = 0$ corresponds to the particle being terminated. Let $\rho = \sum_i i\, p_i$ be the mean of the offspring distribution. We will always assume the supercritical case of $\rho > 1$. The d-dimensional random walk increment is given by $\boldsymbol{\xi}$, and the first particle starts from the origin $\mathbf{0} \in \mathbb{R}^d$. Typically, in this paper, a bold symbol refers to a vector.

Our goal is to study the asymptotic of the FPTs of the BRW. For $x \in \mathbb{R}$, we let B_x denote a ball of radius one[1] centered at $(x, 0, \ldots, 0)$ in $\mathbb{R}^d$. We let V_n denote the collection (i.e. set) of particle locations at time step n, and $\{\boldsymbol{\eta}_{v,n}(k)\}_{0 \leq k \leq n}$ denotes the d-dimensional random walk that leads to $v \in V_n$. We define the FPT τ_x of the BRW to B_x, that is,

$$\tau_x = \min\{ n \geq 0 : \exists\, v \in V_n,\ \boldsymbol{\eta}_{v,n}(n) \in B_x \}.$$

In what follows, we will separately analyze the cases depending on whether the law of $\boldsymbol{\xi}$ is spherically symmetric or not.[2] In both cases, we impose the following assumptions on the BRW:

(A1) the offspring distribution has a finite second moment, i.e., $\sum_i i^2 p_i < \infty$;

(A2) the law of $\boldsymbol{\xi}$ is integrable and centered, i.e., $\mathbb{E}[\boldsymbol{\xi}] = \mathbf{0}$.

---

[1]Replacing the radius one by any fixed positive constant will not change the analysis; the same proof goes through.
[2]A probability distribution on $\mathbb{R}^d$ is *spherically symmetric* if it is invariant under any orthonormal transformation. Equivalent definitions can be found in Theorem 2.5 of [26].

For a stochastic process $\{T_x\}_{x \geqslant 0}$ and a positive deterministic sequence $\{t_x\}_{x \geqslant 0}$, we write $T_x = O_{\mathbb{P}}(t_x)$ if the collection $\{T_x/t_x\}_{x \geqslant 0}$ is tight, and $T_x = o_{\mathbb{P}}(t_x)$ if $T_x/t_x \to 0$ in probability.

1.1.1 Spherically symmetric jumps

Denote the first coordinate of $\boldsymbol{\xi}$ by ξ, which is a real-valued random variable. We introduce the large deviation rate function

$$I(x) := \sup_{\lambda > 0} \Big(\lambda x - \log \phi_\xi(\lambda) \Big), \tag{1}$$

where $\phi_\xi(\lambda) := \mathbb{E}[e^{\lambda \xi}]$ is the moment generating function for ξ. In this case, we make the following assumptions on $\boldsymbol{\xi}$:

(A3) the law of $\boldsymbol{\xi}$ is spherically symmetric in $\mathbb{R}^d$, and its radial component R satisfies $\mathbb{P}(R = 0) < 1$;

(A4) $\log \rho \in (\mathrm{ran}I)^\circ$, where $(\mathrm{ran}I)^\circ$ is the interior of the range of I. In other words, there exists $c_1 > 0$ such that $I(c_1) = \log \rho$. It can be shown that $c_1 \in (\mathrm{ran}(\log \phi_\xi)')^\circ$.

Note that under assumptions (A2) and (A3), I is C^1 and convex in its domain, and attains a unique global minimum at $x = 0$.[3] We will also see from Lemma 24 below that (A3) implies that both $\boldsymbol{\xi}$ and ξ are non-lattice if $d \geqslant 2$.[4] A consequence of (A4) is that the number of particles forms a supercritical Galton-Watson process. Denote by $c_2 := I'(c_1)$ the value of λ at which the supremum in (1) is taken for $x = c_1$. In particular, (A4) implies that ϕ_ξ is well-defined in a neighborhood of c_2. The constants c_1 and c_2 here depend only on the offspring distribution (through ρ) and the increment distribution (through $\phi_\xi(\lambda)$) and will be fixed throughout this paper.

1.1.2 Non-spherically symmetric jumps

Let

$$I(\mathbf{x}) := \sup_{\boldsymbol{\lambda} \in \mathbb{R}^d} \Big(\boldsymbol{\lambda} \cdot \mathbf{x} - \log \phi_{\boldsymbol{\xi}}(\boldsymbol{\lambda}) \Big) = \sup_{\boldsymbol{\lambda} \in \mathbb{R}^d} \Big(\boldsymbol{\lambda} \cdot \mathbf{x} - \log \mathbb{E}[e^{\boldsymbol{\lambda} \cdot \boldsymbol{\xi}}] \Big) \tag{2}$$

denote the large deviation rate function for $\boldsymbol{\xi}$. We impose the following assumptions:

(A5) there exist $m \in \mathbb{N}$ and $r > 1$ such that $\boldsymbol{\xi}_1 + \cdots + \boldsymbol{\xi}_m$ has a density in $L^r(\mathbb{R}^d)$, where $\{\boldsymbol{\xi}_i\}_{1 \leqslant i \leqslant m}$ are i.i.d. copies of $\boldsymbol{\xi}$;

(A6) $\log \rho \in (\mathrm{ran}I(\cdot, \mathbf{0}))^\circ$, where $I(\cdot, \mathbf{0})$ refers to the function I with the last $d-1$ variables fixed at zero; let $\widehat{c}_1 > 0$ satisfy $I(\widehat{c}_1, \mathbf{0}) = \log \rho$, then $(\widehat{c}_1, \mathbf{0}) \in (\mathrm{ran}\nabla \log \phi_{\boldsymbol{\xi}})^\circ$.

Denote by $\mathbf{c}_2 = \nabla I((\widehat{c}_1, \mathbf{0}))$, which is the value of $\boldsymbol{\lambda}$ where the supremum (2) is attained at $\mathbf{x} = (\widehat{c}_1, \mathbf{0})$. The intuition behind the vector $\mathbf{c}_2$ is explained by Figure 1 below. The assumption (A5) is crucial for the non-degeneracy of the jumps. More precisely, (A5) implies that the support of $\boldsymbol{\xi}$ cannot be contained in a proper subspace of $\mathbb{R}^d$ and that $\boldsymbol{\xi}$ is non-lattice. Both conditions are necessary for the BRW to be able to reach the target ball B_x. A consequence of (A6) is that $\phi_{\boldsymbol{\xi}}$ is well-defined in a neighborhood of $\mathbf{c}_2$. We will see in Proposition 21 that (A3) and (A4) together imply (A6) with $\widehat{c}_1 = c_1$.

---

[3]See e.g., Theorem 26.3 of [44].
[4]We say the law of $\boldsymbol{\xi}$ is *non-lattice* if there is no $\mathbf{y} \in \mathbb{R}^d$ such that the support of $\boldsymbol{\xi} + \mathbf{y}$ is contained in a discrete subgroup of $\mathbb{R}^d$.
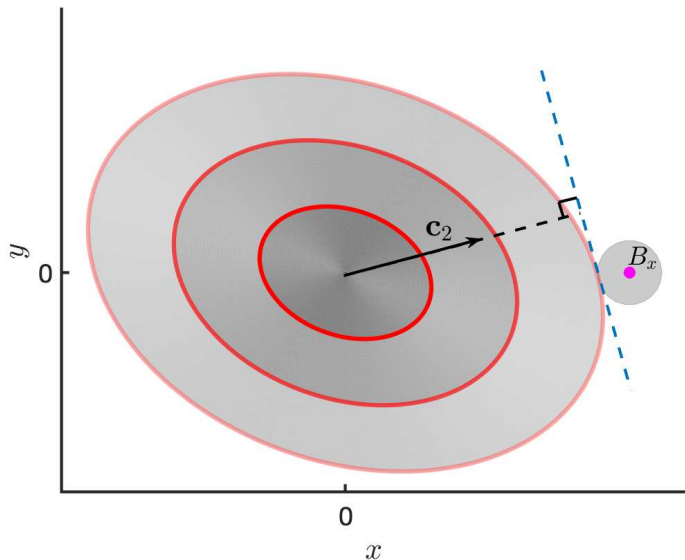
Figure 1: A two-dimensional visualization of the vector $\mathbf{c}_2$. The range of a (non-spherically symmetric) BRW roughly grows linearly in time (illustrated with the shaded ellipses); at time τ_x, the ellipse is nearly tangent to the target ball B_x (shaded disk). Roughly speaking, the vector $\mathbf{c}_2$ is normal to the tangent line.

1.2 Main results

In the following, we fix a dimension $d \geqslant 1$. Consider first the case of spherically symmetric jumps. Let us define

$$A(x) := \frac{x}{c_1} + \frac{d+2}{2c_2c_1} \log x. \tag{3}$$

For a finite set G, we use $\#G$ to denote its cardinality. Denote by $S := \{\forall n \geqslant 1, \ \#V_n > 0\}$ the *survival event* that the underlying branching process survives at all times. Since $\rho > 1$, we have $p := \mathbb{P}(S) > 0$.

Theorem 1. *Assume (A1)–(A4). Conditional upon survival, it holds that*

$$\tau_x = A(x) + O_{\mathbb{P}}(\log \log x) = \frac{x}{c_1} + \frac{d+2}{2c_2c_1} \log x + O_{\mathbb{P}}(\log \log x). \tag{4}$$

In other words, the collection $\{(\tau_x - A(x))/\log \log x\}_{x>0}$ is tight.

A natural question is whether the $O_{\mathbb{P}}(\log \log x)$ can be replaced by a $O_{\mathbb{P}}(1)$, which we later formulate as Conjecture 1. We discuss a few pieces of evidence that support this conjecture, and further aspects will be discussed in Section 5. First, this is the case for $d = 1$ (Proposition 7 below) and for the branching Brownian motion in $\mathbb{R}^d$ (Theorem 1 of [54]). A second piece of evidence is the following tightness result on first passage times that is also central to our proof of Theorem 1.

Theorem 2. *For $r \in (0,1)$ and $x > 0$, let $t_x^{(r)}$ be the r-th quantile of the first passage time τ_x, i.e., $t_x^{(r)} = \inf\{t : \mathbb{P}(\tau_x \leqslant t \mid S) \geqslant r\}$. Assume (A1), (A2), and (A6). There exist constants $C, c > 0$ independent of x such that for each $y \in [0, x]$,*

$$\mathbb{P}\left(|\tau_x - t_x^{(1/2)}| > y \mid S\right) \leqslant Ce^{-cy}. \tag{5}$$

In particular, conditioned upon survival, the collection of laws $\{\tau_x - t_x^{(r)}\}_{x>0}$ is tight for each $r \in (0, 1)$.

The analog of Theorem 2 for the maximum of BRW traces back to [21] and [41], which assumed uniformly bounded jumps. The case of unbounded jumps was later resolved by [15] and [16] using a different approach. Theorem 2 generalizes these results to the first passage times of BRW in arbitrary dimensions. In this setting, for the jumps, we only require mean zero and assumption (A6). Since (A3) and (A4) together imply (A6), Theorem 2 applies in the spherically symmetric case as well.

Extending these results to the case of non-spherically symmetric jumps, we define

$$\widehat{A}(x) := \frac{x}{\widehat{c}_1} + \frac{d+2}{2\,\widehat{c}_1 \partial_{x_1} I(\widehat{c}_1, \mathbf{0})} \, \log x, \tag{6}$$

where we recall the definition of $\widehat{c}_1$ from assumption (A6).

Theorem 3. *Assume (A1), (A2), (A5), and (A6). Conditional upon survival,*

$$\tau_x = \widehat{A}(x) + o_{\mathbb{P}}(\log x) = \frac{x}{\widehat{c}_1} + \frac{d+2}{2\widehat{c}_1 \partial_{x_1} I(\widehat{c}_1, \mathbf{0})} \, \log x + o_{\mathbb{P}}(\log x). \tag{7}$$

In other words, the collection $(\tau_x - \widehat{A}(x))/\log x$ converges to 0 in probability as $x \to \infty$.

There are certain instances for non-spherically symmetric jumps where the asymptotic (7) can be improved. Such a class includes product measures and non-degenerate centered elliptical distributions on $\mathbb{R}^d$.[5] These will be addressed in detail in Section 3.2.

An immediate corollary of Theorems 2 and 3 is the following strong law of large numbers for the first passage times.

Corollary 4. *Assume (A1), (A2), (A5), and (A6). Conditioned upon survival,*

$$\frac{\tau_x}{x} \to \frac{1}{\widehat{c}_1} \qquad a.s.$$

As yet another consequence of Theorems 2 and 3 (while slightly modifying their proofs), the full range of the BRW forms a dense subset of $\mathbb{R}^d$, complementing Corollary 2.5 of [43] that the full range of the BBM is dense.

Corollary 5. *Assume (A1), (A2), (A5), and (A6). Conditioned upon survival, the set*

$$R := \bigcup_{n \geqslant 0} \bigcup_{v \in V_n} \{\boldsymbol{\eta}_{v,n}(n)\}$$

is dense in $\mathbb{R}^d$ almost surely.

In Section 4, we test Theorems 1 and 3 numerically with a *path purging* algorithm that significantly improves the computational efficiency for capturing the extremal behavior of a BRW.

1.3 Proof strategy

In this subsection, we discuss the high-level intuition of the strategy we will follow, building from the case of $d = 1$. Throughout we condition on the survival event. We first assume the jump is

---

[5]We say that a centered random variable $\boldsymbol{\xi}$ follows a *non-degenerate elliptical distribution* if $\boldsymbol{\xi} \overset{\text{law}}{=} TU$ where U is spherically symmetric and T is an invertible linear transformation.

spherically symmetric. For one-dimensional BRW, we expect the following parity relation between maximum and FPT. Recall that the asymptotic for the maximum M_n at time n is

$$M_n = m_n + O_{\mathbb{P}}(1), \quad \text{where} \quad m_n := c_1 n - \frac{3}{2c_2} \log n. \tag{8}$$

Inverting (8) leads to the asymptotic (e.g., using the Lambert W function)

$$n = \frac{M_n}{c_1} + \frac{3}{2c_2 c_1} \log M_n + O_{\mathbb{P}}(1).$$

It is then natural to guess that for one-dimensional BRW,

$$\tau_x = m_1(x) + O_{\mathbb{P}}(1), \quad \text{where} \quad m_1(x) := \frac{x}{c_1} + \frac{3}{2c_2 c_1} \log x, \tag{9}$$

which we will prove in Proposition 7 below in Section 2.1. Recall (3). Let us decompose $A(x) = m_1(x) + m_2(x)$, where

$$m_2(x) := A(x) - m_1(x) = \frac{d-1}{2c_2 c_1} \log x. \tag{10}$$

For a general dimension $d \geqslant 1$, we shall explain the extra term $m_2(x)$.

The main difficulty in analyzing the first passage times in higher dimensions is that the second moment method does not immediately apply if one counts the number of particles in the target ball B_x at a certain time. To bypass this difficulty, we prove separately the upper and lower bounds for τ_x.

A key step in these bounds is Proposition 8, which keeps track of the number of particles that are near the frontier (say, a distance of $\log n$ away) for a one-dimensional BRW. Roughly speaking, there are around $xe^{c_2 x}$ many particles that are of a distance x behind the frontier, where $x \ll \log n$. This amounts to generalizing the proof of the asymptotic of the maximum (8); see [13]. A notable distinction is that unfortunately, one cannot conclude from the second moment method the lower bound for the number of particles near the frontier with overwhelming probability. Instead, one only concludes the lower bound with a uniformly positive probability. This gap can be filled using Theorem 2, which shows that a uniformly positive probability would already be sufficient.

A particularly nice feature of the spherical symmetry of the jumps is that the displacements in different directions are approximately independent (likewise, the same result Theorem 1 holds if we instead assume independent centered jumps in the d directions). In particular, at time n, one roughly expects a proportion $O(n^{-(d-1)/2})$ of the particles that lie within unit distance from the origin in the last $d-1$ dimensions. In other words, one expects that the first passage event is realized around the time when $x^{(d-1)/2}$ particles reach a distance of x for the $\mathbb{R}^d$-valued BRW projected onto the first dimension; see Figure 2 below. On the other hand, by (8), the extra time of $m_2(x)$ from (10) pushes the frontier ahead by a distance of approximately $c_1 m_2(x) = (d-1)(\log x)/(2c_2)$, which amounts to having approximately $x^{(d-1)/2}$ extra particles in view of Proposition 8.

If the jump distribution is not spherically symmetric, applying similar arguments above (with Proposition 8 replaced by Theorem B of [11] and Proposition 10 by Proposition 17; we omit the details here) will lead to

$$\tau_x = \frac{x}{\hat{c}_1} + O_{\mathbb{P}}(\log x), \tag{11}$$

where the corresponding upper and lower bounds for τ_x are separated by an order of $\log x$. To pin down the coefficient of the logarithm correction term, we extend the arguments for the maximum of one-dimensional BRW ([13]) with a modified barrier. The modified barrier event lies in $\mathbb{R}^d$ and hence results on random walks in cones ([23]) are applicable. Note that the same proof also applies for the spherically symmetric case, but would yield less precise asymptotics compared to (4) (an $o_{\mathbb{P}}(\log x)$ instead of an $O_{\mathbb{P}}(\log \log x)$).
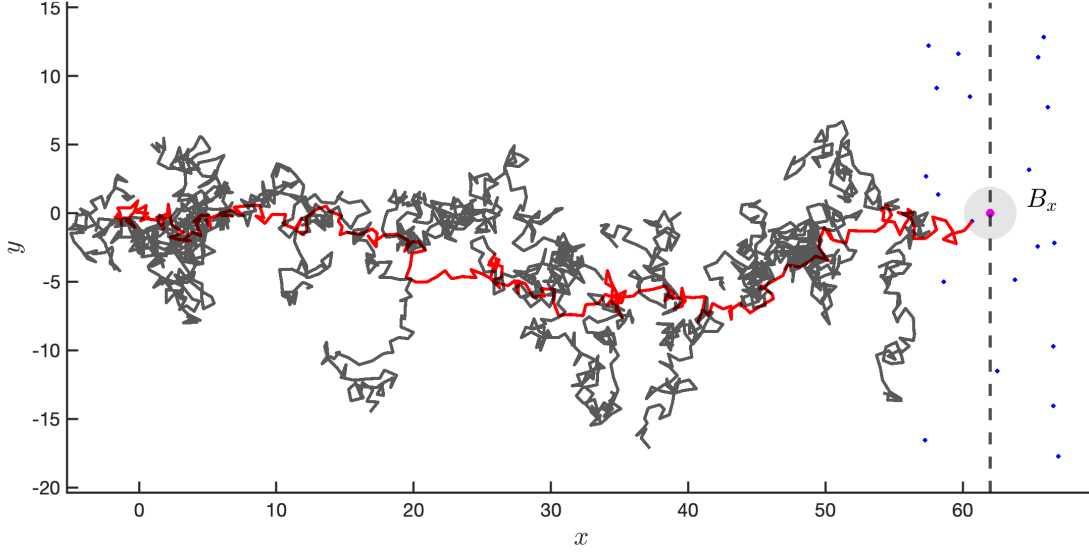
6

Figure 2: Schematic description of the heuristic for spherically symmetric jumps: around a proportion $x^{-(d-1)/2}$ of the particles that reach roughly x far in the first coordinate (represented by blue dots) lie in the ball B_x. The path in red represents (part of) the trajectory that realizes the FPT, whereas the other branches are shown in black.

2 Proofs of main results

This section is devoted to the proofs of the main results. We will first analyze the simple case of $d = 1$ in Section 2.1, where a precise asymptotic of the FPT follows directly from inverting the maxima asymptotic (8). We then collect the necessary tools in Sections 2.2.1 and 2.2.2, and prove Theorems 1 and 2 in Sections 2.2.3 and 2.3 respectively. The case of non-spherically symmetric jumps (Theorem 3) will be addressed in Section 2.4, where we also summarize preliminary results on random walks that are constrained to stay in a cone. Section 2.5 presents the proofs of Corollaries 4 and 5.

We begin by introducing some necessary notations. Following [13], we let $\{\boldsymbol{\eta}_{v,n}(k)\}_{k=0,\dots,n}$ denote the d-dimensional random walk where $\boldsymbol{\eta}_{v,n}(k)$ is the position of the k-th generation for the path in the tree leading to the vertex v on level n, denoted by $v \in V_n$. We also let $\eta_{v,n}(k)$ denote the first coordinate of $\boldsymbol{\eta}_{v,n}(k)$. When $d = 1$, these notions coincide. For deterministic quantities or functions A, B, We use Vinogradov's symbol $A \ll B$ (or $A = O(B)$) to denote $|A| \leqslant CB$ with some constant $C > 0$ that depends only on the distribution of $\boldsymbol{\xi}$ and the dimension d. We write $A \asymp B$ if $A \ll B \ll A$. The notation $\|\cdot\|$ always denotes the Euclidean norm. We use $B_{\mathbf{x}}(r)$ to denote the shifted ball of radius r centered at $\mathbf{x} \in \mathbb{R}^d$, $B_{\mathbf{x}} = B_{\mathbf{x}}(1)$, and $B_x = B_{(x,\mathbf{0})}$ for $x \in \mathbb{R}$.

2.1 Warming up in dimension one

For $\beta > 0$ and $n \in \mathbb{N}$ we define the event

$$G_{n,\beta} := \bigcup_{v \in V_n} \bigcup_{0 \leqslant k \leqslant n} \left\{ \eta_{v,n}(k) \geqslant \frac{km_n}{n} + \beta + \frac{4}{c_2}(\log(\min(k, n-k)))_+ \right\}.$$

Lemma 6 (Lemma 2.4 of [13])**.** *Assume conditions (A1), (A2), and (A4), and assume that the jump distribution is non-lattice. It holds that $\mathbb{P}(G_{n,\beta}) \ll \beta e^{-c_2\beta}$.*

7

Proposition 7. *Suppose that a one-dimensional BRW satisfies conditions (A1), (A2), and (A4), and that the jump distribution is non-lattice. We have*

$$\tau_x = m_1(x) + O_{\mathbb{P}}(1) = \frac{x}{c_1} + \frac{3}{2c_2 c_1} \log x + O_{\mathbb{P}}(1).$$

Proof. For the lower bound, we fix $\varepsilon > 0$ and pick β large enough such that $\mathbb{P}(G_{n,\beta}) < \varepsilon$, by means of Lemma 6. In other words, with a probability larger than $1 - \varepsilon$, all trajectories of the BRW stay below the curve

$$k \mapsto \frac{k m_n}{n} + \beta + \frac{4}{c_2}(\log(\min(k, n - k)))_+.$$

It is clear that with $C_1 = C_1(\beta) > 0$ chosen large enough,

$$\max_{1 \leqslant k \leqslant \tau_x - C_1} \left(\frac{k m_{\tau_x}}{\tau_x} + \beta + \frac{4}{c_2}(\log(\min(k, \tau_x - k)))_+ \right) \leqslant x.$$

This yields $\mathbb{P}(\tau_x \leqslant m_1(x) - C_1) \leqslant \mathbb{P}(G_{n,\beta}) < \varepsilon$, proving the lower bound.

For the upper bound, we fix $\varepsilon > 0$. Pick $C_2 > 0$ such that $\mathbb{P}(M_{m_1(x)+C_2} > x) > 1 - \varepsilon/3$ by (8). Using the same analysis for the lower bound, there exists $C_3 > 0$ such that $\mathbb{P}(M_{m_1(x)+C_2} > x + C_3) < \varepsilon/3$. Therefore, with probability $> 1 - 2\varepsilon/3$, there is $v \in V_{m_1(x)+C_2}$ such that $x \leqslant \eta_{v,m_1(x)+C_2}(m_1(x) + C_2) \leqslant x + C_3$. Since the distribution of ξ is non-lattice, by the Theorem 2 of [42], the random walk initiated from v is recurrent, and hence there exists $C_4 > 0$ such that the descendants of v reaches $[x - 1/2, x + 1/2]$ with probability $> 1 - \varepsilon/3$ within time C_4. Combining the above leads to $\mathbb{P}(\tau_x > m_1(x) + C_2 + C_4) < \varepsilon$, as desired. $\square$

2.2 Proof of Theorem 1

We now present the proof of Theorem 1 for $d \geqslant 2$, conditioned on Theorem 2. The proof of Theorem 2 will be the focus of the next section. The following auxiliary results are necessary as explained in our proof strategy.

2.2.1 The number of particles near maximum

The goal of this section is to prove the following result on the number of particles near the frontier of the BRW, which might be of independent interest.

Proposition 8. *Assume (A1)–(A4). There exists $L > 0$ depending only on the law of the BRW such that the following holds conditioned upon survival. Given any $\varepsilon > 0$, there exists $C > 0$ independent from n and x such that uniformly for n large enough and for $x \in [2, \sqrt{n}]$,*

$$\mathbb{P}\left(\#\{v \in V_n : \eta_{v,n}(n) \geqslant m_n - x\} > Cxe^{c_2 x} \mid S\right) < \varepsilon \tag{12}$$

and

$$\mathbb{P}\left(\#\{v \in V_n : \eta_{v,n}(n) \geqslant m_n - x\} > \frac{1}{C}xe^{c_2 x} \mid S\right) > \frac{1}{L}. \tag{13}$$

Remark 1. Proposition 8 may be compared against [11, 45], where it is proven that for a location that is $\Omega(n)$ away from m_n, the number of particles therein has the same order as its expectation. Proposition 8 shows that this is not the case for locations that are $O(\sqrt{n})$ near m_n. On the other hand, the number of particles that are $O(\sqrt{n})$ away from the origin can be described by the BRW central limit theorem [10]. An interesting question would be to analyze the phase transitions between those regimes. We also expect that (13) holds if we replace its right-hand side by $1 - \varepsilon$, but the current version suffices for our purpose.

Following the setup of Section 2.2 of [13], we introduce λ_n which is the value of λ where the supremum of (1) is attained with $x = m_n/n$. By (15) of [13], it holds $0 \leqslant c_2 - \lambda_n \ll (\log n)/n$. Let $\mathbb{Q}^{(n)}$ be defined by

$$\frac{d\mathbb{P}}{d\mathbb{Q}^{(n)}} := e^{-\lambda_n(\eta_{v,n}(n)-m_n)-nI(m_n/n)} \asymp n^{3/2} \rho^{-n} e^{-\lambda_n(\eta_{v,n}(n)-m_n)}. \tag{14}$$

It follows that under $\mathbb{Q}^{(n)}$, $\{\eta_{v,n}(k) - km_n/n\}_{k=0,\ldots,n}$ is a mean zero random walk. Let

$$Q_{n,\beta} := \left\{ v \in V_n : \text{ for any } 0 \leqslant k \leqslant n, \ \eta_{v,n}(k) < \frac{km_n}{n} + \beta + \frac{4}{c_2}(\log(\min(k, n-k)))_+ \right\}$$

and define

$$g(n, \beta, x) := \mathbb{E}[\#\{v \in Q_{n,\beta} : \eta_{v,n}(n) \geqslant m_n - x\}]. \tag{15}$$

We first reduce the proof of the upper bound (12) into proving the following estimate on $g(n, \beta, x)$.

Proposition 9. *Assume (A1)–(A4). Uniformly in $x \in [2, \sqrt{n}]$,*

$$g(n, \beta, x) \ll \beta(x + \beta)e^{c_2 x}.$$

Proof of (12) *given Proposition 9.* For fixed $C, \varepsilon > 0$, we pick β large enough so that $\mathbb{P}(G_{n,\beta}) < \varepsilon/2$ by Lemma 6. On the event $G_{n,\beta}^c$, $Q_{n,\beta} = V_n$. By Proposition 9, uniformly for n and $x \in [2, \sqrt{n}]$,

$$\mathbb{E}\left[\#\{v \in V_n : \eta_{v,n}(n) \geqslant m_n - x\} \mathbb{1}_{G_{n,\beta}^c}\right] \ll \beta(x + \beta)e^{c_2 x}.$$

Markov's inequality then yields that for some $L > 0$,

$$\mathbb{P}\left(G_{n,\beta}^c \cap \{\#\{v \in V_n : \eta_{v,n}(n) \geqslant m_n - x\} > L\beta(x + \beta)e^{c_2 x}\}\right) < \frac{\varepsilon}{2}.$$

This yields (12). $\qquad\square$

Proof of Proposition 9. Consider a fixed $v \in V_n$. We have

$$\begin{aligned}
&\mathbb{E}[\#\{v \in Q_{n,\beta} : \eta_{v,n}(n) \geqslant m_n - x\}] \\
&= \rho^n \mathbb{P}(v \in Q_{n,\beta}, \ \eta_{v,n}(n) \geqslant m_n - x) \\
&= \sum_{j=0}^{x+\beta-1} \rho^n \mathbb{P}(v \in Q_{n,\beta}, \ \eta_{v,n}(n) \in [m_n + \beta - j - 1, m_n + \beta - j)).
\end{aligned}$$

By substituting $i = \beta - j$, the latter probability is precisely $\chi_{n,n}^{\mathbb{P}}(i)$ defined below (15) in [13]. It is proved in (18) therein the estimate

$$\chi_{n,n}^{\mathbb{P}}(i) \ll \beta(\beta - i + 2)\rho^{-n} e^{-c_2 i},$$

where our assumption $x \leqslant \sqrt{n}$ is used. This leads to

$$\mathbb{E}[\#\{v \in Q_{n,\beta} : \eta_{v,n}(n) \geqslant m_n - x\}] \ll \sum_{j=0}^{x+\beta} \beta(j+2)e^{-c_2(\beta-j)} \ll \beta(x + \beta)e^{c_2 x},$$

completing the proof. $\qquad\square$

Proof of (13). We apply the second moment method. Define

$$P_n := \left\{ v \in V_n : \text{ for any } 0 \leqslant k \leqslant n, \ \eta_{v,n}(k) < \frac{km_n}{n} \right\}.$$

For $x \in [2, \sqrt{n}]$ and $v \in V_n$, define the event

$$H_{v,n}(x) := \{ v \in P_n, \ \eta_{v,n}(n) \in [m_n - x, m_n - x + 1) \}$$

and $\Delta_{n,x} := \sum_{v \in V_n} \mathbb{1}_{H_{v,n}(x)}$. It follows that

$$\#\{ v \in V_n : \eta_{v,n}(n) \geqslant m_n - x \} \geqslant \Delta_{n,x}. \tag{16}$$

Let us first compute the first moment $\mathbb{E}[\Delta_{n,x}]$. Using (14), we have

$$\mathbb{P}(H_{v,n}(x)) \gg n^{3/2} \rho^{-n} \mathbb{E}_{\mathbb{Q}} \left[e^{-\lambda_n(\eta_{v,n}(n) - m_n)} \mathbb{1}_{H_{v,n}(x)} \right]$$

$$\gg n^{3/2} \rho^{-n} \mathbb{E}_{\mathbb{Q}} \left[e^{c_2 x} \mathbb{1}_{H_{v,n}(x)} \right] = n^{3/2} \rho^{-n} e^{c_2 x} \mathbb{Q}(H_{v,n}(x)).$$

By the ballot theorem, in the form of Lemma 2.1 of [13], We have

$$\mathbb{Q}(H_{v,n}(x)) \gg n^{-3/2} \max(x - 1, 1) \gg n^{-3/2} x.$$

Combining the above yields

$$\mathbb{E}[\Delta_{n,x}] = \rho^n \mathbb{P}(H_{v,n}(x)) \gg x e^{c_2 x}. \tag{17}$$

Next we estimate the second moment $\mathbb{E}[\Delta_{n,x}^2]$. Observe that by (A1) (see (29) of [13]),

$$\mathbb{E}[\Delta_{n,x}^2] \ll \mathbb{E}[\Delta_{n,x}] + \rho^n \sum_{s=1}^{n} \rho^s \mathbb{P}(H_{v,n}(x) \cap H_{w,n}(x) \text{ for } v \sim_s w), \tag{18}$$

where $v \sim_s w$ means that the distance of v and w in the genealogical tree is equal to $2s$. To bound the latter probability we condition on the location of the common ancestor of v and w, $\eta_{v,n}(n - s)$. We have for $v \sim_s w$,

$$\mathbb{P}(H_{v,n}(x) \cap H_{w,n}(x))$$
$$\leqslant \sum_{j=0}^{\infty} \mathbb{P} \left(\eta_{v,n}(k) \leqslant \frac{km_n}{n} \text{ for } k \leqslant n - s, \ \eta_{v,n}(n - s) - \frac{(n-s)m_n}{n} \in [-(j+1), -j) \right)$$

$$\times \left(\sup_{y \in [-(j+1), -j)} P(n, s, x, y) \right)^2,$$

where

$$P(n, s, x, y) := \mathbb{P} \left(\eta_{v,n}(n - s + \ell) \leqslant \frac{(n - s + \ell)m_n}{n} \text{ for } 0 \leqslant \ell \leqslant s, \right.$$

$$\left. \eta_{v,n}(n) \in [m_n - (x + 1), m_n - x) \mid \eta_{v,n}(n - s) - \frac{(n-s)m_n}{n} = y \right).$$

10

With the same change of measure and ballot theorem argument as above, we get

$$\mathbb{P}\left(\eta_{v,n}(k) \leqslant \frac{km_n}{n} \text{ for } k \leqslant n - s, \ \eta_{v,n}(n-s) - \frac{(n-s)m_n}{n} \in [-(j+1), -j)\right)$$

$$\ll e^{3(n-s)(\log n)/(2n)}\rho^{-(n-s)}e^{\lambda_n j}$$

$$\times (\max(1, n-s))^{-3/2}\max\left((m_{n-s} - (n-s)m_n/n + j), 1\right)$$

$$\ll \rho^{-(n-s)}e^{\lambda_n((3/(2c_2))((n-s)(\log n)/n - \log(n-s)) + j)}$$

$$\times \max\left(\left(j + \frac{(n-s)\log n}{n} - \log(n-s)\right), 1\right)$$

$$\ll \rho^{-(n-s)}je^{c_2 j - 3\log(n-s)/2 + 3(n-s)\log n/(2n)},$$

where in the last step we use $0 \leqslant c_2 - \lambda_n \ll (\log n)/n$. In a similar manner, (while using (6) of [13]), we obtain the estimate

$$P(n, s, x, y) \ll \rho^{-s}jxe^{c_2(x-j)+3((s\log n)/n - \log s)/2}.$$

In conclusion, we arrive at

$$\mathbb{E}[\Delta_{n,x}^2] \ll xe^{c_2 x} + \sum_{s=1}^{n}\sum_{j=0}^{\infty}je^{c_2 j - 3\log(n-s)/2 + 3(n-s)\log n/(2n)}$$

$$\times \left(jxe^{c_2(x-j)+3((s\log n)/n - \log s)/2}\right)^2$$

$$\ll xe^{c_2 x} + x^2 e^{2c_2 x}\sum_{s=1}^{n}\frac{n^{3/2}e^{(3s\log n)/(2n)}}{(\max(1, n-s))^{3/2}s^3} \ll x^2 e^{2c_2 x}, \tag{19}$$

where in the last step we argue similarly as in the proof of Lemma 2.7 of [13]. Combined with (16), (17), and the Paley-Zygmund inequality leads to (13). $\qquad\square$

2.2.2 Large deviation computations

In the following, we let $\{(X_j, \mathbf{Y}_j)\}_{j\in\mathbb{N}}$ be a sequence of i.i.d. random variables with the same distribution as $\boldsymbol{\xi}$, and $(X, \mathbf{Y}) = \sum_{j=1}^{n}(X_j, \mathbf{Y}_j)$. Recall the definition of m_n from (8).

Proposition 10. *Fix a large constant $C > 0$ and a dimension $d \geqslant 1$. Suppose that the $\mathbb{R}^d$-valued random variable $\boldsymbol{\xi}$ satisfies conditions (A2)–(A4) from Section 1.1.*

(i) Uniformly for any positive sequence $u(n) \leqslant C\sqrt{n}$ and any $c = c(n) \in [-C\log n, C\log n]$,

$$\mathbb{P}(\mathbf{Y} \in B_{\mathbf{0}}(u(n)) \mid X \geqslant m_n + c) \ll u(n)^{d-1}n^{-(d-1)/2}. \tag{20}$$

(ii) Fix $a(n) \to \infty$ and $a(n) = o(\log n)$. Uniformly for $\mathbf{y}$ with $\|\mathbf{y}\| \leqslant C\sqrt{n}$ and $c \in [-C\log n, C\log n]$,

$$\mathbb{P}(\mathbf{Y} \in B_{\mathbf{y}}(1) \mid X \in [m_n + c, m_n + c + a(n)]) \gg n^{-(d-1)/2}. \tag{21}$$

(iii) For any $\varepsilon > 0$, there exists $K > 0$ such that uniformly for $c \in [-C\log n, C\log n]$,

$$\mathbb{P}(\mathbf{Y} \notin B_{\mathbf{0}}(K\sqrt{n}) \mid X \geqslant m_n + c) < \varepsilon. \tag{22}$$

The proof of Proposition 10 is deferred to the appendix.

2.2.3 Proof of Theorem 1 conditioned on Theorem 2

Recall (3). Let us define

$$A_1(x) := A(x) + \frac{2}{c_2 c_1} \log\log x, \quad A_2(x) := A(x) - \frac{1}{c_2 c_1} \log\log x.$$

In view of Theorem 2, it suffices to prove that for any $a(x)$ that tends to infinity, and any $\varepsilon > 0$, there exists $N > 0$ such that for any $x \geqslant N$,

$$\mathbb{P}(\tau_x > A_1(x) + a(x)) \leqslant 1 - \frac{1}{L} \tag{23}$$

and

$$\mathbb{P}(\tau_x < A_2(x) - a(x)) < \varepsilon. \tag{24}$$

In the following, we fix $\varepsilon > 0$ and $a(x)$.

Proof of the upper bound. Define

$$\alpha(x) := \frac{d-1}{2c_2} \log x - \frac{1}{c_2} \log\log x + c_1 a(x)$$

and

$$W_n := \left\{ v \in V_n : \boldsymbol{\eta}_{v,n}(n) \in \left(m_n - \alpha(x), m_n \right) \times B_{\mathbf{0}}(\sqrt{x}) \right\}.$$

We first prove that for x large enough and some large constant L_1 to be determined,

$$\mathbb{P}\big(\#W_{(\log x)^2} < L_1 x^{(d-1)/2}\big) < 1 - \frac{1}{L}. \tag{25}$$

Indeed, let

$$\widetilde{W}_n := \left\{ v \in V_n : \eta_{v,n}(n) \in \left(m_n - \alpha(x), m_n \right) \right\}.$$

It follows from Proposition 8 that for x large enough, $\mathbb{P}(\#\widetilde{W}_n < L_1 x^{(d-1)/2}) < 1 - 1/L$. Moreover, a union bound gives that out of $L_1 x^{(d-1)/2}$ particles in $\widetilde{W}_n$, the probability that the last $d-1$ coordinates fall outside $B_{\mathbf{0}}(\sqrt{x})$ for some of them is at most $L_1 x^{(d-1)/2} \mathbb{P}(\|\mathbf{Z}\| \geqslant \sqrt{x}/\log x) = o(1)$, where $\mathbf{Z}$ denotes a standard Gaussian random variable on $\mathbb{R}^{d-1}$. This proves (25).

The plan is to evolve (on the event of (25)) the $L_1 x^{(d-1)/2}$ particles in $W_{(\log x)^2}$ (independently) so that at least one of those families has a descendant landing in B_x at time $A_1(x) + a(x)$ with high probability, where we recall that B_x is the ball of radius one centered at $(x, \mathbf{0})$. Let us now fix $w \in W_{(\log x)^2}$. Define $\tau'(x) := A_1(x) + a(x) - (\log x)^2$ and

$$p(w) := \mathbb{P}\left(\boldsymbol{\eta}_{v,\tau'(x)}(\tau'(x)) \in B_x - \boldsymbol{\eta}_{w,(\log x)^2}((\log x)^2) \text{ for some } v \in V_{A_1(x)+a(x)-(\log x)^2} \right),$$

which is the probability that at least one descendant of w lands in B_x at time $A_1(x) + a(x)$. Here and later, we will without loss of generality assume that quantities that tend to infinity are integers. By independence and by picking the constant L_1 in (25) large enough, it suffices to prove $p(w) \gg x^{-(d-1)/2}$ for $w \in W_{(\log x)^2}$. Since $\eta_{w,(\log x)^2}((\log x)^2) \in (m_{(\log x)^2} - \alpha(x), m_{(\log x)^2})$ and $m_{\tau'(x)} \geqslant x - (m_{(\log x)^2} - \alpha(x))$, by the main result of [39],

$$\mathbb{P}\left(\eta_{v,\tau'(x)}(\tau'(x)) \in B_x(1) - \eta_{w,(\log x)^2}((\log x)^2) \text{ for some } v \in V_{A_1(x)+a(x)-(\log x)^2} \right) \geqslant \frac{1}{L}.$$

Conditionally on existence, applying Proposition 10(ii) to the particle $v \in V_{A_1(x)+a(x)-(\log x)^2}$ with $\eta_{v,\tau'(x)}(\tau'(x)) \in B_x(1) - \eta_{w,(\log x)^2}((\log x)^2)$ leads to $p(w) \geqslant x^{-(d-1)/2}/L$. This establishes (23).

Proof of the lower bound. Our goal is to bound $\mathbb{P}(\tau_x \leqslant A_2(x) - a(x))$ from above. According to Propositions 8 and 10, at time $A_2(x)$ there will be less than $O(1)$ number of particles that lie in $[x, \infty) \times B_{\mathbf{0}}(1)$ in expectation. We need to argue that other particles that stayed in B_x earlier but currently do not stay in $[x, \infty) \times B_{\mathbf{0}}(1)$ do not contributed asymptotically. This is because most of the $O(1)$ particles will arrive not much earlier than $m_1(x) + m_2(x)$, and will not move quickly.

Consider $n = A_2(x)$. Recall that $\varepsilon > 0$ and $a(x) \to \infty$ are fixed. We define $T_{v,n}(x) := \min\{0 \leqslant k \leqslant n : \eta_{v,n}(k) \geqslant x\}$ and for an integer $j \in [a(x), m_2(x)]$,

$$V_{n,j} := \left\{ v \in V_n : T_{v,n}(x) = A_2(x) - j \right\}.$$

It follows from Lemma 6 that for any $\varepsilon > 0$, there exists L_2 large enough such that

$$\mathbb{P}(\tau_x \leqslant A_2(x) - a(x))$$

$$\leqslant \mathbb{P}\left(\bigcup_{j=a(x)}^{m_2(x)+L_2} \bigcup_{v \in V_{n,j}} \bigcup_{T_{v,n}(x) \leqslant k \leqslant n} \left\{ \eta_{v,n}(k) \in B_x \right\} \right) + \mathbb{P}\left(M_{A_2(x)-m_2(x)-L_2} \geqslant x \right)$$

$$\leqslant \sum_{j=a(x)}^{m_2(x)+L_2} \mathbb{P}\left(\bigcup_{v \in V_{n,j}} \bigcup_{T_{v,n}(x) \leqslant k \leqslant n} \left\{ \eta_{v,n}(k) \in B_x \right\} \right) + \frac{\varepsilon}{4}. \tag{26}$$

Let us further decompose the event $\{\eta_{v,n}(k) \in B_x\}$ depending on the location at the first time the random walk $\{\eta_{v,n}(k)\}_{0 \leqslant k \leqslant n}$ reaches $[x, \infty) \times \mathbb{R}^{d-1}$, whether $\eta_{v,n}(T_{v,n}(x)) \in [x, \infty) \times B_{\mathbf{0}}(c_3 j)$ with some $c_3 > 0$ to be determined.

First, with j fixed,

$$\mathbb{P}\left(\bigcup_{v \in V_{n,j}} \bigcup_{T_{v,n}(x) \leqslant k \leqslant n} \left\{ \eta_{v,n}(k) \in B_x, \ \eta_{v,n}(T_{v,n}(x)) \in [x, \infty) \times B_{\mathbf{0}}(c_3 j) \right\} \right)$$

$$\leqslant \mathbb{P}\left(\bigcup_{v \in V_{n,j}} \left\{ \eta_{v,n}(A_2(x) - j) \in [x, \infty) \times B_{\mathbf{0}}(c_3 j) \right\} \right)$$

$$= \mathbb{P}\left(\bigcup_{v \in V_{A_2(x)-j}} \left\{ \eta_{v,A_2(x)-j}(A_2(x) - j) \in [x, \infty) \times B_{\mathbf{0}}(c_3 j) \right\} \right).$$

Using Proposition 8 (while adjusting the proof suitably), we see that for some $L_3 > 0$,

$$\mathbb{P}\left(\#\{v \in V_{A_2(x)-j} : \eta_{v,A_2(x)-j}(A_2(x) - j) \geqslant x\} \geqslant L_3 j^2 e^{-c_2 c_1 j} n^{(d-1)/2} \right) < \frac{\varepsilon}{10 j^2}.$$

Applying Proposition 10(iii) and the union bound then yields

$$\mathbb{P}\left(\bigcup_{v \in V_{A_2(x)-j}} \left\{ \eta_{v,A_2(x)-j}(A_2(x) - j) \in [x, \infty) \times B_{\mathbf{0}}(c_3 j) \right\} \right) \leqslant \frac{\varepsilon}{10 j^2} + L_3 j^2 e^{-c_2 c_1 j} (c_3 j)^{d-1}.$$

Second, using that $j \geqslant a(x)$ and $a(x) \to \infty$ as $x \to \infty$, as well as Theorem 3.2 of [29], we have for

13

some $c_3, c_4 > 0$ large and independent of x, such that for all x large enough,

$$
\mathbb{P}\left(\bigcup_{v \in V_{n,j}} \bigcup_{T_{v,n}(x) \leqslant k \leqslant n} \{ \boldsymbol{\eta}_{v,n}(k) \in B_x, \ \boldsymbol{\eta}_{v,n}(T_{v,n}(x)) \in [x, \infty) \times B_{\mathbf{0}}(c_3 j)^c \} \right)
$$

$$
\leqslant \mathbb{P}\left(\max_{v \in V_j} \| \boldsymbol{\eta}_{v,j}(j) \| \geqslant (c_3 - 1)j \right)
$$

$$
\leqslant d\, \mathbb{P}\left(\max_{v \in V_j} |\eta_{v,j}(j)| \geqslant \frac{(c_3 - 1)j}{\sqrt{d}} \right)
$$

$$
\leqslant e^{-j/c_4}.
$$

Inserting the above estimates to (26) leads to

$$
\mathbb{P}(\tau_x \leqslant A_2(x) - a(x)) \leqslant \frac{\varepsilon}{4} + \sum_{j=a(x)}^{\infty} \left(\frac{\varepsilon}{10 j^2} + L_3 j^2 e^{-c_2 c_1 j} (c_3 j)^{d-1} + e^{-j/c_4} \right).
$$

For any $\varepsilon > 0$ and $a(x) \to \infty$, the right-hand side is $< \varepsilon$ for x large enough. This completes the proof of the lower bound (24).

2.3 Tightness of first passage times: proof of Theorem 2

We build upon ideas from [41], while in contrast, we do not assume uniform boundedness of the increments.[6] In the following, we write $C_1, C_2, C_3, \ldots$ as large constants and $c_3, c_4, \ldots$ as small constants that may depend on the law of $\boldsymbol{\xi}$ and the branching rate.[7] Here, we do not assume that $\boldsymbol{\xi}$ is spherically symmetric. We start with a few preparatory lemmas.

Lemma 11. *Assume (A6). For any $C_1 > 0$ and $c_3 > 0$, there exists $C_2 > 0$ such that the following holds for k large enough: let $(X_j, \mathbf{Y}_j)$ be a sequence of i.i.d. random variables with the same distribution as $\boldsymbol{\xi}$, $C > C_2$, and $(X, \mathbf{Y}) = \sum_{j=1}^{Ck} (X_j, \mathbf{Y}_j)$. Then uniformly for $k > 0$ and $\mathbf{z} \in \mathbb{R}^d$ such that $\| \mathbf{z} \| \leqslant C_1 k$,*

$$
\mathbb{P}((X, \mathbf{Y}) \in B_{\mathbf{z}}) \gg e^{-c_3 k}.
$$

The proof of Lemma 11 is deferred to the appendix.

Lemma 12. *There exist some $C_3, c_4 > 0$ such that for $k \geqslant 1$ large enough,*

$$
\mathbb{P}(\#\{v \in V_{C_3 k} : \eta_{v, C_3 k}(C_3 k) \geqslant k\} < e^{c_4 k} \mid S) \ll e^{-c_4 k}.
$$

Proof. We first claim that for any $c_5 > 0$, there exist $c_6, c_7 > 0$ such that uniformly for n large enough,

$$
\mathbb{P}(\#\{v \in V_n : \| \boldsymbol{\eta}_{v,n}(n) \| \leqslant c_5 n\} < e^{c_6 n} \mid S) \ll e^{-c_7 n}. \tag{27}
$$

Indeed, given c_5, let $c_8 = c_5/(6 d \bar{c}_1)$ where here $\bar{c}_1$ stands for the maximum linear speed of the BRW across all d directions, so that by Theorem 3.2 of [29],[8]

$$
\mathbb{P}\left(\exists v \in V_{c_8 n}, \ \| \boldsymbol{\eta}_{v, c_8 n}(c_8 n) \| > \frac{c_5 n}{2} \right) \ll e^{-c_7 n}. \tag{28}
$$

---

[6]In fact, we expect that our approach also leads to a concentration result for the maximum of BRW without assuming uniform boundedness of the increments.

[7]Not to confuse with the fixed constants c_1, c_2 defined in Section 1.1.1.

[8]Theorem 3.2 of [29] is stated in the Schröder case, where $p_0 + p_1 > 0$. However, the proof for the upper bound of the large deviation probability does not rely on this assumption; see Remark 3.3(b) therein.

Conditioning upon the survival event S, with an exponentially small probability we have that $\#V_{c_8 n} \geqslant e^{c_9 n}$. Evidently, each $v \in V_{c_8 n}$ will lead independently with probability $> 1/2$ to a particle $w \in V_n$ where $\|\boldsymbol{\eta}_{w,n}(n) - \boldsymbol{\eta}_{w,n}(c_8 n)\| \leqslant c_5 n/2$. This along with standard binomial estimates proves (27).

We continue to the proof. We apply first (27) with $n = k$ and $c_5 = 1$. Outside a set of exponentially small probability, we may consider a subcollection of $e^{c_6 k}$ particles in $\{v \in V_k : \|\boldsymbol{\eta}_{v,k}(k)\| \leqslant k\}$, which we label by v_j, $1 \leqslant j \leqslant e^{c_6 k}$. We next evolve the particles v_j independently and check if they reach k in the x-direction. It suffices to consider the presence of the particles evolved from v_j, $1 \leqslant j \leqslant e^{c_6 k}$ in a ball of radius one and at most $k + k + 1 \leqslant 3k$ apart (from the locations of v_j). By Lemma 11 applied with $C_1 = 3$ and $c_3 = c_6/2$, for each j we have for some $C_2 > 0$, with w_j denoting the location of (a fixed descendant of) v_j after time $C_2 k$, that

$$\mathbb{P}(w_j \in [k, \infty) \times \mathbb{R}^{d-1}) \geqslant \mathbb{P}(w_j \in B_{(k+1,\mathbf{0})}) \gg e^{-c_6 k/2}. \tag{29}$$

Note that such a fixed path exists for at least half of v_j's with overwhelming probability. In conclusion, we evolved at least $e^{c_6 k}/2$ many particles independently at time k for time $C_2 k$, each resulting in a probability of $\gg e^{-c_6 k/2}$ to land in $[k, \infty) \times \mathbb{R}^{d-1}$. A standard binomial estimate yields that with overwhelming probability, there exist $e^{c_6 k/4}$ particles at time $(C_2 + 1)k$ present in $[k, \infty) \times \mathbb{R}^{d-1}$ (conditional on survival). This completes the proof. $\qquad\square$

Lemma 13. *There exist constants $c_{10}, C_4 > 0$ such that for x large enough,*

$$\mathbb{P}(\tau_x \geqslant C_4 x \mid S) \ll e^{-c_{10} x}.$$

Proof. The proof is almost the same as Lemma 12, by replacing k by x and using the second inequality of (29) only. $\qquad\square$

In the proof of Theorem 2, we wish to use Lemma 13 to bound probabilities of the form $\mathbb{P}(\tau_x \geqslant \alpha, \ \tau_y \leqslant \beta)$, where $\beta < \alpha$. This is intuitively possible since we may run the particle v_y that reaches B_y and bound the probability that it never reaches B_x in a time longer than $\alpha - \beta$, where we have used the strong Markov property of the random walk. There is, however, a subtle issue for the survival of the branching process initiated by the particle v_y, as this event is almost independent of the survival event S we are conditioning on (in the proof of Theorem 2). We will bypass this difficulty by excluding this termination event a priori, or by first having an exponential number e^{cn} of particles in B_y instead of a single particle v_y.

Recall our notations beginning Section 2 and denote by $q = 1 - p = 1 - \mathbb{P}(S)$.

Proof of Theorem 2. Pick C_3 and c_4 as in Lemma 12 and let $\delta \in (e^{-c_{13}/(2C_3)}, 1)$ for some $c_{13} > 0$ to be determined. For $x, y > 0$, let

$$s = s(x, y) := \inf\{t \geqslant 0 : \mathbb{P}(\tau_x \leqslant t \mid S) \geqslant \delta^y\}.$$

It follows that $\mathbb{P}(\tau_x < s - 1 \mid S) \leqslant \delta^y$ and

$$\mathbb{P}(\tau_x > s + 1) \leqslant 1 - p\delta^y \leqslant \exp(-p\delta^y). \tag{30}$$

Let $k = y/C_3$ and set $C_5 \gg a + C_6 + C_7$, all to be determined. Denote by $\tau_{x+k}^{(c_{11})}$ the waiting time until $e^{c_{11} k}$ many particles reach the ball B_{x+k}. Now note that

$$\begin{aligned}
\mathbb{P}(\tau_x > s - 1) \leqslant \ & \mathbb{P}(s - 1 < \tau_x < s + C_5 k) + \mathbb{P}\left(\tau_{x+k}^{(c_{11})} > s + C_3 k + 1 + C_6 C_7 k\right) \\
& + \mathbb{P}\left(\tau_x \geqslant s + C_5 k, \ \tau_{x+k}^{(c_{11})} \leqslant s + C_3 k + 1 + C_6 C_7 k\right).
\end{aligned} \tag{31}$$

15

We first decompose the second event $\tau_{x+k}^{(c_{11})} > s + C_3k + 1 + C_6C_7k$ of (31) depending on the number of particles at time C_3k that lie beyond k in the x-coordinate. On one hand, by Lemma 12,

$$\mathbb{P}(\#\{v \in V_{C_3k} : \eta_{v,C_3k}(C_3k) \geqslant k\} < e^{c_4k}) \leqslant q + o(e^{-c_4k}).$$

On the other hand, on the event that $\#\{v \in V_{C_3k} : \eta_{v,C_3k}(C_3k) \geqslant k\} \geqslant e^{c_4k}$, we may label e^{c_4k} many particles by $\{v_i\}_{1 \leqslant i \leqslant e^{c_4k}}$. Our goal is to show that the probability that fewer than $e^{c_{11}k}$ of $\{v_i\}_{1 \leqslant i \leqslant e^{c_4k}}$ reach B_x in the period of time $[C_3k, C_3k + s + 1 + C_6C_7k]$ is small. Note that the trajectories for each i in the period $[C_3k, C_3k + s + 1 + C_6C_7k]$ are independent and equivalent in law to a BRW trajectory initiated at $\boldsymbol{\eta}_{v,C_3k}(C_3k)$. A technical difficulty arises, however, as the events of reaching B_x are not monotone in x in dimensions $d \geqslant 2$. It would be nice if we had stochastic monotonicity for τ_x in x (as in the one-dimensional case), but this appears out of our reach. Instead, we use the following argument to bypass this technical difficulty.

Recall that each $\eta_{v_i,C_3k}(C_3k) \geqslant k$, and the notation $\mathbf{x} = (x, 0 \ldots, 0) \in \mathbb{R}^d$. By a standard large deviation estimate on the last $d-1$ coordinates of $\boldsymbol{\eta}_{v_i,C_3k}(C_3k)$ and using the triangle inequality, we may assume that $\|\mathbf{x} - \boldsymbol{\eta}_{v_i,C_3k}(C_3k)\| \leqslant x + C_6k$ for all $1 \leqslant i \leqslant e^{c_4k}$ by losing an event of probability $O(e^{-c_{12}k})$ with c_4 chosen small enough. Similarly, we may assume $\eta_{v_i,C_3k}(C_3k) \leqslant C_8k$ with an overwhelming probability of at least $1 - O(e^{-c_{12}k})$ (since there are e^{c_4k} many particles in total and we can apply a union bound). As a consequence,

$$\mathbb{P}\left(\forall 1 \leqslant i \leqslant e^{c_4k}, \ x - C_8k \leqslant \|\mathbf{x} - \boldsymbol{\eta}_{v_i,C_3k}(C_3k)\| \leqslant x + C_6k\right) \geqslant 1 - O(e^{-c_{12}k}).$$

For each $1 \leqslant i \leqslant e^{c_4k}$, consider the set of $\mathbf{x}_i$'s such that $\|\mathbf{x}_i - \boldsymbol{\eta}_{v_i,C_3k}(C_3k)\| = x$. Selecting $C_6 > C_8$ and on a set of overwhelming probability of at least $1 - O(e^{-c_{12}k})$, we can select $\mathbf{x}_i$ such that $\|\mathbf{x} - \mathbf{x}_i\| \leqslant C_8k$. For $\mathbf{z} \in \mathbb{R}^d$, we denote by $\tau_{\mathbf{z}}$ the FPT to the ball of radius one centered at $\mathbf{z}$. By Lemma 13 applied with $C_4 = C_7/2$ and picking C_7 large (and without loss of generality, a $p/2$ proportion of the e^{c_4k} many particles do not terminate, by losing an exponentially small probability),

$$\mathbb{P}\left(\#\{1 \leqslant i \leqslant \frac{p}{2}e^{c_4k}, \ \tau_{\mathbf{x}-\boldsymbol{\eta}_{v_i,C_3k}(C_3k)} \geqslant s + 1 + C_6C_7k\} \geqslant e^{c_{11}k}\right)$$

$$\leqslant \mathbb{P}\left(\#\{1 \leqslant i \leqslant \frac{p}{2}e^{c_4k}, \ \tau_{\mathbf{x}_i-\boldsymbol{\eta}_{v_i,C_3k}(C_3k)} \geqslant s + 1\} \geqslant e^{c_{11}k}\right) + O(e^{-c_{12}k})$$

$$\quad + \mathbb{P}\left(\exists 1 \leqslant i \leqslant \frac{p}{2}e^{c_4k}, \ \tau_{\mathbf{x}-\boldsymbol{\eta}_{v_i,C_3k}(C_3k)} \geqslant s + 1 + C_6C_7k, \ \tau_{\mathbf{x}_i-\boldsymbol{\eta}_{v_i,C_3k}(C_3k)} < s + 1\right)$$

$$\leqslant \mathbb{P}\left(\#\{1 \leqslant i \leqslant \frac{p}{2}e^{c_4k}, \ \tau_{\mathbf{x}_i-\boldsymbol{\eta}_{v_i,C_3k}(C_3k)} \geqslant s + 1\} \geqslant e^{c_{11}k}\right)$$

$$\quad\quad + \sum_{1 \leqslant i \leqslant e^{c_4k}} \mathbb{P}\left(\tau_{C_6k+1}^{(i)} > C_6C_7k\right) + O(e^{-c_{12}k})$$

$$\leqslant \mathbb{P}\left(\mathrm{Bin}(\frac{p}{2}e^{c_4k}, \mathbb{P}(\tau_x > s + 1)) \geqslant e^{c_{11}k}\right) + e^{c_4k}e^{-c_{10}k} + O(e^{-c_{12}k}).$$

Here and later, we use $\mathrm{Bin}(n,p)$ to denote a generic binomial random variable with n trials and success probability p. By choosing c_4 and then c_{11} small enough and using (30), we obtain that the above satisfies

$$\mathbb{P}\left(\mathrm{Bin}\left(\frac{p}{2}e^{c_4k}, \mathbb{P}(\tau_x > s + 1)\right) \geqslant e^{c_{11}k}\right) + e^{(c_4-c_{10})k} + O(e^{-c_{12}k})$$

$$\leqslant \exp(-p\delta^y e^{c_{13}k}) + O(e^{-c_{14}k}).$$

Inserting back to (31) and using Lemma 13 on the second term of (31), we have

$$\mathbb{P}(\tau_x > s - 1) \leqslant q + \exp(-p\delta^y e^{c_{13}k}) + O(e^{-c_{14}k}) + \mathbb{P}(s - 1 < \tau_x \leqslant s + C_5k). \tag{32}$$

Since $\delta \in (e^{-c_{13}/(2C_3)}, 1)$ and $k = y/C_3$, for y large enough we have $\exp(-p\delta^y e^{c_{13}k}) \leqslant \delta^y$. Thus, we conclude from (32) that

$$\mathbb{P}(\tau_x > s + C_5k) \leqslant q + O(e^{-c_{15}y}). \tag{33}$$

For $q > 0$, using the fact that $\mathbb{P}(\#V_n > 0 \mid S^c) = o(e^{-c_{16}n})$ (this is a consequence of Theorem 13.3 of [5]), we obtain

$$\begin{aligned}
\mathbb{P}(\tau_x > s + C_5 k) &\geqslant p\,\mathbb{P}(\tau_x > s + C_5 k \mid S) + q\,\mathbb{P}(\tau_x > y,\ \#V_y = 0 \mid S^c) \\
&\geqslant p\,\mathbb{P}(\tau_x > s + C_5 k \mid S) + q(1 - o(e^{-c_{16}y})) - C_9 \mathbb{P}(\tau_x \leqslant y,\ S^c).
\end{aligned} \tag{34}$$

Again by using $\mathbb{P}(\#V_n > 0 \mid S^c) = o(e^{-c_{16}n})$, we have for some c_{17} to be determined later,

$$\mathbb{P}(\tau_x \leqslant y,\ S^c) \leqslant o(e^{-c_{17}c_{16}y}) + \mathbb{P}(\tau_x \leqslant c_{17}y).$$

Recall that M_n denotes the maximum in the first direction of the BRW at time step n. Using Lemma 5.2 of [29], it holds that for $\widetilde{M}_j$ denoting the maxima of $\#V_j$ many independent random walks,

$$\mathbb{P}(\tau_x \leqslant c_{17}y) \leqslant \sum_{j=1}^{c_{17}y} \mathbb{P}(M_j \geqslant x) \leqslant \sum_{j=1}^{c_{17}y} \mathbb{P}(\widetilde{M}_j \geqslant x).$$

It is straightforward to show using a Chebyshev argument that uniformly for $y \leqslant x$, with c_{17} small enough, the right-hand side is bounded by $e^{-c_{18}y}$. Inserting into (34) yields

$$\mathbb{P}(\tau_x > s + C_5 k) \geqslant p\,\mathbb{P}(\tau_x > s + C_5 k \mid S) + q - C_{10}e^{-c_{19}y}.$$

Using (33), it then holds that
$$\mathbb{P}(\tau_x > s + C_5 k \mid S) \ll e^{-c_{19}y}.$$

We conclude that for $y \leqslant x$ large enough,

$$\mathbb{P}(s - 1 \leqslant \tau_x \leqslant s + C_5 k \mid S) \geqslant 1 - \delta^y - O(e^{-c_{19}y}) \geqslant 1 - O(e^{-c_{20}y}).$$

Therefore, for some $y_0 > 0$ large enough, $\mathbb{P}(s - 1 \leqslant \tau_x \leqslant s + C_5 y_0 / C_3 \mid S) > 4/5$, so that $s - 1 \leqslant t_x^{(1/2)} \leqslant s + C_5 y_0 / C_3$. For $y > 2C_5 y_0 / C_3$, we then have

$$\mathbb{P}\left(|\tau_x - t_x^{(1/2)}| > y \mid S \right) \leqslant 1 - \mathbb{P}\left(s - 1 \leqslant \tau_x \leqslant s + \frac{y}{2} + \frac{C_5 y_0}{C_3} \mid S \right) \leqslant O(e^{-c_{20}C_3 y/(2C_5)}).$$

This establishes (5). $\qquad\square$

2.4 Proof of Theorem 3

First, let us recall our notations. Let $\Lambda(\boldsymbol{\lambda}) = \log \phi_{\boldsymbol{\xi}}(\boldsymbol{\lambda}) = \log \mathbb{E}[e^{\boldsymbol{\lambda} \cdot \boldsymbol{\xi}}]$ and $I(\mathbf{x})$ be its Fenchel-Legendre transform, which coincides with the large deviation rate function for the jump distribution $\boldsymbol{\xi}$. The unique positive solution to $I(\widehat{c}_1, \mathbf{0}) = \log \rho$ is given by $\widehat{c}_1$. Our goal is to prove that $\tau_x = \widehat{A}(x) + o_{\mathbb{P}}(\log x)$, where

$$\widehat{A}(x) := \frac{x}{\widehat{c}_1} + \left(\frac{d+2}{2\widehat{c}_1 \partial_{x_1} I(\widehat{c}_1, \mathbf{0})} \right) \log x.$$

2.4.1 Random walks in cones

For estimating τ_x, we need a few preliminary results on random walks and Brownian motion in cones. The intuitive reason why random walks in cones come into play is explained in Figure 3. In this section, we try to be self-contained and summarize only the necessary results. We refer to [6, 19, 20, 23] for more background on this topic.
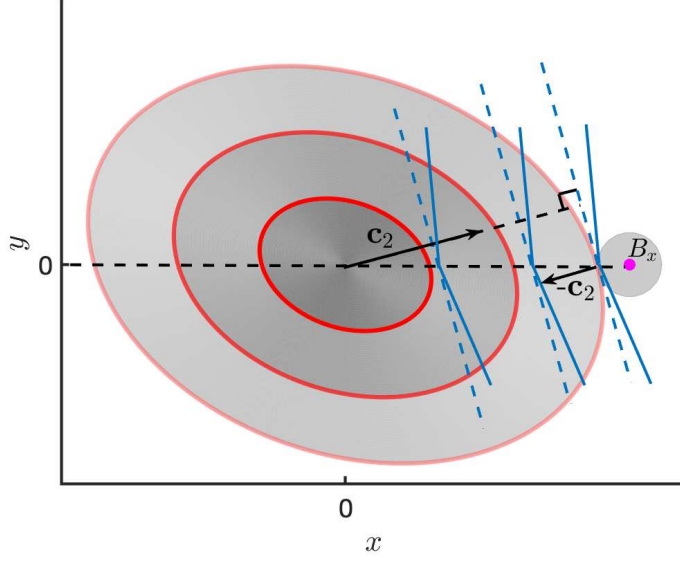
Figure 3: Visualization of how random walks in cones are related to estimating first passage times. The first passage time τ_x is roughly the time when the growing range (illustrated with the shaded ellipses) of the BRW becomes tangent with the target ball B_x. In our proof, we will construct moving cones (that are *circular* centered in the direction $-\mathbf{c}_2$ in the sense of (35); the boundaries are indicated by the blue lines) and show that with high probability, the range of the BRW always lies within the moving cones.

We first fix some notations. Let $\{\mathbf{U}_i\}_{i \geqslant 1}$ be i.i.d. copies of $\mathbf{U} = (U_1, \ldots, U_d)$ in $\mathbb{R}^d$ and $\mathbf{S}_n = \mathbf{U}_1 + \cdots + \mathbf{U}_n$ denote its partial sum. Consider an open connected subset $\mathfrak{S} \subseteq \mathbb{R}^d$ and let $\mathbb{K}$ be the cone generated by the rays emanating from $\mathbf{0} \in \mathbb{R}^d$ that go through $\mathfrak{S}$. For $\mathbf{z} \in \mathbb{K}$, let

$$\tau_{\mathbf{z}}^{\mathbb{K}} = \min\{n \geqslant 1 : \mathbf{z} + \mathbf{S}_n \notin \mathbb{K}\}$$

be the first passage time of the random walk $\{\mathbf{S}_n\}_{n \geqslant 1}$ to $\mathbb{K}^c$, starting from $\mathbf{z} \in \mathbb{K}$. Similarly, let $\tau_{\mathbf{z}}^{\mathbb{K},\mathrm{BM}}$ be the first passage time of the standard d-dimensional Brownian motion to $\mathbb{K}^c$, starting from $\mathbf{z} \in \mathbb{K}$. We will frequently consider cones that are *circular* (in the sense of [20]), that is, of the form

$$\mathbb{K}_\alpha(\mathbf{v}) := \{\mathbf{z} \in \mathbb{R}^d : 0 \leqslant \theta < \alpha\}, \tag{35}$$

where $\alpha \in (0, \pi), \mathbf{0} \neq \mathbf{v} \in \mathbb{R}^d$, and $\theta = \theta(\mathbf{z}, \mathbf{v})$ is the angle between nonzero vectors $\mathbf{z}$ and $\mathbf{v}$ in $\mathbb{R}^d$. In the case $\mathbf{v} = (1, \mathbf{0})$, we may write $\mathbb{K}_\alpha = \mathbb{K}_\alpha((1, \mathbf{0}))$.

The following result combines several results of [23].

Lemma 14. *Suppose that $\mathbf{U}$ is centered, non-lattice, and $\mathbb{E}[U_i U_j] = \mathbb{1}_{\{i=j\}}$. Let $\mathbf{x}, \mathbf{y} \in \mathbb{K}$. There exist positive functions $V, \widetilde{V}$ satisfying $\max(V(\mathbf{x}), \widetilde{V}(\mathbf{x})) \leqslant C(1 + \|\mathbf{x}\|)^p$ and*

$$\mathbb{P}(\mathbf{x} + \mathbf{S}_n \in B_{\mathbf{y}}, \tau_{\mathbf{x}}^{\mathbb{K}} > n) \ll \frac{V(\mathbf{x})\widetilde{V}(\mathbf{y})}{n^{p+d/2}}, \tag{36}$$

where $p = p(\mathbb{K})$ is such that $\mathbb{P}(\tau_{\mathbf{x}}^{\mathbb{K},\mathrm{BM}} > t) \asymp t^{-p/2}$. Moreover, for fixed $\mathbf{y} \in \mathbb{K}$,

$$\mathbb{P}(\mathbf{x} + \mathbf{S}_n \in B_{\mathbf{y}}, \tau_{\mathbf{x}}^{\mathbb{K}} > n) \asymp \frac{V(\mathbf{x})}{n^{p+d/2}}. \tag{37}$$

Proof. The asymptotic (37) is a direct consequence of (1), Theorem 6, and Lemma 13(b) of [23]. On the other hand, (36) follows from the derivation of Lemmas 27 and 28 therein, along with their Theorem 1. The results in [23] are stated in the lattice case but non-lattice walks can be proved in the same way. □

Remark 2. As commented in [23], if $\mathbf{U}$ does not satisfy $\mathbb{E}[U_i U_j] = \mathbb{1}_{\{i=j\}}$ but has a positive definite covariance matrix, then the same results hold with $\mathbb{K}$ replaced by $M\mathbb{K} = \{M\mathbf{x} : \mathbf{x} \in \mathbb{K}\}$, where $\mathbf{V} = M\mathbf{U}$ satisfies $\mathbb{E}[V_i V_j] = \mathbb{1}_{\{i=j\}}$.

The exponent $p(\mathbb{K})$ is in general not explicit and is related to the smallest eigenvalue of the Laplace-Beltrami operator; see [23]. However, it suffices for our purpose to understand the behavior of $p(\mathbb{K})$ when $\mathbb{K}$ is a circular cone that is close to a half-space. The following result shows that $p \approx 1$ if the circular cone $\mathbb{K}$ is close enough to a half-space.

Lemma 15 (Section 3 of [19]). *Let $\alpha \in (0, \pi)$ and $\mathbf{z} \in \mathbb{K}_\alpha$ be arbitrary. There exists a decreasing and continuous[9] function $p_{\alpha,d}$ such that $p_{\pi/2,d} = 1$ and*

$$\mathbb{E}[(\tau_{\mathbf{z}}^{\mathbb{K}_\alpha})^{q/2}] < \infty \quad \Longleftrightarrow \quad q\alpha < \frac{\pi}{2}.$$

The following ballot-type lemma provides bounds when the cone is shifted by a function of time, borrowing ideas from Lemma 2.3 of [13]. We do not attempt the sharpest bound here, but state the minimal result needed for our purpose.

Lemma 16. *Let $\mathbf{v} \in \mathbb{R}^d$, $\mathbf{v} \neq \mathbf{0}$, and $\alpha \in (0, \pi)$. Let $p = p(\mathbb{K}_\alpha(-\mathbf{v}))$ be the exponent defined in Lemma 14, $p' < p$, and $C > 0$. Then uniformly in $\mathbf{z}, \mathbf{y} \in \mathbb{R}^d$ such that $\mathbf{z} - \mathbf{y} \in \mathbb{K}_\alpha(-\mathbf{v})$ and $n \in \mathbb{N}$,*

$$\mathbb{P}(\mathbf{S}_n \in B_{\mathbf{z}}; \forall 1 \leqslant k \leqslant n, \mathbf{S}_k - \mathbf{y} - C\log(\min(k, n-k))_+ \mathbf{v} \in \mathbb{K}_\alpha(-\mathbf{v}))$$
$$\ll \frac{(1 + \|\mathbf{z}\|)^p (1 + \|\mathbf{y}\|)^p}{n^{p'+d/2}}. \tag{38}$$

Moreover, uniformly for $n \in \mathbb{N}$, $1 \leqslant \ell \leqslant n$, and $\mathbf{z}, \mathbf{y} \in \mathbb{R}^d$ with $\mathbf{z} - \mathbf{y} \in C\log(\min(\ell, n-\ell))_+ \mathbf{v} + \mathbb{K}_\alpha(-\mathbf{v})$,

$$\mathbb{P}(\mathbf{S}_\ell \in B_{\mathbf{z}}; \forall 1 \leqslant k \leqslant \ell, \mathbf{S}_k - \mathbf{y} - C\log(\min(k, n-k))_+ \mathbf{v} \in \mathbb{K}_\alpha(-\mathbf{v}))$$
$$\ll \frac{(1 + \|\mathbf{z}\|)^p (1 + \|\mathbf{y}\|)^p}{n^{p'+d/2}}. \tag{39}$$

Proof. Let $\tilde{\mathbf{y}} = -\mathbf{y} - (C\log n)\mathbf{v}$ and $\tilde{\mathbf{z}} = -\mathbf{z} - (C\log n)\mathbf{v}$. We have by using (36) of Lemma 14,

$$\mathbb{P}(\mathbf{S}_n \in B_{\mathbf{z}}; \forall 1 \leqslant k \leqslant n, \mathbf{S}_k - \mathbf{y} - C\log(\min(k, n-k))_+ \mathbf{v} \in \mathbb{K}_\alpha(-\mathbf{v}))$$
$$\leqslant \mathbb{P}(\mathbf{S}_n \in B_{\mathbf{z}}; \forall 1 \leqslant k \leqslant n, \mathbf{S}_k - \mathbf{y} - (C\log n)\mathbf{v} \in \mathbb{K}_\alpha(-\mathbf{v}))$$
$$= \mathbb{P}(\tilde{\mathbf{y}} + \mathbf{S}_n \in B_{\tilde{\mathbf{z}}}; \forall 1 \leqslant k \leqslant n, \tilde{\mathbf{y}} + \mathbf{S}_k \in \mathbb{K}_\alpha(-\mathbf{v}))$$
$$\ll \frac{(1 + \|\tilde{\mathbf{y}}\|)^p (1 + \|\tilde{\mathbf{z}}\|)^p}{n^{p+d/2}}$$
$$\ll \frac{(1 + \|\mathbf{y}\|)^p (1 + \|\mathbf{z}\|)^p}{n^{p'+d/2}}.$$

This completes the proof of (38). The proof of (39) is similar and is omitted here for brevity. □

---

[9]The continuity is not explicitly stated in [19], but follows from the continuity of the hypergeometric function.

2.4.2 A large deviation computation

In the following, we work with a sequence $\{\boldsymbol{\xi}_j = (X_j, \mathbf{Y}_j)\}_{j \in \mathbb{N}}$ of i.i.d. centered random variables in $\mathbb{R}^d$ that are not necessarily spherically symmetric but have densities belonging to $L^r(\mathbb{R}^d)$ for some $r > 1$ and a positive definite covariance matrix V. Let $\Lambda(\boldsymbol{\lambda}) = \log \phi_{\boldsymbol{\xi}}(\boldsymbol{\lambda}) = \log \mathbb{E}[e^{\boldsymbol{\lambda} \cdot \boldsymbol{\xi}}]$ and $I(\mathbf{x})$ be its Fenchel-Legendre transform.[10] Let also $\mathbf{c}_2 = \nabla I(x/n, \mathbf{0})$. By definition,

$$\Lambda(\mathbf{c}_2) = \mathbf{c}_2 \cdot \left(\frac{x}{n}, \mathbf{0} \right) - I \left(\frac{x}{n}, \mathbf{0} \right). \tag{40}$$

We have the following counterpart of Proposition 10. The proof is similar and deferred to the appendix.

Proposition 17. *Fix a large constant $C > 0$ and a dimension $d \geqslant 1$. Suppose that the $\mathbb{R}^d$-valued random variable $\boldsymbol{\xi}$ satisfies conditions (A2), (A5), and (A6) from Section 1.1. Let $(X_j, \mathbf{Y}_j)$ be a sequence of i.i.d. random variables with the same distribution as $\boldsymbol{\xi}$ and $(X, \mathbf{Y}) = \sum_{j=1}^n (X_j, \mathbf{Y}_j)$. Fix $\overline{c} > \underline{c} > 0$.*

(i) *Uniformly for any $c = c(n) \in [-C \log n, C \log n]$ and any $x \in [\underline{c}\, n, \overline{c}\, n]$,*

$$\mathbb{P}(\mathbf{Y} \in B_{\mathbf{0}}(1) \mid X \geqslant x + c) \ll e^{-nI(x/n,\mathbf{0}) - \mathbf{c}_2 \cdot (c,\mathbf{0})} n^{-(d-1)/2} e^{nI(x/n) + cI'(x/n)}. \tag{41}$$

(ii) *Fix $a(n) \to \infty$ and $a(n) = o(\log n)$. Uniformly for $\mathbf{y}$ with $\|\mathbf{y}\| \leqslant C\sqrt{n}$ and $c \in [-C \log n, C \log n]$, and any $x \in [\underline{c}\, n, \overline{c}\, n]$,*

$$\mathbb{P}(\mathbf{Y} \in B_{\mathbf{y}}(1) \mid X \in [x + c, x + c + a(n)])$$
$$\gg e^{-nI(x/n,\mathbf{0}) - \mathbf{c}_2 \cdot (c,\mathbf{y})} n^{-(d-1)/2} e^{nI(x/n) + cI'(x/n)}. \tag{42}$$

2.4.3 Upper bound of Theorem 3

Let $\varepsilon > 0$ be arbitrary and

$$n = \widehat{A}_1(x) = \frac{x}{\widehat{c}_1} + \left(\frac{d + 2 + 2\varepsilon}{2\widehat{c}_1 \partial_{x_1} I(\widehat{c}_1, \mathbf{0})} \right) \log x. \tag{43}$$

Our goal is to show that the event $\{\tau_x > \widehat{A}_1(x)\}$ has a small probability. Define $\mathbf{c}_2 := \nabla I(x/n, \mathbf{0})$. By definition,

$$\Lambda(\mathbf{c}_2) = \mathbf{c}_2 \cdot \left(\frac{x}{n}, \mathbf{0} \right) - I \left(\frac{x}{n}, \mathbf{0} \right). \tag{44}$$

We set up a barrier event. Pick $\alpha \in (0, \pi/2)$ such that $p_{\alpha,d} = 1 + \varepsilon$; cf., Lemma 15. Recall (35). Let

$$\widehat{P}_n := \left\{ v \in V_n : \ \forall 0 \leqslant k \leqslant n, \ \boldsymbol{\eta}_{v,n}(k) - \frac{k\mathbf{x}}{n} \in \mathbb{K}_\alpha(-\mathbf{c}_2) \right\}$$

and let

$$\widehat{\Delta}_{n,x} := \sum_{v \in \widehat{P}_n} \mathbb{1}_{\widehat{H}_{v,n}(x)} := \sum_{v \in \widehat{P}_n} \mathbb{1}_{\{\boldsymbol{\eta}_{v,n}(n) \in B_x\}} \tag{45}$$

be the number of particles lying in the target B_x at time $n = \widehat{A}_1(x)$ that is controlled by the barrier given by $\widehat{P}_n$. The quantity $\widehat{\Delta}_{n,x}$ serves as a lower bound for the number of particles lying in the target B_x at time n. As a consequence, it suffices to show that

$$\mathbb{P}(\widehat{\Delta}_{n,x} \geqslant 1) > 1/C \tag{46}$$

---

[10]Not to confuse with the functions Λ and I for the first coordinate which is defined on $\mathbb{R}$, since we may assume $d \geqslant 2$.

for some $C > 0$ independent of x, which in turn gives the upper bound for τ_x by Theorem 2. We compute first and second moments for $\widehat{\Delta}_{n,x}$. Define

$$\widehat{Q}_n := \{v \in V_n : \ \forall 0 \leqslant k \leqslant n, \boldsymbol{\eta}_{v,n}(k) \in \mathbb{K}_\alpha(-\mathbf{c}_2)\}.$$

Applying the change of measure (71) and using a similar change of measure argument as in the proof of Proposition 17, we have

$$\mathbb{E}[\widehat{\Delta}_{n,x}] = \rho^n \mathbb{P}\left(v \in \widehat{P}_n, \boldsymbol{\eta}_{v,n}(n) \in B_x\right) \gg \rho^n e^{-nI(x/n,\mathbf{0})} \mathbb{Q}\left(v \in \widehat{Q}_n, \|\boldsymbol{\eta}_{v,n}(n)\| \leqslant 1\right).$$

By (37) of Lemma 14,

$$\mathbb{Q}\left(v \in \widehat{Q}_n, \|\boldsymbol{\eta}_{v,n}(n)\| \leqslant 1\right) \asymp n^{-p_{\alpha,d}-d/2} \asymp x^{-p_{\alpha,d}-d/2}.$$

Here, we may without loss of generality replace $\mathbf{c}_2$ by $M\mathbf{c}_2$ where $\mathbb{E}_{\mathbb{Q}}[(M\boldsymbol{\xi})_i(M\boldsymbol{\xi})_j] = \mathbb{1}_{\{i=j\}}$. This does not affect the analysis since an invertible linear transformation of a half-space containing $\mathbf{0}$ is a half-space containing $\mathbf{0}$ and we are in a regime where $p_{\alpha,d} \approx 1$; see Remark 2. Using (43), we have the approximation

$$nI\left(\frac{x}{n},\mathbf{0}\right) = \frac{\log \rho}{\widehat{c}_1}x + \left(\frac{(d+2+2\varepsilon)\log \rho}{2\widehat{c}_1 \partial_{x_1} I(\widehat{c}_1,\mathbf{0})} - (d+2+2\varepsilon)\right)\log x + O(1).$$

These altogether lead to

$$\mathbb{E}[\widehat{\Delta}_{n,x}] \gg \rho^n \rho^{-x/\widehat{c}_1} \rho^{-(d+2+2\varepsilon)\log x/(2\widehat{c}_1 \partial_{x_1} I(\widehat{c}_1,\mathbf{0}))} x^{(d+2+2\varepsilon)/2} x^{-p_{\alpha,d}-d/2} = 1, \tag{47}$$

where we used $p_{\alpha,d} = 1 + \varepsilon$.

Next, we compute the second moment of $\widehat{\Delta}_{n,x}$, following the same type of arguments that lead to (19). Recall the definition of $\widehat{H}_{v,n}(x)$ from (45). By conditioning on the closest integer point to $\boldsymbol{\eta}_{v,n}(n-s) \in (n-s)\mathbf{x}/n + \mathbb{K}_\alpha(-\mathbf{c}_2)$ and applying (36) of Lemma 14, for $v \sim_s w$ (recall that this means that the distance of v and w in the genealogical tree is equal to $2s$),

$$\mathbb{P}\left(\widehat{H}_{v,n}(x) \cap \widehat{H}_{w,n}(x)\right)$$

$$\ll \sum_{\mathbf{y} \in \mathbb{K}_\alpha(-\mathbf{c}_2) \cap \mathbb{Z}^d} \left(e^{-(n-s)I(x/n,\mathbf{0})-\mathbf{c}_2 \cdot \mathbf{y}}(\max(1,n-s))^{-p_{\alpha,d}-d/2}(1+\|\mathbf{y}\|)^{2p_{\alpha,d}}\right)$$

$$\times \left(e^{-sI(x/n,\mathbf{0})-\mathbf{c}_2 \cdot (-\mathbf{y})}s^{-p_{\alpha,d}-d/2}(1+\|\mathbf{y}\|)^{2p_{\alpha,d}}\right)^2$$

$$\ll e^{-(n+s)I(x/n,\mathbf{0})}(\max(1,n-s))^{-p_{\alpha,d}-d/2}s^{-2p_{\alpha,d}-d} \sum_{\mathbf{y} \in \mathbb{K}_\alpha(-\mathbf{c}_2) \cap \mathbb{Z}^d} e^{\mathbf{c}_2 \cdot \mathbf{y}}(1+\|\mathbf{y}\|)^{6p_{\alpha,d}}.$$

By the definition (35),

$$\sum_{\mathbf{y} \in \mathbb{K}_\alpha(-\mathbf{c}_2) \cap \mathbb{Z}^d} e^{\mathbf{c}_2 \cdot \mathbf{y}}(1+\|\mathbf{y}\|)^{6p_{\alpha,d}} = \sum_{j=0}^{\infty} \sum_{\substack{\mathbf{y} \in \mathbb{K}_\alpha(-\mathbf{c}_2) \cap \mathbb{Z}^d \\ \mathbf{y} \cdot (-\mathbf{c}_2) \in [j,j+1]}} e^{\mathbf{c}_2 \cdot \mathbf{y}}(1+\|\mathbf{y}\|)^{6p_{\alpha,d}}$$

$$\ll \sum_{j=0}^{\infty} e^{-j}(1+j)^{12dp_{\alpha,d}} \ll 1.$$

As a consequence and using $p_{\alpha,d} = 1 + \varepsilon$, for $v \sim_s w$,

$$\mathbb{P}\left(\widehat{H}_{v,n}(x) \cap \widehat{H}_{w,n}(x)\right) \ll e^{-(n+s)I(x/n,\mathbf{0})}(\max(1,n-s))^{-p_{\alpha,d}-d/2}s^{-2p_{\alpha,d}-d}$$

$$\ll \rho^{-(n+s)}x^{(p_{\alpha,d}+d/2)(n+s)/n}(\max(1,n-s))^{-p_{\alpha,d}-d/2}s^{-2p_{\alpha,d}-d}.$$

21

Therefore, by assumption (A1),

$$\mathbb{E}[\widehat{\Delta}_{n,x}^2] \ll \mathbb{E}[\widehat{\Delta}_{n,x}] + \sum_{s=1}^{n} \rho^{n+s}\mathbb{P}\left(\widehat{H}_{v,n}(x) \cap \widehat{H}_{w,n}(x) \text{ for } v \sim_s w\right)$$

$$\ll 1 + \sum_{s=1}^{n} x^{(p_{\alpha,d}+d/2)(n+s)/n}(\max(1, n-s))^{-p_{\alpha,d}-d/2}s^{-2p_{\alpha,d}-d} \ll 1,$$

where in the last step we argue similarly as in the proof of Lemma 2.7 of [13]. Combined with (47) and the Paley-Zygmund inequality leads to (46). This proves the desired upper bound.

2.4.4 Lower bound of Theorem 3

Fix $\varepsilon > 0$. Define

$$\widehat{A}_2(x) := \frac{x}{\widehat{c}_1} + \frac{(d+2-2\varepsilon)\log x}{2\,\widehat{c}_1\partial_{x_1}I(\widehat{c}_1, \mathbf{0})}.$$

Our goal is to show that the event $\{\tau_x \leqslant \widehat{A}_2(x)\}$ has a small probability. We first consider the number M_x of particles in B_x at time $n = \widehat{A}_2(x)$. We find $\alpha \in (0, \pi)$ such that $p_{\alpha,d} = 1 - \varepsilon/2$. Recalling $\mathbf{c}_2 = \nabla I(x/n, \mathbf{0}) \in \mathbb{R}_+ \times \mathbb{R}^{d-1}$, we may assume that $\mathbf{x}/n \notin \mathbb{K}_\alpha(-\mathbf{c}_2)$. We may decompose $M_x = \sum_{\ell=0}^{n} M_{x,\ell}$, where

$$M_{x,n} := \#\left\{v \in V_n : \boldsymbol{\eta}_{v,n}(n) \in B_x; \forall 0 \leqslant k \leqslant n, \boldsymbol{\eta}_{v,n}(k) \in \frac{k\mathbf{x}}{n} + \boldsymbol{\psi}_n(k) + \mathbb{K}_\alpha(-\mathbf{c}_2)\right\},$$

where $\boldsymbol{\psi}_n(k) := (C_1 + C_2(\log(\min(k, n-k)))_+)\,\mathbf{c}_2$ for some $C_1, C_2 > 0$ to be determined, and for $0 \leqslant \ell \leqslant n-1$,

$$M_{x,\ell} := \#\Bigg\{v \in V_n : \boldsymbol{\eta}_{v,n}(n) \in B_x; \forall 0 \leqslant k \leqslant \ell, \boldsymbol{\eta}_{v,n}(k) \in \frac{k\mathbf{x}}{n} + \boldsymbol{\psi}_n(k) + \mathbb{K}_\alpha(-\mathbf{c}_2),$$
$$\boldsymbol{\eta}_{v,n}(\ell+1) \notin \frac{(\ell+1)\mathbf{x}}{n} + \boldsymbol{\psi}_n(\ell+1) + \mathbb{K}_\alpha(-\mathbf{c}_2)\Bigg\}.$$

Denote by $\{\mathbf{S}_n\}_{n\geqslant 1}$ the random walk with i.i.d. increments $\{\boldsymbol{\xi}_i\}_{i\geqslant 1}$. Applying the change of measure (71) and (38) of Lemma 16, we have

$$\mathbb{E}[M_{x,n}] = \rho^n\mathbb{P}\left(\mathbf{S}_n \in B_x; \forall 0 \leqslant k \leqslant n, \mathbf{S}_k \in \frac{k\mathbf{x}}{n} + \boldsymbol{\psi}_n(k) + \mathbb{K}_\alpha(-\mathbf{c}_2)\right)$$
$$= \rho^n e^{-nI(x/n, \mathbf{0})}\mathbb{Q}\left(\mathbf{S}_n \in B_{\mathbf{0}}; \forall 0 \leqslant k \leqslant n, \mathbf{S}_k \in \boldsymbol{\psi}_n(k) + \mathbb{K}_\alpha(-\mathbf{c}_2)\right)$$
$$\ll \rho^n e^{-nI(x/n, \mathbf{0})}n^{-p_{\alpha,d}+\varepsilon/4-d/2}(1 + C_1)^{p_{\alpha,d}}$$
$$\ll x^{-\varepsilon/4}.$$

Next, we bound $\mathbb{E}[M_{x,\ell}]$ for $0 \leqslant \ell \leqslant n-1$. We condition on the location of $\boldsymbol{\eta}_{v,n}(\ell)$, which may take value in $\ell\mathbf{x}/n + \boldsymbol{\psi}_n(\ell) + \mathbb{K}_\alpha(-\mathbf{c}_2)$. This leads to

$$\mathbb{E}[M_{x,\ell}] \leqslant \rho^\ell \sum_{\substack{\mathbf{u}\in\mathbb{Z}^d \\ \mathbf{u}-(\ell\mathbf{x}/n+\boldsymbol{\psi}_n(\ell))\in\mathbb{K}_\alpha(-\mathbf{c}_2)}} \mathbb{P}\left(\forall 0 \leqslant k \leqslant \ell, \mathbf{S}_k \in \frac{k\mathbf{x}}{n} + \boldsymbol{\psi}_n(k) + \mathbb{K}_\alpha(-\mathbf{c}_2), \mathbf{S}_\ell \in B_{\mathbf{u}}\right)$$
$$\times \mathbb{P}\left(\boldsymbol{\xi} \in -\mathbf{u} + \frac{(\ell+1)\mathbf{x}}{n} + \boldsymbol{\psi}_n(\ell+1) + (\mathbb{K}_\alpha(-\mathbf{c}_2))^c\right). \tag{48}$$

We bound separately the above two probabilities for a fixed ℓ. Applying the change of measure (71) and (39) of Lemma 16,

$$\mathbb{P}\left(\forall 0 \leqslant k \leqslant \ell,\ \mathbf{S}_k \in \frac{k\mathbf{x}}{n} + \boldsymbol{\psi}_n(k) + \mathbb{K}_\alpha(-\mathbf{c}_2),\ \mathbf{S}_\ell \in B_{\mathbf{u}}\right)$$

$$= e^{-\ell I(\mathbf{u}/\ell)}\mathbb{Q}\left(\forall 0 \leqslant k \leqslant \ell,\ \mathbf{S}_k \in \boldsymbol{\psi}_n(k) + \mathbb{K}_\alpha(-\mathbf{c}_2),\ \mathbf{S}_\ell \in B_{\mathbf{u}-\ell\mathbf{x}/n}\right).$$

$$\ll e^{-\ell I(\mathbf{u}/\ell)}n^{-p_{\alpha,d}+\varepsilon/4-d/2}(1+C_1)^{p_{\alpha,d}}$$

$$\ll \rho^{-\ell}x^{-\varepsilon/4}\left\|\mathbf{u} - \frac{\ell\mathbf{x}}{n}\right\|e^{-\mathbf{c}_2\cdot(\mathbf{u}-\ell\mathbf{x}/n)}.$$

On the other hand, since the moment generating function of $\boldsymbol{\xi}$ exists in a neighborhood of $\mathbf{c}_2$, we have for some $\delta > 0$,

$$\mathbb{P}(\boldsymbol{\xi} \in B_x) \ll e^{-(1+\delta)\mathbf{c}_2\cdot\mathbf{x}}.$$

It follows that for $\mathbf{v} \in \mathbb{K}_\alpha(-\mathbf{c}_2)$,

$$\mathbb{P}\left(\boldsymbol{\xi} \in -\mathbf{v} + \frac{\mathbf{x}}{n} + \boldsymbol{\psi}_n(\ell+1) - \boldsymbol{\psi}_n(\ell) + (\mathbb{K}_\alpha(-\mathbf{c}_2))^c\right)$$

$$\ll \exp\left(-(1+\delta)\inf_{\mathbf{y}\in(\mathbb{K}_\alpha(-\mathbf{c}_2))^c}\left(\mathbf{c}_2 \cdot \left(\mathbf{y} - \mathbf{v} + \frac{\mathbf{x}}{n} + \boldsymbol{\psi}_n(\ell+1) - \boldsymbol{\psi}_n(\ell)\right)\right)\right)$$

$$\leqslant \exp\left(-(1+\delta)\inf_{\mathbf{y}\in(\mathbb{K}_\alpha(-\mathbf{c}_2))^c}(\mathbf{c}_2 \cdot (\mathbf{y} - \mathbf{v}))\right),$$

where we have used that $(x/n, \mathbf{0}) \in (\mathbb{K}_\alpha(-\mathbf{c}_2))^c$.

Inserting the above estimates back to (48) with the change of variable $\mathbf{v} = \mathbf{u} - (\ell\mathbf{x}/n + \boldsymbol{\psi}_n(\ell))$ leads to

$$\mathbb{E}[M_{x,\ell}] \ll \sum_{\mathbf{v}\in\mathbb{Z}^d\cap\mathbb{K}_\alpha(-\mathbf{c}_2)} x^{-\varepsilon/4}\|\mathbf{v} + \boldsymbol{\psi}_n(\ell)\| e^{-\mathbf{c}_2\cdot(\mathbf{v}+\boldsymbol{\psi}_n(\ell))}$$

$$\times \exp\left(-(1+\delta)\inf_{\mathbf{y}\in(\mathbb{K}_\alpha(-\mathbf{c}_2))^c}(\mathbf{c}_2 \cdot (\mathbf{y}-\mathbf{v}))\right)$$

$$\ll x^{-\varepsilon/4}\sum_{j=0}^{\infty}e^{-(1+\delta)j}\sum_{\substack{\mathbf{v}\in\mathbb{Z}^d\cap\mathbb{K}_\alpha(-\mathbf{c}_2)\\ \inf_{\mathbf{y}\in(\mathbb{K}_\alpha(-\mathbf{c}_2))^c}(\mathbf{c}_2\cdot(\mathbf{y}-\mathbf{v}))\in[j,j+1]}}\|\mathbf{v} + \boldsymbol{\psi}_n(\ell)\| e^{-\mathbf{c}_2\cdot(\mathbf{v}+\boldsymbol{\psi}_n(\ell))}$$

$$\ll x^{-\varepsilon/10}\max(1,\min(\ell,n-\ell))^{-C_2\|\mathbf{c}_2\|^2}$$

$$\times \sum_{j=0}^{\infty}e^{-(1+\delta)j}\sum_{\substack{\mathbf{v}\in\mathbb{Z}^d\cap\mathbb{K}_\alpha(-\mathbf{c}_2)\\ \inf_{\mathbf{y}\in(\mathbb{K}_\alpha(-\mathbf{c}_2))^c}(\mathbf{c}_2\cdot(\mathbf{y}-\mathbf{v}))\in[j,j+1]}}\|\mathbf{v}\| e^{-\mathbf{c}_2\cdot\mathbf{v}},$$

where the last step follows from the definition of $\boldsymbol{\psi}_n(\ell)$ and $\boldsymbol{\psi}_n(\ell) \ll \log n \asymp \log x$. By elementary geometry and properties of $\mathbb{K}_\alpha(-\mathbf{c}_2)$, for $j \geqslant 0$,

$$\sum_{\substack{\mathbf{v}\in\mathbb{Z}^d\cap\mathbb{K}_\alpha(-\mathbf{c}_2)\\ \inf_{\mathbf{y}\in(\mathbb{K}_\alpha(-\mathbf{c}_2))^c}(\mathbf{c}_2\cdot(\mathbf{y}-\mathbf{v}))\in[j,j+1]}}\|\mathbf{v}\| e^{-\mathbf{c}_2\cdot\mathbf{v}} \ll e^{-j}.$$

Together we obtain

$$\mathbb{E}[M_{x,\ell}] \ll x^{-\varepsilon/10}\max(1,\min(\ell,n-\ell))^{-C_2\|\mathbf{c}_2\|^2}\sum_{j=0}^{\infty}e^{-\delta j}$$

$$\ll x^{-\varepsilon/10}\max(1,\min(\ell,n-\ell))^{-C_2\|\mathbf{c}_2\|^2}.$$

We thus conclude with the following bound on $\mathbb{E}[M_x]$:

$$\mathbb{E}[M_x] = \sum_{\ell=0}^{n} M_{x,\ell} \ll x^{-\varepsilon/4} + x^{-\varepsilon/10} \sum_{\ell=0}^{n-1} \max(1, \min(\ell, n-\ell))^{-C_2 \|\mathbf{c}_2\|^2}.$$

With C_2 picked large enough, we have $\mathbb{E}[M_x] \ll x^{-\varepsilon/10}$.

The rest of the arguments is standard. Recall the weaker bound of (11). If a particle reaches B_x at time $n - t$ where $t = O(\log x)$, the probability that one of its descendants is located in B_x at time n is $\gg (\log x)^{-d}$. On the other hand, $\mathbb{E}[M_x] \ll x^{-\varepsilon/10} = o((\log x)^{-d})$. As a consequence, with probability $o(1)$, there is no particle present in B_x before time $n = \widehat{A}_2(x)$. This finishes the proof of the lower bound of Theorem 3.

2.5 Proofs of remaining corollaries

2.5.1 Proof of Corollary 4

Denote by $\tau_x(r)$ the first passage time of the BRW to a sphere of radius $r > 0$ centered at $\mathbf{x} = (x, \mathbf{0})$. Note that for a fixed $r > 0$, the proofs of Theorems 2 and 3 go through.

Using elementary geometry, we have

$$\inf_{x \in [n, n+1]} \tau_x(1) \geqslant \max(\tau_n(2), \tau_{n+1}(2)) \tag{49}$$

and

$$\sup_{x \in [n, n+1/2]} \tau_x(1) \leqslant \min\left(\tau_n\left(\frac{1}{2}\right), \tau_{n+1/2}\left(\frac{1}{2}\right)\right). \tag{50}$$

Applying Theorems 2 and 3, we have for any fixed $\varepsilon > 0$,

$$\sum_{n \in \mathbb{N}} \mathbb{P}\left(\left|\frac{\tau_n(2)}{n} - \frac{1}{\widehat{c}_1}\right| > \varepsilon\right) + \sum_{n \in \mathbb{N}} \mathbb{P}\left(\left|\frac{\tau_{n/2}(1/2)}{n/2} - \frac{1}{\widehat{c}_1}\right| > \varepsilon\right) < \infty.$$

By Borel–Cantelli lemma,

$$\frac{\tau_n(2)}{n} \to \frac{1}{\widehat{c}_1} \quad \text{and} \quad \frac{\tau_{n/2}(1/2)}{n/2} \to \frac{1}{\widehat{c}_1} \quad \text{a.s.}$$

Applying (49) and (50) then yields that almost surely,

$$\limsup_{x \to \infty} \frac{\tau_x(1)}{x} \leqslant \frac{1}{\widehat{c}_1} \leqslant \liminf_{x \to \infty} \frac{\tau_x(1)}{x}.$$

This proves Corollary 4.

2.5.2 Proof of Corollary 5

Denote by E_n the event that the range

$$R_n = \{\boldsymbol{\eta}_{v,n}(n) : v \in V_n\}$$

is not $(1/n)$-dense in $(n/C_2)B_{\mathbf{0}}$ (meaning that there exists $\mathbf{y} \in (n/C_2)B_{\mathbf{0}}$ with $(\mathbf{y} + (1/n)B_{\mathbf{0}}) \cap R_n = \emptyset$). To bound $\mathbb{P}(E_n)$ from above, we need a slight modification of Theorem 3 that applies for balls of radii $1/n$ instead of a fixed radius. By performing the same analysis in Section 2.4.3, we have the upper bound

$$\tau_{x,1/n} \leqslant \frac{x}{\widehat{c}_1} + C_1 \log \log x + O_{\mathbb{P}}(1),$$

24

where $\tau_{x,1/n}$ denotes the first passage time to a ball of radius $1/n$ centered at $(x, \mathbf{0})$ and $C_1 > 0$. We need to show that $\widehat{c}_1$ is large in all directions. For a $d \times d$ orthogonal matrix Θ, denote by $\widehat{c}_1(\Theta)$ the $\widehat{c}_1$ corresponding to the BRW with jump distribution $\Theta\boldsymbol{\xi}$. It follows from assumption (A5) that the domain of I contains a neighborhood of $\mathbf{0}$, and hence

$$\inf_{\Theta} \widehat{c}_1(\Theta) > 0. \tag{51}$$

Applying (51) gives that there exists $C_2 > 0$, so that $\tau_{x,1/n,\boldsymbol{\theta}} \leqslant (C_2/2)x + O_{\mathbb{P}}(1)$ in all directions $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$. Therefore, by applying Theorem 2,

$$\mathbb{P}(E_n) \leqslant \sum_{j=1}^{n/C_2} Cj^{d-1} \sup_{\boldsymbol{\theta} \in \mathbb{S}^{d-1}} \mathbb{P}(\tau_{j,1/n,\boldsymbol{\theta}} \geqslant n) \leqslant \sum_{j=1}^{n/C_2} Cj^{d-1} e^{-n/C_3} \ll e^{-n/C_4}$$

for some $C_3, C_4 > 0$. Thus $\sum_n \mathbb{P}(E_n) < \infty$. The rest arguments follow in the same way as Corollary 2.5 of [43].

Remark 3. We expect that using a similar argument, one can prove *uniform* versions of (4) and (7) under stronger assumptions on I. For instance,

$$\sup_{\Theta \in \mathbb{S}^{d-1}} \frac{1}{\log x} \Big| \tau_{\Theta\mathbf{x}} - \Big(\frac{x}{\widehat{c}_1(\Theta)} + \frac{d+2}{2\widehat{c}_1(\Theta)\partial_{x_1} I_{\Theta}(\widehat{c}_1(\Theta), \mathbf{0})} \log x \Big) \Big| \to 0 \quad \text{in probability},$$

where I_{Θ} is the large deviation rate function of $\Theta\boldsymbol{\xi}$.

3 Extensions and applications

3.1 The delayed branching model

In this section, we revisit the applications of BRW to modeling polymer networks. In particular, the FPT for the BRW to reach a unit ball at a distance of x can approximate the shortest path (SP) between cross-links placed at a distance x away from each other in a polymer network. Experimentally, obtaining the SP statistics is infeasible. However, extracting the SP statistics from computational simulations of polymer networks using coarse-grained molecular dynamics (CGMD) has been shown promising [50, 51]. As argued in [54], the SP statistics can be reproduced by FPT statistics of a small variation of the classical BRW model (the reason of which will be explained below), which we refer to as the *delayed branching* model.

Consider the classical BRW model satisfying $p_3 > 0$ and $p_0 + p_1 + p_3 = 1$ (i.e., only branching instances are into three particles) with a spherically symmetric jump distribution (for simplicity). The choice of branching into three particles is to be consistent with the real polymer network being modeled where branching represents the cross-link (bond) between two polymer chains. The first particle would correspond to the continuation of the first polymer chain beyond the point of cross-link whereas, the other two particles would correspond to the segments of the second polymer chain (see [54] for a detailed description).

The classical BRW assumes that the three particles emerge from the same point. However, the more realistic picture states that the first and second chains are displaced at the point of branching due to the presence of the cross-link. The *delayed branching* model is a modification of such a BRW process, where at each branching event (in which the particle was supposed to give rise to three offsprings), a particle branches into two, and one of the two descendants branches again into two particles in the next time step.[11] That is, the branching event (into three particles) is broken into two pieces, one of which is *delayed* by one step. Thus, each branching event consists of two sub-events,

---

[11]A particle may die in each of the two branching sub-events.

which we call types I and II respectively. Here, type I branching corresponds to the branching of the parent, and type II branching corresponds to the branching of one of the two children born in the type I sub-event.
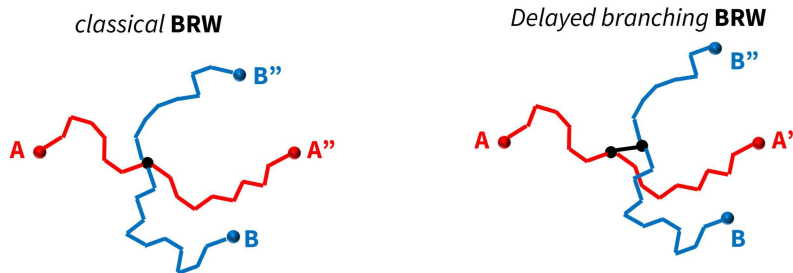

Figure 4: BRW tree representations of polymer networks in the *classical* and the *delayed branching* regimes. The red path from A to A'' represents the first polymer chain and the blue path from B to B'' represents the second chain that cross-links with the first. In the BRW tree genealogy, the particles A'', B, and B'' are the descendants of the particle starting from A.

We first derive the analogue for ρ (the parameter governing the growth of the number of particles) in the classical BRW. By construction, the expected number $\tilde{N}_n$ of particles in the delayed branching model at time n satisfies

$$\tilde{N}_{n+1} = p_1 \tilde{N}_n + p_3 \tilde{N}_n + 2 p_3 \tilde{N}_{n-1}(1 - p_0). \tag{52}$$

The characteristic polynomial of the recursive equation (52) is given by

$$x^2 - (p_1 + p_3)x - 2p_3(1 - p_0). \tag{53}$$

Denote the unique positive zero of (53) by

$$\tilde{\rho} = \frac{1 - p_0}{2} + \sqrt{\frac{(1 - p_0)^2}{4} + 2p_3(1 - p_0)}. \tag{54}$$

In particular, $\tilde{N}_n \asymp \tilde{\rho}^n$.

Let $\tilde{c}_1$ satisfy $I(\tilde{c}_1) = \log \tilde{\rho}$ where the rate function I is defined in (1), and $\tilde{c}_2 = I'(\tilde{c}_1)$. We define $\tilde{A}(x)$ similarly as in (3) with c_1, c_2 replaced by $\tilde{c}_1$, $\tilde{c}_2$. The parameters p_3 and p_0 are denoted by $\tilde{\kappa}$ and $\tilde{\nu}$ in [54] respectively, where $\tilde{\kappa}$ represents the *branching rate* of the BRW and $\tilde{\nu}$ represents the *death rate* of the branches to account for the finite length of the polymer chains in the CGMD simulation. Let $\tilde{\tau}_x$ be the first passage time to B_x.

Theorem 18. *Assume (A3) and (A4) with ρ replaced by $\tilde{\rho}$. Conditioned upon survival, we have the asymptotic*

$$\tilde{\tau}_x = \tilde{A}(x) + O_{\mathbb{P}}(\log \log x) = \frac{x}{\tilde{c}_1} + \frac{d + 2}{2 \tilde{c}_2 \tilde{c}_1} \log x + O_{\mathbb{P}}(\log \log x). \tag{55}$$

In other words, the collection $\{(\tilde{\tau}_x - \tilde{A}(x))/\log \log x\}_{x > 0}$ is tight.

Theorem 18 partially resolves Conjecture 1 of [54]. We sketch its proof below, which will be very similar to the proof of Theorem 1. A key ingredient in the proof of Theorem 1 is the analysis of the expected number N_n of particles present at time n, conditioned upon survival. If the branching events are independent across layers, then $N_n = \rho^n$. On the other hand, for the delayed branching

BRW, due to the two-step branching regime, the branching events are not independent. The following two lemmas perform a similar analysis related to the number of particles for the delayed branching model. In particular, the analogue for our expected number of descendants for each particle, ρ, becomes $\tilde{\rho}$ for the delayed branching model.

Lemma 19 (Lemma 5 of [54]). *Conditioned upon survival, the expected number of particles N_n at time n satisfies $N_n \asymp \tilde{\rho}^n$.*

Lemma 20. *For $1 \leqslant s \leqslant n$, let*

$$N_{n,s} := \mathbb{E}[\#\{(v,w) \in V_n^2 : v \sim_s w\}]$$

be the expected number of pairs of particles at level n whose distance in the genealogical tree is equal to $2s$. Conditioned upon survival, $N_{n,s} \asymp \tilde{\rho}^{n+s}$ uniformly in $1 \leqslant s \leqslant n$.

Proof. By Lemma 19, the expected number of choices of the common ancestor u of v, w is $N_{n-s} \asymp \tilde{\rho}^{n-s}$. We now fix u. Let S_u denote the survival event of the underlying branching process initiated from u. Let $N_{u,s}$ and $N'_{u,s}$ denote the number of particles initiated from u in s steps conditioned on S and S_u respectively. By a coupling argument, if u initiates a branching sub-event of type I,

$$\mathbb{E}[N_{u,s}\mathbb{1}_S] \geqslant \mathbb{E}[N'_{u,s}\mathbb{1}_{S_u}] \asymp N_s \asymp \tilde{\rho}^s.$$

Moreover, $\mathbb{E}[N_{u,s}\mathbb{1}_S]$ can be controlled by the expected number of particles without conditioning upon survival, which is $\asymp \tilde{\rho}^s$. Otherwise, if u initiates a branching sub-event of type II, an extra factor of 2 exists but will be absorbed into the asymptotics. Therefore we conclude that $\mathbb{E}[N_{u,s} \mid S] \asymp \tilde{\rho}^s$ (that $\mathbb{P}(S) \in (0,1)$ has been proven in Lemma 5 of [54]), and hence $N_{n,s} \asymp \tilde{\rho}^{n-s}(\tilde{\rho}^s)^2 = \tilde{\rho}^{n+s}$. $\qquad\square$

We briefly identify the modifications required for the proof of Theorem 18. Note that the only difference between the classical BRW and delayed branching BRW is the branching structure, i.e., the underlying tree that describes the genealogy of the particles. Given that Theorem 2 holds, the proof of Theorem 1 depends on such structure only through (17) and (18). The corresponding modifications can be made by replacing ρ by $\tilde{\rho}$ in view of Lemmas 19 and 20. The proof of Theorem 2 depends on large deviation estimates for the maximum of delayed branching BRW in (28). Nevertheless, this applies by performing the same proof as in [29] by comparing against the maximum of independent random walks (Theorems 3.1 and 5.2 therein). The other changes needed in Theorem 2 are minor and will be omitted for brevity, as the constants can always be adjusted to take into account the delayed branching property. Otherwise, the proof of Theorem 18 is almost verbatim.

The same arguments apply for non-spherically symmetric jumps, where we instead assume (A1), (A2), (A5), and (A6) and replace Theorem 1 by Theorem 3.

3.2 Finer asymptotics for non-spherically symmetric jumps

We discuss a few cases with non-spherically symmetric jumps where the asymptotic (7) can be improved. The only instance where we explicitly required spherical symmetry of the law of $\boldsymbol{\xi}$ for the proof of Theorem 1 lies in the proof of Proposition 10, where we apply the change of measure (65). More precisely, applying the exponential tilt only in the x-coordinate in (65) does not guarantee that the random variable $\boldsymbol{\xi}$ is centered in other coordinates under the new measure; cf. (71).

In certain nice cases, the non-spherically symmetric case can be analyzed using the same approach. Let I be the large deviation rate function for the jump $\boldsymbol{\xi}$ and $\hat{c}_1$ be such that $I(\hat{c}_1, \mathbf{0}) = \log \rho$. One may also define the constant c_1 that arose in assumption (A4) for a non-spherically symmetric jump $\boldsymbol{\xi}$, through the relation $I^{(1)}(c_1) = \log \rho$ where $I^{(1)}$ is the large deviation rate function for the first marginal of $\boldsymbol{\xi}$. If $\hat{c}_1 = c_1$, we can show using essentially the same proof of Theorem 1 that

$$\tau_x = \frac{x}{\hat{c}_1} + \frac{d+2}{2\hat{c}_1 \partial_{x_1} I(\hat{c}_1, \mathbf{0})} \log x + O_{\mathbb{P}}(\log \log x).$$

The following proposition gives a necessary and sufficient condition for this to happen for all $\rho > 1$.

Proposition 21. *It holds that $\widehat{c}_1 = c_1$ for all $\rho > 1$ if and only if $\mathbb{E}[\xi_i \mid \xi_1] = 0$ for all $2 \leqslant i \leqslant d$, where ξ_i denotes the i-th coordinate of the $\mathbb{R}^d$-valued random variable $\boldsymbol{\xi}$.*

Proof. Since $I^{(1)}$ is strictly increasing on its domain intersecting $(0, \infty)$, an equivalent condition for $\widehat{c}_1 = c_1$ for all ρ is that $I^{(1)}(c_1) = I(c_1, \mathbf{0})$ for all c_1. By definition, this amounts to

$$\sup_{\lambda_1 \in \mathbb{R}} \left(\lambda_1 c_1 - \mathbb{E}[e^{\lambda_1 \xi_1}] \right) = \sup_{\lambda_1, \ldots, \lambda_d \in \mathbb{R}} \left(\lambda_1 c_1 - \mathbb{E}[e^{\boldsymbol{\lambda} \cdot \boldsymbol{\xi}}] \right).$$

Since the range of I' is $\mathbb{R}_+$, this is equivalent to

$$\forall \lambda_1 \in \mathbb{R}, \quad \inf_{\lambda_2, \ldots, \lambda_d \in \mathbb{R}} \mathbb{E}[e^{\boldsymbol{\lambda} \cdot \boldsymbol{\xi}}] = \mathbb{E}[e^{\lambda_1 \xi_1}].$$

By taking partial derivatives with respect to λ_i at $\lambda_i = 0$, $2 \leqslant i \leqslant d$, we obtain that for each $2 \leqslant i \leqslant d$, $\mathbb{E}[\xi_i e^{\lambda_1 \xi_1}] = 0$. Since this holds for all $\lambda_1 \in \mathbb{R}$, by the uniqueness of the Laplace transform, we must have $\mathbb{E}[\xi_i \mid \xi_1] = 0$. The reverse direction is obvious by Jensen's inequality. $\qquad\square$

There is yet a different method to deal with distributions on $\mathbb{R}^d$ that can be transformed into a spherically symmetric distribution under an invertible linear transformation of space. Suppose that $\boldsymbol{\xi} \overset{\text{law}}{=} T(\boldsymbol{\zeta})$ for some spherically symmetric distribution $\boldsymbol{\zeta}$ and invertible linear map T. The FPT for the BRW with jump $\boldsymbol{\xi}$ to B_x is then equivalent to the FPT for the BRW with jump $\boldsymbol{\zeta}$ to $T^{-1}(B_x)$. Since Theorem 1 holds with the radius of the ball replaced by any constant (with the same proof) and T is invertible, we may without loss of generality replace $T^{-1}(B_x)$ by $B_{T^{-1}\mathbf{x}}$, where $\mathbf{x} = (x, \mathbf{0})$. This leads to

$$\tau_x = \frac{\left\| T^{-1}\mathbf{x} \right\|}{\widehat{c}_1(\boldsymbol{\zeta})} + \frac{d+2}{2\widehat{c}_1(\boldsymbol{\zeta})\partial_{x_1} I_{\boldsymbol{\zeta}}(\widehat{c}_1(\boldsymbol{\zeta}), \mathbf{0})} \log x + O_{\mathbb{P}}(\log \log x). \tag{56}$$

Summarizing the discussions above, we arrive at the following.

Theorem 22. *Assume assumption (A1). Suppose that either of the following holds:*

   (i) *the law of $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_d)$ satisfies $\mathbb{E}[\xi_i \mid \xi_1] = 0$ a.s. for all $2 \leqslant j \leqslant d$, and satisfies (A2) and (A4);*

   (ii) *there exists an invertible linear transformation T on $\mathbb{R}^d$ with $\boldsymbol{\xi} \overset{\text{law}}{=} T(\boldsymbol{\zeta})$ for some spherically symmetric distribution $\boldsymbol{\zeta}$ satisfying (A2)–(A4).*

Then the first passage time τ_x has the asymptotic

$$\tau_x = \frac{x}{\widehat{c}_1} + \frac{d+2}{2\widehat{c}_1 \partial_{x_1} I(\widehat{c}_1, \mathbf{0})} \log x + O_{\mathbb{P}}(\log \log x). \tag{57}$$

Proof. By comparing (56) and (57), it suffices to show for case (ii) that

$$\frac{x}{\widehat{c}_1(\boldsymbol{\xi})} = \frac{\left\| T^{-1}\mathbf{x} \right\|}{\widehat{c}_1(\boldsymbol{\zeta})} \tag{58}$$

and that

$$\widehat{c}_1(\boldsymbol{\xi})\partial_{x_1} I_{\boldsymbol{\xi}}(\widehat{c}_1(\boldsymbol{\xi}), \mathbf{0}) = \widehat{c}_1(\boldsymbol{\zeta})\partial_{x_1} I_{\boldsymbol{\zeta}}(\widehat{c}_1(\boldsymbol{\zeta}), \mathbf{0}). \tag{59}$$

Let $\mathbf{e}_1 = (1, \mathbf{0}) \in \mathbb{R}^d$. By definition (2) and since T is invertible, $I_{\boldsymbol{\xi}}(\mathbf{x}) = I_{\boldsymbol{\zeta}}(T^{-1}\mathbf{x})$. By spherical symmetry of $\boldsymbol{\zeta}$,

$$I_{\boldsymbol{\xi}}(\widehat{c}_1(\boldsymbol{\xi}), \mathbf{0}) = I_{\boldsymbol{\zeta}}\left(\frac{T^{-1}(\widehat{c}_1(\boldsymbol{\zeta}), \mathbf{0})}{\|T^{-1}\mathbf{e}_1\|} \right) = I_{\boldsymbol{\zeta}}\left(\frac{\|T^{-1}(\widehat{c}_1(\boldsymbol{\zeta}), \mathbf{0})\|}{\|T^{-1}\mathbf{e}_1\|} \right) = I_{\boldsymbol{\zeta}}(\widehat{c}_1(\boldsymbol{\zeta}), \mathbf{0}),$$

28

proving (58). Denote the radial component of $I_{\boldsymbol{\zeta}}$ by f. By applying the multivariate chain rule, we have

$$\partial_{x_1} I_{\boldsymbol{\xi}}(\widehat{c}_1(\boldsymbol{\xi}), \mathbf{0}) = \nabla I_{\boldsymbol{\zeta}}(T^{-1}(\widehat{c}_1(\boldsymbol{\xi}), \mathbf{0})) \cdot \partial_{x_1} T^{-1}(\widehat{c}_1(\boldsymbol{\xi}), \mathbf{0})$$

$$= f'\left(\left\|T^{-1}(\widehat{c}_1(\boldsymbol{\xi}), \mathbf{0})\right\|\right) \frac{T^{-1}(\widehat{c}_1(\boldsymbol{\xi}), \mathbf{0})}{\left\|T^{-1}(\widehat{c}_1(\boldsymbol{\xi}), \mathbf{0})\right\|} \cdot T^{-1}\mathbf{e}_1$$

$$= f'\left(\widehat{c}_1(\boldsymbol{\zeta})\right) \frac{\widehat{c}_1(\boldsymbol{\xi})\left\|T^{-1}\mathbf{e}_1\right\|^2}{\widehat{c}_1(\boldsymbol{\zeta})}$$

$$= \partial_{x_1} I_{\boldsymbol{\zeta}}(\widehat{c}_1(\boldsymbol{\zeta}), \mathbf{0})\left\|T^{-1}\mathbf{e}_1\right\|.$$

Combined with (58), this establishes (59). $\qquad\square$

Condition (i) handles the case where the law of $\boldsymbol{\xi}$ is given by a product measure. The condition (ii) above is satisfied, e.g., for centered non-degenerate elliptical distributions. In particular, this applies to non-degenerate Gaussian distributions, which we explicitly compute in the following example.

Example 23. Suppose that $\boldsymbol{\xi}$ is centered Gaussian with an invertible covariance matrix $\Sigma = AA^{\top}$. In this case, $\boldsymbol{\xi} = A\boldsymbol{\zeta}$ where $\boldsymbol{\zeta}$ is standard Gaussian. In this case, $I_{\boldsymbol{\xi}}(\mathbf{x}) = \mathbf{x}^{\top}\Sigma^{-1}\mathbf{x}/2$, so that $\widehat{c}_1 = \sqrt{2\log\rho/(\Sigma^{-1})_{11}}$ and $\partial_{x_1}I(\widehat{c}_1, \mathbf{0}) = \sqrt{2(\log\rho)(\Sigma^{-1})_{11}}$. By Theorem 22, we have

$$\tau_x = \frac{x}{\sqrt{2\log\rho/(\Sigma^{-1})_{11}}} + \frac{d+2}{4\log\rho}\log x + O_{\mathbb{P}}(\log\log x).$$

One can check indeed that this also follows from (56), using $\left\|A^{-1}\mathbf{x}\right\| = x\sqrt{(\Sigma^{-1})_{11}}$ where $\mathbf{x} = (x, \mathbf{0})$.

4 Numerical analysis

In this section, we test Theorem 3 numerically. Since tracking the locations of all particles in a BRW is time-consuming and memory-intensive, we introduce a path-purging algorithm that removes particles that are highly unlikely to have descendants that realize the first passage time.[12] A preliminary version of the algorithm for spherically symmetric jumps can be found in Section 2.3 of [54].

Recall that I denotes the large deviation rate function of the jump distribution $\boldsymbol{\xi}$. In the path-purging algorithm, we purge along the normal vector

$$\mathbf{n} = \nabla I(\widehat{c}_1, \mathbf{0}). \tag{60}$$

That is, we retain particles that have the $q_c/2$ largest inner product with $\mathbf{n}$, where q_c is a prefixed large number. To intuitively justify this, suppose that a particle is found at $\mathbf{x}' = (x', \mathbf{0})$ where $x' < x$. The expected time for the BRW initiated at $\mathbf{x}'$ to arrive at B_x is around $(x - x')/\widehat{c}_1$. In order for another particle located at $\mathbf{x}' + \mathbf{v}$ to reach B_x within roughly the same amount of time, we need

$$e^{-(x-x')I(\widehat{c}_1, \mathbf{0})/\widehat{c}_1} = e^{-(x-x')I(\widehat{c}_1(\mathbf{x}-\mathbf{x}'-\mathbf{v})/(x-x'))/\widehat{c}_1}.$$

For small $\mathbf{v}$, this holds if and only if $\mathbf{v} \cdot \nabla I(\widehat{c}_1, \mathbf{0}) = 0$, independent of the choice of x'.

If $\boldsymbol{\xi}$ is spherically symmetric, or satisfies condition (i) of Theorem 22, the normal vector $\mathbf{n}$ is then parallel to the direction towards the termination site $\mathbf{x} = (x, \mathbf{0})$.

In our implementation, we take $q_c = 9000$ and consider BRW where particles only branch into three (i.e., $\{i > 1 : p_i > 0\} = \{3\}$) in dimension $d = 3$. At times, we also consider the delayed

---

[12] An efficient algorithm that simulates the tip of one-dimensional BBM has been proposed in [18]. Unfortunately, the arguments therein do not seem to carry over to general BRWs in higher dimensions.
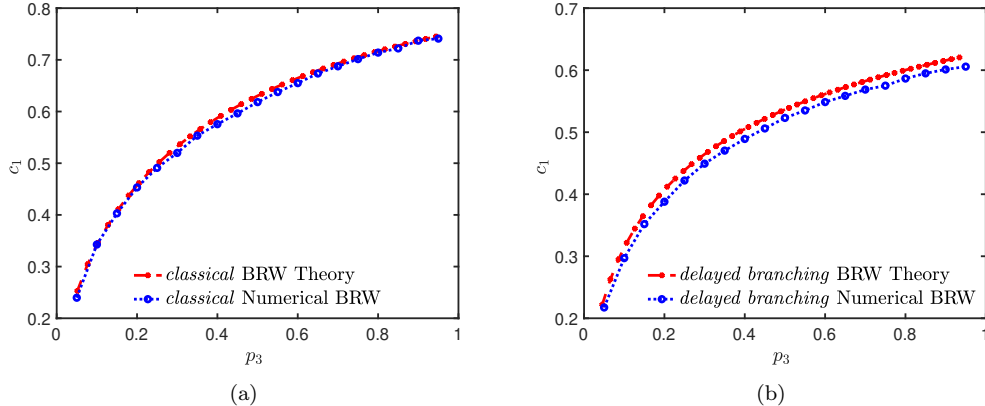
Figure 5: The numerical 3D BRW obtained c_1 compared against the reference calculations at different p_3 for the (a) classical BRW and (b) delayed branching BRW with $p_0 = 0.004$. The mean of the FPT at each x is computed from 2000 samples.

branching model from Section 3.1. We compute the first passage times for offspring distributions with $p_3 \in \{0.05\ell : 1 \leqslant \ell \leqslant 19\}$ and $p_1 = 1 - p_3$ (meaning $p_0 = 0$, unless otherwise stated) while varying x and the distribution of $\boldsymbol{\xi}$. We showcase the following two cases:

- In Section 4.1, we discuss the case where $\boldsymbol{\xi}$ is uniformly distributed on $\mathbb{S}^2$.

- In Section 4.2, we deal with Gaussian distributions with a general covariance matrix.

Finally, we note that the main result of [30] shows that there rarely exists an infinite path in a one-dimensional BRW that always stays above the curve $n \mapsto (c_1 - \varepsilon)n$ for $\varepsilon > 0$ small. While this suggests that our path purging algorithm cannot be backed up by theory for a fixed q_c as $x \to \infty$, the simulations work effectively well in the regime where q_c is quite large compared to x.

4.1 Uniform distribution on the sphere

Consider when $\boldsymbol{\xi}$ is uniformly distributed in $\mathbb{S}^2$, i.e., we are in the spherically symmetric case. The statistic that we compare is the coefficient of the linear term for the estimation of τ_x in (4). For any given $p_3 \in \{0.05\ell : 1 \leqslant \ell \leqslant 19\}$, in the numerical BRW simulation, we calculate the FPT distribution at different offset distances, $x \in \{0.25L_x, 0.5L_x, 0.75L_x, L_x\}$, where $L_x = 65.5$. The mean FPT $\mathbb{E}[\tau_x]$ is then fit to a function of the form

$$\mathbb{E}[\tau_x] = \frac{x}{c_1} + B \log x + C. \tag{61}$$

The numerical implementation validates the theoretical predictions to the classical and delayed branching BRW models, as shown in Figure 5. This also indicates that the numerical approximation of path purging does not affect the linear coefficient of the scaling behavior of the FPT distribution, and can be used as a suitable approximation for carrying out the numerical BRW simulations. The error between the theoretical and numerical implementation can be attributed to the limited range of x over which the fit of (61) is executed. Figure 5 suggests that *classical* model requires a smaller range of x for the FPT evaluation for a higher fidelity estimation of c_1 in comparison to the *delayed branching* model. Figure 6(a) shows the FPT distribution at different x. Figure 6(b) shows the linear increase in the mean FPT (negligible logarithmic effect), whereas the Figure 6(c) validates the asymptotic in (55) by showing a negligible change in the standard deviation of FPT distribution with the increase in the offset distance, x.
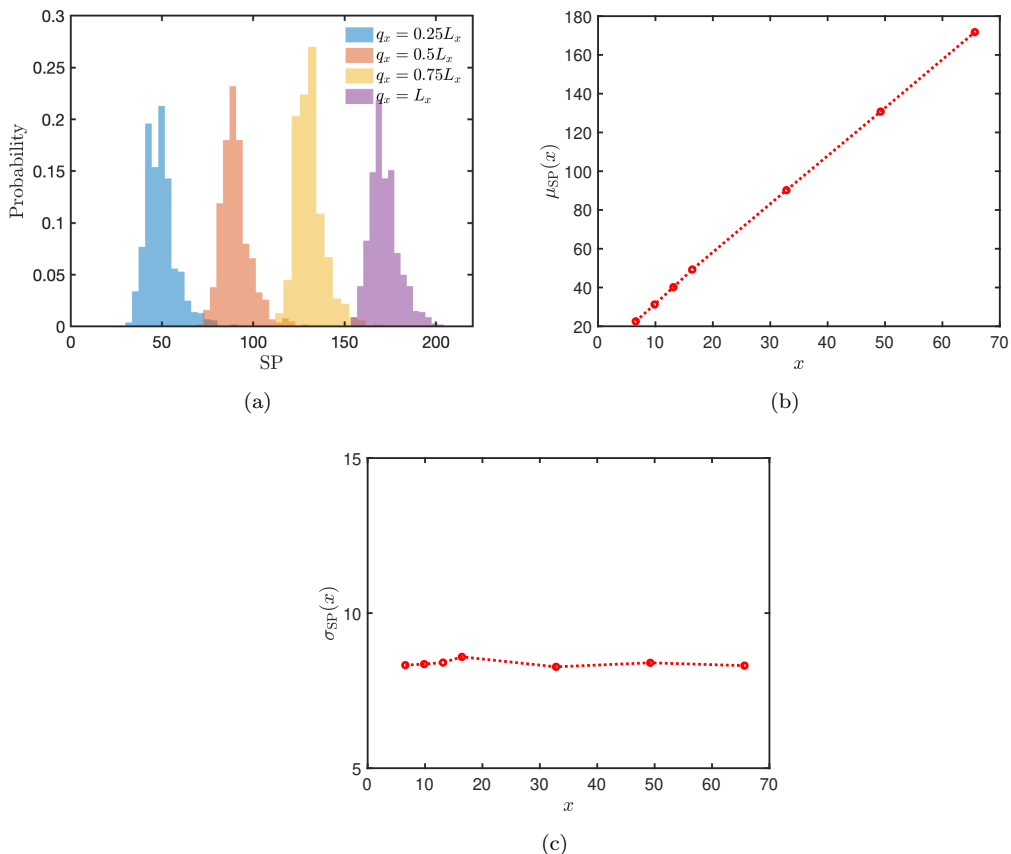
30

(a)

(b)

(c)

Figure 6: (a) The FPT distributions at different values of x for the classical BRW. The (b) mean and (c) standard deviation of the SP as a function of x, respectively. The data points correspond to $x = 0.1L_x,\ 0.15L_x,\ 0.2L_x,\ 0.25L_x,\ 0.5L_x,\ 0.75L_x,\ L_x$ and $p_3 = 0.0856$ for the classical BRW, where $L_x = 65.5$. The mean of the FPT at each x is computed from 500 samples.

4.2 Gaussian distributions

Suppose now that the jump $\boldsymbol{\xi}$ is centered Gaussian with a positive definite covariance Σ. We have $\Lambda(\boldsymbol{\lambda}) = \boldsymbol{\lambda}^\top \Sigma \boldsymbol{\lambda}/2$ and $I(\mathbf{x}) = \mathbf{x}^\top \Sigma^{-1} \mathbf{x}/2$. In particular, $\mathbf{c}_2 = \nabla I(x/n, \mathbf{0}) = ((\Sigma^{-1})_{11} x/n, \ldots, (\Sigma^{-1})_{1d} x/n)$ and $\Lambda(\mathbf{c}_2) = (x/n)^2 (\Sigma^{-1})_{11}/2$. By (60), the normal vector $\mathbf{n}$ is a constant multiple of $\Sigma^{-1} \mathbf{e}_1$, where $\mathbf{e}_1 = (1, \mathbf{0}) \in \mathbb{R}^d$. If Σ is diagonal (independent jumps), the normal vector $\mathbf{n}$ is then parallel to the direction of the termination site $\mathbf{x}$. This is the case in our previous work [54] as opposed to the general case of non-spherically symmetric and dependent jumps that we will discuss in this section.

We look at three specific examples: (i) symmetric and independent (S.I.) jumps (presented earlier [54]), (ii) non-symmetric and independent (N-S.I.) jumps, and (iii) non-symmetric and dependent (N-S.D.) jumps. The corresponding covariances for the three cases are given by

$$\Sigma_{\text{S.I.}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \Sigma_{\text{N-S.I.}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1.5 & 0 \\ 0 & 0 & 0.5 \end{bmatrix}, \quad \Sigma_{\text{N-S.D.}} = \begin{bmatrix} 1 & 0.5 & 0.25 \\ 0.5 & 1.5 & 0.5 \\ 0.25 & 0.5 & 0.5 \end{bmatrix}.$$

We follow the same procedures as Section 4.1 that estimate the statistic $\widehat{c}_1$ in (7) through fitting
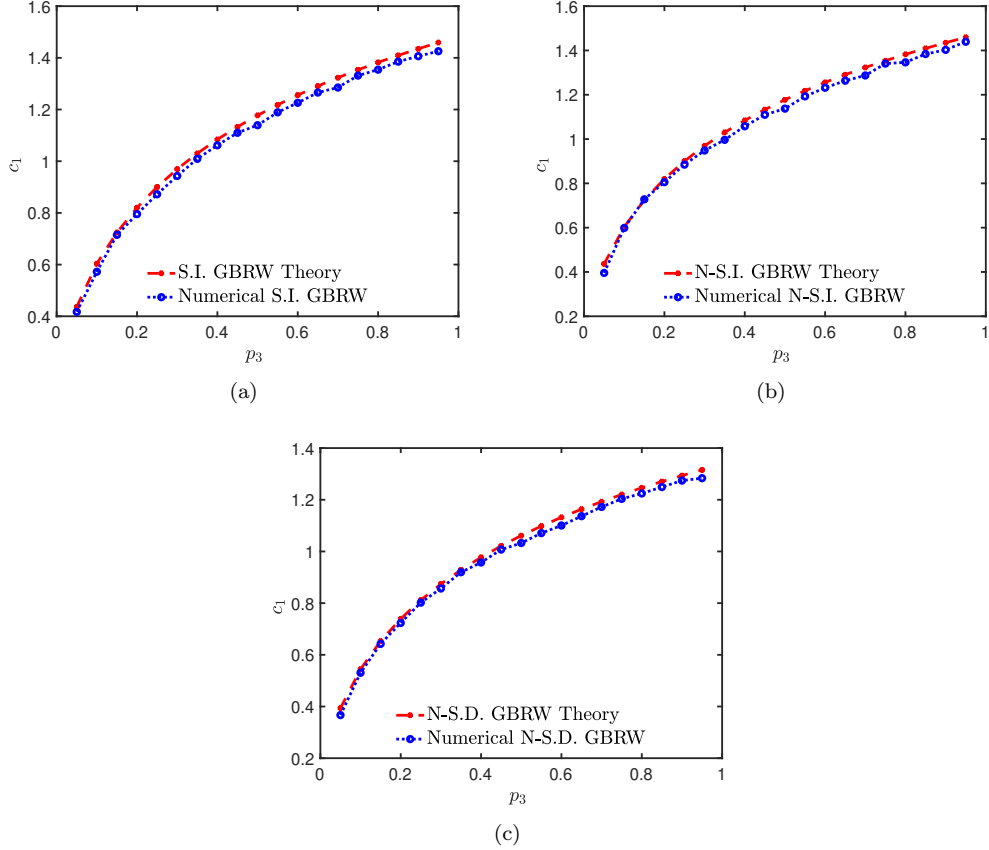
31

Figure 7: The numerically obtained $\widehat{c}_1$ compared against the theoretical $\widehat{c}_1$ for Gaussian BRW with (a) symmetric and independent (S.I.) jumps, (b) non-symmetric and independent (N-S.I.) jumps, and (c) non-symmetric and dependent (N-S.D.) jumps. The mean of the FPT at each x is computed from 10000 samples.

(61). Note that at any given p_3, the $\widehat{c}_1$ is a function of $(\Sigma^{-1})_{11}$. As a result, we expect the c_1 of the Gaussian BRW with S.I. and N-S.I. jumps to be identical as confirmed by our numerical results in Figures 7(a) and 7(b). We use $p_0 = 0$ for the comparison between the numerical and theoretical results in this section. In the case of the N-S.D. jumps, the $(\Sigma^{-1})_{11}$ is larger and we expect the $\widehat{c}_1$ to be smaller. Figure 7(c) shows the agreement of the numerical implementation in being able to capture the lower c_1 at every p_3. The agreement of the numerical results could be improved by incorporating a larger number of paths for the computation of the mean FPT or by increasing the range of x over which the mean FPT is computed to execute the fit in (61). To ascertain that 500 samples are sufficient to capture the mean, we present the mean FPT $\mathbb{E}[\tau_x]$ at different values of p_3 at $x = 65.5$ for the Gaussian BRW with N-S.I. and N-S.D. jumps. Figure 8 confirms that the statistics from 500 independent samples are sufficient to capture the mean of the distribution. This confirms that the approximation of c_1 can be improved by fitting the FPT values over a larger range of x.
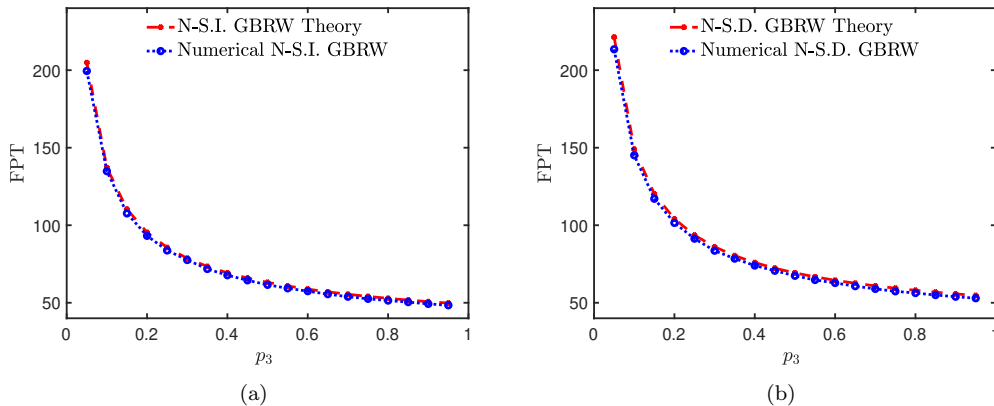
Figure 8: The numerically obtained expected FPT compared against the theoretical asymptotic (7) for Gaussian BRW with (a) non-symmetric and independent (N-S.I.) jumps and (b) non-symmetric and dependent (N-S.D.) jumps. The mean of the FPT is calculated at $x = L_x$, where $L_x = 65.5$, from 10000 samples.

5 Concluding remarks

In this paper, we studied the first passage times of an $\mathbb{R}^d$-valued branching random walk (BRW) to a shifted unit ball. The first passage time asymptotics consists of a linear term and a logarithm correction term, as a function of the Euclidean distance of the target ball from the origin. We discussed extensions of this result to the delayed branching BRW model. As an immediate application, we obtained theoretical predictions of the shortest path statistics for polymer networks consisting of long chains and random cross-links. We conclude with a conjecture on finer asymptotics of the first passage times.

The first passage times of the branching Brownian motion, the continuous-time sibling of the BRW, possess better tractability due to its connection to the Fisher-KPP equation. It is proven in Section 4.1 of [54] the finer asymptotic for the first passage times of a standard branching Brownian motion that

$$\tau_x = \frac{x}{\sqrt{2}} + \frac{d+2}{4} \log x + O_{\mathbb{P}}(1), \tag{62}$$

where the $O_{\mathbb{P}}(1)$ is tight. We refer the readers to [54] for the proof and a detailed discussion. In view also of the tightness result of Theorem 2, it is natural to postulate the following conjecture for BRW. Recall the constant $\widehat{c}_1$ from Section 1.2.

Conjecture 1. Assume (A1)–(A4), or (A1), (A2), (A5), and (A6). Conditioned upon survival, the first passage time for BRW in dimension $d \geqslant 2$ to B_x is given by

$$\tau_x = A(x) + O_{\mathbb{P}}(1) = \frac{x}{\widehat{c}_1} + \frac{d+2}{2\,\widehat{c}_1 \partial_{x_1} I(\widehat{c}_1, \mathbf{0})} \log x + O_{\mathbb{P}}(1),$$

where the $O_{\mathbb{P}}(1)$ is tight.

Conjecture 1 is also supported by our numerical simulations; see Figure 8.

The main technical obstruction to improving upon $o_{\mathbb{P}}(\log x)$ lies in the application of ballot theorems—the cone $\mathbb{K}$ in Sections 2.4.3 and 2.4.4 cannot be simply replaced by a half-space as certain sums would not converge (if this was the case, Conjecture 1 would hold). One possible

33

approach is to develop bounds for random walks in time-dependent (and possibly non-circular) cones, where $\mathbb{K} = \mathbb{K}_n$ may depend on n, the number of steps in the random walk. The $o_{\mathbb{P}}(\log x)$ may be improved if one applies such estimates in the case where $\mathbb{K}_n \to \mathbb{K}_{\pi/2}(-\mathbf{c}_2)$ with a certain rate.

acknowledgement

Data Availability

The open source code for the branching random walk calculations can be accessed from PolyBranchX.

References

[1] Louigi Addario-Berry and Bruce Reed. Minima in branching random walks. *Annals of Probability*, 37(3):1044–1079, 2009.

[2] Elie Aïfdékon. Convergence in law of the minimum of a branching random walk. *Annals of Probability*, 41(3A):1362–1426, 2013.

[3] Louis-Pierre Arguin, David Belius, and Paul Bourgade. Maximum of the characteristic polynomial of random unitary matrices. *Communications in Mathematical Physics*, 349:703–751, 2017.

[4] Louis-Pierre Arguin, David Belius, and Adam J Harper. Maxima of a randomized riemann zeta function, and branching random walks. *The Annals of Applied Probability*, 27(1):178–215, 2017.

[5] Krishna B Athreya and Peter E Ney. *Branching Processes*. Courier Corporation, 2004.

[6] Rodrigo Bañuelos and Robert G Smits. Brownian motion in cones. *Probability Theory and Related Fields*, 108:299–319, 1997.

[7] Julien Berestycki, Yujin H Kim, Eyal Lubetzky, Bastien Mallein, and Ofer Zeitouni. The extremal point process of branching Brownian motion in $\mathbb{R}^d$. *ArXiv preprint arXiv:2112.08407. To appear, Ann. Probab.*, 2021.

[8] Viktor Bezborodov and Nina Gantert. The maximal displacement of radially symmetric branching random walk in $\mathbb{R}^d$. *ArXiv preprint arXiv:2309.14738*, 2023.

[9] Rabi N Bhattacharya and Jayanta K Ghosh. On the validity of the formal edgeworth expansion. *The Annals of Statistics*, 6(2):434–451, 1978.

[10] JD Biggins. The central limit theorem for the supercritical branching random walk, and related results. *Stochastic Processes and their Applications*, 34(2):255–274, 1990.

[11] John D Biggins. Growth rates in the branching random walk. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 48(1):17–34, 1979.

[12] Anton Bovier. *Gaussian Processes on Trees: From Spin Glasses to Branching Brownian Motion*, volume 163. Cambridge University Press, 2017.

[13] Maury Bramson, Jian Ding, and Ofer Zeitouni. Convergence in law of the maximum of nonlattice branching random walk. *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques*, 52(4):1897–1924, 2016.

[14] Maury Bramson, Jian Ding, and Ofer Zeitouni. Convergence in law of the maximum of the two-dimensional discrete gaussian free field. *Communications on Pure and Applied Mathematics*, 69(1):62–123, 2016.

[15] Maury Bramson and Ofer Zeitouni. Tightness for the minimal displacement of branching random walk. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(07):P07010, 2007.

[16] Maury Bramson and Ofer Zeitouni. Tightness for a family of recursion equations. *Annals of Probability*, 37(2):615–653, 2009.

[17] Maury D Bramson. Maximal displacement of branching Brownian motion. *Communications on Pure and Applied Mathematics*, 31(5):531–581, 1978.

[18] Éric Brunet, Anh Dung Le, Alfred H Mueller, and Stéphane Munier. How to generate the tip of branching random walks evolved to large times. *Europhysics Letters*, 131(4):40002, 2020.

[19] DL Burkholder. Exit times of brownian motion, harmonic majorization, and hardy spaces. *Advances in Mathematics*, 26(2):182–205, 1977.

[20] R Dante DeBlassie. Exit times from cones in $\mathbb{R}^n$ of brownian motion. *Probability Theory and Related Fields*, 74(1):1–29, 1987.

[21] Frederik Michel Dekking and Bernard Host. Limit distributions for minimal displacement of branching random walks. *Probability Theory and Related Fields*, 90:403–426, 1991.

[22] Amir Dembo and Ofer Zeitouni. *Large Deviations Techniques and Applications*. Springer, 1998.

[23] Denis Denisov and Vitali Wachtel. Random walks in cones. *Annals of Probability*, 43(3):992–1044, 2015.

[24] Jian Ding and Ofer Zeitouni. Extreme values for two-dimensional discrete gaussian free field. *Annals of Probability*, 42(4):1480–1515, 2014.

[25] Elizaveta Ermakova, Polina Makhmutova, and Elena Yarovaya. Branching random walks and their applications for epidemic modeling. *Stochastic Models*, 35(3):300–317, 2019.

[26] Kai Wang Fang. *Symmetric Multivariate and Related Distributions*. CRC Press, 2018.

[27] William Feller. *An Introduction to Probability Theory and its Applications, Volume 2*, volume 81. John Wiley & Sons, 1991.

[28] Ronald Aylmer Fisher. The wave of advance of advantageous genes. *Annals of Eugenics*, 7(4):355–369, 1937.

[29] Nina Gantert and Thomas Höfelsauer. Large deviations for the maximum of a branching random walk. *Electronic Communications in Probability*, 23(34):1–12, 2018.

[30] Nina Gantert, Yueyun Hu, and Zhan Shi. Asymptotics for the survival probability in a killed branching random walk. In *Annales de l'IHP Probabilités et Statistiques*, volume 47, pages 111–129, 2011.

[31] Adam J Harper. Moments of random multiplicative functions, I: Low moments, better than squareroot cancellation, and critical multiplicative chaos. In *Forum of Mathematics, Pi*, volume 8. Cambridge University Press, 2020.

[32] Yujin H Kim, Eyal Lubetzky, and Ofer Zeitouni. The maximum of branching Brownian motion in $\mathbb{R}^d$. *The Annals of Applied Probability*, 33(2):1515–1568, 2023.

[33] Yujin H Kim and Ofer Zeitouni. The shape of the front of multidimensional branching brownian motion. *ArXiv preprint arXiv:2401.12431*, 2024.

[34] Andrei Kolmogorov, Ivan Petrovskii, and Nikolai Piskunov. Étude de léquation de la diffusion avec croissance de la quantité de matière et son application à un problème biologigue. *Moscow Univ. Bull. Ser. Internat. Sect. A*, 1:1, 1937.

[35] Wolfgang König. Branching random walks in random environment: a survey. *ArXiv preprint arXiv:2010.06942*, 2020.

[36] Mark Kot, Jan Medlock, Timothy Reluga, and D Brian Walton. Stochasticity, invasions, and branching random walks. *Theoretical Population Biology*, 66(3):175–184, 2004.

[37] Andreas E Kyprianou. Asymptotic radial speed of the support of supercritical branching and super-brownian motion in $\mathbb{R}^d$. *Markov Processes and Related Fields*, 11:145–156, 2005.

[38] Steven P Lalley and Thomas Sellke. A conditional limit theorem for the frontier of a branching brownian motion. *The Annals of Probability*, pages 1052–1061, 1987.

[39] Thomas Madaule. Convergence in law for the branching random walk seen from its tip. *Journal of Theoretical Probability*, 30:27–63, 2017.

[40] Bastien Mallein. Maximal displacement in the d-dimensional branching Brownian motion. *Electronic Communications in Probability*, 20:1–12, 2015.

[41] Colin McDiarmid. Minimal positions in a branching random walk. *The Annals of Applied Probability*, 5:128–139, 1995.

[42] John Mineka and S Silverman. A local limit theorem and recurrence conditions for sums of independent non-lattice random variables. *The Annals of Mathematical Statistics*, 41(2):592–600, 1970.

[43] Mehmet Öz. On the density of branching Brownian motion. *Hacettepe Journal of Mathematics & Statistics*, 52(1):229–247, 2023.

[44] R Tyrrell Rockafellar. *Convex Analysis*, volume 18. Princeton University Press, 1970.

[45] Alain Rouault. Precise estimates of presence probabilities in the branching random walk. *Stochastic Processes and Their Applications*, 44(1):27–39, 1993.

[46] Zhan Shi. *Branching Random Walks*. Volume 2151 of Lecture Notes in Mathematics. Springer, 2015.

[47] Roman Stasiński, Julien Berestycki, and Bastien Mallein. Derivative martingale of the branching brownian motion in dimension d$\geqslant$1. In *Annales de l'Institut Henri Poincare (B) Probabilites et Statistiques*, volume 57, pages 1786–1810, 2021.

[48] Kohei Uchiyama. Spatial growth of a branching process of particles living in $\mathbb{R}^d$. *The Annals of Probability*, 10(4):896–918, 1982.

[49] Frank JS Wang. The convergence of a branching brownian motion used as a model describing the spread of an epidemic. *Journal of Applied Probability*, 17(2):301–312, 1980.

[50] Yikai Yin, Nicolas Bertin, Yanming Wang, Zhenan Bao, and Wei Cai. Topological origin of strain induced damage of multi-network elastomers by bond breaking. *Extreme Mechanics Letters*, 40:100883, 2020.

[51] Yikai Yin, Shaswat Mohanty, Christopher B Cooper, Zhenan Bao, and Wei Cai. Network evolution controlling strain-induced damage and self-healing of elastomers with dynamic bonds. *ArXiv preprint arXiv:2401.11087*, 2024.

[52] Ofer Zeitouni. Branching random walks and gaussian fields. *Probability and statistical physics in St. Petersburg*, 91:437–471, 2016.

[53] Shuxiong Zhang. Large deviation probabilities for the range of a *d*-dimensional supercritical branching random walk. *Applied Mathematics and Computation*, 462:128344, 2024.

[54] Zhenyuan Zhang, Shaswat Mohanty, Jose Blanchet, and Wei Cai. Modeling shortest paths in polymeric networks using spatial branching processes. *Journal of the Mechanics and Physics of Solids*, page 105636, 2024.

A   Large deviation estimates

In this appendix, we collect the deferred proofs of the large deviation estimates (Proposition 10 in Section 2.2.2, Lemma 11 in Section 2.3, and Proposition 17 in Section 2.4.2). We first quote a result on a quantitative version of the central limit theorem for non-lattice distributions.

Lemma 24. *Consider $d \geqslant 2$. Suppose that the radial component R of the spherically symmetric random variable $\boldsymbol{\xi}$ has exponential moments and $\mathbb{P}(R = 0) < 1$. Let $\boldsymbol{\lambda} \in \mathbb{R}^d$ be such that $\mathbb{E}_{\mathbb{P}}[e^{\boldsymbol{\lambda} \cdot \mathbf{X}}]$ exists and define a probability measure $\mathbb{Q}$ by $d\mathbb{Q}/d\mathbb{P}(\mathbf{x}) := e^{\boldsymbol{\lambda} \cdot \mathbf{x}}/\mathbb{E}_{\mathbb{P}}[e^{\boldsymbol{\lambda} \cdot \mathbf{X}}]$ where $\mathbb{P}$ is the law of $\mathbf{X}$. Then the law of $\boldsymbol{\xi}$ under $\mathbb{Q}$ satisfies the strong non-lattice property*

$$\limsup_{|\mathbf{t}| \to \infty} |\mathbb{E}_{\mathbb{Q}}[e^{i\mathbf{t} \cdot \boldsymbol{\xi}}]| < 1. \tag{63}$$

Moreover, consider an i.i.d. sequence $\{\mathbf{X}_n\}_{n \geqslant 1}$ with the same distribution as $\boldsymbol{\xi}$ under $\mathbb{Q}$ with a positive definite covariance matrix V and suppose that $\mathbb{E}_{\mathbb{Q}}[\|\boldsymbol{\xi}\|^3] < \infty$. Then

$$\sup_B \left| \frac{1}{\sqrt{n}} \sum_{j=1}^n (\mathbf{X}_j - \mathbb{E}_{\mathbb{Q}}[\mathbf{X}_j]) - \mu_{\mathbf{0}, V}(B) \right| = o(1),$$

where the supremum is taken over all boxes $B \subseteq \mathbb{R}^d$ and $\mu_{\mathbf{0}, V}$ is the Gaussian measure on $\mathbb{R}^d$ with mean $\mathbf{0}$ and covariance V.

Proof. Let us first assume that $\mathbb{P}(R = 0) = 0$. Denote by $\boldsymbol{\Xi} = (\Xi_1, \ldots, \Xi_d)$ the uniform distribution on $\mathbb{S}^{d-1}$. Suppose that $\boldsymbol{\xi}_1 = R_1 \boldsymbol{\Xi}^{(1)}$ and $\boldsymbol{\xi}_2 = R_2 \boldsymbol{\Xi}^{(2)}$ are i.i.d. copies of $\boldsymbol{\xi}$, and $\boldsymbol{\zeta} = \boldsymbol{\xi}_1 + \boldsymbol{\xi}_2$; since each $\boldsymbol{\xi}_i$ is radially symmetric, R_i and $\boldsymbol{\Xi}^{(i)}$ are independent. We first find the conditional density $z_{R_1, R_2}(r, \boldsymbol{\theta})$ of $\boldsymbol{\zeta}$ under $\mathbb{P}$ given R_1 and R_2, where r and $\boldsymbol{\theta}$ are the radial and polar components. Observe that the conditional distribution of $\boldsymbol{\zeta}$ given R_1 and R_2 is spherically symmetric. We have conditionally,

$$\mathbb{P}(\|\boldsymbol{\zeta}\| \geqslant r \mid R_1, R_2) = \mathbb{P}(\|(R_1, \mathbf{0}) + R_2 \boldsymbol{\Xi}\| \geqslant r)$$
$$= \mathbb{P}\left(R_1^2 + 2R_1 R_2 \Xi_1 + R_2^2 \geqslant r^2 \right) = \mathbb{P}\left(\Xi_1 \geqslant \frac{r^2 - R_1^2 - R_2^2}{2R_1 R_2} \right).$$

By Theorem 2.10 of [26], for $y \in [-1, 1]$ and some $C(d) \geqslant 0$,

$$\mathbb{P}\left(\Xi_1 \geqslant y\right) = C(d) \int_{-1}^{y} (1 - t^2)^{(d-3)/2} \mathrm{d}t.$$

Altogether it follows that for some $C(d) > 0$,

$$z_{R_1, R_2}(r, \boldsymbol{\theta}) = C(d) r \left(1 - \left(\frac{r^2 - R_1^2 - R_2^2}{2R_1 R_2}\right)^2\right)^{\frac{d-3}{2}}, \ r \in [|R_1 - R_2|, R_1 + R_2], \ \boldsymbol{\theta} \in \mathbb{S}^{d-1}.$$

Therefore, the conditional law of $\boldsymbol{\zeta}$ under $\mathbb{P}$, and hence also the unconditional law under $\mathbb{Q}$, has a density in $\mathbb{R}^d$. If $\mathbb{P}(R = 0) > 0$, the laws have densities except for an atom at zero, and the densities are non-trivial since $\mathbb{P}(R = 0) < 1$. By Lemma 4 in Section XV.4 of [27], (63) holds with $\boldsymbol{\zeta}$ in place of $\boldsymbol{\xi}$, and hence (63) holds since $\boldsymbol{\zeta}$ is a sum of two i.i.d. copies of $\boldsymbol{\xi}$. The final claim follows from Theorem 2(b) of [9]. □

Proof of Proposition 10. Let $\Lambda(\lambda) = \log \phi_\xi(\lambda) = \log \mathbb{E}[e^{\lambda \xi}]$. It is clear from (A2) that ξ is non-lattice. The Bahadur-Rao theorem (Theorem 3.7.4 of [22]) then implies that uniformly in $c \in [-C \log n, C \log n]$,

$$\mathbb{P}(X \geqslant m_n + c) \asymp \frac{e^{-nI((m_n + c)/n)}}{\sqrt{n}} \asymp \frac{e^{-nI(c_1) - (c - 3 \log n/2) I'(c_1)}}{\sqrt{n}} = n \rho^{-n} e^{-c_2 c}. \tag{64}$$

By definition $c_2 = I'(c_1)$, the constant $c_2 > 0$ is such that the measure $\mathbb{Q}$ defined by

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}(x) := e^{c_2 x - \Lambda(c_2)} \tag{65}$$

satisfies that under $\mathbb{Q}$, $\{X_i\}_{i \in \mathbb{N}}$ are i.i.d. with mean c_1. The random variable $(\widetilde{X}_i, \mathbf{Y}_i)$ is then centered under $\mathbb{Q}$ where $\widetilde{X}_i = X_i - c_1$. Define $\widetilde{X} := \widetilde{X}_1 + \cdots + \widetilde{X}_n$. It follows that

$$\begin{aligned}
&\mathbb{P}(X \geqslant m_n + c, \mathbf{Y} \in B_{\mathbf{0}}(u(n))) \\
&= e^{n\Lambda(c_2)} \mathbb{E}_{\mathbb{Q}}\left[e^{-c_2 X} \mathbb{1}_{\{X \geqslant m_n + c, \mathbf{Y} \in B_{\mathbf{0}}(u(n))\}}\right] \\
&\asymp e^{n\Lambda(c_2) - m_n c_2} \sum_{j=c+1}^{\infty} e^{-c_2 j} \mathbb{Q}\left(\widetilde{X} \in \left[j - 1 - \frac{3}{2c_2} \log n, j - \frac{3}{2c_2} \log n\right), \mathbf{Y} \in B_{\mathbf{0}}(u(n))\right) \\
&= \rho^{-n} n^{3/2} \sum_{j=c+1}^{\infty} e^{-c_2 j} \mathbb{Q}\left(\frac{(\widetilde{X}, \mathbf{Y})}{\sqrt{n}} \in \left[\frac{j - 1 - \frac{3}{2c_2} \log n}{\sqrt{n}}, \frac{j - \frac{3}{2c_2} \log n}{\sqrt{n}}\right) \times B_{\mathbf{0}}\left(\frac{u(n)}{\sqrt{n}}\right)\right).
\end{aligned}$$

Note that for $j \geqslant \sqrt{n}$, we may bound the probability from above by 1. For $j \leqslant \sqrt{n}$, the sets

$$E_{j,n} := \left[\frac{j - 1 - \frac{3}{2c_2} \log n}{\sqrt{n}}, \frac{j - \frac{3}{2c_2} \log n}{\sqrt{n}}\right) \times B_{\mathbf{0}}\left(\frac{u(n)}{\sqrt{n}}\right)$$

are bounded as $n \to \infty$. Therefore, by Lemma 24, with $\mu_{\mathbf{0},V}$ denoting the centered Gaussian measure with covariance matrix V on $\mathbb{R}^d$,

$$\mathbb{Q}\left(\frac{(\widetilde{X}, \mathbf{Y})}{\sqrt{n}} \in E_{j,n}\right) = (1 + o(1)) \mu_{\mathbf{0},V}(E_{j,n}) \asymp u(n)^{d-1} n^{-d/2}, \tag{66}$$

38

where the last step follows from the multivariate Gaussian density formula and that the positive definite matrix V depends only on the law of $\boldsymbol{\xi}$. Altogether this yields

$$\mathbb{P}(\mathbf{Y} \in B_{\mathbf{0}}(u(n)) \mid X \geqslant m_n) \ll (n\rho^{-n}e^{-c_2 c})^{-1}\rho^{-n}n^{3/2}\sum_{j=c+1}^{\infty}e^{-c_2 j}u(n)^{d-1}n^{-d/2}$$

$$\ll u(n)^{d-1}n^{-(d-1)/2}.$$

This establishes (20).

Using a similar approach, we see that

$$\mathbb{P}(X \in [m_n + c, m_n + c + a(n)]) \asymp n\rho^{-n}e^{-c_2 c}$$

and that

$$\mathbb{P}(\mathbf{Y} \in B_{\mathbf{y}}(1),\ X \in [m_n + c, m_n + c + a(n)])$$
$$\asymp \rho^{-n}n^{3/2}\sum_{j=c+1}^{c+a(n)}e^{-c_2 j}\mathbb{Q}\left(\frac{(\widetilde{X}, \mathbf{Y})}{\sqrt{n}} \in \left[\frac{j-1-\frac{3}{2c_2}\log n}{\sqrt{n}}, \frac{j-\frac{3}{2c_2}\log n}{\sqrt{n}}\right) \times B_{\frac{\mathbf{y}}{\sqrt{n}}}\left(\frac{1}{\sqrt{n}}\right)\right)$$
$$\gg \rho^{-n}n^{-(d-3)/2}e^{-c_2 c}.$$

This proves (21). The same computation also yields

$$\mathbb{P}(\mathbf{Y} \in B_{\mathbf{0}}(K\sqrt{n}),\ X \geqslant m_n + c) \ll \rho^{-n}n^{3/2}n^{-1/2} = n\rho^{-n},$$

where the asymptotic constant does not depend on K. Together with (64) completes the proof of (22). $\qquad\square$

Proof of Lemma 11. We use a similar change of measure argument as in the proof of Proposition 10. Denote by I the large deviation rate function for the $\mathbb{R}^d$-valued random variable $\boldsymbol{\xi}$. For $\widehat{\mathbf{c}}_2 := \nabla I(\mathbf{z}/(Ck))$, we define the tilted measure $\mathbb{Q}$ by $d\mathbb{Q}/d\mathbb{P}(\mathbf{x}) := e^{\widehat{\mathbf{c}}_2 \cdot \mathbf{x} - \Lambda(\widehat{\mathbf{c}}_2)}$, where we recall $\Lambda = \log \phi_{\boldsymbol{\xi}}$ is the log-moment generating function. Under $\mathbb{Q}$, each $\boldsymbol{\xi}_i$ is i.i.d. with mean $\mathbf{z}/(Ck)$. It follows that

$$\mathbb{P}((X, \mathbf{Y}) \in B_{\mathbf{z}}) = e^{Ck\Lambda(\widehat{\mathbf{c}}_2)}\mathbb{E}_{\mathbb{Q}}[e^{-\widehat{\mathbf{c}}_2 \cdot (X, \mathbf{Y})}\mathbb{1}_{\{(X,\mathbf{Y}) \in B_{\mathbf{z}}\}}]$$
$$\gg e^{Ck\Lambda(\widehat{\mathbf{c}}_2) - \widehat{\mathbf{c}}_2 \cdot \mathbf{z}}\mathbb{Q}\left(\left(\frac{X}{\sqrt{Ck}}, \frac{\mathbf{Y}}{\sqrt{Ck}}\right) \in \frac{B_{\mathbf{z}}}{\sqrt{Ck}}\right)$$
$$= e^{Ck\Lambda(\widehat{\mathbf{c}}_2) - \widehat{\mathbf{c}}_2 \cdot \mathbf{z}}\mathbb{Q}\left(\left(\frac{\widetilde{X}}{\sqrt{Ck}}, \frac{\widetilde{\mathbf{Y}}}{\sqrt{Ck}}\right) \in \frac{B_{\mathbf{0}}}{\sqrt{Ck}}\right), \qquad (67)$$

where each $(\widetilde{X}_i, \widetilde{\mathbf{Y}}_i) = (X_i, \mathbf{Y}_i) - \mathbf{z}/(Ck)$ is a centered random variable. Using definition of $\widehat{\mathbf{c}}_2$, we have

$$\Lambda(\widehat{\mathbf{c}}_2) = \nabla I\left(\frac{\mathbf{z}}{Ck}\right) \cdot \frac{\mathbf{z}}{Ck} - I\left(\frac{\mathbf{z}}{Ck}\right). \qquad (68)$$

In addition, Lemma 24 yields

$$\mathbb{Q}\left(\left(\frac{\widetilde{X}}{\sqrt{Ck}}, \frac{\widetilde{\mathbf{Y}}}{\sqrt{Ck}}\right) \in \frac{B_{\mathbf{0}}}{\sqrt{Ck}}\right) \gg (Ck)^{-d/2}. \qquad (69)$$

Combining (67), (68), and (69) yields that

$$\mathbb{P}((X, \mathbf{Y}) \in B_{\mathbf{z}}) \gg e^{-CkI(\mathbf{z}/(Ck))}(Ck)^{-d/2}.$$

39

Recalling that I is convex, C^1 on its domain, and $I(\mathbf{0}) = \nabla I(\mathbf{0}) = 0$ (by Jensen's inequality, $\mathbb{E}[e^{\boldsymbol{\lambda} \cdot \boldsymbol{\xi}}]$ is minimized at $\boldsymbol{\lambda} = \mathbf{0}$ when $\boldsymbol{\xi}$ is centered), we have

$$CI\left(\frac{\mathbf{z}}{Ck}\right) \leqslant C \sup_{\|\mathbf{x}\|_2 = C_1} I\left(\frac{\mathbf{x}}{C}\right) \leqslant C_1 \sup_{\|\mathbf{x}\|_2 = C_1} \left\|\nabla I\left(\frac{\mathbf{x}}{C}\right)\right\|_2 \leqslant \frac{c_3}{2}$$

for $C \geqslant C_2$ as we pick C_2 large enough. This concludes the proof. $\qquad \square$

Proof of Proposition 17. The Bahadur-Rao theorem yields that uniformly in $c \in [-C \log n, C \log n]$,

$$\mathbb{P}(X \geqslant x + c) \asymp n^{-1/2} e^{-nI(x/n) - cI'(x/n)}. \tag{70}$$

Mimicking (65), define the probability measure $\mathbb{Q}$ by

$$\frac{d\mathbb{Q}}{d\mathbb{P}}(\mathbf{x}) := e^{\mathbf{c_2} \cdot \mathbf{x} - \Lambda(\mathbf{c_2})}. \tag{71}$$

It follows that $\mathbb{E}_{\mathbb{Q}}[\boldsymbol{\xi}] = (x/n, \mathbf{0})$. That is, the random variable $(\widetilde{X}_i, \mathbf{Y}_i)$ is then centered under $\mathbb{Q}$ where $\widetilde{X}_i = X_i - x/n$. Define $\widetilde{X} := \widetilde{X}_1 + \cdots + \widetilde{X}_n$. It follows that

$$\mathbb{P}(X \geqslant x + c, \ \mathbf{Y} \in B_{\mathbf{0}}(1))$$

$$= e^{n\Lambda(\mathbf{c_2})} \mathbb{E}_{\mathbb{Q}}\left[e^{-\mathbf{c_2} \cdot (X, \mathbf{Y})} \mathbb{1}_{\{X \geqslant x + c, \ \mathbf{Y} \in B_{\mathbf{0}}(1)\}}\right]$$

$$\asymp e^{n\Lambda(\mathbf{c_2}) - \mathbf{c_2} \cdot (x, \mathbf{0})} \sum_{j=c+1}^{\infty} e^{-\mathbf{c_2} \cdot (j, \mathbf{0})} \mathbb{Q}\left(\widetilde{X} \in [j-1, j), \ \mathbf{Y} \in B_{\mathbf{0}}(1)\right)$$

$$= e^{-nI(x/n, \mathbf{0})} \sum_{j=c+1}^{\infty} e^{-\mathbf{c_2} \cdot (j, \mathbf{0})} \mathbb{Q}\left(\frac{(\widetilde{X}, \mathbf{Y})}{\sqrt{n}} \in \left[\frac{j-1}{\sqrt{n}}, \frac{j}{\sqrt{n}}\right) \times B_{\mathbf{0}}\left(\frac{1}{\sqrt{n}}\right)\right)$$

$$\asymp e^{-nI(x/n, \mathbf{0}) - \mathbf{c_2} \cdot (c, \mathbf{0})} n^{-d/2},$$

where in the last step we apply the same argument that leads to (66) and use that the first coordinate of $\mathbf{c_2}$ is positive by our assumption $x/n > \underline{c} > 0$. Altogether this yields (41).

Using a similar approach, we see that

$$\mathbb{P}(X \in [x + c, x + c + a(n)]) \asymp n^{-1/2} e^{-nI(x/n) - cI'(x/n)}$$

and that

$$\mathbb{P}(\mathbf{Y} \in B_{\mathbf{y}}(1), \ X \in [x + c, x + c + a(n)])$$

$$\asymp e^{n\Lambda(\mathbf{c_2}) - \mathbf{c_2} \cdot (x, \mathbf{y})} \sum_{j=c+1}^{c+a(n)} e^{-\mathbf{c_2} \cdot (j, \mathbf{0})} \mathbb{Q}\left(\frac{(\widetilde{X}, \mathbf{Y})}{\sqrt{n}} \in \left[\frac{j-1}{\sqrt{n}}, \frac{j}{\sqrt{n}}\right) \times B_{\frac{\mathbf{y}}{\sqrt{n}}}\left(\frac{1}{\sqrt{n}}\right)\right)$$

$$\gg e^{-nI(x/n, \mathbf{0}) - \mathbf{c_2} \cdot (c, \mathbf{y})} n^{-d/2}.$$

This proves (42). $\qquad \square$