ARTICLE

Height function localisation on trees

Piet Lammers[1] and Fabio Toninelli[2]

[1]Institut des Hautes Études Scientifiques, Bures sur Yvette, France and [2]Technische Universität Wien, Vienna, Austria
Corresponding author: Fabio Toninelli; Email: fabio.toninelli@tuwien.ac.at

Abstract

We study two models of *discrete height functions*, that is, models of random integer-valued functions on the vertices of a tree. First, we consider the *random homomorphism model*, in which neighbours must have a height difference of exactly one. The local law is uniform by definition. We prove that the height variance of this model is bounded, uniformly over all boundary conditions (both in terms of location and boundary heights). This implies a strong notion of localisation, uniformly over all extremal Gibbs measures of the system. For the second model, we consider directed trees, in which each vertex has exactly one parent and at least two children. We consider the locally uniform law on height functions which are *monotone*, that is, such that the height of the parent vertex is always at least the height of the child vertex. We provide a complete classification of all extremal gradient Gibbs measures, and describe exactly the localisation-delocalisation transition for this model. Typical extremal gradient Gibbs measures are localised also in this case. Localisation in both models is consistent with the observation that the Gaussian free field is localised on trees, which is an immediate consequence of transience of the random walk.

1. Introduction

1.1 Height functions as lattice models

Height functions, in the broadest sense, are random real- or integer-valued functions on the vertices of a connected graph $G = (V, E)$. Let us say for now that G is finite. Each edge $xy \in E$ has an associated potential function V_{xy} (often assumed to be convex, symmetric, and satisfying $V_{xy}(a) \to \infty$ as $|a| \to \infty$) and the density of a configuration is proportional to

$$\exp\left[-H(h)\right]; \quad H(h) := \sum_{xy \in E} V_{xy}(h(y) - h(x)). \tag{1}$$

This well-defines the distribution of the *gradient* of the height function h. We informally think of this construction as follows: each edge has a spring associated to it, which pulls the heights at its two endpoints together. The value $V_{xy}(a)$ is then the energy needed to impose the difference $h(y) - h(x) = a$. One may also introduce boundary conditions by fixing the value of the height on a certain subset $\partial V \subset V$, seen as the 'boundary'. This way, we obtain a well-defined probability distribution of the height profile $h = (h(x))_{x \in V}$ and not just a distribution of height gradients. One can also make sense of height functions on infinite connected graphs, through the standard Dobrushin-Lanford-Ruelle formalism detailed below.

Height functions arise as *lattice models* once we choose the role of the graph G to be played by the square lattice graph $(\mathbb{Z}^d, \mathbb{E})$ or a variant thereof. There is a range of height function models which are each of great physical importance: examples include the *dimer models* (which arise from random tilings), the *six-vertex model* (which is a model of random arrows and relates to *Fortuin-Kasteleyn percolation*), and the *discrete Gaussian free field* (DGFF) (which has continuous heights and quadratic potential functions, and is also known as the *harmonic crystal*). Height functions with convex potential functions on lattices were studied in great generality in the work of Sheffield [26].

A very intriguing issue for height functions is the so-called *localisation-delocalisation transition*. *Localisation* essentially means that, for any given $x \in V$, the variance $\mathrm{Var}\,[h(x)]$ remains bounded uniformly as boundary conditions are taken further and further away; *delocalisation* means that the variance grows infinite. In dimension $d = 1$, all measures are delocalised, with a full description of the scaling limit in terms of Brownian motion. In dimension $d = 2$, measures are expected to either localise, or to delocalise with the Gaussian free field as their scaling limit. Real-valued models should always be delocalised (quite general results are discussed in [22, 28]). Integer-valued models in dimension $d = 2$ are known to localise at low temperature due to the Peierls argument [5], and they are expected to delocalise above the temperature of the roughening transition T_R. There exist many results for specific models in this direction [1, 7, 8, 10, 13, 14, 18, 19, 20], but a general theory of the roughening transition is still lacking. In dimension $d \geq 3$, all reasonable height functions are expected to localise at any temperature. The heuristics behind this is that, in the harmonic case $V_{xy}(a) = a^2$ (which corresponds to the DGFF on G), the variance of $h(x)$ is simply the Green's function $G(x, x)$ of the simple random walk on $\mathbb{Z}^d$, which is finite for $d \geq 3$ due to transience. This has been proved rigorously for the discrete Gaussian model [9, 11] and the solid-on-solid model [2] in dimension three and higher, as well as for the uniformly random Lipschitz function in sufficiently high dimension [23]. Our work focuses on the case where G is a tree. Since trees correspond, in a sense, to a high-dimensional setting, one typically expects localisation. Partial results in this direction, for special boundary conditions, already exist in the literature (see Subsection 1.3).

1.2 Height functions on trees: an apparent contradiction

If G is a tree, a description of the distribution (1) is particularly easy (and the construction can be carried out directly on infinite trees): in that case, the height increments on the edges are independent, with the increment $h(y) - h(x)$ having a density proportional to $\exp[-V_{xy}]$. If all potential functions are the same, then the measure (1) has the property that the variance of $h(y) - h(x)$ is proportional to the graph distance from x to y. Thus, we might think that the variance of $h(x)$ blows up as we put boundary conditions further and further away: the measure would then be *delocalised*. On the other hand, we mentioned above that one expects localisation when the random walk on G is transient: since this is the case on trees with degree strictly larger than two, one expects that the variance should be uniformly bounded: the model would be *localised*. One purpose of this article is to address this apparent contradiction. In fact, we will study a (not-exactly-solvable) model of integer-valued heights on trees, and we will show that the model is localised in a strong sense: the height variance is bounded, uniformly on the choice of the vertex and (for finite trees) on boundary conditions. For infinite trees, localisation holds for any extremal Gibbs measure. In particular, the Gibbs measure described above, with i.i.d. gradients on edges, is *not* extremal; see Remark 2.7 below. The notion of extremality really provides the key to resolving the apparent contradiction, and is closely related to the *reconstructability* of the height distribution given a single sample (see for example [[12], Theorem 7.12] or [[26], Lemma 3.2.3, Statement 1]). We conjecture that some ideas apply in great generality: we expect that, under a strict convexity assumption on the potential, all height functions on trees are uniformly localised, that is, the variance is bounded uniformly over the distance to the boundary and the choice of

boundary conditions. For real-valued height functions the analogous uniform localisation result is an immediate corollary of the Brascamp-Lieb inequality. We will mention exactly the parts in the proof that are missing for general potentials in the integer-valued setting.

1.3 Models under consideration and relations to previous works

The first model under consideration in this article is the model of graph homomorphisms, in which the height at neighbours must always differ by exactly one. The local law is uniform. The model of graph homomorphisms has been considered by several mathematicians, and in particular also on non-lattice graphs. General, rudimentary inequalities were obtained in [3]: for example, it is proven that the variance of the height difference grows at most linearly in the distance between vertices. That article also studies the behaviour of the model on a tree under zero boundary conditions at the leaves. Further bounds on the *maximum* of the random graph homomorphism are derived in [4], by an analogy with the DGFF. The same problem is revisited in [6], which, in particular, relates the bounds on the maximum to the degree of the graph (in relation to its size). These first three papers consider height functions in the very first sense presented in the introduction: on finite graphs, and without boundary. In two later articles, Peled, Samotij, and Yehudayoff [24, 25] consider uniformly random K-Lipschitz functions on a tree, subjected to zero boundary conditions on a designated boundary set $\partial V \subset V$. In either case, the authors obtain extremely tight bounds on the pointwise fluctuations of the random integer-valued functions.

Our presentation differs from the results quoted so far, in the sense that we have as objective to bound the fluctuations of the model uniformly over all boundary conditions. This makes our setup less symmetric: for example, under both free and zero boundary conditions, the law of the model is invariant under a global sign flip. With non-zero boundary conditions this is clearly not the case.

The second model under consideration lives on directed trees which have the property that each vertex has exactly one parent and at least two children. We consider the locally uniform law on height functions which are *monotone* in the sense that the height at the parent always dominates the height at the child. We give a complete classification of Gibbs measures, and describe exactly the (nontrivial) localisation-delocalisation transition. It is possible for this model to delocalise, even though we shall argue that this behaviour is atypical. The potential associated to this model is convex but not strictly convex, which invalidates our heuristic which (we believe) rules out delocalisation on trees. There are several works on potentials which are not (strictly) convex and which have delocalised extremal Gibbs measures [15, 16, 17]. This includes the solid-on-solid model which has the convex potential function $V_\beta(a) := \beta|a|$ associated to it. It is easy to see that such a potential cannot induce uniform localisation (although ruling out general localisation is harder): if one forces the height function to equal some non-constant boundary height function b on the complement of a finite set $\Lambda \subset V$, then one can approximate a continuum model by multiplying the boundary height function b by an extremely large integer k. This necessarily blows up the pointwise variance of the height function at certain vertices in Λ. This suggests that the strictly convex framework introduced below is indeed the correct framework for proving uniform localisation.

2. Definitions and main results

2.1 Graph homomorphisms on trees

We adopt several notations and concepts from the work of Georgii [12] and Sheffield [26], but we will try to be as self-contained as possible.

Definition 2.1 (*Graph homomorphisms*). An irreducible tree is a tree $\mathbb{T} = (\mathbb{V}, \mathbb{E})$ with minimum degree at least three. A height function is a map $h : \mathbb{V} \to \mathbb{Z}$, and a graph homomorphism is a height function $h : \mathbb{V} \to \mathbb{Z}$ such that $h(y) - h(x) \in \{\pm 1\}$ for any edge $xy \in \mathbb{E}$. Write Hom $(\mathbb{T}, \mathbb{Z})$ for the set of graph homomorphisms on $\mathbb{T}$.

Write $\Lambda \subset\subset \mathbb{V}$ to say that Λ is a finite subset of $\mathbb{V}$.

Definition 2.2 (*Uniform graph homomorphisms*). Write $\mathcal{F}$ for the natural product sigma-algebra on $\Omega := \mathbb{V}^{\mathbb{Z}} \supset \mathrm{Hom}\,(\mathbb{T}, \mathbb{Z})$. Write $\mathcal{P}(\Omega, \mathcal{F})$ for the set of probability measures on $(\Omega, \mathcal{F})$. For any $\Lambda \subset\subset \mathbb{V}$ and $b \in \mathrm{Hom}\,(\mathbb{T}, \mathbb{Z})$, let $\gamma_\Lambda(\,\cdot\,, b) \in \mathcal{P}(\Omega, \mathcal{F})$ denote the uniform probability measure on the set

$$\left\{ h \in \mathrm{Hom}\,(\mathbb{T}, \mathbb{Z}) : h|_{\mathbb{V} \smallsetminus \Lambda} = b|_{\mathbb{V} \smallsetminus \Lambda} \right\}.$$

The probability measure $\gamma_\Lambda(\,\cdot\,, b)$ is called the local Gibbs measure in Λ with boundary height function b. The family $\gamma := (\gamma_\Lambda)_{\Lambda \subset\subset \mathbb{V}}$ forms a specification. A Gibbs measure is a measure $\mu \in \mathcal{P}(\Omega, \mathcal{F})$ which is supported on $\mathrm{Hom}\,(\mathbb{T}, \mathbb{Z})$ and which satisfies the Dobrushin-Lanford-Ruelle (DLR) equations in the sense that $\mu \gamma_\Lambda = \mu$ for any $\Lambda \subset\subset \mathbb{V}$. In this case we say that (the distribution of) h is locally uniform in μ. A Gibbs measure is said to be tail-trivial if it is trivial on the tail sigma-algebra $\mathcal{T} \subset \mathcal{F}$. This sigma-algebra is realised as the intersection over $\Lambda \subset\subset \mathbb{V}$ of all sigma-algebras $\mathcal{T}_\Lambda$ which make $h(x)$ measurable for all $x \in \mathbb{V} \smallsetminus \Lambda$. Such measures are also called extremal.

Definition 2.3 (*Gradient measures*). Let $\mathcal{F}^{\nabla}$ denote the smallest sub-sigma-algebra of $\mathcal{F}$ which makes all the height differences $h(y) - h(x)$ measurable. A measure $\mu \in \mathcal{P}(\Omega, \mathcal{F}^{\nabla})$ is called a gradient Gibbs measure whenever it is supported on $\mathrm{Hom}\,(\mathbb{T}, \mathbb{Z})$ and satisfies the DLR equation $\mu \gamma_\Lambda = \mu$ for each $\Lambda \subset\subset \mathbb{V}$. Extremal gradient Gibbs measures are gradient Gibbs measures which are trivial on the gradient tail sigma-algebra $\mathcal{T}^{\nabla} := \mathcal{T} \cap \mathcal{F}^{\nabla}$.

Note that we have defined gradient Gibbs measures μ on the same state space Ω as for Gibbs measures. A point $h \in \Omega$ is a height function; however, the height $h(x)$ at a given vertex is not a μ-measurable function if μ is a gradient measure: only height differences are. In other words, the measure μ returns as samples height functions which are measurable *up to an additive constant*. For the validity of the previous definition, observe that $\mathrm{Hom}\,(\mathbb{T}, \mathbb{Z}) \in \mathcal{F}^{\nabla}$, and that each probability kernel γ_Λ restricts to a probability kernel from $\mathcal{F}^{\nabla}$ to $\mathcal{F}^{\nabla}$.

Remark 2.4 (*Convex interaction potentials*). Uniform graph homomorphisms (and also the uniform monotone functions of the next section) are examples of random height functions which arise from a convex interaction potential. This means that the local Gibbs measure $\gamma_\Lambda(\cdot, b)$ can be written as a tilted version of the counting measure, with the Radon-Nikodym derivative proportional to

$$1_{\{h|_{\mathbb{V} \smallsetminus \Lambda} = b|_{\mathbb{V} \smallsetminus \Lambda}\}} \exp\left[-\sum_{xy \in \mathbb{E}(\Lambda)} V(h(y) - h(x)) \right], \qquad (2)$$

and where the potential function V is convex. Here $\mathbb{E}(\Lambda)$ denotes the set of edges having at least one endpoint in Λ. For graph homomorphisms, V is defined on odd integers, and it equals $V(a) = \infty \times 1_{|a| > 1}$. Sheffield [26] developed a broad theory for the study of height functions arising from convex interaction potentials, which applies to our setting.

Definition 2.5 (*Localisation-delocalisation*). An extremal gradient Gibbs measure μ is said to be localised if it is the restriction of a (non-gradient) Gibbs measure to the gradient sigma-algebra, and delocalised otherwise. If such a Gibbs measure exists, then it may be chosen to be extremal as well (this is a straightforward consequence of extremality of the gradient measure μ). Lemma 8.2.5 of [26] asserts that if μ is an extremal Gibbs measure, then at every vertex $x \in \mathbb{V}$ the law of $h(x)$ is log-concave, and therefore has finite moments of all order. This implies immediately that the height fluctuations of (localised) extremal Gibbs measures are finite.

Theorem 2.6 (Uniform localisation for graph homomorphisms). *Let $\mathbb{T} = (\mathbb{V}, \mathbb{E})$ denote an irreducible tree with a distinguished vertex $x \in \mathbb{V}$. Then the locally uniform graph homomorphism is uniformly localised, in the sense that:*

1. *Any gradient Gibbs measure is the restriction of a Gibbs measure to $\mathcal{F}^{\nabla}$,*

2. *There exists a universal constant C such that any extremal Gibbs measure μ satisfies*

$$\mathrm{Var}_{\mu}\left[h(x)\right] \leq C.$$

With respect to Definition 2.5, the localisation is said to be uniform in this theorem because of the universal upper bound in the second statement. The constant C is also uniform with respect to the choice of the irreducible tree $\mathbb{T}$. Note that the result is false if we drop the requirement that the tree has minimum degree at least three: if the tree is $\mathbb{Z}$, then graph homomorphisms are just trajectories of simple random walks. In this case, most gradient Gibbs measures are delocalised: these are i.i.d. laws on interface gradients, where each $h(x+1) - h(x)$ is a ± 1-valued Bernoulli random variable of some fixed parameter $p \in [0, 1]$. Such measures are localised if and only if $p \in \{0, 1\}$, in which case the configuration is completely frozen.

Let us remark that Theorem 2.6 can be extended (with the same proof) to other convex potentials V with a uniform lower bound on the second derivative, conditional on a technical assumption that is explained in Remark 3.6 below.

Remark 2.7. The uniform localisation statement of the theorem is not in contradiction with the fact that the gradient Gibbs measure μ where all height gradients along edges of the tree are i.i.d. symmetric random variables with values in $\{-1, +1\}$ is such that $\mathrm{Var}_{\mu}\left[h(x) - h(y)\right]$ is proportional to the graph distance between x and y. In fact, it turns out that μ may be viewed as an actual Gibbs measure (that is, not just a gradient Gibbs measure), but also that it is not extremal; this resolves the apparent contradiction mentioned in Section 1.2. Both facts can be shown by defining the following random variable. Let A_k denote the (empirical) average of h on the vertices at distance k from some distinguished root vertex. (This sequence is always well-defined, even though it is not $\mathcal{F}^{\nabla}$-measurable. It is obviously $\mathcal{F}$-measurable.) Then $(A_k)_k$ is a martingale with independent increments (the increments are given by averaging the coin flips on the edges leading to the new vertices). Its limit A_{∞} is called a height offset variable (first defined in the work of Sheffield [26], and discussed in more detail below). Note that A_{∞} is $\mathcal{T}$-measurable, but not $\mathcal{T}^{\nabla}$-measurable. However, its fractional part $\tilde{A}_{\infty} := A_{\infty} - \lfloor A_{\infty} \rfloor$ is also $\mathcal{T}^{\nabla}$-measurable. Its distribution is called the spectrum of A_{∞}. The distributions of the increments of the martingale are known explicitly, and therefore it is easy to see that the spectrum cannot be a Dirac mass. Thus, μ is nontrivial on the gradient tail sigma-algebra, that is, the measure is not extremal: see [[26], Sections 8.4 and 8.7] for details.

2.2 Monotone functions on directed trees

Definition 2.8 (*Directed trees and monotone functions*). By a directed tree we mean a directed tree $\vec{\mathbb{T}} = (\mathbb{V}, \vec{\mathbb{E}})$ such that each vertex has exactly one parent and at least one child. The parent of a vertex $x \in \mathbb{V}$ is written $p(x)$. If we write $xy \in \vec{\mathbb{E}}$, then x is the child and y the parent. A monotone function is a function $f : \mathbb{V} \to \mathbb{Z}$ such that $f(y) - f(x) \geq 0$ for all $xy \in \vec{\mathbb{E}}$. Write $\mathrm{Mon}\,(\vec{\mathbb{T}}, \mathbb{Z})$ for the set of monotone functions.

Notice that although our interest is in trees where each vertex has at least two children, we do not impose this requirement in the definition. In fact, the classification of Gibbs measures works also for trees in which some vertices have a single child.

Definition 2.9 (*Uniform monotone functions*). For any $\Lambda \subset\subset \mathbb{V}$ and $b \in \mathrm{Mon}\,(\vec{\mathbb{T}}, \mathbb{Z})$, let $\gamma_{\Lambda}(\,\cdot\,, b) \in \mathcal{P}(\Omega, \mathcal{F})$ denote the uniform probability measure on the set

$$\left\{ h \in \mathrm{Mon}\,(\vec{\mathbb{T}}, \mathbb{Z}) : h|_{\mathbb{V} \smallsetminus \Lambda} = b|_{\mathbb{V} \smallsetminus \Lambda} \right\}.$$

All other definitions carry over from Definitions 2.2 and 2.3.

Monotone functions fit the framework of convex interaction potentials (Remark 2.4) by setting $V(a) := \infty \times \mathbf{1}_{a<0}$ in the sum in (2), where the sum now runs over *directed* edges.

Definition 2.10 (*Flow, flow measures*). A flow $\varphi : \vec{\mathbb{E}} \to (0, \infty]$ on a directed tree $\vec{\mathbb{T}} = (\mathbb{V}, \vec{\mathbb{E}})$ is a map which satisfies

$$\varphi_{xp(x)} = \sum_{y : \, yx \in \vec{\mathbb{E}}} \varphi_{yx} \quad \forall x \in \mathbb{V}. \tag{3}$$

The associated flow measure is the gradient Gibbs measure μ_φ defined such that

$$(h(y) - h(x))_{xy \in \vec{\mathbb{E}}}$$

is an independent family of random variables, and such that $h(y) - h(x) \sim \text{Geom}(\varphi_{xy})$, where $\text{Geom}(\alpha)$ is the distribution on $\mathbb{Z}_{\geq 0}$ such that the probability of the outcome k is proportional to $e^{-k\alpha}$.

Theorem 2.11 (Complete classification of gradient Gibbs measures). *The map $\varphi \mapsto \mu_\varphi$ is a bijection from the set of flows to the set of extremal gradient Gibbs measures.*

The classification relies on the very particular combinatorial properties of the model; a similar classification is not expected in general.

Theorem 2.12 (Localisation criterion). *A flow measure μ_φ is localised if and only if*

$$\sum_{y \in A(x)} e^{-\varphi_{yp(y)}} < \infty \tag{4}$$

for some $x \in \mathbb{V}$, where $A(x) \subset \mathbb{V}$ denotes the set of ancestors of x.

Remark 2.13.

1. The summability condition is independent of the choice of $x \in \mathbb{V}$, because $|A(x) \Delta A(y)| < \infty$ for any $x, y \in \mathbb{V}$.

2. The previous theorem implies that gradient Gibbs measures are localised in most natural cases. Indeed, if $\vec{\mathbb{T}}$ is a regular tree in which each vertex has $d \geq 2$ children, and if φ is invariant under the automorphisms of $\vec{\mathbb{T}}$ which preserve the generations of the tree, then φ increases by a factor d with each older generation. In that case, it is clear that the localisation criterion is satisfied. Since it is natural to impose extra symmetries on φ, we say that uniform monotone functions are typically localised on trees.

3. On the other hand, a (pathological) example of a delocalised flow measure is obtained by taking φ to be a flow that equals $p > 0$ on a directed infinite ray $\mathcal{R}$ of $\vec{\mathbb{T}}$ and zero elsewhere. In this case, the height increments are i.i.d. $\text{Geom}(p)$ as one walks along $\mathcal{R}$, while the height equals $-\infty$ on vertices outside of $\mathcal{R}$. Strictly speaking this example is not allowed by Definition 2.10 because the flow should be strictly positive, but one can easily construct flows that approximate the finite constant p along $\mathcal{R}$ in the ancestor direction, and such that it is strictly positive (but small) on the remaining edges.

4. If $\mathbb{T} = \mathbb{Z}$, then graph homomorphisms and monotone height functions are in bijection (simply via a 45° rotation of the graph of the height function). This mapping fails when $\mathbb{T}$ is a non-trivial tree. In fact, the results above show that one can have delocalisation for uniform monotone surfaces but not for uniform homomorphisms.

Our results so far concern (gradient) Gibbs measures on infinite trees. We address now the following finite-volume question. Let $\vec{\mathbb{T}} = (\mathbb{V}, \vec{\mathbb{E}})$ be a regular directed tree where each vertex has $d \geq 2$ children. Fix a vertex $v \in \mathbb{V}$, and let D^n denote the set of descendants of v (including v itself), up to distance $n - 1$ from v. Let also ∂D^n denote the set of descendants of v that are at distance

exactly n from it. Let γ_n be the uniform measure on monotone functions on D^{n+1}, subject to the boundary conditions $h(v) = 0$ and $h|_{\partial D^n} \equiv -n$. Theorem 2.12 already suggests that the height will be localised under γ_n, with bounded fluctuations as $n \to \infty$. However, it is not *a priori* clear whether the typical height profile will be dictated by the boundary height 0, which is enforced at the vertex v, or the boundary height $-n$, enforced at the descendants at distance n, or if the height profile will somewhat smoothly interpolate between them. This question is answered by the following result.

Theorem 2.14. *Sample $h|_{D^{n+1}}$ from γ_n. For any fixed descendant x of v, $h(x)$ tends to 0 in probability, as $n \to \infty$. Moreover, for some suitably chosen constant $c(d)$ depending only on d, the event that $h(x) = 0$ for all descendants of v at distance at most $n - c(d) \log n$ from it has probability $1 + o(1)$ as $n \to \infty$.*

The theorem holds true more generally if the boundary height at ∂D^n is fixed to some value $-\lfloor an \rfloor$ for $a > 0$. The fact that the effect of the boundary condition at the single vertex v prevails against the effect of the boundary condition at the d^n vertices of ∂D^n, except at distance $O(\log n)$ from the latter, is perhaps surprising at first sight. On the other hand, note that a positive fraction of the vertices of D^n is actually at distance $O(1)$ from ∂D^n. This property sets the theorem apart from the case that $d = 1$, in which case the height profile is linear.

3. Uniform localisation of graph homomorphisms

Definition 3.1. Let $p : \mathbb{Z} \to [0, 1]$ denote a probability mass function, whose support consists of a consecutive sequence of either odd or even integers. The distribution p is called α-strongly log-concave for some $\alpha \geq 1$ whenever

$$p(k)^2 \geq \alpha p(k - 2) p(k + 2) \quad \forall k \in \mathbb{Z}.$$

The following follows immediately from the previous definition.

Proposition 3.2. *Let $p, q : \mathbb{Z} \to [0, 1]$ denote probability mass functions which are supported on the even or the odd integers, and whose supports are not disjoint. If p and q are α- and β-strongly log-concave respectively, then the normalised probability mass function $\frac{1}{Z} pq$ is $\alpha\beta$-strongly log-concave.*

Remark 3.3. If a probability mass function p is α-strictly log-concave for some $\alpha > 1$, then it has finite moments of all orders and its variance is bounded by a constant $C(\alpha)$ that depends only on α. To see this, suppose that p takes its absolute maximum at some $k_0 \in \mathbb{Z}$. Strict convexity implies

$$\log p(k) - \frac{1}{2}(\log p(k - 1) + \log p(k + 1)) \geq \beta =: \frac{1}{2} \log \alpha > 0,$$

so that

$$p(k_0 + \ell) \leq p(k_0) e^{-\beta(\ell^2 - 1)} \quad \forall \ell \in \mathbb{Z}$$

and finiteness of all moments is obvious. The variance of any real-valued random variable X may be written $\mathrm{Var}\, X = \inf_z \mathbb{E}[(X - z)^2]$, and the uniform bound depending only on α follows by choosing $z = k_0$.

If p and q are probability mass functions, then write $p * q$ for their convolution, which is also a probability mass function. Write $X : \mathbb{Z} \to [0, 1]$ for the probability mass function defined by $2X(k) = \mathbf{1}_{|k|=1}$.

Definition 3.4. Let $\lambda \in [3, 4]$ denote the unique real root of $x^3 - 3x^2 - x - 1$.

Lemma 3.5. *If p is λ^2-strongly log-concave, then $p * X$ is λ-strongly log-concave.*

Remark 3.6. Suppose that W is a convex potential function, and let $Y = e^{-W}$. If there exists a constant $\lambda(W) > 1$ such that $p * Y$ is $\lambda(W)$-strongly log concave whenever p is $\lambda(W)^2$-strongly log concave, then all results in this section generalise to the height function model induced by the potential W. Following [[27], Section 6.2], it is natural to expect that such a constant $\lambda(W)$ exists whenever Y is α-strongly log-concave for some $\alpha > 1$. Note that Y is α-strongly log-concave for some $\alpha > 1$ if and only if W is uniformly strictly convex, that is, a potential of the form $k \mapsto \varepsilon k^2 + V(k)$ where $\varepsilon > 0$ and where V is any convex function.

Proof of Lemma 3.5. Suppose that p is supported on odd integers. Let $q := p * X$. By shifting the domain of q if necessary (which does not modify its log-concavity properties), it suffices to show, without loss of generality, that

$$q(0)^2 \geq \lambda q(-2)q(2),$$

which is equivalent to

$$(p(-1) + p(1))^2 \geq \lambda(p(-3) + p(-1))(p(1) + p(3)).$$

If $p(-1)$ or $p(1)$ equals zero, then it is easy to see that the right hand side must be zero, and we are done. Otherwise, let $\alpha := p(1)/p(-1)$. By multiplying p by a constant if necessary, we suppose that $p(-1) = \alpha\lambda^2$ and $p(1) = \alpha^2\lambda^2$. Now λ^2-strong log-concavity of p implies that $p(-3) \leq 1$ and $p(3) \leq \alpha^3$. For the sake of deriving λ-strong log-concavity, we may take these inequalities for equalities. Thus, it now becomes our goal to demonstrate that

$$(1 + \alpha)^2 \alpha^2 \lambda^4 \geq \lambda \left(1 + \alpha\lambda^2\right) \alpha^2 \left(\lambda^2 + \alpha\right)$$

for any $\alpha > 0$, for the given value of λ. Rearranging gives

$$\alpha^2 - \frac{\lambda^4 - 2\lambda^3 + 1}{\lambda^3 - \lambda^2}\alpha + 1 \geq 0.$$

This parabola in α does not take negative values for positive α whenever

$$-\frac{\lambda^4 - 2\lambda^3 + 1}{\lambda^3 - \lambda^2} \geq -2,$$

that is,

$$(1 - \lambda)(\lambda^3 - 3\lambda^2 - \lambda - 1) \geq 0.$$

This inequality holds trivially true when λ is a zero of the factor on the right. □

Lemma 3.7. *Let $\mathbb{T} = (\mathbb{V}, \mathbb{E})$ denote an irreducible tree and $x \in \mathbb{V}$ some distinguished vertex. Then for any $\Lambda \subset\subset \mathbb{V}$ and $b \in \mathrm{Hom}\,(\mathbb{T}, \mathbb{Z})$, the law of $h(x)$ in the measure $\gamma_\Lambda(\,\cdot\,, b)$ is λ^2-strongly log-concave.*

Proof. The proof idea is as follows. For each $y \in \mathbb{V}$ distinct from x, there is a unique edge incident to y which points in the direction of x. We will let p_y denote the law of $h(y)$ in the measure $\gamma_\Lambda(\,\cdot\,, b)$, except that to define this measure we pretend that this special edge is not there. (In particular, when $y \neq z$, p_y and p_z are marginals of distinct measures on height functions.) This yields a recursive relation (see (5) below) which, in combination with Proposition 3.2 and Lemma 3.6, allows us to derive the asserted concentration. It is important to observe that the Markov property is very strong on trees: removing edges induces independence of the model on the connected components of the remaining graph.

Fix Λ and b. If $x \notin \Lambda$, then the distribution of $h(x)$ is a Dirac mass at $b(x)$, which means in particular that it is λ^2-strongly log-concave, and we are done. Let

$$\Lambda_k := \{y \in \mathbb{V} : d(x, y) < k\},$$

and write $\mathbb{E}_k$ for the set of edges which are incident to Λ_k. Write μ_k for the uniform measure on the set of functions $h \in \mathbb{Z}^{\mathbb{V} \setminus \Lambda_k}$ such that h equals b on the complement of $\Lambda \cup \Lambda_k$, and such
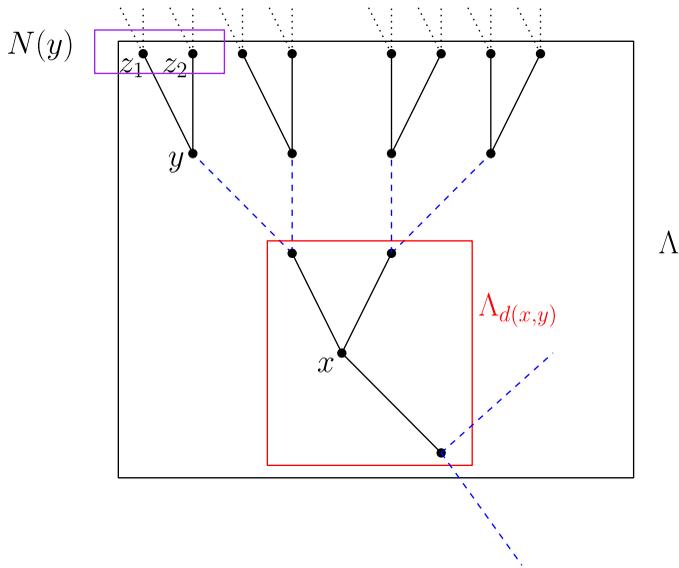
Figure 1. A graphical explanation of the procedure in the proof of Lemma 3.7, in a case where $\mathbb{T}$ is a regular tree of degree 3. Only the portion of the tree inside $\Lambda \cup \Lambda_{d(x,y)}$ is drawn; outside, the configuration coincides with b. In defining the measure μ_k with $k = d(x,y)$ (this measure is defined on the complement of Λ_k), the blue dashed edges are removed and the tree splits into a finite number of connected components. Under μ_k, the height at y is independent of the height on the components that do not contain $N(y) = \{z_1, z_2\}$ (that is, the connected components not containing y). Its law p_y is obtained by conditioning two independent random variables having distributions $p_{z_1} * X$ and $p_{z_2} * X$ to be equal, which gives (5). Here, the definition of X comes from the fact that height the gradients along the edges yz_1 and yz_2 take only values ± 1 with equal likelihood.

that $h(z) - h(y) \in \{\pm 1\}$ for all $yz \in \mathbb{E} \setminus \mathbb{E}_k$. Remark that $\mu_0 = \gamma_\Lambda(\,\cdot\,, b)$. For each $y \in \mathbb{V}$, we let p_y denote the law of $h(y)$ in μ_k where $k := d(x,y)$, see Fig. 1 (this is equivalent to the informal definition of p_y given above). It suffices to prove the claim that p_y is λ^2-strongly log-concave for any $y \in \mathbb{V}$ (the lemma follows by taking $y = x$).

We prove the statement by induction on the distance $k := d(x,y)$. If $d(x,y) \geq d_0$ for a suitable choice of d_0 depending on Λ, then $\Lambda_k \supset \Lambda$. In this case, μ_k is the Dirac measure on $b|_{\mathbb{V} \setminus \Lambda_k}$ and p_y is also a Dirac measure, hence λ^2-strongly log-concave. We prove the full claim by inductively lowering $d(x,y)$ from d_0 to 0. Fix y with $d(x,y) = d < d_0$. Let $N(y)$ denote the neighbours of y which are at distance $d + 1$ from x. Note that $|N(y)| \geq 2$ because the minimum degree of the tree is at least three. Moreover, it is easy to see (as explained in the caption of Fig. 1) that

$$p_y = \frac{1}{Z} \prod_{z \in N(y)} (p_z * X). \tag{5}$$

By induction, all the functions p_z are λ^2-strongly log-concave, so that Proposition 3.2 and Lemma 3.6 imply that p_y is λ^2-strongly log-concave. $\qquad\square$

The proof of Theorem 2.6 now follows from abstract arguments which may appear technical but which are more or less standard. First, we introduce height offset variables.

Definition 3.8 (*Height offset variables,* [26]). A height offset variable is a random variable $H : \Omega \to \mathbb{R} \cup \{\infty\}$ which is $\mathcal{T}$-measurable and satisfies $H(h + a) = H(h) + a$ for all $h \in \Omega$ and $a \in \mathbb{Z}$. The variable H is called a height offset variable for some gradient Gibbs measure μ whenever $\mu(H < \infty) = 1$.

Remark that $\{H < \infty\} \in \mathcal{T}^\nabla$ because the event is unchanged if we add a constant to each height function, which means that it depends only on the height gradients. This is perhaps surprising

because H is not $\mathcal{T}^\nabla$-measurable. Height offset variables capture the following idea. Let μ denote a gradient Gibbs measure. This measure may be turned into a non-gradient measure by a procedure called *anchoring*, that is, by defining $h(x) := 0$ for some fixed vertex $x \in \mathbb{V}$. This preserves the DLR equations $\mu\gamma_\Lambda = \mu$ for all $\Lambda \subset\subset \mathbb{V}$ which do not contain x, but not for sets Λ that do contain x. Thus, we would like instead to *anchor at the tail*. This is exactly what height offset variables do. For height functions on $\mathbb{Z}^d$, height offset variables, when they exist, often arise as the $n \to \infty$ limit of the empirical average of the height function over concentric balls of radius n.

Lemma 3.9. *If there exists a height offset variable H for some gradient Gibbs measure μ, then μ is the restriction of a Gibbs measure to the gradient sigma-algebra.*

Proof. The proof, including more context and background, may be found in Sheffield [26], but we include it here for completeness. Define a new measure $\tilde\mu$ as follows: to sample from $\tilde\mu$, sample first h from μ, then output $\tilde h := h - \lfloor H(h) \rfloor$. Since H is a height offset variable, $\tilde h$ is invariant under replacing h by $h + a$ for any $a \in \mathbb{Z}$. In other words, $\tilde h$ is invariant under the choice of the height function h to represent its gradient. This means that $\tilde\mu$ is well-defined as a probability measure in $\mathcal{P}(\Omega, \mathcal{F})$, and $\tilde h(x)$ is a $\tilde\mu$-measurable function. Since H is tail-measurable, $\tilde\mu$ is a Gibbs measure: the DLR equations hold because the measure is anchored at the tail, as mentioned above. Finally, plainly, the measure $\tilde\mu$ restricted to the gradient sigma-algebra is just μ. $\qquad\square$

We are now ready to prove Theorem 2.6.

Proof of Theorem 2.6. The choice of the vertex x is fixed throughout. Start with the first part. Let μ denote a gradient Gibbs measure. Our goal is to construct a height offset variable. Define Λ_k as in the proof of Lemma 3.7. Let $\mathcal{F}_k^\nabla$ denote the smallest sigma-algebra which makes $h(z) - h(y)$ measurable for all $y, z \in \mathbb{V} \setminus \Lambda_k$. Note that $\left(\mathcal{F}_k^\nabla\right)_{k \geq 0}$ is a backward filtration with $\cap_k \mathcal{F}_k^\nabla = \mathcal{T}^\nabla$. Write X_k for the random variable defined by

$$X_k := \mu\left(h(x) - h(a_k) | \mathcal{F}_k^\nabla\right) - (h(x) - h(a_k)),$$

where each a_k is chosen in the complement of Λ_k. It is straightforward to work out that this definition is independent of the choice of a_k, precisely because the height gradient outside Λ_k is measurable with respect to $\mathcal{F}_k^\nabla$. Thus, X_k tells us, given the current height function h, how much we expect $h(x)$ to increase once we resample all heights at a distance strictly below k from x. Remark that $X_0 = 0$. It suffices to demonstrate that X_k converges almost surely to some $\mathcal{F}^\nabla$-measurable limit X_∞, because

$$H := X_\infty + h(x)$$

is the desired height offset variable in that case (it is clear that it is indeed tail-measurable).

For fixed n, the sequence $(X_k - X_n)_{0 \leq k \leq n}$ is a backward martingale in the backward filtration $\left(\mathcal{F}_k^\nabla\right)_{0 \leq k \leq n}$. By orthogonality of martingale increments, this means that

$$\|X_n\|_2^2 = \|X_k\|_2^2 + \|X_k - X_n\|_2^2$$

in $L^2(\mu)$ for any $0 \leq k \leq n$. But by Lemma 3.7,

$$\|X_n\|_2^2 = \int \mathrm{Var}_{\gamma_{\Lambda_n}(\cdot,b)}[h(x)] d\mu(b) \leq C\left(\lambda^2\right)$$

independently of n, where the constant $C\left(\lambda^2\right)$ comes from Remark 3.3. Therefore $(X_k)_{k \geq 0}$ is Cauchy in $L^2(\mu)$ and converges to some limit X_∞ in $L^2(\mu)$. This convergence can be upgraded to almost sure convergence by using Doob's upcrossing inequality or the backward martingale convergence theorem, applied to the backward martingale $(X_k - X_\infty)_{k \geq 0}$.

Now focus on the second part. Let μ denote an extremal Gibbs measure. Define Y_k by $X_k + h(x)$; this is simply the expected value of $h(x)$ given the values of h on the complement of

Λ_k. By the first part, Y_k converges almost surely to some limit H. Moreover, since μ is extremal, H is almost surely equal to some constant $a \in \mathbb{R}$. The tower property implies that $\mu(h(x)) = a$. With Fatou's lemma, we may estimate the variance by

$$\mathrm{Var}_\mu [h(x)] \leq \lim_{k \to \infty} \mu \left((h(x) - Y_k)^2 \right) = \lim_{k \to \infty} \|X_k\|_2^2.$$

But $\|X_k\|_2^2 \leq C\left(\lambda^2\right)$ as observed above, which proves the second part of the theorem with $C = C\left(\lambda^2\right)$. (In fact, the inequality on the left in the display turns into an equality with these *a posteriori* bounds, but this is not important to us.) $\qquad\square$

4. Complete classification of Gibbs measures for monotone height functions

Let μ denote an extremal gradient Gibbs measure throughout this section. For $x \in \mathbb{V}$, let $D(x) \subset \mathbb{V}$ denote the set of descendants of x, which by convention includes x. For $k \geq 0$, let $D_k(x)$ denote the descendants which are at a distance k from x, and let $E_k(x)$ denote the set of edges $yz \in \vec{\mathbb{E}}$ with $y \in D_k(x)$. This means that $E_0(x)$ contains one edge, which connects x with its parent $p(x)$. Define a sequence of random variables $(X_k)_{k \geq 0}$ by

$$X_k := \min_{yz \in E_k(x)} h(z) - h(y).$$

For convenience we shall write $\nabla h|_V$ for the map

$$V \times V \to \mathbb{Z}, \ (y, z) \mapsto h(z) - h(y)$$

for any $V \subset \mathbb{V}$.

Lemma 4.1. *Under μ, the sequence $(X_k)_{k \geq 0}$ is i.i.d. with distribution $\mathrm{Geom}(\alpha)$ for some $\alpha \in (0, \infty]$. Moreover, this sequence is independent of $\nabla h|_{\mathbb{V} \smallsetminus D(x)}$.*

Proof. For lightness, we drop the argument x in the notations $D(x)$ and $D_k(x)$. We first claim that conditional on $\nabla h|_{\mathbb{V} \smallsetminus D}$, the distribution of the sequence $(X_k)_{k \geq 0}$ satisfies the following two criteria:

1. It is exchangeable,
2. Conditional on $X_0 + \cdots + X_n = a$, the distribution of $(X_0, \ldots, X_n)$ is uniform in

$$\left\{ x \in \mathbb{Z}_{\geq 0}^{\{0, \cdots, n\}} : x_0 + \cdots + x_n = a \right\}.$$

   In fact, the second criterion clearly implies the first, which is why we focus on proving it. First, let $D^n := \cup_{0 \leq k < n} D_k$. Write μ' for the measure μ conditioned on $\nabla h|_{\mathbb{V} \smallsetminus D^n}$. Since μ is a gradient Gibbs measure and since μ' is only conditioned on the complement of D^n, μ' still satisfies the DLR equation $\mu' \gamma_{D^n} = \mu'$. Suppose now that we also condition μ' on $\nabla h|_{D_k}$ for every $0 \leq k < n$. This means that we are adding the information on the height difference within each of sets D_k, *but not on the height gradients between a vertex in D_k and a vertex in $D_{k'}$ for $k \neq k'$*. Observe that this means that the gradient of the height function h is completely determined if we also added the information of the variables $X_0, \ldots, X_n$. The random variables $X_0, \ldots, X_n$ are nonnegative, and their sum can be determined from the information we conditioned on (that is $\nabla h|_{\mathbb{V} \smallsetminus D^n}$ and $\nabla h|_{D_k}$ for every $0 \leq k < n$). The values of these random variables are not constrained in any other way. Since the specification induces local uniformity, the second criterion follows.

   Continue working in the measure μ conditioned on $\nabla h|_{\mathbb{V} \smallsetminus D}$. De Finetti's theorem implies that the sequence $(X_k)_{k \geq 0}$ is i.i.d. once we condition on information in the tail of $(X_k)_{k \geq 0}$. Since μ is extremal, this tail is trivial, which means that the sequence $(X_k)_{k \geq 0}$ is i.i.d.. Moreover, it is easy to see that the second criterion implies that the distribution of X_0 is $\mathrm{Geom}(\alpha)$ for some $\alpha \in (0, \infty]$. In fact, taking $n = 1$ and letting $U(x) := \log \mathbb{P}(X_0 = x)$, the second criterion implies that

$U(x) + U(a - x)$ is constant for $0 \le x \le a$, so that $U(x + 1) - U(x) = U(a - x) - U(a - x - 1)$ for $0 \le x < a$. Since a is arbitrary, U is affine and therefore X_0 is a geometric random variable.

To prove the full lemma, we must demonstrate that this parameter α does not depend on the information in $\nabla h|_{\mathbb{V} \smallsetminus D(x)}$ that we conditioned on. Indeed, if α did depend on this information, then this would imply again that the tail sigma-algebra of $(X_k)_{k \ge 0}$ is not trivial, contradicting extremality of μ. $\square$

Lemma 4.2. *There exists a map $\varphi : \vec{\mathbb{E}} \to (0, \infty]$ such that under the extremal gradient measure μ, the gradients $(h(y) - h(x))_{xy \in \vec{\mathbb{E}}}$ are independent with distribution $h(y) - h(x) \sim \mathrm{Geom}\,(\varphi_{xy})$. Moreover, φ is a flow.*

Proof. Let, as above, $p(x)$ denote the parent of a vertex x. The previous lemma implies that there exists some map $\varphi : \vec{\mathbb{E}} \to (0, \infty]$ such that for each $x \in \mathbb{V}$, we have:

1. $h(p(x)) - h(x) \sim \mathrm{Geom}\,(\varphi_{xp(x)})$,
2. $h(p(x)) - h(x)$ is independent of $\nabla h|_{\mathbb{V} \smallsetminus D(x)}$.

This implies readily that the family $(h(y) - h(x))_{xy \in \vec{\mathbb{E}}}$ is independent. It suffices to demonstrate that φ is necessarily a flow. The flow condition (3) for a fixed vertex $x \in \mathbb{V}$ follows immediately from the definition of the probability kernel $\gamma_{\{x\}}$, which induces local uniformity. In particular, it is important to observe that the minimum of n independent random variables having the distributions $(\mathrm{Geom}\,(\alpha_k))_{1 \le k \le n}$ respectively, is precisely $\mathrm{Geom}\,(\alpha_1 + \cdots + \alpha_n)$, so that the sequence $(X_k)_k$ is i.i.d. if and only if φ is a flow. $\square$

Proof of Theorem 2.11. By the previous lemma it suffices to demonstrate that for each flow φ, the measure μ_φ is an extremal gradient Gibbs measure. The reasoning at the end of the previous proof implies that μ_φ satisfies $\mu_\varphi \gamma_{\{x\}} = \mu_\varphi$ for any $x \in \mathbb{V}$. In other words, for any vertex x, the measure μ_φ is invariant under the 'heat bath' Glauber update where the value of the height at the single vertex x is resampled according to the specification, and conditional on the heights of all other vertices. On the other hand, the state space of configurations coinciding with some fixed boundary condition on the complement of a fixed finite domain, is connected under updates at a single vertex. This follows from a standard application of the Kirszbraun theorem, see for example [[21], Proof of Lemma 13.1]. It is well-known that this implies the DLR equation $\mu_\varphi \gamma_\Lambda = \mu_\varphi$ for fixed Λ, simply because the measure on the left may be approximated by composing μ_φ with many probability kernels of the form $(\gamma_{\{x\}})_{x \in \Lambda}$. In other words, μ_φ is a gradient Gibbs measure. It suffices to show that it is also extremal. If μ_φ was not extremal, then the previous lemma implies that it decomposes as a combination of other flow measures. But since a geometric distribution cannot be written as the non-trivial convex combination of geometric distributions, we arrive at a contradiction. $\square$

Proof of Theorem 2.12. Recall that every extremal gradient Gibbs measure is a flow measure. Let φ denote a flow and consider the flow measure μ_φ. Fix $x_0 \in \mathbb{V}$, and define a sequence $(x_k)_{k \ge 0} \subset \mathbb{V}$ by $x_{k+1} := p(x_k)$. Suppose first that the localisation criterion is satisfied, that is, that the sum (4) converges. We observe that $h(x_n) - h(x_0)$ is the sum of n independent Geometric random variables with their respective parameters given by $(\varphi_{x_k p(x_k)})_{0 \le k < n}$. A random variable $X \sim \mathrm{Geom}\,(\alpha)$ satisfies

$$\mathbb{E}(X) = \frac{e^{-\alpha}}{1 - e^{-\alpha}}.$$

Since $\varphi_{x_k p(x_k)}$ is increasing in k and $\varphi_{x_0 p(x_0)} > 0$, we observe that the sequence

$$h(x_n) - h(x_0)$$

converges almost surely as $n \to \infty$ if the localisation criterion is satisfied. (By looking at the Laplace transform, it is also easy to see that the sequence $h(x_n) - h(x_0)$ diverges almost surely

otherwise, but this fact is not used in this proof.) If this sequence converges, then $\lim_{n\to\infty} h(x_n)$ is a height offset variable, so that μ_φ is localised.

Finally, we must prove the converse statement. Assume that the sum in the localisation criterion diverges; our goal is now to prove delocalisation, which is achieved through a slightly different route. By Definition 2.5, it suffices to prove that

$$\lim_{\Lambda\uparrow\mathbb{V}} \int \mathrm{Var}_{\gamma_\Lambda(\,\cdot\,,b)}\,[h(x_0)]d\mu(b) = \infty.$$

The integral is increasing in Λ due to orthogonality of martingale increments. Write $\Lambda = \Lambda_{n,k} := D^{n+k}(x_n) \smallsetminus \{x_n\} \subset \mathbb{V}$, which by definition contains the descendants of x_n from generation 1 to $n + k - 1$. We prove a lower bound on the quantity in the previous display, by taking first $k \to \infty$ and then $n \to \infty$. For fixed n, we can essentially repeat the arguments in the proof of Lemma 4.1 to see that the following claim is true.

Claim. *For μ-almost every b, the quenched law of $\nabla h|_{D^{n+1}(x_n)}$ in $\gamma_{\Lambda_{n,k}}(\,\cdot\,,b)$ converges weakly as $k \to \infty$ to the law of $\nabla h|_{D^{n+1}(x_n)}$ in the unconditioned measure μ.*

Proof of the claim. For each fixed vertex $x \in D^{n+1}(x_n)$, introduce the sequence of random variables $(X_\ell)_{\ell\geq 0}$ as above. Note that if x is at distance $r \geq 0$ from x_n, then under $\gamma_{\Lambda_{n,k}}(\,\cdot\,,b)$, the random variable X_ℓ is deterministic whenever $\ell > n + k - r$. Each of these sequences $(X_0, \dots, X_{n+k-r})$ has the same exchangeability and local uniformity properties as in the proof of Lemma 4.1. In addition we know that, under the unconditioned law μ, the empirical average of $X_0, \dots, X_\ell$ tends almost surely to a constant α as $\ell \to \infty$. Since μ is extremal, it follows that for μ-almost every b, the quenched law of the empirical average of $X_0, \dots, X_{n+k-r}$ under the conditional measure $\gamma_{\Lambda_{n,k}}(\cdot, b)$ tends as $k \to \infty$ to a Dirac mass at α. This implies that for every $x \in D^{n+1}(x_n)$, the quenched law of $(X_\ell)_{\ell\geq 0}$ converges almost surely to the law of a sequence of i.i.d. $\mathrm{Geom}(\alpha)$, independent of $\nabla h|_{\mathbb{V}\smallsetminus D(x)}$. This readily implies the claim. $\qquad\square$

In turn, the claim implies that

$$\lim_{k\to\infty} \int \mathrm{Var}_{\gamma_{\Lambda_{n,k}}(\,\cdot\,,b)}\,[h(x_0)]d\mu(b) \geq \mathrm{Var}_\mu\,[h(x_n) - h(x_0)]$$

$$= \sum_{m=0}^{n-1} \frac{e^{-\varphi_{x_m x_{m+1}}}}{(1 - e^{-\varphi_{x_m x_{m+1}}})^2} \geq \sum_{m=0}^{n-1} e^{-\varphi_{x_m x_{m+1}}} \to_{n\to\infty} \infty$$

by divergence of the sum in the localisation criterion. This proves the theorem. $\qquad\square$

5. Monotone functions on finite trees

This section contains the proof of Theorem 2.14. Recall that each vertex has d children. Recall also that $D_n(v)$ denotes the set of n-th generation descendants of some fixed vertex v, and that $D^n(v) := \cup_{0\leq k<n} D_k(v)$. Let $\gamma_{n,k}$ denote the uniform measure on monotone height functions on the set $D^{n+1}(v)$, with boundary conditions $h(v) = 0$ and $h|_{D_n(v)} \equiv -k$ for some fixed $k \geq 0$.

Lemma 5.1. *Let $z_1, \dots, z_d$ denote the children of v. For any $i = 1, \dots, d$, we have*

$$\gamma_{n,k}(h(z_i) = 0) \geq 1 - \left(1 - \frac{1}{k+1}\right)^{d^{n-2}}.$$

Proof. Fix i. Under $\gamma_{n,k}$, the height restricted to the descendants of z_i is independent of the height on the rest of the graph. To match the notations of Section 4, write x instead of z_i. For $k = 0, \dots, n - 1$, define the random variable X_k and the set of directed edges $E_k(x)$ as in Section 4. We know from the proof of Lemma 4.1 that the sequence $X_0, \dots, X_{n-1}$ is exchangeable. Note that

$X_0 = h(v) - h(x)$ and that X_{n-1} is the minimum height gradient between one of the d^{n-1} ancestors of x in $D_n(v)$ and their parent, which belongs to $D_{n-1}(v)$. Given the height on $D_{n-2}(v)$, the height on the vertices in $D_{n-1}(v)$ are independent and each one is uniformly distributed in an interval of size at most $k + 1$. Therefore,

$$\gamma_{n,k}(X_{n-1} = 0) \geq 1 - \left(1 - \frac{1}{k+1}\right)^{d^{n-2}}.$$

Since X_0 has the same distribution as X_{n-1}, the lemma follows. □

Proof of Theorem 2.14. The measure γ_n is nothing but $\gamma_{n,n}$ in the notation of Lemma 5.1. From Lemma 5.1 we have that with $\gamma_{n,n}$ probability at least

$$1 - d\left(1 - \frac{1}{n+1}\right)^{d^{n-2}},$$

we have $h(z) = 0$ for each of the d children of v. Conditioned on this event, the measure $\gamma_{n,n}$ restricted $D^n(v) \smallsetminus \{v\}$ is just the d-fold product measure $\gamma_{n-1,n}^{\otimes d}$. Iterating the argument, the probability that $h(x) = 0$ for every $x \in D^n(v)$ at distance up to m from v has γ_n-probability at least

$$1 - \sum_{j=1}^{m} d^j \left(1 - \frac{1}{n+1}\right)^{d^{n-j-1}}.$$

Given that $d \geq 2$, it is easy to see that this quantity is $o(1)$ as $n \to \infty$, as soon as $m \leq n - c(d)\log n$ for large enough $c(d)$. □

Acknowledgments

References

[1] Aizenman, M., Harel, M., Peled, R. and Shapiro, J. (2022) Depinning in integer-restricted Gaussian fields and BKT phases of two-component spin models, arXiv preprint arXiv: 2110.09498v3.

[2] Bricmont, J., Fontaine, J. R. and Lebowitz, J. L. (1982) Surface tension, percolation, and roughening. *J. Stat. Phys.* **29**(2) 193–203.

[3] Benjamini, I., Häggström, O. and Mossel, E. (2000) On random graph homomorphisms into $\mathbb{Z}$. *J. Comb. Theory, Ser. B* **78**(1) 86–114.

[4] Benjamini, I. and Schechtman, G. (2000) Upper bounds on the height difference of the Gaussian random field and the range of random graph homomorphisms into $\mathbb{Z}$. *Rand. Struct. Algorithms* **17**(1) 20–25.

[5] Brandenberger, R. and Wayne, C. E. (1982) Decay of correlations in surface models. *J. Stat. Phys.* **27**(3) 425–440.

[6] Benjamini, I., Yadin, A. and Yehudayoff, A. (2007) Random graph-homomorphisms and logarithmic degree. *Electron. J. Probab.* **12** 926–950.

[7] Duminil-Copin, H., Glazman, A., Peled, R. and Spinka, Y. (2020) Macroscopic loops in the loop $O(n)$ model at Nienhuis' critical point. *J. Eur. Math. Soc.* **23**(1) 315–347.

[8] Duminil-Copin, H., Karrila, A., Manolescu, I. and Oulamara, M. (2020) Delocalization of the height function of the six-vertex model, arXiv preprint arXiv: 2012.13750.

[9] Fröhlich, J., Israel, R., Lieb, E. H. and Simon, B. (1978) Phase transitions and reflection positivity. I. General theory and long range lattice models. *Commun. Math. Phys.* **62**(1) 1–34.

[10] Fröhlich, J. and Spencer, T. (1981) The Kosterlitz-Thouless transition in two-dimensional abelian spin systems and the Coulomb gas. *Commun. Math. Phys.* **81**(4) 527–602.

[11] Fröhlich, J., Simon, B. and Spencer, T. (1976) Infrared bounds, phase transitions and continuous symmetry breaking. *Commun. Math. Phys.* **50**(1) 79–85.

[12] Georgii, H.-O. (2011) *Gibbs measures and phase transitions, De Gruyter Studies in Mathematics, vol. 9.* Walter de Gruyter, 2nd ed.

[13] Glazman, A. and Manolescu, I. (2021) Uniform Lipschitz functions on the triangular lattice have logarithmic variations. *Commun. Math. Phys.* **381**(3) 1153–1221.

[14] Giuliani, A., Mastropietro, V. and Toninelli, F. L. (2020) Non-integrable dimers: Universal fluctuations of tilted height profiles. *Commun. Math. Phys.* **377**(3) 1883–1959.

[15] Henning, F. and Külske, C. (2021) Coexistence of localized Gibbs measures and delocalized gradient Gibbs measures on trees. *Ann. Appl. Prob.* **31**(5) 2284–2310.

[16] Henning, F. and Külske, C. (2021) Existence of gradient Gibbs measures on regular trees which are not translation invariant, arXiv preprint arXiv: 2102.11899.

[17] Henning, F., Külske, C., Ny, A. L. and Rozikov, U. A. (2019) Gradient Gibbs measures for the SOS model with countable values on a Cayley tree. *Electron. J. Prob.* **24** 1–23.

[18] Kenyon, R. (2001) Dominos and the Gaussian free field. *Ann. Prob.*, 1128–1137.

[19] Lammers, P. (2022) Height function delocalisation on cubic planar graphs. *Prob. Theory Relat. Fields*, **182**(1), 531–550.

[20] Lammers, P. and Ott, S. (2023) Delocalisation and absolute-value-FKG in the solid-on-solid model. *Prob. Theory Relat. Fields*, arXiv preprint arXiv: 2101.05139.

[21] Lammers, P. and Tassy, M. (2020) *Macroscopic behavior of Lipschitz random surfaces,* arXiv preprint arXiv: 2004.15025.

[22] Miłoś, P. and Peled, R. (2015) Delocalization of two-dimensional random surfaces with hard-core constraints. *Commun. Math. Phys.* **340**(1) 1–46.

[23] Peled, R. (2017) High-dimensional Lipschitz functions are typically flat. *Ann. Prob.*, **45**(3),1351–1447.

[24] Peled, R., Samotij, W. and Yehudayoff, A. (2013) Grounded Lipschitz functions on trees are typically flat. *Electron. Commun. Prob.* **18** 1–9.

[25] Peled, R., Samotij, W. and Yehudayoff, A. (2013) Lipschitz functions on expanders are typically flat. *Comb. Prob. Comp.* **22**(4) 566–591.

[26] Sheffield, S. (2005) Random surfaces. *Astérisque* **304** 1–175.

[27] Saumard, A. and Wellner, J. A. (2014) Log-concavity and strong log-concavity: a review. *Stat. Surv.* **8** 45–114.

[28] Velenik, Y. (2006) Localization and delocalization of random interfaces. *Prob. Surv.* **3** 112–169.