



Universiteit  
Leiden  
The Netherlands

## The parabolic Anderson model on Galton-Watson trees

Wang, D.

### Citation

Wang, D. (2024, May 28). *The parabolic Anderson model on Galton-Watson trees*. Retrieved from <https://hdl.handle.net/1887/3754826>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3754826>

**Note:** To cite this publication please use the final published version (if applicable).

---

---

# PART I

---

## PARABOLIC ANDERSON MODEL: ANNEALED

---

---



# Annealed parabolic Anderson model on a regular tree

This chapter is based on the following paper:

F. den Hollander and D. Wang. Annealed parabolic Anderson model on a regular tree. *Markov Process. Related Fields*, 30:105–147, 2024.

## Abstract

We study the total mass of the solution to the parabolic Anderson model on a regular tree with an i.i.d. random potential whose marginal distribution is double-exponential. In earlier work we identified two terms in the asymptotic expansion for large time of the total mass under the *quenched law*, i.e., conditional on the realisation of the random potential. In the present paper we do the same for the *annealed law*, i.e., averaged over the random potential. It turns out that the annealed expansion differs from the quenched expansion. The derivation of the annealed expansion is based on a *new approach* to control the local times of the random walk appearing in the Feynman-Kac formula for the total mass. In particular, we condition on the backbone to infinity of the random walk, truncate and periodise the infinite tree relative to the backbone to obtain a random walk on a finite subtree with a specific boundary condition, employ the large deviation principle for the empirical distribution of *Markov renewal processes* on finite graphs, and afterwards let the truncation level tend to infinity to obtain an asymptotically sharp asymptotic expansion.

## §2.1 Introduction and main results

Section 2.1.1 provides background and motivation, Section 2.1.2 lists notations, definitions and assumptions, Section 2.1.3 states the main theorems, while Section 2.1.4 places these theorems in their proper context.

### §2.1.1 Background and motivation

The *parabolic Anderson model* (PAM) is the Cauchy problem

$$\partial_t u(x, t) = \Delta_{\mathcal{X}} u(x, t) + \xi(x) u(x, t), \quad t > 0, x \in \mathcal{X}, \quad (2.1)$$

where  $t$  is time,  $\mathcal{X}$  is an ambient space,  $\Delta_{\mathcal{X}}$  is a Laplace operator acting on functions on  $\mathcal{X}$ , and  $\xi$  is a random potential on  $\mathcal{X}$ . Most of the literature considers the setting where  $\mathcal{X}$  is either  $\mathbb{Z}^d$  or  $\mathbb{R}^d$  with  $d \geq 1$ , starting with the foundational papers [24], [25], [23] and further developed through a long series of follow-up papers (see the monograph [29] and the survey paper [3] for an overview). More recently, other choices for  $\mathcal{X}$  have been considered as well:

- (I) *Deterministic graphs* (the complete graph [16], the hypercube [5]).
- (II) *Random graphs* (the Galton-Watson tree [12], [13], the configuration model [12]).

Much remains open for the latter class.

The main target for the PAM is a description of *intermittency*: for large  $t$  the solution  $u(\cdot, t)$  of (2.1) concentrates on well-separated regions in  $\mathcal{X}$ , called *intermittent islands*. Much of the literature focusses on a detailed description of the size, shape and location of these islands, and on the profiles of the potential  $\xi(\cdot)$  and the solution  $u(\cdot, t)$  on them. A special role is played by the case where  $\xi$  is an i.i.d. random potential with a *double-exponential* marginal distribution

$$P(\xi(0) > u) = e^{-e^{u/\varrho}}, \quad u \in \mathbb{R}, \quad (2.2)$$

where  $\varrho \in (0, \infty)$  is a parameter. This distribution turns out to be critical, in the sense that the intermittent islands neither grow nor shrink with time, and represents a class of its own.

In the present paper we consider the case where  $\mathcal{X}$  is an *unrooted regular tree*  $\mathcal{T}$ . Our focus will be on the asymptotics as  $t \rightarrow \infty$  of the total mass

$$U(t) = \sum_{x \in \mathcal{T}} u(x, t).$$

In earlier work [12], [13] we were concerned with the case where  $\mathcal{X}$  is a *rooted Galton-Watson tree* in the *quenched* setting, i.e., almost surely with respect to the random tree and the random potential. This work was restricted to the case where the random potential is given by (2.2) and the offspring distribution of the Galton-Watson tree has support in  $\mathbb{N} \setminus \{1\}$  with a sufficiently thin tail. In the present paper our focus will

be on the *annealed* setting, i.e., averaged over the random potential. We derive two terms in the asymptotic expansion as  $t \rightarrow \infty$  of the average total mass

$$\langle U(t) \rangle = \sum_{x \in \mathcal{T}} \langle u(x, t) \rangle,$$

where  $\langle \cdot \rangle$  denotes expectation with respect to the law of the random potential. It turns out that the annealed expansion *differs* from the quenched expansion, even though the same variational formula plays a central role in the two second terms.

The derivation in the annealed setting forces us to follow a different route than in the quenched setting, based on various approximations of  $\mathcal{T}$  that are more delicate than the standard approximation of  $\mathbb{Z}^d$  (see [11, Chapter VIII]). This is the reason why we consider regular trees rather than Galton-Watson trees, to which we hope to return later. A key tool in the analysis is the large deviation principle for the empirical distribution of *Markov renewal processes* on finite graphs derived in [31], which is recalled in Appendix A.1.

## §2.1.2 The PAM on a graph

### Notations and definitions

Let  $G = (V, E)$  be a *simple connected undirected* graph, either finite or countably infinite, with an arbitrarily designated vertex  $\mathcal{O}$ . Let  $\Delta_G$  be the Laplacian on  $G$ , i.e.,

$$(\Delta_G f)(x) = \sum_{\substack{y \in V: \\ \{x, y\} \in E}} [f(y) - f(x)], \quad x \in V, f: V \rightarrow \mathbb{R},$$

which acts along the edges of  $G$ . Let  $\xi := (\xi(x))_{x \in V}$  be a random potential attached to the vertices of  $G$ , taking values in  $\mathbb{R}$ . Our object of interest is the non-negative solution of the Cauchy problem with localised initial condition,

$$\begin{aligned} \partial_t u(x, t) &= (\Delta_G u)(x, t) + \xi(x)u(x, t), & x \in V, t > 0, \\ u(x, 0) &= \delta_{\mathcal{O}}(x), & x \in V. \end{aligned} \quad (2.3)$$

The quantity  $u(x, t)$  can be interpreted as the amount of mass at time  $t$  at site  $x$  when initially there is unit mass at  $\mathcal{O}$ . The total mass at time  $t$  is  $U(t) = \sum_{x \in V} u(x, t)$ . The total mass is given by the *Feynman-Kac formula*

$$U(t) = \mathbb{E}_{\mathcal{O}} \left( e^{\int_0^t \xi(X_s) ds} \right), \quad (2.4)$$

where  $X = (X_t)_{t \geq 0}$  is the continuous-time random walk on the vertices  $V$  with jump rate 1 along the edges  $E$ , and  $\mathbb{P}_{\mathcal{O}}$  denotes the law of  $X$  given  $X_0 = \mathcal{O}$ . Let  $\langle \cdot \rangle$  denote expectation with respect to  $\xi$ . The quantity of interest in this paper is the average total mass at time  $t$ :

$$\langle U(t) \rangle = \left\langle \mathbb{E}_{\mathcal{O}} \left( e^{\int_0^t \xi(X_s) ds} \right) \right\rangle. \quad (2.5)$$

### Assumption on the potential

Throughout the paper we assume that the random potential  $\xi = (\xi(x))_{x \in V}$  consists of i.i.d. random variables with a marginal distribution whose cumulant generating function

$$H(u) = \log \left\langle e^{u\xi(\mathcal{O})} \right\rangle \quad (2.6)$$

satisfies the following:

**Assumption 2.A.** [Asymptotic double-exponential potential]

There exists a  $\varrho \in (0, \infty)$  such that

$$\lim_{u \rightarrow \infty} uH''(u) = \varrho. \quad (2.7)$$

■

**Remark 2.1.1.** [Double-exponential potential] A special case of (2.7) is when  $\xi(\mathcal{O})$  has the double-exponential distribution in (2.2), in which case

$$H(u) = \log \Gamma(\varrho u + 1)$$

with  $\Gamma$  the gamma function. ♠

By Stirling's approximation, (2.7) implies

$$H(u) = \varrho u \log(\varrho u) - \varrho u + o(u), \quad u \rightarrow \infty. \quad (2.8)$$

Assumption 2.A is more than enough to guarantee existence and uniqueness of the non-negative solution to (2.3) on any discrete graph with at most exponential growth (as can be inferred from the proof in [24], [25] for the case  $G = \mathbb{Z}^d$ ). Since  $\xi$  is assumed to be i.i.d., we have from (2.5) that

$$\langle U(t) \rangle = \mathbb{E}_{\mathcal{O}} \left( \exp \left[ \sum_{x \in V} H(\ell_t(x)) \right] \right), \quad (2.9)$$

where

$$\ell_t(x) = \int_0^t \mathbb{1}\{X_s = x\} ds, \quad x \in V, t \geq 0,$$

is the local time of  $X$  at vertex  $x$  up to time  $t$ .

### Variational formula

The following *characteristic variational formula* is important for the description of the asymptotics of  $\langle U(t) \rangle$ . Denote by  $\mathcal{P}(V)$  the set of probability measures on  $V$ . For  $p \in \mathcal{P}(V)$ , define

$$I_E(p) = \sum_{\{x,y\} \in E} \left( \sqrt{p(x)} - \sqrt{p(y)} \right)^2, \quad J_V(p) = - \sum_{x \in V} p(x) \log p(x), \quad (2.10)$$

and set

$$\chi_G(\varrho) = \inf_{p \in \mathcal{P}(V)} [I_E(p) + \varrho J_V(p)], \quad \varrho \in (0, \infty). \quad (2.11)$$

The first term in (2.11) is the quadratic form associated with the Laplacian, which is the large deviation rate function for the *empirical distribution*

$$L_t = \frac{1}{t} \int_0^t \delta_{X_s} \, ds = \frac{1}{t} \sum_{x \in V} \ell_t(x) \delta_x \in \mathcal{P}(V) \quad (2.12)$$

(see e.g. [11, Section IV]). The second term in (2.11) captures the second order asymptotics of  $\sum_{x \in V} H(tp(x))$  as  $t \rightarrow \infty$  via (2.8) (see e.g. [11, Section VIII]).

### Reformulation

The following lemma pulls the leading order term out of the expansion and shows that the second order term is controlled by the large deviation principle for the empirical distribution.

**Lemma 2.1.2.** [*Key object for the expansion*] *If  $G = (V, E)$  is finite, then*

$$\langle U(t) \rangle = e^{H(t) + o(t)} \mathbb{E}_{\mathcal{O}} \left( e^{-\varrho t J_V(L_t)} \right), \quad t \rightarrow \infty.$$

where  $J_V$  is the functional in (2.10) and  $L_t$  is the empirical distribution in (2.12).

*Proof.* Because  $\sum_{x \in V} \ell_t(x) = t$ , we can rewrite (2.9) as

$$\begin{aligned} \langle U(t) \rangle &= \mathbb{E}_{\mathcal{O}} \left( \exp \left[ \sum_{x \in V} H(\ell_t(x)) \right] \right) \\ &= e^{H(t)} \mathbb{E}_{\mathcal{O}} \left( \exp \left\{ t \sum_{x \in V} \frac{1}{t} \left[ H\left(\frac{\ell_t(x)}{t} t\right) - \frac{\ell_t(x)}{t} H(t) \right] \right\} \right). \end{aligned}$$

Assumption 2.A implies that  $H$  has the following scaling property (see [23]):

$$\lim_{t \rightarrow \infty} \frac{1}{t} [H(ct) - cH(t)] = \varrho c \log c \quad \text{uniformly in } c \in [0, 1].$$

Hence the claim follows.  $\square$

### §2.1.3 The PAM on an unrooted regular tree: annealed total mass for large times and key variational formula

In this section we specialise to the case where  $G = \mathcal{T} = (E, V)$ , an unrooted regular tree of degree  $d + 1$  with  $d \geq 2$  (see Fig. 2.1). The main theorem of our paper is the following expansion.



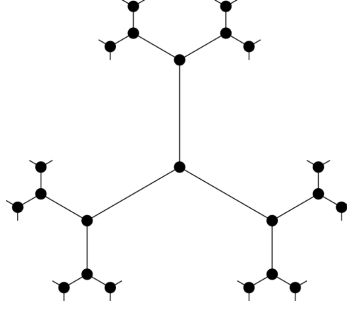


Figure 2.1: An unrooted tree with degree 3 (= offspring size 2).

**Theorem 2.1.3.** [Growth rate of the total mass] For any  $d \geq 4$ , subject to Assumption 2.A,

$$\frac{1}{t} \log \langle U(t) \rangle = \varrho \log(\varrho t) - \varrho - \chi_{\mathcal{T}}(\varrho) + o(1), \quad t \rightarrow \infty, \quad (2.13)$$

where  $\chi_{\mathcal{T}}(\varrho)$  is the variational formula in (2.11) with  $G = \mathcal{T}$ .

The proof of Theorem 2.1.3 is given in Sections 2.2–2.4 and makes use of technical computations collected in Appendices A.2–A.3.

The main properties of the key quantity

$$\chi_{\mathcal{T}}(\varrho) = \inf_{p \in \mathcal{P}(V)} [I_E(p) + \varrho J_V(p)], \quad \varrho \in (0, \infty), \quad (2.14)$$

are collected in the following theorem (see Fig. 2.2).

**Theorem 2.1.4.** [Properties of the variational formula] For any  $d \geq 2$  the following hold:

(a) The infimum in (2.14) may be restricted to the set

$$\mathcal{P}_{\mathcal{O}}^{\downarrow}(V) = \{p \in \mathcal{P}(V) : \arg\max p = \mathcal{O}, p \text{ is non-increasing along paths to infinity}\}. \quad (2.15)$$

(b) For every  $\varrho \in (0, \infty)$ , the infimum in (2.14) restricted to  $\mathcal{P}_{\mathcal{O}}^{\downarrow}(V)$  is attained, every minimiser  $\bar{p}$  is such that  $\bar{p} > 0$  on  $V$ , and  $\partial S_R = \sum_{\partial B_R(\mathcal{O})} \bar{p}(x)$ ,  $R \in \mathbb{N}_0$ , satisfies

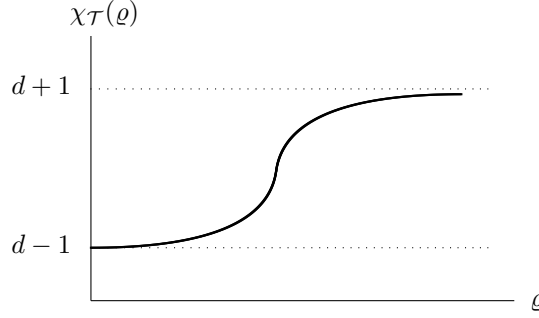
$$\sum_{R \in \mathbb{N}_0} \partial S_R \log(R+1) \leq \frac{d+1}{\varrho},$$

where  $B_R(\mathcal{O})$  is the ball of radius  $R$  centred at  $\mathcal{O}$ .

(c) The function  $\varrho \mapsto \chi_{\mathcal{T}}(\varrho)$  is strictly increasing and globally Lipschitz continuous on  $(0, \infty)$ , with

$$\lim_{\varrho \downarrow 0} \chi_{\mathcal{T}}(\varrho) = d-1, \quad \lim_{\varrho \rightarrow \infty} \chi_{\mathcal{T}}(\varrho) = d+1.$$

The proof of Theorem 2.1.4 is given in Section 2.5 (see Fig. 2.2).


 Figure 2.2: Qualitative plot of  $\rho \mapsto \chi_T(\rho)$ .

### §2.1.4 Discussion

1. Theorem 2.1.3 identifies the scaling of the total mass up to and including terms that are exponential in  $t$ . The first two terms in the right-hand side of (2.13) are the same as those of  $\frac{1}{t}H(t)$  (recall (2.8)). The third term is a correction that comes from the cost for  $X$  in the Feynman-Kac formula in (2.4) to create an *optimal local time profile* somewhere in  $\mathcal{T}$ , which is captured by the minimiser(s) of the variational formula in (2.14).
2. For the quenched model on a *rooted Galton-Watson tree*  $\mathcal{GW}$  we found in [12], [13] that

$$\frac{1}{t} \log U(t) = \rho \log \left( \frac{\rho t \vartheta}{\log \log t} \right) - \rho - \chi(\rho) + o(1), \quad t \rightarrow \infty, \quad \mathbf{P} \times \mathfrak{P}\text{-a.s.}, \quad (2.16)$$

where  $\mathbf{P}$  is the law of the potential,  $\mathfrak{P}$  is the law of  $\mathcal{GW}$ ,  $\vartheta$  is the logarithm of the mean of the offspring distribution, and

$$\chi_S(\rho) = \inf_{S \subset \mathcal{GW}} \chi_S(\rho) \quad (2.17)$$

with  $\chi_S(\rho)$  given by (2.11) and the infimum running over all subtrees of  $\mathcal{GW}$ . This result was shown to be valid as soon as the offspring distribution has support in  $\mathbb{N} \setminus \{1\}$  (i.e., all degrees are at least 3) and has a sufficiently thin tail. The extra terms in (2.16) come from the cost for  $X$  in the Feynman-Kac formula in (2.4) to travel in a time of order  $o(t)$  to an optimal finite subtree with an optimal profile of the potential, referred to as *intermittent islands*, located at a distance of order  $\rho t / \log \log t$  from  $\mathcal{O}$ , and to subsequently spend most of its time on that subtree. In this cost the parameter  $\vartheta$  appears, which is absent in (2.13). It was shown in [12] that if  $\rho \geq 1/\log(d_{\min} + 1)$ , with  $d_{\min}$  the minimum of the support of the offspring distribution, then the infimum in (2.17) is attained at the unrooted regular tree with degree  $d_{\min} + 1$ , i.e., the *minimal unrooted regular tree contained in  $\mathcal{GW}$* , for which  $\vartheta = \log d_{\min}$ . Possibly the bound on  $\rho$  is redundant.

3. In view of Lemma 2.1.2 and the fact that Assumption 2.A implies (2.8), we see that the proof of Theorem 2.1.3 amounts to showing that, on  $\mathcal{T} = (V, E)$ ,

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E}_{\mathcal{O}} \left( e^{-\varrho t J_V(L_t)} \right) = -\chi_{\mathcal{T}}(\varrho).$$

We achieve this by deriving asymptotically matching upper and lower bounds. These bounds are obtained by truncating  $\mathcal{T}$  outside a ball of radius  $R$ , to obtain a finite tree  $\mathcal{T}_R$ , deriving the  $t \rightarrow \infty$  asymptotics for finite  $R$ , and letting  $R \rightarrow \infty$  afterwards. For the lower bound we can use the standard truncation technique based on *killing*  $X$  when it exits  $\mathcal{T}_R$  and applying the large deviation principle for the empirical distribution of *Markov processes* on finite graphs derived in [15]. For the upper bound, however, we cannot use the standard truncation technique based on *periodisation* of  $X$  beyond radius  $R$ , because  $\mathcal{T}$  is an expander graph (see [29, Chapter IV] for a list of known techniques on  $\mathbb{Z}^d$  and  $\mathbb{R}^d$ ). Instead, we follow a route in which  $\mathcal{T}$  is approximated in successive stages by a version of  $\mathcal{T}_R$  with a *specific boundary condition*, based on monitoring  $X$  relative to its backbone to infinity. This route allows us to use the large deviation principle for the empirical distribution of *Markov renewal processes* on finite graphs derived in [31], but we need the condition  $d \geq 4$  to *control* the specific boundary condition in the limit as  $R \rightarrow \infty$  (see Remark 2.4.1 for more details). The reason why the approximation of  $\mathcal{T}$  by finite subtrees is successful is precisely because in the parabolic Anderson model the *total mass tends to concentrate on intermittent islands*.

4. Theorem 2.1.4 shows that, modulo translations, the optimal strategy for  $L_t$  as  $t \rightarrow \infty$  is to be close to a minimiser of the variational formula in (2.14) restricted to  $\mathcal{P}_{\mathcal{O}}^{\downarrow}(V)$ . Any minimiser is centred at  $\mathcal{O}$ , strictly positive everywhere, non-increasing in the distance to  $\mathcal{O}$ , and rapidly tending to zero. The following questions remain open:

- (1) Is the minimiser  $\bar{p}$  unique modulo translation?
- (2) Does  $\bar{p}(x)$  satisfy  $\lim_{|x| \rightarrow \infty} |x|^{-1} \log \bar{p}(x) = -\infty$ , with  $|x|$  the distance between  $x$  and  $\mathcal{O}$ ?
- (3) Is  $\bar{p}$  radially symmetric?
- (4) Is  $\varrho \mapsto \chi_{\mathcal{T}}(\varrho)$  analytic on  $(0, \infty)$ ?

We expect the answer to be yes for (1) and (2), and to be no for (3) and (4).

## §2.2 Proof of the main theorem: lower bound

In this section we prove the lower bound in Theorem 2.1.3, which is standard and straightforward. In Section 2.2.1 we obtain a lower bound in terms of a variational formula by *killing* the random walk when it exits  $\mathcal{T}_R$ . In Section 2.2.2 we derive the lower bound of the expansion by letting  $R \rightarrow \infty$  in the variational formula.

### §2.2.1 Killing and lower variational formula

Fix  $R \in \mathbb{N}$ . Let  $\mathcal{T}_R$  be the subtree of  $\mathcal{T} = (V, E)$  consisting of all the vertices that are within distance  $R$  of the root  $\mathcal{O}$  and all the edges connecting them. Put  $V_R = V_R(\mathcal{T}_R)$  and  $E_R = E(\mathcal{T}_R)$ . Let  $\tau_R = \inf\{t \geq 0: X_t \notin V_R\}$  denote the first time that  $X$  exits  $\mathcal{T}_R$ . It follows from (2.9) that

$$\langle U(t) \rangle \geq \mathbb{E}_{\mathcal{O}} \left( \exp \left[ \sum_{x \in V_R} H(\ell_t(x)) \right] \mathbb{1}\{\tau_R > t\} \right).$$

Since  $\mathcal{T}_R$  is finite, Lemma 2.1.2 gives

$$\langle U(t) \rangle \geq e^{H(t)+o(t)} \mathbb{E}_{\mathcal{O}} \left[ e^{-\varrho t J_V(L_t)} \mathbb{1}\{\tau_R > t\} \right]$$

with  $J_V$  the functional defined in (2.10). As shown in [22] (see also [25]), the family of sub-probability distributions  $\mathbb{P}_{\mathcal{O}}(L_t \in \cdot, \tau_R > t)$ ,  $t \geq 0$ , satisfies the LDP on  $\mathcal{P}^R(V) = \{p \in \mathcal{P}(V): \text{supp}(p) \subset V_R\}$  with rate function  $I_E$ , with  $I_E$  the functional defined in (2.10). This is the *standard* LDP for the empirical distribution of *Markov processes*. Therefore, by Varadhan's Lemma,

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E}_{\mathcal{O}} \left[ e^{-\varrho t J_V(L_t)} \mathbb{1}\{\tau_R > t\} \right] = -\chi_R^-(\varrho)$$

with

$$\chi_R^-(\varrho) = \inf_{p \in \mathcal{P}^R(V)} [I_E(p) + \varrho J_V(p)], \quad (2.18)$$

where we use that  $p \mapsto J_V(p)$  is bounded and continuous (in the discrete topology) on  $\mathcal{P}^R(V)$ . Note that

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{P}_{\mathcal{O}}(\tau_R > t) = - \inf_{p \in \mathcal{P}^R(V)} I_E(p) < 0,$$

which is non-zero because any  $p \in \mathcal{P}^R(V)$  is non-constant on  $V$ . The expression in (2.18) is the same as (2.11) with  $G = \mathcal{T}$ , except that  $p$  is restricted to  $V_R$ .

### §2.2.2 Limit of the lower variational formula

Clearly,  $R \mapsto \chi_R^-(\varrho)$  is non-increasing. To complete the proof of the lower bound in Theorem 2.1.3, it remains is to show the following.

**Lemma 2.2.1.**  $\limsup_{R \rightarrow \infty} \chi_R^-(\varrho) \leq \chi_{\mathcal{T}}(\varrho)$ .

*Proof.* Pick any  $p \in \mathcal{P}(V)$  such that  $I_E(p) < \infty$  and  $J_V(p) < \infty$ . Let  $p^R$  be the projection of  $p$  onto  $V_R$ , i.e.,

$$p^R(x) = \begin{cases} p(x), & x \in \text{int}(V_R), \\ \sum_{y \geq x} p(y), & x \in \partial V_R, \end{cases}$$

where  $y \geq x$  means that  $y$  is an element of the progeny of  $x$  in  $\mathcal{T}$ . Since  $p^R \in \mathcal{P}^R(V)$ , we have from (2.18) that  $\chi_R^-(\varrho) \leq I_E(p^R) + \varrho J_V(p^R)$ . Trivially,  $\lim_{R \rightarrow \infty} I_E(p^R) = I_E(p)$  and  $\lim_{R \rightarrow \infty} J_V(p^R) = J_V(p)$ , and so we have  $\limsup_{R \rightarrow \infty} \chi_R^-(\varrho) \leq I_E(p) + \varrho J_V(p)$ . Since this bound holds for arbitrary  $p \in \mathcal{P}(V)$ , the claim follows from (2.18).  $\square$

## §2.3 Proof of the main theorem: upper bound

In this section we prove the upper bound in Theorem 2.1.3, which is more laborious and requires a more delicate approach than the standard periodisation argument used on  $\mathbb{Z}^d$ . In Section 2.3.1 we obtain an upper bound in terms of a variational formula on a version of  $\mathcal{T}_R$  with a specific boundary condition. The argument comes in four steps, encapsulated in Lemmas 2.3.1–2.3.6 below:

- (I) *Condition on the backbone* of  $X$  (Section 2.3.1).
- (II) *Project*  $X$  onto a concatenation of finite subtrees attached to this backbone that are *rooted* versions of  $\mathcal{T}_R$  (Section 2.3.1).
- (III) *Periodise* the projected  $X$  to obtain a *Markov renewal process* on a single finite subtree and show that the periodisation can be chosen such that the local times at the vertices on the *boundary* of the finite subtree are negligible (Section 2.3.1).
- (IV) Use the *large deviation principle* for the empirical distribution of *Markov renewal processes* derived in [31] to obtain a variational formula on a single subtree (Section 2.3.1).

In Section 2.3.2 we derive the upper bound of the expansion by letting  $R \rightarrow \infty$  in the variational formula.

### §2.3.1 Backbone, projection, periodisation and upper variational formula

#### Backbone

For  $r \in \mathbb{N}$ , let  $\tau_r$  be the last time when  $X$  visits  $\partial B_r(\mathcal{O}) = \{x \in V : d(x, \mathcal{O}) = r\}$ , the boundary of the ball of radius  $r$  around  $\mathcal{O}$ . Then the sequence  $\mathcal{B} = (X_{\tau_r})_{r \in \mathbb{N}_0}$  forms the backbone of  $X$ , running from  $\mathcal{O}$  to infinity.

**Lemma 2.3.1.** [*Condition on a backbone*] For every backbone  $\text{bb}$  and every  $t \geq 0$ ,

$$\mathbb{E}_{\mathcal{O}} \left( \exp \left[ \sum_{x \in V(\mathcal{T})} H(\ell_t(x)) \right] \right) = \mathbb{E}_{\mathcal{O}} \left( \exp \left[ \sum_{x \in V(\mathcal{T})} H(\ell_t(x)) \right] \middle| \mathcal{B} = \text{bb} \right).$$

*Proof.* By symmetry, the conditional expectation in the right-hand side does not depend on the choice of  $\text{bb}$ . Indeed, permutations of the edges away from the root do not affect the law of  $\sum_{x \in V(\mathcal{T})} H(\ell_t(x))$ .  $\square$

Turn the one-sided backbone into a two-sided backbone by adding a second backbone from  $\mathcal{O}$  to infinity. By symmetry, the choice of this second backbone is arbitrary, say  $\text{bb}'$ . Redraw  $\mathcal{T}$  by representing  $\text{bb}' \cup \text{bb}$  as  $\mathbb{Z}$  and representing the rest of  $\mathcal{T}$  as a sequence of rooted trees  $\mathcal{T}^* = (\mathcal{T}_x^*)_{x \in \mathbb{Z}}$  hanging off  $\mathbb{Z}$  (see Fig. 2.3). In  $\mathcal{T}_x^*$ , the root sits at  $x$  and has  $d - 1$  downward edges, while all lower vertices have  $d$  downward edges.

Let  $\mathcal{T}^{\mathbb{Z}}$  denote the redrawn tree and  $X^{\mathbb{Z}} = (X_t^{\mathbb{Z}})_{t \geq 0}$  be the random walk on  $\mathcal{T}^{\mathbb{Z}}$ . Furthermore let  $(\ell_t^{\mathbb{Z}}(x))_{x \in \mathcal{T}^{\mathbb{Z}}}$  denote the local times of  $X^{\mathbb{Z}}$  at time  $t$ .

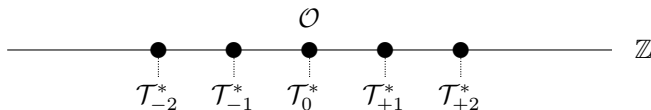


Figure 2.3: Redrawing of  $\mathcal{T}$  as  $\mathcal{T}^{\mathbb{Z}}$ : a two-sided backbone  $\mathbb{Z}$  with a sequence  $\mathcal{T}^* = (\mathcal{T}_x^*)_{x \in \mathbb{Z}}$  of rooted trees hanging off. The upper index  $*$  is used to indicate that the tree is rooted.

**Lemma 2.3.2.** [Representation of  $\mathcal{T}$  as a backbone with rooted trees] For every  $\text{bb}$  and  $t \geq 0$ ,

$$\mathbb{E}_{\mathcal{O}} \left( \exp \left[ \sum_{x \in V(\mathcal{T})} H(\ell_t(x)) \right] \middle| \mathcal{B} = \text{bb} \right) = \mathbb{E}_{\mathcal{O}} \left( \exp \left[ \sum_{x \in V(\mathcal{T}^{\mathbb{Z}})} H(\ell_t^{\mathbb{Z}}(x)) \right] \middle| X_{\infty}^{\mathbb{Z}} = +\infty \right).$$

*Proof.* Simply redraw  $\mathcal{T}$  as  $\mathcal{T}^{\mathbb{Z}}$ . □

Note that  $X^{\mathbb{Z}}$  is a Markov process whose sojourn times have distribution  $\text{EXP}(d + 1)$  and whose steps are drawn uniformly at random from the  $d + 1$  edges that are incident to each vertex.

### Projection

For  $R \in \mathbb{N} \setminus \{1\}$ , cut  $\mathbb{Z}$  into slices of length  $R$ , i.e.,

$$\mathbb{Z} = \cup_{k \in \mathbb{Z}} (z + (kR + I)), \quad I = \{0, 1, \dots, R - 1\},$$

where  $z$  is to be chosen later. Apply the following two maps to  $\mathcal{T}^{\mathbb{Z}}$  (in the order presented):

- (i) For each  $k \in \mathbb{Z}$ , fold  $\mathcal{T}_{z+(kR+(R-1))}^*$  onto  $\mathcal{T}_{z+(k+1)R}^*$  by folding the  $d - 1$  edges downwards from the root on top of the edge in  $\mathbb{Z}$  connecting  $z + (kR + (R - 1))$  and  $z + (k + 1)R$ , and putting the  $d$  infinite rooted trees hanging off each of these  $d - 1$  edges on top of the rooted tree  $\mathcal{T}_{z+(k+1)R}^*$  hanging off  $z + (k + 1)R$ . Note that each of the  $d$  infinite rooted trees is a copy of  $\mathcal{T}_{z+(k+1)R}^*$ .

## 2. Annealed PAM on a regular tree

- (ii) For each  $k \in \mathbb{Z}$  and  $m \in \{0, 1, \dots, R-2\}$ , cut off all the infinite subtrees in  $\mathcal{T}_{z+(kR+m)}^*$  whose roots are at depth  $(R-1) - m$ . Note that the total number of leaves after the cutting equals

$$(d-1) \sum_{m=0}^{R-2} d^{(R-2)-m} = (d-1)d^{R-2} \frac{1-d^{-(R-1)}}{1-d^{-1}} = d^{R-1} - 1,$$

which is the same as the total number of leaves of the rooted tree  $\mathcal{T}_R^*$  of depth  $R-1$  (i.e., with  $R$  generations) minus 1 (a fact we will need below).

By doing so we obtain a *concatenation* of finite units

$$\mathcal{U}_R = (\mathcal{U}_R[k])_{k \in \mathbb{Z}}$$

that are rooted trees of depth  $R-1$  (see Fig. 2.4). Together with the two maps that turn  $\mathcal{T}^{\mathbb{Z}}$  into  $\mathcal{U}_R$ , we apply two maps to  $X^{\mathbb{Z}}$ :

- (i) All excursions of  $X^{\mathbb{Z}}$  in the infinite subtrees that are *folded to the right and on top* are projected accordingly.
- (ii) All excursions of  $X^{\mathbb{Z}}$  in the infinite subtrees that are *cut off* are replaced by a sojourn of  $X^{\mathcal{U}_R}$  in the *tadpoles* that replace these subtrees (see Fig. 2.4)

The resulting path, which we call  $X^{\mathcal{U}_R} = (X_t^{\mathcal{U}_R})_{t \geq 0}$ , is a *Markov renewal process* with the following properties:

- The sojourn times in all the vertices that are not tadpoles have distribution  $\text{EXP}(d+1)$ .
- The sojourn times in all the tadpoles have distribution  $\psi$ , defined as the conditional distribution of the *return time*  $\tau$  of the random walk on the infinite rooted tree  $\mathcal{T}^*$  given that  $\tau < \infty$  (see [30] for a proper definition).
- The transitions into the tadpoles have probability  $\frac{d}{d+1}$ , the transitions out of the tadpoles have probability 1 (because of the condition  $X_{\infty}^{\mathbb{Z}} = +\infty$ ).
- The transitions from  $z + (kR + (R-1))$  to  $z + (k+1)R$  have probability  $\frac{d}{d+1}$ , while the reverse transitions have probability  $\frac{1}{d+1}$ .

Write  $(\ell_t^{\mathcal{U}_R}(x))_{x \in V_{\mathcal{U}_R}}$  to denote the local times of  $X^{\mathcal{U}_R}$  at time  $t$ .

**Lemma 2.3.3.** [Projection onto a concatenation of finite subtrees] For every  $R \in \mathbb{N} \setminus \{1\}$  and  $t \geq 0$ ,

$$\begin{aligned} & \mathbb{E}_{\mathcal{O}} \left( \exp \left[ \sum_{x \in V(\mathcal{T}^{\mathbb{Z}})} H(\ell_t^{\mathbb{Z}}(x)) \right] \middle| X_{\infty}^{\mathbb{Z}} = +\infty \right) \\ & \leq \mathbb{E}_{\mathcal{O}} \left( \exp \left[ \sum_{x \in V(\mathcal{U}_R)} H(\ell_t^{\mathcal{U}_R}(x)) \right] \middle| X_{\infty}^{\mathcal{U}_R} = +\infty \right). \end{aligned}$$

*Proof.* The maps that are applied to turn  $X^{\mathbb{Z}}$  into  $X^{\mathcal{U}_R}$  are such that local times are *stacked on top of each other*. Since  $H$  defined in (2.6) is convex and  $H(0) = 0$ , we have  $H(\ell) + H(\ell') \leq H(\ell + \ell')$  for all  $\ell, \ell' \in \mathbb{N}_0$ , which implies the inequality.  $\square$

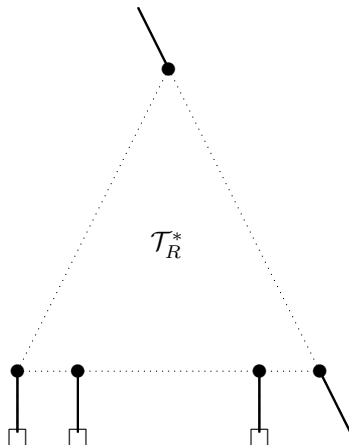


Figure 2.4: A unit in  $\mathcal{U}_R$ . Inside is a rooted tree  $\mathcal{T}_R^*$  of depth  $R-1$ , of which only the root and the leaves are drawn. Hanging off the leaves at depth  $R-1$  from the root are tadpoles, except for the right-most bottom vertex, which has a downward edge that connects to the root of the next unit. The vertices marked by a bullet form the boundary of  $\mathcal{U}_R$ , the vertices marked by a square box form the tadpoles of  $\mathcal{U}_R$ .

## Periodisation

Our next observation is that the condition  $\{X_{\infty}^{\mathcal{U}_R} = +\infty\}$  is *redundant*.

**Lemma 2.3.4.** [Condition redundant] For every  $R \in \mathbb{N} \setminus \{1\}$  and  $t \geq 0$ ,

$$\mathbb{E}_{\mathcal{O}} \left( \exp \left[ \sum_{x \in V(\mathcal{U}_R)} H(\ell_t^{\mathcal{U}_R}(x)) \right] \mid X_{\infty}^{\mathcal{U}_R} = +\infty \right) = \mathbb{E}_{\mathcal{O}} \left( \exp \left[ \sum_{x \in V(\mathcal{U}_R)} H(\ell_t^{\mathcal{U}_R}(x)) \right] \right).$$

*Proof.* The event  $\{X_{\infty}^{\mathcal{U}_R} = +\infty\}$  has probability 1 because on the edges connecting the units of  $\mathcal{U}_R$  (see Fig. 2.4) there is a drift downwards. To see why, note that  $\frac{1}{d+1} < \frac{1}{2} < \frac{d}{d+1}$  because  $d \geq 2$ , and use that a one-dimensional random walk with drift is transient to the right [33].  $\square$

Since  $\mathcal{U}_R$  is periodic, we can *fold*  $X^{\mathcal{U}_R}$  onto a single unit  $\mathcal{W}_R$ , to obtain a Markov renewal process  $X^{\mathcal{W}_R}$  on  $\mathcal{W}_R$  (see Fig. 2.5) in which the transition from the top vertex to the right-most bottom vertex has probability  $\frac{1}{d+1}$ , while the reverse transition has probability  $\frac{d}{d+1}$ . Clearly, the sojourn time distributions are not affected by the folding and therefore remain as above. Write  $(\ell_t^{\mathcal{W}_R}(x))_{x \in V(\mathcal{W}_R)}$  to denote the local times of  $X^{\mathcal{W}_R}$  at time  $t$ .



**Lemma 2.3.5.** *[Periodisation to a single finite subtree] For every  $R \in \mathbb{N} \setminus \{1\}$  and  $t \geq 0$ ,*

$$\mathbb{E}_{\mathcal{O}} \left( \exp \left[ \sum_{x \in V(\mathcal{U}_R)} H(\ell_t^{\mathcal{U}_R}(x)) \right] \right) \leq \mathbb{E}_{\mathcal{O}} \left( \exp \left[ \sum_{x \in V(\mathcal{W}_R)} H(\ell_t^{\mathcal{W}_R}(x)) \right] \right).$$

*Proof.* The periodisation again stacks local time on top of each other.  $\square$

Before we proceed we make a *crucial observation*, namely, we may still choose the shift  $z \in \{0, 1, \dots, R-1\}$  of the cuts of the two-sided backbone  $\mathbb{Z}$  (recall Fig. 2.3). We will do so in such a way that the *local time* up to time  $t$  spent in the set  $\partial_{\mathcal{U}_R}$  defined by

$$\begin{aligned} \partial_{\mathcal{U}_R} &= \text{all vertices at the top or at the bottom of a unit in } \mathcal{U}_R \\ &= \text{all vertices marked by } \bullet \text{ in Fig. 2.4} \end{aligned} \quad (2.19)$$

is at most  $t/R$ . After the periodisation these vertices are mapped to the set  $\partial_{\mathcal{W}_R}$  defined by

$$\begin{aligned} \partial_{\mathcal{W}_R} &= \text{all vertices at the top or at the bottom of } \mathcal{W}_R \\ &= \text{all vertices marked by } \bullet \text{ in Fig. 2.5.} \end{aligned}$$

**Lemma 2.3.6.** *[Control on the time spent at the boundary] For every  $R \in \mathbb{N} \setminus \{1\}$  and  $t \geq 0$ ,*

$$\begin{aligned} \mathbb{E}_{\mathcal{O}} \left( \exp \left[ \sum_{x \in V(\mathcal{U}_R)} H(\ell_t^{\mathcal{U}_R}(x)) \right] \right) \\ \leq \mathbb{E}_{\mathcal{O}} \left( \exp \left[ \sum_{x \in V(\mathcal{W}_R)} H(\ell_t^{\mathcal{W}_R}(x)) \right] 1_{\left\{ \frac{1}{t} \sum_{x \in \partial_{\mathcal{W}_R}} \ell_t^{\mathcal{W}_R}(x) \leq 1/R \right\}} \right). \end{aligned}$$

*Proof.* For different  $z$  the sets of vertices making up  $\partial_R$  correspond to *disjoint* sets of vertices in  $\mathcal{T}^{\mathbb{Z}}$  (see Fig. 2.4). Since  $\sum_{x \in \mathcal{T}^{\mathbb{Z}}} \ell_t^{\mathbb{Z}}(x) = t$  for all  $t \geq 0$ , it follows that there exists a  $z$  for which  $\sum_{x \in \partial_R} \ell_t^{\mathbb{Z}}(x) \leq t/R$ . Therefore the upper bound in Lemma 2.3.3 can be strengthened to the one that is claimed.  $\square$

## Upper variational formula

Lemmas 2.3.1–2.3.6 provide us with an upper bound for the average total mass (recall ((2.9)) on the *infinite* tree  $\mathcal{T}$  in terms of the same quantity on the *finite* tree-like unit  $\mathcal{W}_R$  with a *specific boundary condition*. Along the way we have paid a price: the sojourn times in the tadpoles are *no longer* exponentially distributed, and the transition probabilities into and out of the tadpoles and between the top vertex and the right-most bottom vertex are *biased*. We therefore need the large deviation principle for the empirical distribution of Markov renewal processes derived in [31], which we can now apply to the upper bound.

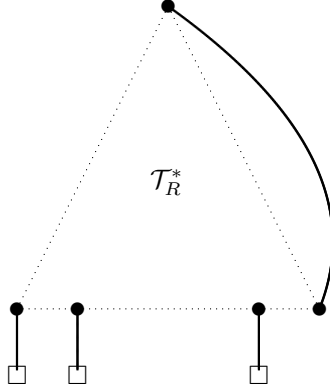


Figure 2.5: A unit  $\mathcal{W}_R$  with the top vertex and the right-most bottom vertex connected by an edge.

Since  $\mathcal{W}_R$  is finite, Lemma 2.1.2 gives

$$\langle U(t) \rangle \leq e^{H(t)+o(t)} \mathbb{E}_{\mathcal{O}} \left( e^{-\varrho J_V(\mathcal{W}_R)(L_t^{\mathcal{W}_R})} 1_{\{L_t^{\mathcal{W}_R}(\partial \mathcal{W}_R) \leq 1/R\}} \right)$$

with  $J_V$  the functional defined in (2.10). The following lemma controls the expectation in the right-hand side.

**Lemma 2.3.7.** [Scaling of the key expectation] For every  $R \in \mathbb{N} \setminus \{1\}$ ,

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E}_{\mathcal{O}} \left( e^{-\varrho t J_V(\mathcal{W}_R)(L_t^{\mathcal{W}_R})} 1_{\{L_t^{\mathcal{W}_R}(\partial \mathcal{W}_R) \leq 1/R\}} \right) = -\chi_R^+(\varrho),$$

where

$$\chi_R^+(\varrho) = \inf_{\substack{p \in \mathcal{P}(V(\mathcal{W}_R)) : \\ p(\partial \mathcal{W}_R) \leq 1/R}} \left\{ I_{E(\mathcal{W}_R)}^\dagger(p) + \varrho J_V(\mathcal{W}_R)(p) \right\}, \quad (2.20)$$

with

$$I_{E(\mathcal{W}_R)}^\dagger(p) = \inf_{\beta \in (0, \infty)} \inf_{q \in \mathcal{P}(V(\mathcal{W}_R))} [\hat{K}(\beta q) + \tilde{K}(p | \beta q)], \quad (2.21)$$

where

$$\hat{K}(\beta q) = \sup_{\hat{q} \in \mathcal{P}(V(\mathcal{W}_R))} \sum_{x \in V(\mathcal{W}_R)} \beta q(x) \log \left( \frac{\hat{q}(x)}{\sum_{y \in V(\mathcal{W}_R)} \pi_{x,y} \hat{q}(y)} \right), \quad (2.22)$$

$$\tilde{K}(p | \beta q) = \sum_{x \in V(\mathcal{W}_R)} \beta q(x) (\mathcal{L}\lambda_x) \left( \frac{p(x)}{\beta q(x)} \right), \quad (2.23)$$

with

$$(\mathcal{L}\lambda_x)(\alpha) = \sup_{\theta \in \mathbb{R}} [\alpha \theta - \lambda_x(\theta)], \quad \alpha \in [0, \infty), \quad (2.24)$$

$$\lambda_x(\theta) = \log \int_0^\infty e^{\theta \tau} \psi_x(d\tau), \quad \theta \in \mathbb{R}, \quad (2.25)$$

where  $\psi_x = \psi$  when  $x$  is a tadpole,  $\psi_x = \text{EXP}(d+1)$  when  $x$  is not a tadpole, and  $\pi_{x,y}$  is the transition kernel of the discrete-time Markov chain on  $V(\mathcal{W}_R)$  embedded in  $X^{\mathcal{W}_R}$ .

*Proof.* Apply the large deviation principle derived in [31], which we recall in Proposition A.1.1 in Appendix A.1.  $\square$

The expression in (2.20) is similar to (2.11) with  $G = \mathcal{W}_R$ , expect that the rate function  $I_{E(\mathcal{W}_R)}$  in (2.21) is more involved than the rate function  $I_E$  in (2.10).

### §2.3.2 Limit of the upper variational formula

The prefactor  $e^{H(t)+o(1)}$  in Lemma 2.1.2 accounts for the terms  $\varrho \log(\varrho t) - \varrho$  in the right-hand side of (2.13) (recall 2.8). In view of Lemma 2.3.7, in order to complete the proof of the upper bound in Theorem 2.1.3 it suffices to prove the following lemma.

**Lemma 2.3.8.** *For any  $d \geq 4$ ,  $\liminf_{R \rightarrow \infty} \chi_R^+(\varrho) \geq \chi\tau(\varrho)$ .*

*Proof.* The proof is given in Section 2.4 and relies on two steps:

- Show that, for  $d \geq 4$ ,

$$I_{E(\mathcal{W}_R)}^+(p) \geq I_{E(\mathcal{W}_R)}^+(p) + O(1/R) \quad (2.26)$$

with  $I_{E(\mathcal{W}_R)}^+$  a rate function similar to the *standard rate function*  $I_{E(\mathcal{W}_R)}$  given by (2.10).

- Show that,  $d \geq 2$ ,

$$\hat{\chi}_R^+(\varrho) = \inf_{\substack{p \in \mathcal{P}(V(\mathcal{W}_R)): \\ p(\partial \mathcal{W}_R) \leq 1/R}} \left\{ I_{E(\mathcal{W}_R)}^+(p) + \varrho J_{V(\mathcal{W}_R)}(p) \right\}$$

satisfies

$$\liminf_{R \rightarrow \infty} \hat{\chi}_R^+(\varrho) \geq \chi\tau(\varrho). \quad (2.27)$$

$\square$

## §2.4 Analysis of the upper variational formula

In this section we carry out the proof of the claims in Section 2.3.2, namely, we settle (2.26) in Section 2.4.1 and (2.27) in Section 2.4.2. The computations carried out in Appendix A.3 guide us along the way.

## §2.4.1 Identification of the rate function for the local times on the truncated tree

To identify the rate function  $I_{E(\mathcal{W}_R)}^\dagger$  in Lemma 2.3.7, we need to work out the two infima between braces in (2.20). The computation follows the same line of argument as in Appendix A.3, but is more delicate. We will only end up with a lower bound. However, this is sufficient for the upper variational formula.

To simplify the notation we write (recall Fig. 2.5):

$(V_R, E_R)$	=	vertex and edge set of $\mathcal{W}_R$ <i>without the tadpoles</i> ,
$\mathcal{O}$	=	<i>top</i> vertex of $V_R$ ,
$\star$	=	<i>right-most bottom</i> vertex of $V_R$ ,
$\partial V_R$	=	set of vertices at the <i>bottom</i> of $V_R$ ,
$\square$	=	set of <i>tadpoles</i> ,
$\square_x$	=	tadpole attached to $x \in \partial V_R \setminus \star$ .

Note that  $\partial V_R$  consists of  $\star$  and the vertices to which the tadpoles are attached. Note that  $\text{int}(V_R) = V_R \setminus \partial V_R$  *includes*  $\mathcal{O}$ .

1. Inserting (A.9) in Appendix A.2 into (2.22)–(2.23), we get

$$I_{E(\mathcal{W}_R)}^\dagger(p) = (d+1) \sum_{x \in V_R} p(x) + \inf_{\beta \in (0, \infty)} \inf_{q \in \mathcal{P}(V_R)} \sup_{\hat{q} \in \mathcal{P}(V_R)} L(\beta, q, \hat{q} \mid p)$$

with

$$L(\beta, q, \hat{q} \mid p) = -A - B - C - D,$$

where

$$\begin{aligned} A &= \sum_{x \in \text{int}(V_R)} \beta q(x) \left\{ 1 + \log \left( \frac{\sum_{y \sim x} \hat{q}(y)}{\hat{q}(x)} \frac{p(x)}{\beta q(x)} \right) \right\}, \\ B &= \sum_{x \in \partial V_R \setminus \star} \beta q(x) \left\{ 1 + \log \left( \frac{\hat{q}(x^\uparrow) + d\hat{q}(\square_x)}{\hat{q}(x)} \frac{p(x)}{\beta q(x)} \right) \right\}, \\ C &= \beta q(\star) \left\{ 1 + \log \left( \frac{\hat{q}(\star^\uparrow) + d\hat{q}(\mathcal{O})}{\hat{q}(\star)} \frac{p(\star)}{\beta q(\star)} \right) \right\}, \\ D &= \sum_{x \in \square} \beta q(x) \left\{ \log \left( \frac{\hat{q}(x^\uparrow)}{\hat{q}(x)} \right) - (\mathcal{L}\lambda) \left( \frac{p(x)}{\beta q(x)} \right) \right\}, \end{aligned}$$

with  $\mathcal{L}\lambda$  the Legende transform of the cumulant generating function of  $\psi$  (recall (2.25)) and  $x^\uparrow$  the unique vertex to which  $x$  is attached upwards. (Recall that  $y \sim x$  means that  $x$  and  $y$  are connected by an edge in  $E_R$ .) Note that  $A, B, C$  each combine two terms, and that  $A, B, C, D$  depend on  $p$ . We suppress this dependence because  $p$  is fixed.

2. Inserting the parametrisation  $\hat{q} = u/\|u\|_1$  and  $q = v/\|v\|_1$  with  $u, v: V_R \rightarrow (0, \infty)$  and putting  $\beta q = v$ , we may write

$$I_{E(\mathcal{W}_R)}^\dagger(p) = (d+1) \sum_{x \in V_R} p(x) + \inf_{v: V_R \rightarrow (0, \infty)} \sup_{u: V_R \rightarrow (0, \infty)} L(u, v) \quad (2.28)$$

with

$$L(u, v) = -A - B - C - D,$$

where

$$\begin{aligned} A &= \sum_{x \in \text{int}(V_R)} v(x) \left\{ 1 + \log \left( \frac{\sum_{y \sim x} u(y)}{u(x)} \frac{p(x)}{v(x)} \right) \right\}, \\ B &= \sum_{x \in \partial V_R \setminus \star} v(x) \left\{ 1 + \log \left( \frac{u(x^\uparrow) + du(\square_x)}{u(x)} \frac{p(x)}{v(x)} \right) \right\}, \\ C &= v(\star) \left\{ 1 + \log \left( \frac{u(\star^\uparrow) + du(\mathcal{O})}{u(\star)} \frac{p(\star)}{v(\star)} \right) \right\}, \\ D &= \sum_{x \in \square} v(x) \left\{ \log \left( \frac{u(x^\uparrow)}{u(x)} \right) - (\mathcal{L}\lambda) \left( \frac{p(x)}{v(x)} \right) \right\}. \end{aligned} \tag{2.29}$$

Our task is to carry out the supremum over  $u$  and the infimum over  $v$  in (2.28).

**3.** First, we compute the infimum over  $v$  for fixed  $u$ . (Later we will make a judicious choice for  $u$  to obtain a lower bound.) Abbreviate

$$\begin{aligned} A_u(x) &= \frac{\sum_{y \sim x} u(y)}{u(x)} p(x), & x \in \text{int}(V_R), \\ B_u(x) &= \frac{u(x^\uparrow) + du(\square_x)}{u(x)} p(x), & x \in \partial V_R \setminus \star, \\ C_u(\star) &= \frac{u(\star^\uparrow) + du(\mathcal{O})}{u(\star)} p(\star). \end{aligned} \tag{2.30}$$

• For  $z \in V_R$ , the first derivatives of  $L$  are

$$\begin{aligned} z \in \text{int}(V_R): \quad & \frac{\partial L(u, v)}{\partial v(z)} = -\log \left( \frac{A_u(z)}{v(z)} \right), \\ z \in \partial V_R \setminus \star: \quad & \frac{\partial L(u, v)}{\partial v(z)} = -\log \left( \frac{B_u(z)}{v(z)} \right), \\ z = \star: \quad & \frac{\partial L(u, v)}{\partial v(z)} = -\log \left( \frac{C_u(z)}{v(z)} \right), \end{aligned}$$

while the second derivatives of  $L$  equal  $1/v(z) > 0$ . Hence the infimum is uniquely taken at

$$\begin{aligned} x \in \text{int}(V_R): \quad & \bar{v}(x) = A_u(x), \\ x \in V_R \setminus \star: \quad & \bar{v}(x) = B_u(x), \\ x = \star: \quad & \bar{v}(x) = C_u(x). \end{aligned}$$

• For  $z \in \square$ , the computation is more delicate. Define (see (A.6) in Appendix A.2)

$$\mu(\alpha) = \alpha(\mathcal{L}\lambda)'(\alpha) - (\mathcal{L}\lambda)(\alpha).$$

The function  $\mu$  has range  $(-\infty, \log \sqrt{d}]$ , with the maximal value uniquely taken at  $\alpha = \infty$ . Therefore there are two cases.

►  $u(x)/u(x^\uparrow) \leq \sqrt{d}$ : Abbreviate  $\alpha_u(z) = p(z)/v(z)$ . For  $z \in \square$ ,

$$\begin{aligned} \frac{\partial L(u, v)}{\partial v(z)} &= \log \left( \frac{u(z)}{u(z^\uparrow)} \right) + (\mathcal{L}\lambda) \left( \frac{p(z)}{v(z)} \right) - \frac{p(z)}{v(z)} (\mathcal{L}\lambda)' \left( \frac{p(z)}{v(z)} \right) \\ &= \log \left( \frac{u(z)}{u(z^\uparrow)} \right) - \mu(\alpha_u(z)), \\ \frac{\partial^2 L(u, v)}{v(z)^2} &= \frac{p^2(z)}{v^3(z)} (\mathcal{L}\lambda)'' \left( \frac{p(z)}{v(z)} \right) > 0, \end{aligned}$$

where we use that  $\mathcal{L}\lambda$ , being a Legendre transform, is strictly convex. Hence the infimum is uniquely taken at

$$\bar{v}(x) = \frac{p(x)}{\alpha_u(x)}, \quad x \in \square,$$

with  $\alpha_u(x)$  solving the equation

$$\log \left( \frac{u(x)}{u(x^\uparrow)} \right) = \mu(\alpha_u(x)), \quad x \in \square.$$

Since  $\mu'(\alpha) = \alpha(\mathcal{L}\lambda)''(\alpha)$  and  $\mathcal{L}\lambda$  is strictly convex (see Fig. A.3 in Appendix A.2),  $\mu$  is strictly increasing and therefore invertible. Consequently,

$$\alpha_u(x) = \mu^{-1} \left( \log \left( \frac{u(x)}{u(x^\uparrow)} \right) \right), \quad x \in \square.$$

Putting the above formulas together, we arrive at (recall (2.30))

$$\begin{aligned} L(u) &= \inf_{v: V_R \rightarrow (0, \infty)} L(u, v) \\ &= - \sum_{x \in \text{int}(V_R)} A_u(x) - \sum_{x \in \partial V_R \setminus \star} B_u(x) - C_u(\star) + \sum_{x \in \square} D_u(x) \end{aligned} \quad (2.31)$$

with (recall (2.29))

$$\begin{aligned} D_u(x) &= -\frac{p(x)}{\alpha_u(x)} \left[ \log \left( \frac{u(x^\uparrow)}{u(x)} \right) - (\mathcal{L}\lambda)(\alpha_u(x)) \right] \\ &= \frac{p(x)}{\alpha_u(x)} [(\mathcal{L}\lambda)(\alpha_u(x)) - \mu(\alpha_u(x))] \\ &= p(x) (\mathcal{L}\lambda)'(\alpha_u(x)) = p(x) ((\mathcal{L}\lambda)' \circ \mu^{-1}) \left( \log \left( \frac{u(x)}{u(x^\uparrow)} \right) \right). \end{aligned}$$

In (A.8) in Appendix A.2 we show that  $(\mathcal{L}\lambda)' \circ \mu^{-1} = \lambda^{-1}$ . Moreover In (A.14) in Appendix A.2 we show that  $(\lambda^{-1} \circ \log) = S$  with

$$S(\beta) = d + 1 - \beta - \frac{d}{\beta}, \quad \beta \in (0, \sqrt{d}].$$

Since  $S$  has domain  $(0, \sqrt{d}]$ ,  $D_u(x)$  is only defined when  $u(x)/u(x^\uparrow) \leq \sqrt{d}$ , in which case

$$D_u(x) = p(x) S\left(\frac{u(x)}{u(x^\uparrow)}\right), \quad x \in \square. \quad (2.32)$$

►  $u(x)/u(x^\uparrow) > \sqrt{d}$ : In this case  $\frac{\partial L(u,v)}{\partial v(z)} > 0$ , the infimum is uniquely taken at  $\bar{v}(x) = 0$ , and

$$D_u(x) = p(x) (\sqrt{d} - 1)^2 = p(x) S(\sqrt{d}), \quad x \in \square,$$

where we use (A.13). Note that the right-hand side does not depend on  $u$ .

4. Next, we compute the supremum over  $u$ . The first derivatives of  $L$  are

$$\begin{aligned} z \in \text{int}(V_R) \setminus \mathcal{O}: \quad & \frac{\partial L(u)}{\partial u(z)} = \frac{\sum_{y \sim z} u(y)}{u^2(z)} p(z) - \sum_{y \sim z} \frac{1}{u(y)} p(y), \\ z = \mathcal{O}: \quad & \frac{\partial L(u)}{\partial u(\mathcal{O})} = \frac{\sum_{y \sim \mathcal{O}} u(y)}{u(\mathcal{O})^2} p(\mathcal{O}) - \sum_{y: y^\uparrow = \mathcal{O}} \frac{1}{u(y)} p(y) - \frac{d}{u(\star)} p(\star), \\ z = \star: \quad & \frac{\partial L(u)}{\partial u(\star)} = -\frac{1}{u(\mathcal{O})} p(\mathcal{O}) + \frac{u(\star^\uparrow) + du(\mathcal{O})}{u(\star)^2} p(\star), \\ z \in \partial V_R \setminus \star: \quad & \frac{\partial L(u)}{\partial u(z)} = -\frac{1}{u(z^\uparrow)} p(z^\uparrow) + \frac{u(z^\uparrow) + du(\square_z)}{u(z)^2} p(z) \\ & + \left[ \frac{u(\square_z)}{u(z)^2} - \frac{d}{u(\square_z)} \right] p(\square_z) 1_{\left\{ \frac{u(z)}{u(z^\uparrow)} \leq \sqrt{d} \right\}}, \\ z \in \square: \quad & \frac{\partial L(u)}{\partial u(z)} = -\frac{d}{u(z^\uparrow)} p(z^\uparrow) + \left[ -\frac{1}{u(z^\uparrow)} + \frac{du(z^\uparrow)}{u(z)^2} \right] p(z) 1_{\left\{ \frac{u(z)}{u(z^\uparrow)} \leq \sqrt{d} \right\}}. \end{aligned} \quad (2.33)$$

The second derivatives of  $L$  are all  $< 0$ . The first line in (2.33) can be rewritten as

$$\sum_{y \sim z} u(y) \left[ \frac{p(z)}{u^2(z)} - \frac{p(y)}{u^2(y)} \right],$$

which is zero when

$$\bar{u}(x) = \sqrt{p(x)}, \quad x \in V_R. \quad (2.34)$$

Given the choice in (2.34), the fifth line in (2.33) is zero when

$$\bar{u}(x) = \sqrt{\frac{dp(x^\uparrow)p(x)}{dp(x^\uparrow) + p(x)}}, \quad x \in \square. \quad (2.35)$$

Indeed, the derivative is strictly negative when the indicator is 0 and therefore the indicator must be 1. But the latter is guaranteed by (2.34)–(2.35), which imply that

$$\frac{\bar{u}(x)}{\bar{u}(x^\uparrow)} = \sqrt{\frac{dp(x)}{dp(x^\uparrow) + p(x)}} \leq \sqrt{d}, \quad x \in \square.$$

Given the choice in (2.34)–(2.35), also the fourth line in (2.33) is zero. Thus, only the second and third line in (2.33) are non-zero, but this is harmless because  $\mathcal{O}, \star$  carry a negligible weight in the limit as  $R \rightarrow \infty$  because of the constraint  $p(\partial V_R \cup \mathcal{O}) \leq 1/R$  in Lemma 2.3.7 (recall (2.19)).

Inserting (2.34)–(2.35) into (2.31) and using (2.30), (2.32), we get the following lower bound:

$$\begin{aligned}
 & \sup_{u: V_R \rightarrow (0, \infty)} L(u) \\
 & \geq - \sum_{x \in \text{int}(V_R)} A_{\bar{u}}(x) - \sum_{x \in \partial V_R \setminus \star} B_{\bar{u}}(x) - C_{\bar{u}}(\star) + \sum_{x \in \square} D_{\bar{u}}(x) \\
 & = - \sum_{x \in \text{int}(V_R)} \sum_{y \sim x} \sqrt{p(y)p(x)} - \sum_{x \in \partial V_R \setminus \star} \sqrt{p(x)} \left( \sqrt{p(x^\uparrow)} + d \sqrt{\frac{dp(x)p(\square_x)}{dp(x) + p(\square_x)}} \right) \\
 & \quad - \sqrt{p(\star)} \left( \sqrt{p(\star^\uparrow)} + d \sqrt{p(\mathcal{O})} \right) \\
 & \quad + \sum_{x \in \square} p(x) \left( d + 1 - \sqrt{d} \left[ \sqrt{\frac{p(x)}{dp(x^\uparrow) + p(x)}} + \sqrt{\frac{dp(x^\uparrow) + p(x)}{p(x)}} \right] \right).
 \end{aligned}$$

5. Using the relation  $(d+1)p(x) = \sum_{y \sim x} p(x)$ ,  $x \in \text{int}(V_R)$ , we get from (2.28) that

$$I_{E(\mathcal{W}^R)}^\dagger(p) \geq K_R^1(p) + K_R^2(p)$$

with

$$\begin{aligned}
 K_R^1(p) &= \sum_{x \in \text{int}(V_R)} \sum_{y \sim x} \left[ p(x) - \sqrt{p(x)p(y)} \right] \\
 &= \sum_{\{x, y\} \in \widehat{E}_R} \left( \sqrt{p(x)} - \sqrt{p(y)} \right)^2 + \left[ p(\mathcal{O}) - \sqrt{p(\mathcal{O})p(\star)} \right] - \sum_{x \in \partial V_R} \left[ p(x) - \sqrt{p(x)p(x^\uparrow)} \right]
 \end{aligned}$$

and

$$\begin{aligned}
 K_R^2(p) &= \sum_{x \in \partial V_R \setminus \star} \left[ (d+1)p(x) - \sqrt{p(x)} \left( \sqrt{p(x^\uparrow)} + d \sqrt{\frac{dp(x)p(\square_x)}{dp(x) + p(\square_x)}} \right) \right] \\
 & \quad + (d+1)p(\star) - \sqrt{p(\star)} \left( \sqrt{p(\star^\uparrow)} + d \sqrt{p(\mathcal{O})} \right) \\
 & \quad + \sum_{x \in \square} p(x) \left[ d + 1 - \sqrt{d} \left( \sqrt{\frac{p(x)}{dp(x^\uparrow) + p(x)}} + \sqrt{\frac{dp(x^\uparrow) + p(x)}{p(x)}} \right) \right].
 \end{aligned}$$

The first sum in the right-hand side of  $K_R^1(p)$  equals the *standard rate function*  $I_{\widehat{E}_R}(p)$  given by (2.10), with

$$\widehat{E}_R = E_R \setminus \{\mathcal{O}, \star\}$$



the set of edges in the unit  $\mathcal{W}_R$  *without the tadpoles and without the edge  $\{\mathcal{O}, \star\}$*  (i.e.,  $\widehat{E}_R = E(\mathcal{T}_R^*)$ ; recall Fig. 2.4). Rearranging and simplifying terms, we arrive at

$$I_{E(\mathcal{W}_R)}^\dagger(p) \geq I_{\widehat{E}_R}(p) + K_R^3(p) \quad (2.36)$$

with

$$K_R^3(p) = S_{\partial V_R \setminus \star}(p) + S_{\mathcal{O}, \star}(p) + S_{(\partial V_R \setminus \star) \cup \square}(p),$$

where

$$\begin{aligned} S_{\partial V_R \setminus \star}(p) &= d \sum_{x \in \partial V_R \setminus \star} p(x), \\ S_{\mathcal{O}, \star}(p) &= \left( \sqrt{p(\mathcal{O})} - \sqrt{p(\star)} \right)^2 + (d-1) [p(\star) - \sqrt{p(\mathcal{O})p(\star)}], \\ S_{(\partial V_R \setminus \star) \cup \square}(p) &= - \sum_{x \in \partial V_R \setminus \star} p(x) d \sqrt{\frac{dp(\square_x)}{dp(x) + p(\square_x)}} \\ &\quad + \sum_{x \in \partial V_R \setminus \star} p(\square_x) \left( d+1 - \sqrt{d} \left[ \sqrt{\frac{p(\square_x)}{dp(x) + p(\square_x)}} + \sqrt{\frac{dp(x) + p(\square_x)}{p(\square_x)}} \right] \right). \end{aligned} \quad (2.37)$$

6. Since  $\sqrt{p(\mathcal{O})p(\star)} \leq \frac{1}{2}[p(\mathcal{O}) + p(\star)]$ , the boundary constraint  $\sum_{x \in \partial V_R \cup \mathcal{O}} p(x) \leq 1/R$  implies that  $S_{\partial V_R \setminus \star}(p) + S_{\mathcal{O}, \star}(p) = O(1/R)$ . The same constraint implies that the first sum in  $S_{(\partial V_R \setminus \star) \cup \square}(p)$  is  $O(1/R)$ . Hence

$$K_R^3(p) = O(1/R) + \sum_{x \in \partial V_R \setminus \star} p(x) F\left(\frac{p(\square_x)}{p(x)}\right)$$

with

$$F(w) = w \left( d+1 - \sqrt{d} \left[ \sqrt{\frac{w}{d+w}} + \sqrt{\frac{d+w}{w}} \right] \right).$$

The map  $w \mapsto F(w)$  is continuous on  $(0, \infty)$  with

$$F(w) = \begin{cases} -\sqrt{w} + (d+1)w + O(w^{3/2}), & w \downarrow 0, \\ [(d+1) - 2\sqrt{d}]w + O(w^{-1}), & w \rightarrow \infty. \end{cases}$$

From this we see that if  $d \geq 4$ , then there exists a  $C \in (1, \infty)$  such that

$$F(w) + C \geq (1 - \sqrt{w})^2, \quad w \in [0, \infty). \quad (2.38)$$

Hence we have the lower bound

$$\begin{aligned} K_R^3(p) &\geq O(1/R) + \sum_{x \in \partial V_R \setminus \star} p(x) \left[ -C + \left( 1 - \sqrt{\frac{p(\square_x)}{p(x)}} \right)^2 \right] \\ &= O(1/R) + \sum_{x \in \partial V_R \setminus \star} \left( \sqrt{p(x)} - \sqrt{p(\square_x)} \right)^2. \end{aligned}$$

Via (2.36)–(2.37), it follows that

$$I_{E(\mathcal{W}_R)}^\dagger(p) \geq O(1/R) + I_{\widetilde{E}_R}(p), \quad R \in \mathbb{N}, \quad (2.39)$$

with  $I_{\widetilde{E}_R}(p)$  the *standard rate function* given by (2.10), with

$$\widetilde{E}_R = \widehat{E}_R \cup \left[ \cup_{x \in \partial V_R \setminus \star} \{x, \square_x\} \right]$$

the set of edges in the unit  $\widetilde{\mathcal{W}}_R$  that is obtained from the unit  $\mathcal{W}_R$  by *removing the edge*  $\{\mathcal{O}, \star\}$  (i.e.,  $\widetilde{E}_R = E(\widetilde{\mathcal{W}}_R)$ ; recall Fig. 2.5). This completes the proof of (2.26).

**Remark 2.4.1.** The condition  $d \geq 4$  is needed only in (2.38). For  $d = 2, 3$  we have  $F(w) + C \geq \theta_c(1 - \sqrt{w})^2$  with  $\theta_c = d + 1 - 2\sqrt{d} \in (0, 1)$ . Consequently, the edges  $\{x, \square_x\}$ ,  $x \in \partial V_R \setminus \star$ , carry a weight that is smaller than that of the edges in  $\mathcal{T}$ , which may cause the optimal  $p$  to stick to the boundary as  $R \rightarrow \infty$ , in which case we do not have (2.39).  $\spadesuit$

## §2.4.2 Limit of the upper variational formula

Note that

$$\widetilde{\mathcal{W}}_R \subseteq \mathcal{T},$$

with  $\mathcal{T}$  the infinite tree. Consequently,

$$I_{\widetilde{E}_R}(p) = I_{E(\mathcal{T})}(p) - (d-1) \sum_{x \in \partial V_R \setminus \star} p(x), \quad \forall p \in \mathcal{P}(V(\mathcal{T})): \text{supp}(p) = V(\widetilde{\mathcal{W}}_R),$$

where the sum compensates for the contribution coming from the edges in  $\mathcal{T}$  that link the vertices in  $\partial V_R \setminus \star$  to the vertices one layer deeper in  $\mathcal{T}$  that are not tadpoles. Since this sum is  $O(1/R)$ , we obtain (recall (2.20))

$$\begin{aligned} \chi_R^+(\varrho) &= \inf_{\substack{p \in \mathcal{P}(V(\mathcal{W}_R)): \\ p(\partial \mathcal{W}_R) \leq 1/R}} \left\{ I_{E(\mathcal{W}_R)}^\dagger(p) + \varrho J_{V(\mathcal{W}_R)}(p) \right\} \\ &\geq O(1/R) + \inf_{\substack{p \in \mathcal{P}(V(\mathcal{T})): \\ \text{supp}(p) = V(\widetilde{\mathcal{W}}_R), p(\partial \widetilde{\mathcal{W}}_R) \leq 1/R}} \left\{ I_{E(\mathcal{T})}(p) + \varrho J_{V(\mathcal{T})}(p) \right\} \\ &\geq O(1/R) + \chi_{\mathcal{T}}(\rho), \end{aligned}$$

where the last inequality follows after dropping the constraint under the infimum and recalling (2.14). This completes the proof of (2.27).

## §2.5 Analysis of the variational problem on the infinite regular tree

In this Section we prove Theorem 2.1.4. Section 2.5.1 formulates two theorems that imply Theorem 2.1.4, Section 2.5.2 provides the proof of these theorems. Recall the

definition of  $\mathcal{P}(V)$ ,  $I_E(p)$  and  $J_V(p)$  from (2.10). Set

$$\chi_{\mathcal{T}}(\varrho) = \inf_{p \in \mathcal{P}_{\mathcal{O}}(V)} [I_E(p) + \varrho J_V(p)], \quad \varrho \in (0, \infty), \quad (2.40)$$

where  $\mathcal{P}_{\mathcal{O}}(V) = \{p \in \mathcal{P}(V) : \operatorname{argmax} p = \mathcal{O}\}$ . Since  $\mathcal{P}(V)$ ,  $I_E$  and  $J_V$  are invariant under translations, the centering at  $\mathcal{O}$  is harmless.

## §2.5.1 Two properties

**Theorem 2.5.1.** *For every  $\varrho \in (0, \infty)$  the infimum in (2.40) is attained, and every minimiser  $\bar{p}$  is strictly positive, non-increasing along backbones, and such that*

$$\sum_{N \in \mathbb{N}_0} \partial S_R \log(R+1) \leq \frac{d+1}{\varrho}, \quad \partial S_R = \sum_{\partial B_R(\mathcal{O})} \bar{p}(x),$$

where  $B_R(\mathcal{O})$  is the ball of radius  $R$  around  $\mathcal{O}$ .

**Theorem 2.5.2.** *The function  $\varrho \mapsto \chi_{\mathcal{T}}(\varrho)$  is strictly increasing and globally Lipschitz continuous on  $(0, \infty)$ , with  $\lim_{\varrho \downarrow 0} \chi_{\mathcal{T}}(\varrho) = d-1$  and  $\lim_{\varrho \rightarrow \infty} \chi_{\mathcal{T}}(\varrho) = d+1$ .*

Theorems 2.5.1–2.5.2 settle Theorem 2.1.4. Their proof uses the following two lemmas.

**Lemma 2.5.3.** *For every  $\varrho \in (0, \infty)$ , the infimum in (2.40) may be restricted to  $p \in \mathcal{P}_{\mathcal{O}}(V)$  such that  $J_V(p) \leq \frac{d+1}{\varrho}$ .*

*Proof.* Let  $\delta_{\mathcal{O}} \in \mathcal{P}_{\mathcal{O}}(V)$  denote the point measure at  $\mathcal{O}$ . Then, for all  $\varrho \in (0, \infty)$ ,

$$\chi_{\mathcal{T}}(\varrho) \leq I_E(\delta_{\mathcal{O}}) + \varrho J_V(\delta_{\mathcal{O}}) = (d+1) + \varrho \times 0 = d+1.$$

Since  $I_V \geq 0$ , we may restrict the infimum in (2.40) to  $p$  with  $J_V(p) \leq \frac{d+1}{\varrho}$ . □

**Lemma 2.5.4.** *For every  $\varrho \in (0, \infty)$ , there exists a  $c(\varrho) > 0$  such that the infimum in (2.40) may be restricted to  $p \in \mathcal{P}_{\mathcal{O}}(V)$  such that  $J_V(p) \geq c(\varrho)$ .*

*Proof.* Since  $J_V(p) = 0$  if and only if  $p = \delta_{\mathcal{O}}$  is a point measure, it suffices to show that  $\delta_{\mathcal{O}}$  is not a minimiser of  $\chi_{\mathcal{T}}(\varrho)$ . To that end, for  $y \in V$  compute

$$\frac{\partial}{\partial p(y)} [I_E(p) + \varrho J_V(p)] = 1 - \sum_{z \sim y} \sqrt{\frac{p(z)}{p(y)}} - \varrho \log p(y) - \varrho. \quad (2.41)$$

Because  $p(\mathcal{O}) > 0$ , it follows that the right-hand side tends to  $-\infty$  as  $p(y) \downarrow 0$  for every  $y \sim \mathcal{O}$ . Hence, no  $p \in \mathcal{P}_{\mathcal{O}}(V)$  with  $p(y) = 0$  for some  $y \sim \mathcal{O}$  can be a minimiser of (2.40), or be the weak limit point of a minimising sequence. In particular,  $\delta_{\mathcal{O}}$  cannot. □

## §2.5.2 Proof of the two properties

*Proof of Theorem 2.5.1.* First observe that  $\mathcal{P}(V)$  and  $J_V$  are invariant under permutations, i.e., for any  $p \in \mathcal{P}(V)$  and any relabelling  $\pi$  of the vertices in  $V$ , we have  $\pi p \in \mathcal{P}(V)$  and  $J_V(\pi p) = J_V(p)$ . The same does not hold for  $I_E$ , but we can apply permutations such that  $I_E(\pi p) \leq I_E(p)$ .

1. Pick any  $p \in \mathcal{P}(V)$ . Pick any backbone  $\text{bb} = \{x_0, x_1, \dots\}$  that runs from  $x_0 = \mathcal{O}$  to infinity. Consider a permutation  $\pi$  that *reorders* the vertices in  $\text{bb}$  such that  $\{(\pi p)(x)\}_{x \in \text{bb}}$  becomes *non-increasing*. Together with the reordering, transport all the trees that hang off  $\text{bb}$  as well. Since  $\pi p$  is non-increasing along  $\text{bb}$ , while all the edges that do not lie on  $\text{bb}$  have the same neighbouring values in  $p$  and in  $\pi p$ , we have

$$I_E(\pi p) \leq I_E(p). \quad (2.42)$$

Indeed,

$$\frac{1}{2} [I_E(p) - I_E(\pi p)] = \sum_{k \in \mathbb{N}_0} \sqrt{(\pi p)(x_k)(\pi p)(x_{k+1})} - \sum_{k \in \mathbb{N}_0} \sqrt{p(x_k)p(x_{k+1})}, \quad (2.43)$$

where we use that  $p(x_0) = (\pi p)(x_0)$  (because  $p(x_0) \geq p(x_k)$  for all  $k \in \mathbb{N}$ ) and  $\sum_{k \in \mathbb{N}} p(x_k) = \sum_{k \in \mathbb{N}} (\pi p)(x_k)$ . The right-hand side of (2.43) is  $\geq 0$  by the rearrangement inequality for sums of products of two sequences [26, Section 10.2, Theorem 368]. In fact, strict inequality in (2.43) holds unless  $p$  is constant along  $\text{bb}$ . But this is impossible because it would imply that  $p(\mathcal{O}) = 0$  and hence  $p(x) = 0$  for all  $x \in V$ . Thus,  $p$  and  $\text{bb}$  being arbitrary, it follows from (2.42) that any minimiser or minimising sequence must be non-increasing in the distance to  $\mathcal{O}$ . Indeed, if it were not, then there would be a  $\text{bb}$  along which the reordering would lead to a lower value of  $I_E + \varrho J_V$ . Hence we may replace (2.40) by

$$\chi_{\mathcal{T}}(\varrho) = \inf_{p \in \mathcal{P}_{\mathcal{O}}^{\downarrow}(V)} [I_E(p) + \varrho J_V(p)], \quad \varrho \in (0, \infty), \quad (2.44)$$

with  $\mathcal{P}_{\mathcal{O}}^{\downarrow}(V)$  defined in (2.15).

2. Let  $p \in \mathcal{P}_{\mathcal{O}}^{\downarrow}(V)$ . Estimate

$$J_V(p) = \sum_{R \in \mathbb{N}_0} \sum_{x \in \partial B_R(\mathcal{O})} [-p(x) \log p(x)] \geq \sum_{R \in \mathbb{N}_0} \sum_{x \in \partial B_R(\mathcal{O})} \left[ -p(x) \log \left( \frac{1}{R+1} \right) \right],$$

where we use that  $p(x) \leq \frac{1}{R+1}$  for all  $x \in \partial B_R(\mathcal{O})$ . Hence

$$J_V(p) \geq \sum_{R \in \mathbb{N}_0} \partial S_R \log(R+1)$$

with  $\partial S_R = \sum_{x \in \partial B_R(\mathcal{O})} p(x)$ . By Lemma 2.5.3,  $J_V(p) \leq \frac{d+1}{\varrho}$ , and so

$$\sum_{R \in \mathbb{N}_0} \partial S_R \log(R+1) \leq \frac{d+1}{\varrho}. \quad (2.45)$$

The computation in (2.41) shows that any  $p$  for which there exist  $z \sim y$  with  $p(z) > 0$  and  $p(y) = 0$  cannot be minimiser nor a weak limit point of a minimising sequence. Hence all minimisers or weak limit points of minimising sequences are strictly positive everywhere.

**3.** Take any minimising sequence  $(p_n)_{n \in \mathbb{N}}$  of (2.44). By (2.45),  $\lim_{R \rightarrow \infty} \sum_{x \notin B_R(\mathcal{O})} p_n(x) = 0$  uniformly in  $n \in \mathbb{N}$ , and so  $(p_n)_{n \in \mathbb{N}}$  is tight. By Prokhorov's theorem, tightness is equivalent to  $(p_n)_{n \in \mathbb{N}}$  being relatively compact, i.e., there is a subsequence  $(p_{n_k})_{k \in \mathbb{N}}$  that converges weakly to a limit  $\bar{p} \in \mathcal{P}_{\mathcal{O}}^{\downarrow}(V)$ . By Fatou's lemma, we have  $\liminf_{k \rightarrow \infty} I_E(p_{n_k}) \geq I_E(\bar{p})$  and  $\liminf_{k \rightarrow \infty} J_V(p_{n_k}) \geq J_V(\bar{p})$ . Hence

$$\chi_{\mathcal{T}}(\varrho) = \lim_{k \rightarrow \infty} [I_E(p_{n_k}) + \varrho J_V(p_{n_k})] \geq I_E(\bar{p}) + \varrho J_V(\bar{p}).$$

Hence  $\bar{p}$  is a minimiser of (2.44).  $\square$

*Proof of Theorem 2.5.2.* The proof uses approximation arguments.

**1.** We first show that  $\varrho \mapsto \chi_{\mathcal{T}}(\varrho)$  is strictly increasing and globally Lipschitz. Pick  $\varrho_1 < \varrho_2$ . Let  $\bar{p}_{\varrho_1}$  be any minimiser of (2.40) at  $\varrho_1$ , i.e.,

$$\chi_{\mathcal{T}}(\varrho_1) = I_E(\bar{p}_{\varrho_1}) + \varrho_1 J_V(\bar{p}_{\varrho_1}).$$

Estimate

$$\begin{aligned} [I_E(\bar{p}_{\varrho_1}) + \varrho_1 J_V(\bar{p}_{\varrho_1})] &= [I_E(\bar{p}_{\varrho_1}) + \varrho_2 J_V(\bar{p}_{\varrho_1})] - (\varrho_2 - \varrho_1) J_V(\bar{p}_{\varrho_1}) \\ &\geq \chi_{\mathcal{T}}(\varrho_2) - (\varrho_2 - \varrho_1) J_V(\bar{p}_{\varrho_1}) \geq \chi_{\mathcal{T}}(\varrho_2) - (\varrho_2 - \varrho_1) \frac{d+1}{\varrho_1}, \end{aligned}$$

where we use Lemma 2.5.3. Therefore

$$\chi_{\mathcal{T}}(\varrho_2) - \chi_{\mathcal{T}}(\varrho_1) \leq (\varrho_2 - \varrho_1) \frac{d+1}{\varrho_1}.$$

Similarly, let  $\bar{p}_{\varrho_2}$  be any minimiser of (2.40) at  $\varrho_2$ , i.e.,

$$\chi_{\mathcal{T}}(\varrho_2) = I_E(\bar{p}_{\varrho_2}) + \varrho_2 J_V(\bar{p}_{\varrho_2}).$$

Estimate

$$\begin{aligned} [I_E(\bar{p}_{\varrho_2}) + \varrho_2 J_V(\bar{p}_{\varrho_2})] &= [I_E(\bar{p}_{\varrho_2}) + \varrho_1 J_V(\bar{p}_{\varrho_2})] + (\varrho_2 - \varrho_1) J_V(\bar{p}_{\varrho_2}) \\ &\geq \chi_{\mathcal{T}}(\varrho_1) + (\varrho_2 - \varrho_1) J_V(\bar{p}_{\varrho_2}) \geq \chi_{\mathcal{T}}(\varrho_1) + (\varrho_2 - \varrho_1) c(\varrho_2), \end{aligned}$$

where we use Lemma 2.5.4. Therefore

$$\chi_{\mathcal{T}}(\varrho_2) - \chi_{\mathcal{T}}(\varrho_1) \geq c(\varrho_2)(\varrho_2 - \varrho_1).$$

**2.** Because  $\chi_{\mathcal{T}}(\varrho) \leq d + 1$  for all  $\varrho \in (0, \infty)$ , it follows that  $\lim_{\varrho \rightarrow \infty} \chi_{\mathcal{T}}(\varrho) \leq d + 1$ . To obtain the reverse inequality, let  $\bar{p}_{\varrho}$  be any minimiser of (2.44) at  $\varrho$ . By Lemma 2.5.3, we may assume that  $J_V(\bar{p}_{\varrho}) \leq \frac{d+1}{\varrho}$ . Hence  $\lim_{\varrho \rightarrow \infty} J_V(\bar{p}_{\varrho}) = 0$ , and consequently  $\lim_{\varrho \rightarrow \infty} \bar{p}_{\varrho} = \delta_{\mathcal{O}}$  weakly. Therefore, by Fatou's lemma,  $\lim_{\varrho \rightarrow \infty} \chi_{\mathcal{T}}(\varrho) = \lim_{\varrho \rightarrow \infty} [I_E(\bar{p}_{\varrho}) + \varrho J_V(\bar{p}_{\varrho})] \geq \liminf_{\varrho \rightarrow \infty} I_E(\bar{p}_{\varrho}) \geq I_E(\delta_{\mathcal{O}}) = d + 1$ .

3. To prove that  $\lim_{\varrho \downarrow 0} \chi_{\mathcal{T}}(\varrho) \leq d - 1$ , estimate

$$\chi_{\mathcal{T}}(\varrho) \leq \inf_{\substack{p \in \mathcal{P}_{\mathcal{O}}^{\downarrow}(V) \\ \text{supp}(p) \subseteq B_R(\mathcal{O})}} [I_E(p) + \varrho J_V(p)], \quad R \in \mathbb{N}_0.$$

Because

$$\sup_{\substack{p \in \mathcal{P}_{\mathcal{O}}^{\downarrow}(V) \\ \text{supp}(p) \subseteq B_R(\mathcal{O})}} J_V(p) = J_V(p_R) = \log |B_R(\mathcal{O})|, \quad R \in \mathbb{N}_0,$$

with

$$p_R(x) = \begin{cases} |B_R(\mathcal{O})|^{-1}, & x \in B_R(\mathcal{O}), \\ 0, & \text{else,} \end{cases}$$

it follows that

$$\lim_{\varrho \downarrow 0} \chi_{\mathcal{T}}(\varrho) \leq \inf_{\substack{p \in \mathcal{P}_{\mathcal{O}}^{\downarrow}(V) \\ \text{supp}(p) \subseteq B_R(\mathcal{O})}} I_E(p) \leq I_E(p_R), \quad R \in \mathbb{N}_0.$$

Compute (recall (2.10)) ,

$$I_E(p_R) = \frac{|\partial B_{R+1}(\mathcal{O})|}{|B_R(\mathcal{O})|}, \quad R \in \mathbb{N}_0.$$

Inserting the relations

$$\begin{aligned} |\partial B_R(\mathcal{O})| &= \begin{cases} 1, & R = 0, \\ (d+1)d^{R-1}, & R \in \mathbb{N}, \end{cases} \\ |B_R(\mathcal{O})| &= \sum_{R'=0}^R |\partial B_{R'}(\mathcal{O})| = 1 + \frac{d+1}{d-1}(d^R - 1), \quad R \in \mathbb{N}_0, \end{aligned}$$

we get

$$I_E(p_R) = (d-1) \frac{(d+1)d^R}{(d+1)d^R - 2}.$$

Hence  $\lim_{R \rightarrow \infty} I_E(p_R) = d - 1$ , and so  $\lim_{\varrho \downarrow 0} \chi_{\mathcal{T}}(\varrho) \leq d - 1$ .

4. To prove that  $\lim_{\varrho \downarrow 0} \chi_{\mathcal{T}}(\varrho) \geq d - 1$ , note that because  $J_V \geq 0$  we can estimate

$$\lim_{\varrho \downarrow 0} \chi_{\mathcal{T}}(\varrho) \geq \inf_{p \in \mathcal{P}_{\mathcal{O}}^{\downarrow}(V)} I_E(p).$$

It therefore suffices to show that

$$\inf_{p \in \mathcal{P}_{\mathcal{O}}^{\downarrow}(V)} I_E(p) \geq d - 1,$$

i.e.,  $(p_R)_{R \in \mathbb{N}_0}$  is a minimising sequence of the infimum in the left-hand side. The proof goes as follows. Write (recall (2.10))

$$\begin{aligned} I_E(p) &= \frac{1}{2} \sum_{\substack{x, y \in V \\ x \sim y}} \left( \sqrt{p(x)} - \sqrt{p(y)} \right)^2 \\ &= \frac{1}{2} \sum_{\substack{x, y \in V \\ x \sim y}} \left[ p(x) + p(y) - 2\sqrt{p(x)p(y)} \right] = (d+1) - \sum_{\substack{x, y \in V \\ x \sim y}} \sqrt{p(x)p(y)}. \end{aligned}$$

Since  $\mathcal{T}$  is a tree, each edge can be labelled by the end-vertex that is farthest from  $\mathcal{O}$ . Hence the sum in the right-hand side can be written as

$$\sum_{x \in V \setminus \mathcal{O}} 2\sqrt{p(x)p(x^\downarrow)},$$

where  $x^\downarrow$  is the unique neighbour of  $x$  that is closer to  $\mathcal{O}$  than  $x$ . Since  $2\sqrt{p(x)p(x^\downarrow)} \leq p(x) + p(x^\downarrow)$ , it follows that

$$\sum_{x \in V \setminus \mathcal{O}} 2\sqrt{p(x)p(x^\downarrow)} \leq \sum_{x \in V \setminus \mathcal{O}} p(x) + \sum_{x \in V \setminus \mathcal{O}} p(x^\downarrow) = [1 - p(\mathcal{O})] + 1.$$

Therefore

$$I_E(p) \geq d - 1 + p(\mathcal{O}),$$

which settles the claim. □

