On the almost eigenvectors of random regular graphs

Ágnes Backhausz   Balázs Szegedy

July 19, 2016

Abstract

Let $d \geq 3$ be fixed and G be a large random d-regular graph on n vertices. We show that if n is large enough then the entry distribution of every almost eigenvector v of G (with entry sum 0 and normalized to have length $\sqrt{n}$) is close to some Gaussian distribution $N(0, \sigma)$ in the weak topology where $0 \leq \sigma \leq 1$. Our theorem holds even in the stronger sense when many entries are looked at simultaneously in small random neighborhoods of the graph. Furthermore, we also get the Gaussianity of the joint distribution of several almost eigenvectors if the corresponding eigenvalues are close. Our proof uses graph limits and information theory. Our results have consequences for factor of i.i.d. processes on the infinite regular tree.

1   Introduction

Let $d \geq 3$ and let $G(n, d)$ denote the random d-regular graph on n vertices (see e.g. the monograph [10]). Equivalently, we can think of $G(n, d)$ as a random model of symmetric $0 - 1$ matrices in which the row sums are conditioned to be d. It is expected that the spectral properties of $G(n, d)$ are closely related to random matrix theory; however, many questions in the area are still open. It is well known that the spectral measure of $G(n, d)$ converges to the so called Kesten–McKay measure in the weak topology as n goes to infinity. This gives an approximate semicircle law if d is large. A famous result by J. Friedman solves Alon's second eigenvalue conjecture showing that $G(n, d)$ is almost Ramanujan [22]. Much less is known about the scaled eigenvalue spacing and about the structure of the eigenvectors. Recent results in the area include [12, 33] on the second eigenvalue; [6, 7, 19, 20] on eigenvalue spacing, local semicircle law and functional limit theorems; [2, 15, 26] on the delocalization of the eigenvectors.

In the present paper we study approximate eigenvectors or shortly: *almost eigenvectors* of $G(n, d)$ i.e. vectors that satisfy the eigenvector equation $(A - \lambda I)v = 0$ with some small error.

Almost eigenvectors are not necessarily close to proper eigenvectors. They are much more general objects. For example any linear combination of eigenvectors with eigenvalues in the interval $[\lambda - \varepsilon, \lambda + \varepsilon]$ is an almost eigenvector with error depending on ε. In general a vector is an almost eigenvector if and only if its spectral measure is close to a Dirac measure in the weak topology.

We show that despite of this generality, almost eigenvectors of $G(n, d)$ have a quite rigid structure if n is big. Our main result implies that every almost eigenvector (with entry sum 0 and normalized to have length $\sqrt{n}$) has an entry distribution close to some Gaussian distribution $N(0, \sigma)$ in the weak topology where $0 \leq \sigma \leq 1$. Note that if $\sigma = 0$, then the ℓ^2-weight of the vector is concentrated on a small fraction of the vertices. Such vectors are called localized. Our main result holds even in a stronger sense where joint distributions are considered using the local structure of the graph. In some sense our result is best possible since there are examples for both localized and delocalized almost eigenvectors (see chapter 3). Note that proper eigenvectors are conjectured to be delocalized.

The issue of eigenvector Gaussianity goes back to random matrix theory. It is not hard to show that in the GUE (Gaussian Unitary Ensemble) random matrix model every eigenvector has a near Gaussian entry distribution. It is much harder to analyze the random model when the elements of the matrix are chosen from a non Gaussian distribution. Nevertheless Gaussianity of the eigenvectors is proved under various conditions for generalized Wigner matrices [13, 37] and also for various other models (see e.g. [9] and [8, 34] for recent surveys). Sparser models are harder to analyze [30]. This paper deals with the sparsest case, where the matrix is the adjacency matrix of a random d-regular graph with some fixed d. In this case there is a stronger meaning of eigenvector Gaussianity. For example it is natural to ask about the Gaussianity of the joint distribution of the entries at the two endpoints of a randomly chosen edge in the graph. More generally one can look at the joint distribution of the entries in random balls of radius r. Our Gaussianity results for almost eigenvectors are established in this strong sense. Furthermore, it makes sense to study the joint distribution of the entries of several almost eigenvectors. More precisely, a k-tuple of almost eigenvectors can be interpreted as a function from the vertices to $\mathbb{R}^k$. This function evaluated at a randomly chosen vertex gives a probability distribution on $\mathbb{R}^k$ that we define as the joint distribution of the entries. It is an interesting question whether such joint distributions are Gaussian. We prove this if k is fixed, n is large and the eigenvalues corresponding to the almost eigenvectors are close to each other. Moreover, we also prove this result when the joint distribution of many eigenvectors is considered in random neighborhoods.

The proof of our main theorem is based on the so-called local-global graph limits [11, 28]. However, to keep the paper self-contained, we use a slightly simplified framework (see section 5) optimized for this particular problem. We relate the properties of random regular graphs to random processes on the infinite d-regular tree T_d. Most of the work is done in this convenient

limiting framework. An invariant process on T_d is a joint distribution of random variables $\{X_v\}_{v \in T_d}$ labeled by the vertices of T_d such that it is invariant under the automorphism group of T_d. A special class of invariant processes, called *typical processes*, was introduced in [3]. Roughly speaking, an invariant process is typical if it can be obtained as the Benjamini–Schramm limit of colored random regular graphs. There is a correspondence between the properties of typical processes on T_d and the properties of large random d-regular graphs.

Our main theorem is equivalent to the statement that if a typical process satisfies the eigenvector equation at every vertex of T_d and has finite variance (at every vertex), then the process is jointly Gaussian. Note that such Gaussian eigenvector processes on T_d are completely characterized and there is a unique one for each possible eigenvalue. A key ingredient in our proof is a general entropy inequality for typical processes. It implies that typical eigenvector processes obey another inequality involving differential entropy. From here we finish the proof using heat propagation on the space of typical eigenvector processes combined with DeBrujin's identity for Fisher information. Gaussianity will follow from the fact that heat propagation converges to a Gaussian distribution.

A well studied subclass of typical processes is the class of factor of i.i.d. processes. These processes appeared first in ergodic theory but they are also relevant in probability theory, combinatorics, statistical physics and in computer science. Not every typical process is factor of i.i.d.; this follows from the results of Gamarnik and Sudan [25]; see also Rahman and Virág [36]. Despite of recent progress in the area [4, 5, 14, 18, 16, 24, 27, 28, 31, 32, 35], a satisfying understanding of factor of i.i.d. processes is only available in the case $d = 2$ [23], which is basically equivalent to the framework of classical ergodic theory of $\mathbb{Z}$ actions. Our results imply that if an invariant process (with finite variance) is in the weak closure of factor of i.i.d. processes and satisfies the eigenvector equation then the process is Gaussian. This answers a question of B. Virág.

Outline of the paper. In Chapter 2, we formulate the main results for finite random d-regular graphs. Chapter 3 and 4 contain general statements about invariant processes, eigenvector processes, entropy and almost eigenvectors. Chapter 5 provides the translation of our main result to the infinite setting using typical eigenvector processes on the infinite d-regular tree. In Chapter 6, we prove a necessary condition for a process to be typical in the form of an entropy inequality. In Chapter 7 we reduce the limiting form of the main theorem to a special family of eigenvector processes called smooth eigenvector processes. Chapter 8 gives a differential entropy inequality for smooth eigenvector processes. In Chapter 9, we calculate the eigenvalues of special submatrices of the covariance matrices of eigenvector processes corresponding to balls around vertices and edges on the tree. In Chapter 10, we use the results from Chapter 9 to prove that among smooth typical eigenvector processes the Gaussian minimizes the differential entropy formula from Chapter 8. On the other hand, in Chapter 11, we show that the Gaussian eigenvector process maximizes the same formula. Moreover, we finish the proof of the main result in Chapter 11.

3

2 The main theorem

In this chapter we state the main theorem first in a simpler but weaker form and later in the strong form. An ε-almost eigenvector of a matrix $A \in \mathbb{R}^{n \times n}$ with eigenvalue λ is a vector $v \in \mathbb{R}^n$ such that $\|v\|_2 = 1$ and $\|Av - \lambda v\|_2 \leq \varepsilon$. To every vector v in $\mathbb{R}^n$ we associate a probability distribution $\mathrm{distr}(v)$ on $\mathbb{R}$ obtained by choosing a uniform random entry from v. If $\|v\|_2 = 1$, then the second moment of $\mathrm{distr}(v)$ is $1/n$. Thus, to avoid degeneracy in this case, it is more natural to consider $\mathrm{distr}(\sqrt{n}v)$ whose second moment is 1. We will compare probability distributions on $\mathbb{R}$ using an arbitrary but fixed metrization of the weak convergence of probability measures. Our theorem in a weak form says the following

Theorem 1 (Weak form of main theorem) *For every $\varepsilon > 0$ there are constants N, δ such that if G is a random d-regular graph on $n \geq N$ vertices, then with probability at least $1 - \varepsilon$ the following holds. We have that every δ-almost eigenvector v of G (with entry sum 0) has the property that $\mathrm{distr}(\sqrt{n}v)$ is at most ε-far from some Gaussian distribution $N(0, \sigma)$ in the weak topology where $0 \leq \sigma \leq 1$.*

Note that if $\mathrm{distr}(\sqrt{n}v)$ is close to the degenerate distribution $N(0,0)$ then most of the ℓ^2 weight of v is concentrated on $o(n)$ points. Such vectors are called localized. In general, if σ is smaller than, 1 then some of the ℓ^2 weight is concentrated on $o(n)$ vertices and the rest is Gaussian.

To formulate our main theorem in the strong form we need some more notation. Recall that T_d denotes the infinite d-regular tree and o is a distinguished vertex called *root* in T_d. We will denote the vertex set $V(T_d)$ of T_d by V_d. For two vertices in a graph we write $v \sim w$ if they are connected to each other. Let $[n]$ denote the set $\{1, 2, \ldots, n\}$ and let G be a d-regular graph on the vertex set $[n]$. We denote by $\mathrm{Hom}^*(T_d, G)$ the set of all covering maps from T_d to G. In other words $\mathrm{Hom}^*(T_d, G)$ is the set of maps $\phi : V_d \to V(G)$ such that for every vertex $v \in V_d$ the neighbors of v are mapped bijectively to the neighbors of $\phi(v)$. The set $\mathrm{Hom}^*(T_d, G)$ has a natural probability measure. We first choose the image of o uniformly at random in $V(G)$. Then we recursively extend the map ϕ to larger and larger neighborhoods of o in a random (conditionally independent) way preserving the local bijectivity. It is easy to see that this probability distribution is independent from the choice of o.

Let X be a topological space and $f : [n] \to X$ be a function. We define the probability distribution $\mathrm{distr}^*(f, G)$ on X^{V_d} as the distribution $f \circ \phi$ where ϕ is a random covering in $\mathrm{Hom}^*(T_d, G)$. In other words $\mathrm{distr}^*(f, G)$ is a random lift of f to T_d using a random covering of G with T_d. By regarding vectors $v \in \mathbb{R}^n$ as functions form $[n]$ to $\mathbb{R}$ it makes sense to use $\mathrm{distr}^*(v, G)$.

To formulate our main theorem we need the concepts of eigenvector processes and Gaussian waves on T_d (see [21]). An *eigenvector process* with eigenvalue λ is a joint distribution $\{X_v\}_{v \in V_d}$ of real valued random variables with variance 1 such that it is $\mathrm{Aut}(T_d)$ invariant and satisfies the

eigenvector equation

$$\sum_{v \sim o} X_v = \lambda X_o \qquad (1)$$

with probability 1. Note that the group invariance implies that the eigenvector equation is satisfied at every vertex on T_d. We call an eigenvector process *trivial* if $\mathbb{E}(X_o) \neq 0$. Notice that the triviality of an eigenvector process implies that $\lambda = d$. This follows by taking expectation in (1) and using the invariance of the process. Furthermore, trivial eigenvector processes are constants in the sense that $X_v = X_w$ holds with probability one for every pair of vertices v, w in V_d. A *Gaussian wave* is an eigenvector process whose joint distribution is Gaussian. It is proved in [21] that for every $-d \leq \lambda \leq d$ there is a unique Gaussian wave Ψ_λ with eigenvalue λ.

Let us choose a fix metrization of the weak topology on $\mathbb{R}^{V_d}$. Our main theorem on random d-regular graphs is the following.

Theorem 2 (Main theorem) *For every $\varepsilon > 0$ there exist constants N, δ such that if G is a random d-regular graph on $n \geq N$ vertices then with probability at least $1 - \varepsilon$ the following holds. For every δ-almost eigenvector v of G (with entry sum 0) has the property that $\mathrm{distr}^*(\sqrt{n}v, G)$ is at most ε-far (in the weak topology) from some Gaussian wave Ψ_λ with $|\lambda| \leq 2\sqrt{d-1}$.*

Corollary 2.1 *For every $\varepsilon > 0$ and $k \in \mathbb{N}$ there exist constants N, δ such that if G is a random d-regular graph on $n \geq N$ vertices then with probability at least $1 - \varepsilon$ the following holds. For every k-tuple of δ-almost eigenvectors $Q = (v^{(1)}, v^{(2)}, \ldots, v^{(k)})$ of G (with entry sum 0) corresponding to eigenvalues $\lambda_1, \ldots, \lambda_k$ with the property $|\lambda_i - \lambda_j| < \delta$ we have that $\mathrm{distr}^*(\sqrt{n}Q, G)$ is at most ε-far (in the weak topology) from some Gaussian wave Ψ_λ with $|\lambda| \leq 2\sqrt{d-1}$.*

3 Preliminaries

Invariant processes

For a separable metric space Y we denote by $I_d(Y)$ the set of Borel probability measures on Y^{V_d} that are invariant under the automorphisms of the tree. More precisely, for every $\tau \in \mathrm{Aut}(T_d)$ (not necessarily fixing the root), the probability measure on Y^{V_d} is required to be invariant under the natural $Y^{V_d} \to Y^{V_d}$ map induced by τ. Note that eigenvector processes are in $I_d(\mathbb{R})$. If $\mu \in I_d(Y)$ and $F \subseteq V_d$, then we denote by μ_F the marginal distribution of μ at F. We can equivalently think of $\mu \in I_d(Y)$ as a joint distribution $\{X_v\}_{v \in V_d}$ of Y-valued random variables that is invariant under the automorphism group of T_d. In this language μ_F is the same as the joint distribution $\{X_v\}_{v \in F}$. If both F and Y are finite then μ_F is a probability distribution on the finite set Y^F. In this case we denote the entropy of μ_F by $\mathbb{H}(F)$. By invariance of μ the quantity $\mathbb{H}(F)$ depends only on the isomorphism class of F.

We will consider convergence in $I_d(Y)$ with respect to the weak topology. Since the weak topology is metrizable we can always choose a fixed metrization of it in advance.

Almost eigenvectors

Let $A \in \mathbb{R}^{n \times n}$ be a matrix. An ε-almost eigenvector of A (with eigenvalue λ) is a vector $v \in \mathbb{R}^n$ such that $\|v\|_2 = 1$ and $\|Av - \lambda v\|_2 \leq \varepsilon$.

Lemma 3.1 *Let A be the adjacency matrix of a d-regular graph on n vertices. Let $\lambda_2(A)$ denote the second largest (in absolute value) eigenvalue of A. Let v be an ε-almost eigenvector with eigenvalue λ such that the entry sum of v is 0. Then $|\lambda| \leq \lambda_2(A) + \varepsilon$.*

Proof. Since v has 0 entry sum, we can write $v = \sum a_i v_i$, where each v_i is a nonconstant eigenvector of A with eigenvalue λ_i. It follows that $(A - \lambda I)v = \sum a_i(\lambda_i - \lambda)v_i$, where $|\lambda_i| \leq \lambda_2(A)$. Thus, $\varepsilon^2 \geq \|(A - \lambda I)v\|_2^2 = \sum a_i^2(\lambda_i - \lambda)^2$. Suppose that $|\lambda| \geq \lambda_2(A)$ (otherwise the statement is trivial). Then $|\lambda_i - \lambda| \geq |\lambda| - \lambda_2(A)$. Therefore $\varepsilon^2 \geq (\sum a_i^2)(|\lambda| - \lambda_2(A))^2$, which completes the proof by using that $\sum a_i^2 = \|v\|_2^2 = 1$. $\qquad\square$

As we mentioned in the introduction, we will give examples for both localized and delocalized almost eigenvectors on essentially large girth d-regular graphs. (Note that a graph is called essentially large girth if most vertices are not contained in short cycles.) The purpose of these examples is to show that our results on the almost eigenvectors of random regular graphs are best possible in the sense that all $0 \leq \sigma \leq 1$ can indeed occur in the statement.

We will need some preparation. For $k \geq 1$ and $x \in [-1, 1]$ let

$$f(k, x) = \frac{1}{\sqrt{d(d-1)^{k-1}}} q_k(x), \tag{2}$$

where

$$q_k(x) = \sqrt{\frac{d-1}{d}} U_k(x) - \frac{1}{\sqrt{d(d-1)}} U_{k-2}(x); \qquad U_k(\cos \vartheta) = \frac{\sin((k+1)\vartheta)}{\sin \vartheta}. \tag{3}$$

($U_k(x)$ is the Chebyshev polynomial of the second kind.) Let $g_\lambda : V_d \to \mathbb{R}$ be the function defined by

$$g_\lambda(v) = f(|v|, \lambda/(2\sqrt{d-1})),$$

where $|v|$ denotes the distance of v and the root o. It is easy to see (and well known) that g_λ satisfies the eigenvector equation with eigenvalue λ at every vertex v (for $\lambda \in [-2\sqrt{d-1}, 2\sqrt{d-1}]$).

Now let G be a d-regular graph such that there is a vertex $w \in V(G)$ with the property that the shortest cycle containing w has length at least $2n$. We define the function $g'_\lambda : V(G) \to \mathbb{R}$ by $g'_\lambda(v) = f(d(w, v), \lambda/2\sqrt{d-1})$ if $d(v, w) < n$ and 0 otherwise. It is easy to see that $u = g'_\lambda/\|g'_\lambda\|_2$ is an almost eigenvector with eigenvalue λ with error tending to 0 as $n \to \infty$. Furthermore, if

6

$|V(G)|$ is much larger than n, then u is close to the constant 0 distribution in the weak topology. Thus we obtain examples for completely localized almost eigenvectors for d-regular graphs (for all $\lambda \in [-2\sqrt{d-1}, 2\sqrt{d-1}]$), which corresponds to the case $\sigma = 0$.

We switch to the delocalized example. In [27], the authors construct eigenvector processes on T_d for every $\lambda \in [-2\sqrt{d-1}, 2\sqrt{d-1}]$ that are weak limits of factor of i.i.d. processes. These processes have the property that they can be arbitrarily well approximated on any essential large girth d-regular graph. It is easy to see that these approximations are almost eigenvectors that are completely delocalized. This corresponds to the case $\sigma = 1$. Finally, every σ occurs by mixing completely localized and completely delocalized almost eigenvectors corresponding to the same λ.

4 Eigenvectors and eigenvector processes on the tree

For a vertex set $F \subseteq V_d$ we denote by $B_k(F)$ the neighborhood of radius k around F. Let $F \subseteq V_d$ be a subset of the vertices of the tree, and let $f \in \mathbb{R}^F$. We say that v satisfies the eigenvector equation with eigenvalue λ if for every $v \in F$ with $B_1(v) \subseteq F$ we have that $\lambda f(v) = \sum_{w \sim v} f(w)$. It is clear that for a fixed λ these vectors form a linear subspace of $\mathbb{R}^F$ that we denote by $W_\lambda(F)$. We will need a formula for $\dim W_\lambda(F)$ for a family of special finite sets F.

Given F, we say that $F_0 \subseteq F$ is a basis if for all $f \in \mathbb{R}^{F_0}$ and $\lambda \in \mathbb{R}$ the subspace $W_\lambda(F)$ contains exactly one extension of f to F. It is clear that if F_0 is a basis in F, then $\dim W_\lambda(F) = |F_0|$ for all $\lambda \in \mathbb{R}$.

Lemma 4.1 *Let F_0 be a basis of a path-connected set F. Suppose that $v \in F$ and $|F \cap B_1(v)| = 2$. Furthermore, let $D \subseteq B_1(v)$ such that $|D| = d - 2$ and $D \cap F = \emptyset$. Then $F_0 \cup D$ is a basis of $F \cup B_1(v)$.*

Proof. Let f be a function from $F_0 \cup D$ to $\mathbb{R}$. By assumption, we have that $f|_{F_0}$ extends to a unique function $\tilde{f}$ on F. Now using the eigenvector equation at v, we obtain a unique value for $B_1(v) \setminus (D \cup F)$. Note that the connectivity of F implies that the function constructed this way is in $W_\lambda(F \cup B_1(v))$. $\qquad\square$

We will use C to denote the star $B_1(o)$ and we will use e to denote a distinguished edge in T_d. It is clear that if v is a neighbor of o, then $C \setminus \{v\}$ is a basis of C. Similarly e is a basis of itself. Using Lemma 4.1 and induction, we obtain that

$$\dim W_\lambda(B_k(C)) = |\partial B_k(C)| = d(d-1)^k; \qquad \dim W_\lambda(B_k(e)) = |\partial B_k(e)| = 2(d-1)^k \quad (4)$$

holds for every $\lambda \in \mathbb{R}$ and $k \in \mathbb{N}$, where ∂F denotes the boundary of a set F.

Recall that an invariant process $\{X_v\}_{v \in V_d}$ is an eigenvector process with eigenvalue λ if it satisfies the eigenvector equation (1) with probability 1 and $\mathrm{Var}(X_v) = 1$. Notice that if F is any

vertex set in T_d, then $\{X_v\}_{v \in F}$ is supported on $W_\lambda(F)$. In the rest of this chapter we investigate the joint distribution $\{X_v\}_{v \in S}$ in a non-trivial eigenvector process where S is one of $B_k(C)$ or $B_k(e)$ for some $k \in \mathbb{N}$. Our goal is to find uncorrelated linear combinations of the variables $\{X_v\}_{v \in S}$ with the property that they linearly generate every random variable in $\{X_v\}_{v \in S}$. Observe that since the covariance matrix of $\{X_v\}_{v \in S}$ has rank at most $\dim(W_\lambda(S)) = |\partial S|$, it is enough to find that many uncorrelated linear combinations with nontrivial variance to guarantee that they generate everything.

Let $Y = (Y_1, Y_2, \ldots, Y_n)$ be an n-dimensional distribution with 0 mean such that the covariance matrix is of full rank. We can always find a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ with the property that the covariance matrix of TY is the identity matrix. The probability distribution TY is unique up to an orthogonal transformation on $\mathbb{R}^n$. We call it the standardized version of Y.

Let $\mu \in I_d(\mathbb{R})$ represented by an invariant joint distribution $\{X_v\}_{v \in V_d}$ of random variables. Assume that $\mathbb{E}(X_o) = 0$. With every directed edge (v, w) of T_d we associate a probability distribution on $\mathbb{R}^{d-2}$ defined up to an orthogonal transformation. Let $\{v_i\}_{i=1}^{d-1}$ be the set of neighbors of w different from v. We denote by $A_{v,w}$ the standardized version of $\{X_{v_i} - X_{v_{d-1}}\}_{i=1}^{d-2}$. It is easy to see that (the orthogonal equivalence class of) $A_{v,w}$ does not depend on the labeling of the vectors v_i.

We introduce the symmetric relation r on directed edges of T_d such that $((x, y), (v, w)) \in r$ if and only if the unique shortest path connecting y and w contains at least one of x and v. The next lemma implies that if $((x, y), (v, w)) \in r$ then $A_{x,y}$ and $A_{v,w}$ are uncorrelated and so if μ is Gaussian then $A_{x,y}$ and $A_{v,w}$ are independent.

Lemma 4.2 *Let (v, w) be a directed edge in T_d. Let $\{v_i\}_{i=1}^{d-1}$ be the set of neighbors of w different from v. Let $X = \sum_{j=1}^{k} a_j X_{u_j}$ be a linear combination such that the shortest path connecting w and u_j contains v for every $1 \le j \le k$. Then $\mathbb{E}(X(X_{v_i} - X_{v_{d-1}})) = 0$ holds for $1 \le i \le d-2$.*

Proof. The condition on the vertices u_j guarantees that for every fix j the distance of u_j from v_i does not depend on i. Using this and the automorphism invariance of μ it follows that $\mathbb{E}(X_{u_j}(X_{v_i} - X_{v_{d-1}})) = 0$ holds for every $1 \le j \le k$ and $1 \le i \le d-2$. By linearity of expected value the proof is complete. $\square$

Let $S \subset V_d$ be either $B_k(C)$ or $B_k(e)$ for some $k \in \mathbb{N}$. Assume that μ is an eigenvector process. Let $p \in S$ be such that it has distance at least one from the boundary ∂S. Let D denote the set of directed edges (v, w) inside S with the following three properties: (a) the unique shortest path connecting p and w contains v, (b) $p \notin \{v, w\}$, (c) $B_1(w) \subset S$. Let B_p denote the joint distribution $(X_v - X_p)_{v \sim p}$. We denote by $\mathcal{Q}(S, p, \mu)$ the joint distribution of the random variables $\{A_{v,w}\}_{(v,w) \in D}$ and B_p. Lemma 4.2 implies that the components of $\mathcal{Q}(S, p, \mu)$ are uncorrelated multidimensional random variables. It is easy to see that if $|\lambda| < d$, then the correlation matrix of each such multidimensional random variable is of full rank. This implies that if $|\lambda| < d$, then

8

$\mathcal{Q}(S, p, \mu)$ provides a linear basis for $\{X_v\}_{v \in S}$. By counting dimensions, this yields the following corollary.

Corollary 4.1 *Let either $S = B_k(C)$ or $S = B_k(e)$ and μ an eigenvector process with eigenvalue $|\lambda| < d$. Then we have that*

$$\langle \mathrm{supp}\,(\mu_S) \rangle_{\mathbb{R}} = W_\lambda(S).$$

5   Typical processes and the limiting form of the main theorem

Now we describe a limiting form of our main theorem using typical processes on T_d. Typical processes on T_d were first introduced in [3] to study the properties of random regular graphs via ergodic theory. Here we use a slightly different definition which extends the original notion to processes that take values in a separable, metrizable space Y.

Definition 5.1 *Let Y be separable, metrizable topological space and let μ be a Borel probability distribution on Y^{V_d}. We say that μ is a typical process if there is a growing sequence of natural numbers $\{n_i\}$ with the following property. Assume that $\{G_i\}_{i=1}^n$ is a random graph sequence whose elements G_i are independently chosen random d-regular graphs on n_i vertices. Then with probability 1 there are maps $f_i : V(G_i) \to Y$ such that the distributions $\mathrm{distr}^*(f_i, G_i)$ are converging to μ in the weak topology.*

Our main theorem in the limit setting is the following.

Theorem 3 (Limiting form of the main theorem) *If μ is a nontrivial typical eigenvector process with eigenvalue λ, then $|\lambda| \leq 2\sqrt{d-1}$ and μ is the Gaussian wave Ψ_λ.*

Using the fact that weak limits of factor of i.i.d processes are typical (see [3]) we obtain the next corollary which answers a question of B. Virág.

Corollary 5.1 *If μ is a nontrivial eigenvector process that is a weak limit of factor of i.i.d. processes, then μ is a Gaussian wave with eigenvalue $|\lambda| \leq 2\sqrt{d-1}$.*

Note that Corollary 5.1 implies that if many eigenvector processes corresponding to the same eigenvalue are coupled in a way that the coupling is a weak limit of factor of i.i.d. processes, then its distribution is jointly Gaussian.

We emphasize that the first part of the statement in Theorem 3, namely that $|\lambda| \leq 2\sqrt{d-1}$ is a consequence of Friedman's theorem [22]. We prove this implication in Lemma 5.1. In addition, the main goal of this chapter is to show that Theorem 3 implies Theorem 2.

Lemma 5.1 *If μ is a nontrivial typical eigenvector process with eigenvalue λ, then $|\lambda| \leq 2\sqrt{d-1}$.*

Proof. Our first goal is to prove that there exists a sequence of d-regular graphs $\{G_i\}_{i=1}^\infty$ and vectors $\{f_i : V(G_i) \to \mathbb{R}\}_{i=1}^\infty$ such that (a) $|\lambda_2(G_i)| \to 2\sqrt{d-1}$, where $\lambda_2(G)$ denotes the second largest (in absolute value) eigenvalue of a finite graph G; (b) $\mathrm{distr}^*(f_i, G_i) \to \mu$ in the weak topology. Using that μ is typical, there exists a sequence $\{n_i\}_{i=1}^\infty$ of growing natural numbers such that with probability 1, if $\{G_i'\}_{i=1}^\infty$ is a sequence of independent random d-regular graphs with $|V(G_i')| = n_i$, then there exists a sequence of functions $\{f_i'\}_{i=1}^\infty$ satisfying (b). Friedman's theorem [22] implies that with probability 1 we can choose a further subsequence satisfying (a).

We fix an arbitrary $\varepsilon > 0$. Let X_o be a random variable with distribution μ_o. Since $\mathbb{E}(X_o^2) = 1$, we can find $k > 0$ such that k is a continuity point of the cumulative distribution function of X_o and $k^2 \mathbb{P}(|X_o| > k) < \varepsilon$. Furthermore, if k is sufficiently large, we can also assume that $\mathbb{E}([X_o]_k^2) \geq 1/2$ and $|\mathbb{E}([X_o]_k)| \leq \varepsilon$, where $[x]_k = \max(\min(x, k), -k)$. Using that $\mathrm{distr}^*(f_i, G_i)$ converges to the eigenvector process μ, we obtain that if i is large enough, then (i) $k^2 \mathbb{P}(\mathrm{distr}(|f_i|) > k) < 2\varepsilon$; (ii) $\mathbb{P}(\mathrm{distr}(|(G_i - \lambda I)f_i|) \geq \varepsilon) \leq \varepsilon/k^2$; (iii) $\mathbb{E}(\mathrm{distr}([f_i]_k^2)) \geq 1/3$; (iv) $c_i = |\mathbb{E}(\mathrm{distr}([f_i]_k))| \leq 2\varepsilon$. Note that (i) implies that $f_i \neq [f_i]_k$ holds on a vertex set of density at most $2\varepsilon/k^2$. It follows that $(G_i - \lambda I)[f_i]_k \neq (G_i - \lambda I)f_i$ holds on a vertex of density at most $(d+1)2\varepsilon/k^2$. Furthermore, we have that $\|(G_i - \lambda I)[f_i]_k\|_\infty \leq (d + |\lambda|)k$. Putting all this together, we obtain that for i large enough

$$\|(G_i - \lambda I)([f_i]_k - c_i)\|_2 \leq \sqrt{\varepsilon^2 n_i + (2d+3)(\varepsilon/k^2)\cdot n_i(d+|\lambda|)^2 k^2} + |d - \lambda|c_i\sqrt{n_i}$$

$$\leq \sqrt{n_i}\left(\sqrt{\varepsilon^2 + (2d+3)\varepsilon(d+|\lambda|)^2} + 2|d-\lambda|\varepsilon\right).$$

Let $v_i = ([f_i]_k - c_i)/\|[f_i]_k - c_i\|_2$. Using that (iii) implies that $\|[f_i]_k\|_2^2 \geq 1/3n_i$, we obtain that for an appropriate choice of small enough ε and large enough i the quantity $\|(G_i - \lambda I)v_i\|_2$ is arbitrarily small. Thus, by using (a) and Lemma 3.1, we get that $|\lambda| \leq 2\sqrt{d-1}$. $\square$

Proposition 5.1 *Theorem 3 implies Theorem 2.*

We need a few notions and lemmas. Let $\mathcal{P}$ denote the set of Borel probability distributions μ on $\mathbb{R}^{V_d}$ which have a second moment bounded from above by 1 at each coordinate. By tightness of $\mathcal{P}$, we have that $\mathcal{P}$ is compact with respect to the weak topology of measures. Let m be a fixed metrization of the weak topology on $\mathcal{P}$. Let us denote by $\mathcal{T}$ the set of closed subsets in $\mathcal{P}$, and by d_H the the Hausdorff metric on $\mathcal{T}$. We have that d_H induces a compact topology on $\mathcal{T}$. Let us define the distance m^* for d-regular graphs in the following way. If G_1 and G_2 are d-regular graphs then $m^*(G_1, G_2)$ is the infimum of the numbers δ with the property the if $f_1 : V(G_1) \to \mathbb{R}$, $f_2 : V(G_2) \to \mathbb{R}$ are arbitrary functions with $\mathbb{E}(\mathrm{distr}(f_1^2)), \mathbb{E}(\mathrm{distr}(f_2^2)) \leq 1$ then there are functions $f_1' : V(G_2) \to \mathbb{R}$, $f_2' : V(G_1) \to \mathbb{R}$ with $\mathbb{E}(\mathrm{distr}(f_1'^2)), \mathbb{E}(\mathrm{distr}(f_2'^2)) \leq 1$ such that

$$m(\mathrm{distr}^*(f_1, G_1), \mathrm{distr}^*(f_1', G_2)) \leq \delta \quad , \quad m(\mathrm{distr}^*(f_2, G_2), \mathrm{distr}^*(f_2', G_1)) \leq \delta.$$

We can describe the metric m^* in terms of the metric d_H as follows. For a graph G, let

$$S(G) = \{\mathrm{distr}^*(f, G) | f \in \mathbb{R}^{V(G)}, \mathbb{E}(\mathrm{distr}(f^2)) \leq 1\}.$$

We have that $m^*(G_1, G_2) = d_H(S(G_1), S(G_2))$.

Definition 5.2 *We say that a finite d-regular graph G is ε-typical for some $\varepsilon > 0$ if with probability at least $1 - \varepsilon$ a random d-regular graph G' on $|V(G)|$ vertices has the property that $m^*(G, G') < \varepsilon$.*

Lemma 5.2 *For every $\varepsilon > 0$ and $T \in \mathcal{T}$ there exists $n(\varepsilon)$ such that for all $N > n(\varepsilon)$ there exists a value c satisfying*

$$\mathbb{P}(|d_H(S(G), T) - c| > \varepsilon) < \varepsilon. \tag{5}$$

Proof. The proof relies on a certain continuity property of the metric m^* with respect to small changes in a graph. More precisely, we show that for every ε_2 if N is large enough, then for every pair G, G' of d-regular graphs on the vertex set $[N]$ satisfying $|E(G)\Delta E(G')| \leq 4$ we have that $m^*(G, G') < \varepsilon_2$. The significance of the number 4 comes from the fact that d-regular graphs can be transformed into each other through a sequence of operations in which two independent edges $(u_1, v_1), (u_2, v_2)$ are replaced by $(u_1, v_2), (u_2, v_1)$. To prove the continuity property, we show that if N is large enough, then the inequality

$$m(\mathrm{distr}^*(f, G), \mathrm{distr}^*(f, G')) < \varepsilon_2$$

holds for every function $f : [N] \to \mathbb{R}$ with $\mathbb{E}(f^2) \leq 1$. Let k be an arbitrary integer. Observe that the marginal distribution $\mathrm{distr}^*(f, G)|_{B_k(o)}$ can be obtained from the distribution of $\mathbb{R}$-colored neighborhoods of radius k of a random vertex in G (and the analogous statement holds for G'). Since $|E(G)\Delta E(G')|$ intersects such a neighborhood with probability tending to 0 as $|V(G)| \to \infty$, we have that the distance between $\mathrm{distr}^*(f, G)|_{B_k(o)}$ and $\mathrm{distr}^*(f, G')|_{B_k(o)}$ converges to 0 in any metrization of the weak topology of $\mathbb{R}^{B_k(o)}$. This implies the desired continuity property.

We obtain by the above statement and the triangle inequality that if $|E(G)\Delta E(G')| \leq 4$ and N is large enough, then $|d_H(S(G), T) - d_H(S(G'), T)| < \varepsilon_2$. It is well-known that graph parameters on random d-regular graphs that satisfy this Lipschitz property are concentrated around their mean; see [38, Theorem 2.19]. $\square$

Lemma 5.3 *For every $\varepsilon > 0$ there exists $n(\varepsilon)$ such that for every $N > n(\varepsilon)$ with probability at least $1 - \varepsilon$ a random d-regular graph on N vertices is ε-typical.*

Proof. Let M be a finite $\varepsilon/2$-net in $\mathcal{T}$. If G is a random d-regular graph on N vertices, then there exists $T \in M$ with the property that

$$\mathbb{P}(d_H(S(G), T) < \varepsilon/2) \geq 1/|M|. \tag{6}$$

We apply Lemma 5.2 with $\varepsilon' \leq \varepsilon/4$. Combining inequalities (5) and (6), we obtain that for $\varepsilon' < 1/|M|$ and N large enough $|c| \leq 3/4\varepsilon$. Then applying (5) again, the proof is complete. $\qquad\square$

Now we enter the proof of Lemma 5.1.

Proof. We go by contradiction. If Theorem 2 fails then there is a growing sequence of natural numbers $\{n_i\}_{i=1}^{\infty}$, $\varepsilon > 0$ and a sequence $\{\delta_i\}_{i=1}^{\infty}$ with $\lim_{i\to\infty} \delta_i = 0$ such that the following holds. If G is a random d-regular graph on n_i vertices then we have with probability at least ε that there is an δ_i-almost eigenvector v of G (with entry sum 0) such that $\mathrm{distr}^*(v, G)$ is at least ε-separated from any Gaussian wave in the weak topology.

From Lemma 5.3 we obtain that there is a sequence $\{\varepsilon'_i\}_{i=1}^{\infty}$ with $\lim_{i\to\infty} \varepsilon'_i = 0$ such that a random d-regular graph on n_i vertices is ε'_i-typical with probability at least $1 - \varepsilon'_i$ for every i. There exists an index j such that for all $i \geq j$ we have $\varepsilon'_i < \varepsilon$. This implies that for all $i \geq j$ we can choose a graph G_i on n_i vertices such that G_i is ε'_i-typical and there exists a δ_i-almost eigenvector f_i of G_i (with entry sum 0) satisfying that $\mathrm{distr}^*(\sqrt{n_i}f_i, G_i)$ is at least ε-separated from any Gaussian wave in the weak topology.

By choosing a subsequence we can assume (by abusing the notation) that $\mathrm{distr}^*(\sqrt{n_i}f_i, G_i)$ weakly converges to some measure $\mu \in \mathcal{P}$. It is clear that μ is a nontrivial eigenvector process which is at least ε-separated from any Gaussian wave in the weak topology. To get a contradiction it remains to show that μ is typical.

Again by choosing a subsequence we can assume that $\sum_{i=1}^{\infty} \varepsilon'_i < \infty$. Let $\{G'_i\}_{i=1}^{\infty}$ be such that G'_i is a random d-regular graph on n_i vertices and the terms of the sequence are independent. It follows from the Borel–Cantelli lemma that almost surely all but finitely many indices i satisfy that $m^*(G'_i, G_i) \leq \varepsilon'_i$. For such indices we can find f'_i with $m(\mathrm{distr}^*(f'_i, G'_i), \mathrm{distr}^*(\sqrt{n_i}f_i, G_i)) \leq \varepsilon'_i$. We obtain that $\mathrm{distr}^*(f'_i, G'_i)$ converges to μ showing that μ is typical. $\qquad\square$

6 Entropy inequality for typical processes

Let X be a separable metric space; let F be a finite set and let $\mathcal{P}(F)$ denote the set of probability distributions on F equipped with the topology generated by total variation distance. A continuous discretization of X is a continuous function $\phi : X \to \mathcal{P}(F)$. If $\alpha \in X^V$ is an X-coloring of a finite or countable set V, then we denote by $\phi * \alpha$ the probability distribution on F^V obtained by independently choosing an element from F for each $v \in V$ with distribution $\phi(\alpha(v))$. If μ is a probability distribution on X^V, then we denote by $\phi * \mu$ the probability distribution obtained by first taking a μ random element $\alpha : V \to X$ and then in a second round of randomization we take $\phi * \alpha$. The main result of this chapter is the next entropy inequality for typical processes.

Theorem 4 *If $\mu \in I_d(X)$ is a typical process and $\phi : X \to \mathcal{P}(F)$ is a continuous discretization,*

then the process $\phi * \mu$ satisfies the next entropy inequality.

$$\mathbb{H}(B_k(C)) - (d/2)\mathbb{H}(B_k(e)) \geq \mathbb{E}_{\mu_o}(\mathbb{H}(\phi(x))).$$

Before proving Theorem 4 we need some preparation. Let us fix a metrization of the weak topology on $I_d(X)$. For a finite d-regular graph G let $\omega(G)$ denote the infimum of the numbers $\varepsilon > 0$ for which it is true that at least $1 - \varepsilon$ fraction of the vertices of G are not contained in a cycle of length at most $\lfloor 1/\varepsilon \rfloor$. The quantity $\omega(G)$ measures how similar the graph G is to the tree T_d in the Benjamini–Schramm metric. Throughout this chapter, G is always assumed to be finite.

The operator distr^* maps X-colored d-regular graphs $(\alpha \in X^{V(G)}, G)$ to invariant processes in $I_d(X)$. Using this correspondence and the metric on $I_d(X)$ we define the distance of an X-colored graph (α, G) and a process $\mu \in I_d(X)$ as $\omega(G)$ plus the distance of $\mathrm{distr}^*(\alpha, G)$ and μ in I_d. Note that if $(\alpha \in X^{V(G)}, G)$ is an X-colored d-regular graph, then $(\phi * \alpha, G)$ is a probability distribution on F-valued colorings of the vertices of G. In this case $\mathrm{distr}^*(\phi * \alpha, G)$ is a probability distribution on $I_d(X)$, while $\phi * \mathrm{distr}^*(\alpha, G)$ is a single element in $I_d(X)$.

Proposition 6.1 *Let $\mu \in I_d(X)$ be an invariant process and ϕ is a continuous discretization of X. Then for every $\varepsilon > 0$ there is $\delta > 0$ such that if a colored d-regular graph $(\alpha \in X^{V(G)}, G)$ is at most of distance δ from μ, then with probability at least $1 - \varepsilon$ we have that $(\phi * \alpha, G)$ is at most of distance ε from $\phi * \mu$.*

The proof of Proposition 6.1 relies on the next technical lemma.

Lemma 6.1 *Let $\mu \in I_d(X)$ be an invariant process. Let $r \in \mathbb{N}$, $B = B_r(o) \subset V_d$ and let $\beta : F^B \to \mathbb{R}$ be any automorphism invariant function. Then for every $\varepsilon > 0$ there is $\delta > 0$ such that if a colored d-regular graph $(\alpha \in X^{V(G)}, G)$ is at most of distance δ from μ then with probability at least $1 - \varepsilon$ we have that*

$$|W|^{-1} \sum_{v \in W} \beta(\gamma|_{B_r(v)})$$

is at most ε-far from $\mathbb{E}(\beta(\phi * \mu_B))$ where $\gamma = \phi * \alpha$ and $W = \{v : v \in V(G), B_r(v) \simeq B\}$.

Proof. Assume first that $(\alpha \in X^{V(G)}, G)$ is an arbitrary d-regular X-colored graph whose distance is δ' from μ and let γ, W as in the statement of the lemma. For a vertex $v \in W$ let Y_v be the random variable with value $\beta(\gamma|_{B_r(v)})$. Let $g : X^B \to \mathbb{R}$ be the function defined by $g(h) = \mathbb{E}(\beta(\phi * h))$. We have for $v \in W$ that $\mathbb{E}(Y_v)$ is equal to $g(\alpha|_{B_r(v)})$. Let $Y = |W|^{-1} \sum_{v \in W} Y_v$. It follows that

$$\mathbb{E}(Y) = |W|^{-1} \sum_{v \in W} g(\alpha|_{B_r(v)}) = \int_{X^B} g \, d\nu_G, \tag{7}$$

where ν_G describes the probability distribution of the isomorphism classes of $\alpha|_{B_r(v)}$ where $v \in W$ is a uniform random point. Using the fact that g is a continuous function we obtain that if δ' is

13

small enough then the right hand side of (7) is at most $\varepsilon/2$ far from $\int_{X^B} g \, d\mu_B = \mathbb{E}(\beta(\phi * \mu_B))$. Observe that Y_v and Y_w are independent if v and w have distance at least $2r + 1$ in G. It follows that there are at most $|B_{2r+1}(o)||W|$ correlated pairs in $\{Y_v\}_{v \in W}$. We obtain that the variance of Y is at most $|B_{2r+1}(o)|^{1/2}|W|^{-1/2} \max |\beta|$. We use that $\omega(G)$ goes to 0 as δ' goes to 0 and thus $|W|$ tends to infinity. This implies that if δ' is sufficiently small then the variance of Y is at most $\varepsilon^2/3$. Now by Chebyshev's inequality we have that $\mathbb{P}(|Y - \mathbb{E}(Y)| \geq \varepsilon/2) \leq \varepsilon^2$. It follows that $\mathbb{P}(|Y - \mathbb{E}(\beta(\phi * \mu_B))| \geq \varepsilon) \leq \varepsilon^2$ which completes the proof. $\qquad\square$

We continue with the proof of Proposition 6.1.

Proof. Let δ' be an arbitrary positive number. Let $(\alpha \in X^{V(G)}, G)$ be an X-colored d-regular graph of distance δ' from μ. Let $(\gamma \in F^{V(G)}, G)$ be chosen according to the probability distribution $(\phi * \alpha, G)$. Let $\varepsilon' > 0$, $r = \lfloor 1/\varepsilon' \rfloor$ and $W = \{v : v \in V(G), B_r(v) \simeq B\}$. It follows from Lemma 6.1 that there is $c = c(\varepsilon') > 0$ such that if $\delta' < c$ then the condition of Lemma 6.1 holds for (γ, G) simultaneously for every $0 - 1$ valued β with probability at least $1 - \varepsilon'$. (Here we use the fact that there are finitely many such functions β.) If c is small enough, then it also guarantees that $|W|/|V(G)| \geq 1 - \varepsilon'$. Now it is clear that if ε' is small enough, then these properties imply that (γ, G) is at most ε far from $\phi * \mu$. $\qquad\square$

For the next lemma let $R(d, n)$ denote the number of d-regular graphs on the vertex set $[n]$. In case d is odd we will always assume that the number of vertices is even.

Lemma 6.2 *Let F be a finite set and $\mu \in I_d(F)$. Let $N(n, \varepsilon)$ denote the number of d-regular F-colored graphs on the vertex set $[n]$ whose distance from μ is at most ε. Assume that $\{n_i\}_{i=1}^\infty$ is a growing sequence of natural numbers. Then for every $\varepsilon > 0$ and $k \in \mathbb{N}$ we have that*

$$\mathbb{H}(B_k(C)) - (d/2)\mathbb{H}(B_k(e)) \geq \lim_{\varepsilon \to 0} \left(\limsup_{i \to \infty} n_i^{-1} \log(N(n_i, \varepsilon)/R(d, n_i)) \right). \qquad (8)$$

Proof. First we prove the statement for $k = 0$. In this case we use a formula from [3] that approximates the number $N'(n, \varepsilon)$ of F-colored graphs on $[n]$ in which the statistics of colored 1-neighborhoods is ε-close to μ_C. We have that

$$N'(n, \varepsilon) = R(d, n)\mathbb{H}(C)^{(n(1+o(1))}\mathbb{H}(e)^{-(dn/2)(1+o(1))}$$

where $o(1)$ is a quantity which goes to 0 when first $n \to \infty$ and then $\varepsilon \to 0$. This implies that the right hand side of (8) is equal to the left hand side when $N(n_i, \varepsilon)$ is replaced by $N'(n_i, \varepsilon)$. Now we use that for every $\varepsilon > 0$ there exists $\varepsilon' > 0$ such that $N'(n_i, \varepsilon) \geq N(n_i, \varepsilon')$ holds for all i. This finishes the proof of the first part.

The idea of the proof in case of $k > 0$ is to generate a new process from μ in which the color of every vertex $v \in V_d$ is replaced by the isomorphism type of the colored neighborhood of v of radius k. The main difficulty in this approach is that the isomorphism type describes the neighborhood

only up to automorphisms which leads to extra constants in the entropy formulas. To control these constants (and to eventually get rid of them) we add some extra randomness to the process.

Let us introduce new processes $\mu_{r,m}$ on T_d for every $k, m \in \mathbb{N}$. If $r = 0$, then $\mu_{0,m}$ denotes the $F \times [m]$ valued process in which we generate a μ-random coloring on T_d and then we add a second coordinate from $[m]$ to every vertex independently and uniformly. In general $\mu_{r,m}$ denotes the process obtained from $\mu_{0,m}$ by coloring $v \in V_d$ with the isomorphism class of the coloring of $B_r(v)$ in $\mu_{0,m}$. Let $a_{r,m,k}$ denote the left hand side (resp. $b_{r,m}$ denote the right hand side) of (8) evaluated for $\mu_{r,m}$ and k. It is easy to see that $b_{r,m}$ does not depend on r. On the other hand the independence of the two coordinates implies that $b_{0,m} = b_{0,1} + \log m$. All together this means that $b_{r,m} = b_{0,1} + \log m$. It is clear that $a_{0,m,k} = a_{0,1,k} + \log m$. From the case $k = 0$ we have that $a_{r,m,0} \geq b_{r,m}$. We can write this as

$$a_{0,1,r} + (a_{0,m,r} - a_{0,1,r}) + (a_{r,m,0} - a_{0,m,r}) \geq b_{0,1} + \log m$$

and thus $a_{0,1,r} + c_{r,m} \geq b_{0,1}$ holds for every m where $c_{r,m} = a_{r,m,0} - a_{0,m,r}$. Since the inequality $a_{0,1,r} \geq b_{0,1}$ is equivalent to the statement of the lemma for $k = r$ it remains to show that $\lim_{m \to \infty} c_{r,m} = 0$ holds for every r.

Let t_r denote the size of the automorphism group of the rooted $d - 1$-regular tree of depth r. We claim that $\mathbb{H}_{\mu_{0,m}}(B_r(C)) - \mathbb{H}_{\mu_{r,m}}(C) = d \log t_r + o(1)$ and that $\mathbb{H}_{\mu_{0,m}}(B_r(e)) - \mathbb{H}_{\mu_{r,m}}(e) = 2 \log t_r + o(1)$ as $m \to \infty$. It is clear that this claim implies $c_{r,m} = o(1)$. We show the proof of the first claim. (The proof of the second one is almost identical.) If m is large enough then in the process $\mu_{0,m}$ restricted to $B_{r+1}(o)$ all labels are different with probability converging to 1. In such a case knowing the isomorphism classes of the colored neighborhoods of radius $r - 1$ of vertices in C is equivalent with knowing the colored version of $B_{r+1}(o)$ up to an isomorphism that fixes C. The stabilizer of C in the automorphism group of $B_{r+1}(o)$ is the d-th power of the automorphism group of the rooted $d - 1$ regular tree of depth r. Thus the entropy loss of $\mathbb{H}_{\mu_{r,m}}(C)$ compared to $\mathbb{H}_{\mu_{0,m}}(B_r(C))$ is converging to $d \log t_r$. $\qquad\square$

Now we arrived to the proof of Theorem 4.

Proof. According to Lemma 6.2 it is enough to show that the process $\phi * \mu$ satisfies

$$\lim_{\varepsilon \to 0} \left(\limsup_{i \to \infty} n_i^{-1} \log(N(n_i, \varepsilon)/R(d, n)) \right) \geq \mathbb{E}_{\mu_o}(\mathbb{H}(\phi(x)))$$

for some growing sequence $\{n_i\}_{i=1}^{\infty}$ of natural numbers. For $n \in \mathbb{N}, \varepsilon > 0$ let $a(n, \varepsilon)$ denote the number of d-regular graphs G on $[n]$ with the property that there exists an X-coloring α of $[n]$ such that (G, α) is of distance at most ε from μ. The fact that μ is typical is equivalent to the fact that there is a sequence $\{n_i\}_{i=1}^{\infty}$ such that $\lim_{i \to \infty} a(n_i, \varepsilon)/R(d, n_i) = 1$ holds for every $\varepsilon > 0$.

From Proposition 6.1 we obtain that for every $\varepsilon_2 > 0$ there is $\varepsilon > 0$ such that if a graph G has an X coloring α of distance at most ε from μ, then with probability at least $1 - \varepsilon_2$ we have that

$(\phi * \alpha, G)$ is of distance ε_2 from $\phi * \mu$. It follows that G has at least $\exp(\mathbb{H}(\phi * \alpha, G) + o(1))$ F-colorings of distance at most ε_2 from $\phi * \mu$. On the other hand, we have $|V(G)|^{-1}\mathbb{H}(\phi * \alpha, G) = |V(G)|^{-1} \sum_{v \in V(G)} \mathbb{H}(\phi(\alpha(v)))$, which converges to $\mathbb{E}_{\mu_o}(\mathbb{H}(\phi(x)))$ as α converges to μ.

New let $N(n, \varepsilon)$ defined as in Lemma 6.2 for the process $\phi * \mu$. From the above observations we obtain that $n_i^{-1} \log(N(n_i, \varepsilon)/R(d, n_i))$ can be estimated from below by $\mathbb{E}_{\mu_o}(\mathbb{H}(\phi(x))) - o(1)$ as n_i goes to infinity. Then Lemma 6.2 finishes the proof. $\qquad\square$

7 Smooth eigenvector processes

Let μ be an eigenvector process. If $F \subseteq V_d$ is a finite set then the distribution of μ, when restricted to F, is concentrated on the subspace $W_\lambda(F)$ (recall Chapter 4). We denote by $\mathbb{D}_{\mathrm{sp}}(F, \mu)$ the differential entropy of μ_F measured inside this subspace using the Euclidean structure inherited from $\mathbb{R}^F$. We say that μ is *smooth* if $\mathbb{D}_{\mathrm{sp}}(B_k(C))$ and $\mathbb{D}_{\mathrm{sp}}(B_k(e))$ are finite for every k. In this chapter we reduce Theorem 3 to smooth eigenvector processes. The reduction will rely on the following statement.

Proposition 7.1 *Let $\{X_v\}_{v \in V_d}$ be a typical eigenvector process with eigenvalue λ and $\{Y_v\}_{v \in V_d}$ the unique Gaussian wave Ψ_λ with eigenvalue λ. Then the independent sum $\{X_v + aY_v\}_{v \in V_d}$ is smooth for $a > 0$.*

Proof. Let S be one of $B_k(C)$ and $B_k(e)$. Both $\{X_v\}_{v \in S}$ and $\{Y_v\}_{v \in S}$ are supported on $W_\lambda(S)$. Moreover, by Corollary 4.1, we obtain that the support of $\Psi_\lambda|_S$ is equal to $W_\lambda(S)$. Using Lemma 12.3 inside the space $W_\lambda(S)$, we get that the differential entropy of the joint distribution $\{X_v + aY_v\}_{v \in S}$ is finite. $\qquad\square$

Assume that Theorem 3 holds for smooth typical eigenvector processes. Using Proposition 7.1 we obtain that if $\{X_v\}_{v \in V_d}$ is an arbitrary typical eigenvector process then $\{X_v + aY_v\}_{v \in V_d}$ is smooth for all $a > 0$. In addition, by Proposition 14.1 and Proposition 14.2, $\{X_v + aY_v\}_{v \in V_d}$ is typical. By our assumption, we obtain the Gaussianity of $\{X_v + aY_v\}_{v \in V_d}$ for all $a > 0$. This implies the Gaussianity of $\{X_v\}_{v \in V_d}$ by going to 0 with a.

8 Entropy Inequality for typical eigenvector processes

In this chapter we prove the next theorem.

Theorem 5 *Let $\mu \in I_d(\mathbb{R})$ be a smooth typical eigenvector process. Then*

$$\mathbb{D}_{\mathrm{sp}}(B_k(C)) - \frac{d}{2}\mathbb{D}_{\mathrm{sp}}(B_k(e)) \geq 0$$

holds for every $k \geq 0$ integer.

To prove the above theorem we will need some preparation. For $a \in \mathbb{N}$ let us define the continuous discretization $t_{0,a}$ of $\mathbb{R}$ in the following way. If $x > a$ (resp. $x < -a$), then $t_{0,a}(x) = a$ (resp. $t_{0,a}(x) = -a$) with probability 1. Otherwise let $t_{0,a}(x)$ denote the probability distribution that takes $\lfloor x * a \rfloor / a$ with probability $1 + \lfloor x * a \rfloor - x * a$ and takes $1/a + \lfloor x * a \rfloor / a$ with probability $x * a - \lfloor x * a \rfloor$. For $\sigma > 0$ we define the discretization $t_{\sigma,a}$ by $t_{\sigma,a}(x) = t_{0,a}(x + \sigma N)$ where N is a random variable with standard normal distribution. We denote by $t_{\sigma,a}^n$ the continuous discretization of $\mathbb{R}^n$ obtained by the coordinatewise independent application of $t_{\sigma,a}$.

Lemma 8.1 *Let X be a random variable with values in $\mathbb{R}^n$ with finite variance, i.e. $\mathbb{E}(\|X\|_2^2) < \infty$. Then we have for every fixed $\sigma > 0$ that*

$$\mathbb{H}(t_{\sigma,a}^n(X)) = n \log a + \mathbb{D}(X + \sigma M) + o(1)$$

as $a \to \infty$, where M is independent of X and has standard normal distribution on $\mathbb{R}^n$.

The main difficulty of the proof of Lemma 8.1 comes from the fact that the support of X is not necessarily compact. We have to treat a situation where we refine and increase the interval of discretization simultaneously.

Proof. By Lemma 12.3, the finite variance of X guarantees that $\mathbb{D}(X + \sigma M)$ exists and is a finite quantity. Let $S_a = \{r/a | r \in \mathbb{Z}^n, \|r\|_\infty \le a^2\}$ and let $S_a' = \{x | x \in S_a, \|x\|_\infty < a\}$. For $x \in S_a$ let $p_a(x)$ denote the probability of x in the distribution $t_{\sigma,a}^n(X)$. We have that

$$\mathbb{H}(t_{\sigma,a}^n(X)) = \sum_{x \in S_a} -p_a(x) \log p_a(x).$$

Let q_a denote the quantity $\mathbb{P}(t_{\sigma,a}^n(X) \in S_a \setminus S_a') = \sum_{x \in S_a \setminus S_a'} p_a(x)$. The construction of $t_{\sigma,a}^n$ shows that

$$q_a \le \mathbb{P}(\|X + \sigma M\|_\infty \ge a - a^{-1}).$$

By Chebyshev's inequality we have that

$$\mathbb{P}(\|X + \sigma M\|_\infty \ge a - a^{-1}) = O(a^{-2})$$

and so $q_a = O(a^{-2})$. It follows that

$$\sum_{x \in S_a \setminus S_a'} -p_a(x) \log p_a(x) \le -q_a \log q_a + q_a \log |S_a \setminus S_a'| \le -q_a \log q_a + q_a \log |S_a|$$

$$= -q_a \log q_a + q_a n \log(2a^2 + 1) = o(1).$$

For $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ let $g_a(x) = \prod_{i=1}^n \max\{1 - a|x_i|, 0\}$. For every $x \in S_a'$ we have that $p_a(x) = \int_{z \in \mathbb{R}^n} g_a(z) f(x - z)$ where f is the density function of $X + \sigma M$ on $\mathbb{R}^n$. Using that $a^n \int_{\mathbb{R}^n} g_a = 1$ we have that $a^n p_a(x)$ is a weighted average of the values of f in an L^∞-ball of radius

$1/a$. It follows that for every $x \in S'_a$ there is $\alpha(x) \in \mathbb{R}^n$ with the property that $\|x - \alpha(x)\|_\infty \le 1/a$ and $a^n p_a(x) = f(\alpha(x))$ (using that f is continuous). Now we have that

$$\sum_{x \in S'_a} -p_a(x) \log p_a(x) = (n \log a) \sum_{x \in S'_a} p_a(x) - a^{-n} \sum_{x \in S'_a} f(\alpha(x)) \log f(\alpha(x)). \qquad (9)$$

It follows from $q_a = O(a^{-2})$ that

$$(n \log a) \sum_{x \in S'} p_a(x) = n(\log a)(1 - q_a) = n \log a + o(1).$$

It remains to bound the second part of (9). From the equation

$$\int_{z \in [-a+a^{-1}, a]^n} -f(z) \log f(z) = \int_{z \in [0, a^{-1}]^n} \sum_{x \in S'_a} -f(x+z) \log f(x+z) \qquad (10)$$

we obtain that there is a fixed $\gamma \in [0, a^{-1}]^n$ such that

$$a^{-n} \sum_{x \in S'_a} -f(x+\gamma) \log f(x+\gamma) \qquad (11)$$

is equal to the left hand side of (10). On the other hand, the left hand side of (10) is equal to $\mathbb{D}(X + \sigma M) + o(1)$. It remains to show that

$$a^{-n} \sum_{x \in S'_a} \big(f(\alpha(x)) \log f(\alpha(x)) - f(\beta(x)) \log f(\beta(x)) \big) = o(1) \qquad (12)$$

where $\beta(x) = x + \gamma$. (Note that α, β, γ all depend on a.) We will use that $\|\alpha(x) - \beta(x)\|_\infty \le 2/a$ holds for every $x \in S'_a$ and thus $r(x) := \|\alpha(x) - \beta(x)\|_2 \le 2\sqrt{n}/a$. Let $t_x = f(\alpha(x))/f(\beta(x))$. We have by Lemma 12.4 that for every $\varepsilon > 0$ if a is big enough, then $t_x \ge 1 - \varepsilon$ (resp. $t_x^{-1} \ge 1 - \varepsilon$) provided that $f(\beta(x)) > c$ (resp. $f(\alpha(x)) > c$) where $c = \sigma^{-n/2} \exp(-a/(32n\sigma^2))$. This implies that for every $a > 0$ we can choose $\varepsilon = \varepsilon(a)$ such that $\lim_{a \to \infty} \varepsilon(a) = 0$ and the previous property holds with ε. We will assume that a is so large that $\varepsilon(a) < 1/3$.

Let T_1 denote the sum of the terms in (12) where $f(\beta(x)) \le 2c$ and let T_2 denote the sum of the remaining terms. According to Lemma 12.4 either $f(\alpha(x)) < c$ or $t_x^{-1} \ge 1 - \varepsilon$. If $f(\beta(x)) \le 2c$, then we have $f(\alpha(x)) \le 2c/(1-\varepsilon) \le 3c$ in both cases. It follows that $T_1 \le 3c \log(3c)(2a^2 + 1)^n / a^n = o(1)$.

Now we estimate T_2. From now on we assume that $f(\beta(x)) > 2c$. By Lemma 12.4 we obtain $t_x \ge 1 - \varepsilon$ holds and thus $f(\alpha(x)) \ge (1-\varepsilon)2c > c$. This implies again by Lemma 12.4 that $t_x^{-1} \ge 1 - \varepsilon$ and so $|1 - t_x| \le 2\varepsilon$. We have that

$$f(\alpha(x)) \log f(\alpha(x)) - f(\beta(x)) \log f(\beta(x)) = f(\alpha(x)) \log t_x - (1 - t_x) f(\beta(x)) \log f(\beta(x)).$$

Using that $f(\alpha(x)) = a^n p_a(x)$ and $|\log t_x| \le -\log(1 - 2\varepsilon)$ we get that

$$\left| a^{-n} \sum_{x \in S'_a, f(\beta(x)) > 2c} f(\alpha(x)) \log t_x \right| \le \sum_{x \in S_a} p_a(x)(-\log(1 - 2\varepsilon)) = -\log(1 - 2\varepsilon) = o(1).$$

It is now clear that the following claim finishes the proof of the lemma.

Claim: $a^{-n} \sum_{x \in S'_a} |f(\beta(x)) \log f(\beta(x))| = O(1)$.

By Lemma 12.2 there is $b \in \mathbb{R}^+$ such that $f(x) < 1$ whenever $\|x\|_\infty > b-1$. Let $B = [-b, b]^n$. By the finiteness of B we have that

$$a^{-n} \sum_{x \in S'_a \cap B} -\beta(x) \log(\beta(x)) = \int_B -f \log f + o(1) \tag{13}$$

and that

$$a^{-n} \sum_{x \in S'_a \cap B} |\beta(x) \log(\beta(x))| = \int_B |f \log f| + o(1) \tag{14}$$

It follows from (13) and the property that the left hand side of (10) is equal to (11) that

$$a^{-n} \sum_{x \in S'_a \setminus B} -\beta(x) \log(\beta(x)) = \int_{\overline{B}} -f \log f + o(1). \tag{15}$$

By (14) and (15) we get that

$$a^{-n} \sum_{x \in S'_a} |\beta(x) \log(\beta(x))| = \int_{\mathbb{R}^n} |f \log f| + o(1) = O(1). \qquad \square$$

In the following lemma, first we calculate the entropy of the random variable $t_{\sigma,a}(x)$ for every fixed x, and then we take the expectation of this quantity with respect to the distribution of a random variable X.

Lemma 8.2 *Let X be a real valued random variable with finite variance and with distribution ν. Then $\mathbb{E}_\nu(\mathbb{H}(t_{\sigma,a}(x))) = \log a + \mathbb{D}(N(0, \sigma)) + o(1)$ for every fixed $\sigma > 0$ as $a \to \infty$.*

Proof. Let ν_a denote the conditional distribution of X when $|X| < a/2$ and let ν'_a denote the conditional distribution when $|X| \geq a/2$. We have that

$$\mathbb{E}_\nu(\mathbb{H}(t_{\sigma,a}(x))) = \mathbb{P}(|X| < a/2)\mathbb{E}_{\nu_a}(\mathbb{H}(t_{\sigma,a}(x))) + \mathbb{P}(|X| \geq a/2)\mathbb{E}_{\nu'_a}(\mathbb{H}(t_{\sigma,a}(x))).$$

By Chebyshev's inequality we obtain that $\mathbb{P}(|X| \geq a/2) = O(a^{-2})$. It follows from the trivial uniform bound $\mathbb{H}(t_{\sigma,a}(x)) \leq \log(a^2 + 1)$ that $\mathbb{P}(|X| \geq a/2)\mathbb{E}_{\nu'_a}(\mathbb{H}(t_{\sigma,a}(x))) = o(1)$. Similarly by $\mathbb{P}(|X| < a/2) = 1 - O(a^{-2})$ we obtain that

$$\mathbb{P}(|X| < a/2)\mathbb{E}_\nu(\mathbb{H}(t_{\sigma,a}(x))) = \mathbb{E}_{\nu_a}(\mathbb{H}(t_{\sigma,a}(x))) + o(1).$$

It is now enough to prove that if $|x| \leq a/2$ then $\mathbb{H}(t_{\sigma,a}(x)) = \log a + \mathbb{D}(N(0, \sigma)) + o(1)$, where the $o(1)$ error term does not depend on x but tends to 0 as $a \to \infty$.

Lemma 8.1 implies that $\mathbb{H}(t_{\sigma,a}(0)) = \log a + \mathbb{D}(N(0, \sigma)) + o(1)$; this is the $X = 0$ case. Next, suppose that $x \in S_a = \{r/a | r \in \mathbb{Z}, \|r\|_\infty \leq a^2\}$ and $0 < x \leq a/2$. Notice that if $y \in S_a$ and $-a < y < a - x$, then $\mathbb{P}(t_{\sigma,a}(0) = y) = \mathbb{P}(t_{\sigma,a}(x) = y + x)$, because the distance of 0 and x is a multiple

of the distance of the points in the grid S_a that we used for discretization. Hence the difference $\mathbb{H}(t_{\sigma,a}(x)) - \mathbb{H}(t_{\sigma,a}(0))$ contains only terms corresponding to $|y - x| > a/2, y \in S_a$ in the first entropy expression (for x) and $|y| > a/2, y \in S_a$ in the second one (for 0). The facts that $t_{\sigma,a}$ is supported on a set of at most $a^2 + 1$ elements and that the probability that a Gaussian distribution is farther from its expectation than $a/2$ is $O(\exp(-a^2)/2)$ imply that $\mathbb{H}(t_{\sigma,a}(x)) - \mathbb{H}(t_{\sigma,a}(0)) = o(1)$ uniformly in $0 < x < a/2$ as $a \to \infty$, when $x \in S_a$. A similar argument works for $-a/2 < x < 0$ if x is an element of S_a.

Finally, let $x \in [-a/2, a/2]$ arbitrary, and $\overline{x}$ be the closest element of S_a to x. As it is well known (e.g. as a consequence of Pinsker's inequality), the total variation distance of $N(x, \sigma)$ and $N(\overline{x}, \sigma)$ is of order $O(1/a)$ provided $|x - \overline{x}| \leq 1/a$. By choosing an appropriate coupling of these two distributions and applying the same discretization, it follows that $d_{\mathrm{TV}}(t_{\sigma,a}(x), t_{\sigma,a}(\overline{x})) = O(1/a)$. Applying Theorem 17.3.3. of [17] we obtain that

$$|\mathbb{H}(t_{\sigma,a}(x)) - \mathbb{H}(t_{\sigma,a}(\overline{x}))| \leq -d_{\mathrm{TV}}(t_{\sigma,a}(x), t_{\sigma,a}(\overline{x})) \log \frac{d_{\mathrm{TV}}(t_{\sigma,a}(x), t_{\sigma,a}(\overline{x}))}{a^2 + 1} = o(1),$$

and the error term does not depend on x. This concludes the proof. $\qquad\square$

We finish this chapter with the proof of Theorem 5. Let $\mu_{\sigma,a}$ denote the process in which we pointwise discretize μ using $t_{\sigma,a}$ (using the notation of Chapter 5, we have $\mu_{\sigma,a} = t_{\sigma,a} * \mu$) and let μ_σ denote the process obtained from μ by adding σ times the i.i.d normal distribution. By Lemma 8.1 and $|B_k(C)| - (d/2)|B_k(e)| = 1$ we obtain that

$$\mathbb{H}(B_k(C), \mu_{\sigma,a}) - \frac{d}{2}\mathbb{H}(B_k(e), \mu_{\sigma,a}) = \log a + \mathbb{D}(B_k(C), \mu_\sigma) - \frac{d}{2}\mathbb{D}(B_k(e), \mu_\sigma) + o(1).$$

By Theorem 4 and the typicality of μ we get that

$$\mathbb{H}(B_k(C), \mu_{\sigma,a}) - \frac{d}{2}\mathbb{H}(B_k(e), \mu_{\sigma,a}) \geq \mathbb{E}_{\mu_o}(\mathbb{H}(t_{\sigma,a}(x))).$$

Using the previous formulas, Lemma 8.2 and the limit $a \to \infty$ we obtain that

$$\mathbb{D}(B_k(C), \mu_\sigma) - \frac{d}{2}\mathbb{D}(B_k(e), \mu_\sigma) \geq D(N(0, \sigma)) \tag{16}$$

for every $\sigma > 0$.

Let $S \subset V_d$ be either $B_k(C)$ or $B_k(e)$. We denote by $\mu_{S,\sigma}$ the probability measure obtained by convolving the measure μ_S with the standard normal distribution on $W_\lambda(S)$ (for the definition of $W_\lambda(S)$ see chapter 4), where λ is the eigenvalue corresponding to μ. Observe that the standard normal distribution on $\mathbb{R}^S$ is the independent sum of the standard normal distribution on $W_\lambda(S)$ and on $W_\lambda(S)^\perp$. Then by using that $\dim W_\lambda(S)^\perp = |S| - |\partial S|$, we have

$$\mathbb{D}(S, \mu_\sigma) = \mathbb{D}_{\mathrm{sp}}(\mu_{S,\sigma}) + (|S| - |\partial S|)\mathbb{D}(N(0, \sigma)).$$

Using this formula for $S = B_k(C)$ and $S = B_k(e)$ in (16) together with $|\partial B_k(C)| = (d/2)|\partial B_k(e)|$, we obtain that

$$\mathbb{D}_{\mathrm{sp}}(\mu_{B_k(C),\sigma}) - \frac{d}{2}\mathbb{D}_{\mathrm{sp}}(\mu_{B_k(e),\sigma}) \geq 0.$$

If $\sigma \to 0$, then we obtain the statement of Theorem 5. $\qquad\qquad\qquad\qquad\square$

9 Eigenvalues of the covariance operator

The main goal of this section is to prove the following statement on the differential entropies of Gaussian waves. Recall the definition of $\mathbb{D}_{\mathrm{sp}}$ from Chapter 7. In addition, by $\det_{\mathrm{sp}}$ we mean the product of the non-zero eigenvalues of a matrix.

Theorem 6 *Let Ψ_λ be a Gaussian wave function on the d-regular tree with $\lambda \in [-2\sqrt{d-1}, 2\sqrt{d-1}]$. Then we have*
$$\mathbb{D}_{\mathrm{sp}}(B_k(C), \Psi_\lambda) - \frac{d}{2}\mathbb{D}_{\mathrm{sp}}(B_k(e), \Psi_\lambda) \to 0 \qquad (k \to \infty).$$

Let Σ_k be the covariance matrix of the joint distribution of Ψ_λ restricted to the ball $B_k(C)$, and Σ'_k be the similar covariance matrix on $B_k(e)$. The differential entropy of a multivariate normal random variable with covariance matrix Σ of rank m is given by $\frac{1}{2}\log\left((2\pi e)^m \det_{\mathrm{sp}}\Sigma\right)$ if we measure differential entropy inside the support of the variable. Equation (4) and Corollary 4.1 imply that the rank of Σ_k is $d/2$ times the rank of Σ'_k. Hence we need to prove that

$$\log\det_{\mathrm{sp}}\Sigma_k - \frac{d}{2}\log\det_{\mathrm{sp}}\Sigma'_k \to 0 \qquad (k \to \infty). \tag{17}$$

Notice that if s is an eigenvalue of both Σ_k and Σ'_k, and its multiplicity in the first case is $d/2$ times its multiplicity in the second case, then it is canceled out in the difference. In order to find the eigenvalues that do not cancel out, we decompose both $\mathbb{R}^{|B_k(C)|}$ and $\mathbb{R}^{|B_k(e)|}$ as a union of orthogonal subspaces that are invariant under the corresponding covariance operators.

First we need some notation. We will use the genealogical labeling of the vertices in $B_k(C)$ and in $B_k(e)$ (in this section, we will not distinguish vertices and labels). In $B_k(C)$, the root gets label $\emptyset$, and we put the labels on the vertices such that the labels of neighbors differ only in the last coordinate (i.e. $1, 2, \ldots, d$ are the neighbors of the root; $11, 12, \ldots, 1(d-1)$ are the further neighbors of 1, and so on). For a vertex v the length of its label is denoted by $|v|$, which is its distance from $\emptyset$. For a vertex v and a sequence y, by vy we mean the vertex with the label obtained by concatenating v and y. We say that v is an ancestor of w (denoted by $v \to w$), if $w = vy$ for some $y \neq \emptyset$. As for $B_k(e)$, we use a similar notation, but keeping track of symmetry with respect to the central edge e. The endpoints of e have labels $\emptyset$ and $\emptyset'$. The descendants of $\emptyset$ have labels $1, \ldots, d-1$, their descendants have labels $11, 12, \ldots, 1(d-1), 21, \ldots$ and so on. Similarly, the descendants of $\emptyset'$ have labels $1', \ldots, (d-1)'$, their descendants have labels $11', 12', \ldots, 1(d-1)', 21', \ldots$ and so on. Note that $|v|$ still denotes the length of the label.

We assign a linear subspace to each vertex in $B_k(C) \setminus \partial B_k(C)$ and $B_k(e) \setminus \partial B_k(e)$. Fix $v \in B_k(C) \setminus \partial B_k(C)$. Let $\mathcal{E}_v$ be the elements $\alpha \in \mathbb{R}^{B_k(C)}$ for which the following hold. (i)

$\alpha_w = 0$ if v is not an ancestor of w. (ii) Suppose that $1 \leq j \leq (d-1)$ and y, z are labels with $|y| = |z|$. Then $\alpha_{vjy} = \alpha_{vjz}$. (To put it in another way, for descendants of v, the value of α depends only on the first coordinate after v and the distance from v.) (iii) We have $\sum_{y:|y|=r} \alpha_{vy} = 0$ for $r \geq 1$.

In addition, we introduce the following subspace:

$$\mathcal{G} = \{\alpha \in \mathbb{R}^{B_k(C)} : \alpha_v = \alpha_w \text{ if } |v| = |w|\}.$$

We will also refer to $\mathcal{E}'_v$ (when $v \in B_k(e) \setminus \partial B_k(e)$), which are linear subspaces of $\mathbb{R}^{B_k(e)}$ defined similarly. The definition of the complement subspace is somewhat different:

$$\mathcal{G}'_1 = \{\alpha \in \mathbb{R}^{B_k(e)} : \alpha_v = \alpha_w \text{ if } |v| = |w|\}.$$

$$\mathcal{G}'_2 = \{\alpha \in \mathbb{R}^{B_k(e)} : \alpha_v = \alpha_w \text{ if } \emptyset \to v, \emptyset \to w, |v| = |w|; \ \alpha_{v'} = -\alpha_v \text{ if } v = \emptyset \text{ or } \emptyset \to v\}.$$

Lemma 9.1 *The following hold for the linear subspaces defined above.*

(a) $\mathcal{E}_v, v \in B_k(C) \setminus \partial B_k(C)$ and $\mathcal{G}$ are invariant under Σ_k. Similarly, $\mathcal{E}'_v, v \in B_k(e) \setminus \partial B_k(e)$ and $\mathcal{G}'_1, \mathcal{G}'_2$ are invariant under Σ'_k.

(b) $\mathcal{E}_v, v \in B_k(C) \setminus \partial B_k(C)$ and $\mathcal{G}$ are pairwise orthogonal. Similarly, $\mathcal{E}'_v, v \in B_k(e) \setminus \partial B_k(e)$ and $\mathcal{G}'_1, \mathcal{G}'_2$ are pairwise orthogonal.

(c) $\mathbb{R}^{B_k(C)} = \mathcal{G} + \sum_{v \in B_k(C)} \mathcal{E}_v$ and $\mathbb{R}^{B_k(e)} = \mathcal{G}'_1 + \mathcal{G}'_2 + \sum_{v \in B_k(e)} \mathcal{E}'_v$.

Proof. Before going into the proof, we note that we will only use the property that every entry (i, j) of Σ_k depends only on the distance of i and j. (a) First observe $\mathcal{G}$ consists of all vectors that are invariant under the full automorphism group of $B_k(C)$. This property is preserved by Σ_k and thus $\mathcal{G}$ is invariant under Σ_k.

Take any $\alpha \in \mathcal{E}_v$. For $\Sigma_k \alpha$, property (ii) is preserved because the entries of Σ_k depend only on the distance of the two corresponding vertices (as it is the covariance matrix of an invariant random process). Putting this together with the third property we get that (i) also holds for $\Sigma_k \alpha$. Property (iii) means orthogonality to $\mathcal{G}$; using the invariance of $\mathcal{G}$, this will also be satisfied by $\Sigma_k \alpha$, which is thus in $\mathcal{E}_v$. Similar arguments work for the other two linear subspaces.

(b) Fix $v_1, v_2 \in B_k(C)$. If none of them is ancestor of the other one, then the support of any vector of $\mathcal{E}_{v_1}$ is disjoint from the support of any vector in $\mathcal{E}_{v_2}$, which implies orthogonality. If $v_1 \to v_2$, then the value of a vector in $\mathcal{E}_{v_1}$ is the same at all vertices of type $v_2 y$ with $|y|$ fixed. Multiplying this by the values of a vector in $\mathcal{E}_{v_2}$ and summing this up for different ys (of fixed length) we get 0, because of property (iii). This implies the orthogonality. The other cases are similar; we omit the details.

(c) The dimensions of these subspaces are as follows.

$$\dim \mathcal{E}_v = (k - |v| + 1)(d - 2) \qquad \text{for } \emptyset \neq v \in B_k(C) \setminus \partial B_k(C);$$

$$\dim \mathcal{E}_\emptyset = (k+1)(d-1); \qquad \dim \mathcal{G} = k+2;$$

$$\dim \mathcal{E}'_v = (k-|v|+1)(d-2) \qquad \text{for } \emptyset \neq v \in B_k(e) \setminus \partial B_k(e);$$

$$\dim \mathcal{G}'_1 = k+1; \qquad \dim \mathcal{G}'_2 = k+1.$$

The following equalities are easy to check by induction on k:

$$1 + (k+1)d + \sum_{j=1}^{k} d(d-1)^{j-1}(k+1-j)(d-2) = |B_k(C)|;$$

$$2(k+1) + 2\sum_{j=1}^{k-1}(d-1)^{j-1}(k-j)(d-2) = |B_k(e)|.$$

Hence the sum of the dimension of the linear subspaces $\mathcal{E}_v$ and $\mathcal{G}$ is equal to the dimension of the space $\mathbb{R}^{B_k(C)}$. Since the subspaces are pairwise orthogonal by part (b) of the lemma, this implies that the sum must be equal to $\mathbb{R}^{B_k(C)}$. A similar argument works for $B_k(e)$. $\qquad\square$

For the following lemma, recall the definition of $f(k, x)$ from equation (2). Furthermore, the calculation about this recurrence relation in [1] imply that if we take $\lambda = 2\sqrt{d-1}x$, then the covariance of the values at distance k in the Gaussian wave Ψ_λ is equal to $f(k, x)$.

Lemma 9.2 *Let $f(k, x)$ be defined by equation (2). We define*

$$l(k, x) = 1 + \sum_{j=1}^{k-1}(d-2)(d-1)^{j-1}f(2j, x).$$

Then the eigenvalues of Σ_k corresponding to $\mathcal{E}_\emptyset$ and $\mathcal{G}$ are as follows.

$$s_1(k, x) = 1 + \sum_{j=1}^{k}(d-1)^j f(2j, x) + \sum_{j=1}^{k} l(j, x) \quad \text{with multiplicity } 1;$$

$$s_2(k, x) = \left(1 - s_1(k, x) + \sum_{j=1}^{k} dl(j, x)\right)/(d-1) \quad \text{with multiplicity } d-1.$$

The eigenvalues of Σ'_k corresponding to $\mathcal{G}'_1$ and $\mathcal{G}'_2$ are as follows.

$$s_3(k, x) = \sum_{j=1}^{k} l(j, x) + (d-1)^{j-1}f(2j-1, x) \quad \text{with multiplicity } 1;$$

$$s_4(k, x) = \sum_{j=1}^{k} l(j, x) - (d-1)^{j-1}f(2j-1, x) \quad \text{with multiplicity } 1.$$

Proof. Using the notation from Chapter 4, let $\mathcal{S}_k = W_\lambda(B_k(C))$ and $\mathcal{S}'_k = W_\lambda(B_k(e))$. First we consider $\mathcal{G}$. Since $f(k, x)$ is the covariance of the values at distance k in a (nontrivial) Gaussian wave, we have by linearity that every row of Σ_k is in $\mathcal{S}_k$. It follows that $\text{Im}(\Sigma_k) \subseteq \mathcal{S}_k$. On the other hand, by the previous lemma, $\mathcal{G}$ is invariant under Σ_k. Notice that $\mathcal{G} \cap \mathcal{S}_k$ is one dimensional: given the value at the root, the common value of its neighbors is determined (even for $\lambda = 0$), and this

23

can be continued. It follows that $\Sigma_k|_{\mathcal{G}}$ has rank one, and the eigenvalue corresponding to $\mathcal{G}$ can be obtained by calculating the trace of the matrix of $\Sigma_k|_{\mathcal{G}}$ in an arbitrary basis. By choosing the basis of the indicator functions of the spheres of radius $0, 1, \ldots, k$ around the root, elementary calculation shows that this eigenvalue is equal to $s_1(k, x)$.

The three eigenvalues corresponding to the invariant subspaces $\mathcal{E}_\emptyset$, $\mathcal{G}'_1$ and $\mathcal{G}'_2$ can be obtained by similar arguments, by identifying the image of Σ_k (or Σ'_k) restricted to the given subspace and calculating the trace of its matrix. In the second case, $\mathcal{E}'_\emptyset \cap \mathcal{S}_k$ has dimension $d-1$: 0 is assigned to the root; the values of the d neighbors of $\emptyset$ have to sum up to 0, but there are no other conditions; given these values, all the others are uniquely determined. The only nonzero eigenvalue has multiplicity $d-1$, which makes it possible to calculate it based on the trace of the matrix. As for the last two cases, $\mathcal{G}'_1 \cap \mathcal{S}'_k$ and $\mathcal{G}'_2 \cap \mathcal{S}'_k$ both have dimension 1 again: given the value at $\emptyset$, the value at $\emptyset'$ has to be the same or the opposite. Then the eigenvalue equation and the equality conditions in $\mathcal{G}'_1$ and $\mathcal{G}'_2$ uniquely determine all the other values. By choosing appropriate bases in these subspaces, it is straightforward to obtain the eigenvalues in the lemma. $\qquad\square$

Lemma 9.3 *Using the notation of the previous lemma, for every $x \in [-1, 1]$, we have*

$$\log s_1(k, x) + (d-1)\log s_2(k, x) - \frac{d}{2}\log s_3(k, x) - \frac{d}{2}\log s_4(k, x) \to 0 \qquad (k \to \infty).$$

Proof. First we calculate the middle term of $s_1(k, x)$. Using equation (2), we obtain that

$$T(k, x) := \sum_{j=1}^{k} (d-1)^j f(2j, x) = \sum_{j=1}^{k} \frac{(d-1)^j}{\sqrt{d(d-1)^{2j-1}}} q_{2j}(x),$$

where the polynomials q are defined by equation (3). Straightforward calculation shows that with $x = \cos\vartheta$ we have

$$T(k, x) = \frac{d-1}{d} U_{2k}(x) + \frac{d-2}{d\sin\vartheta} \mathrm{Im}\frac{e^{3i\vartheta}(e^{(2k-2)i\vartheta} - 1)}{e^{2i\vartheta} - 1} - \frac{1}{d},$$

if $\sin\vartheta \neq 0$ and $e^{2\vartheta} \neq 1$. (We deal with the exceptional cases at the end of the proof.) Using this formula, we obtain that

$$l(k, x) = 1 - \frac{d-2}{(d-1)d} + \frac{d-2}{d} U_{2k-2}(x) + \frac{(d-2)^2}{d(d-1)\sin\vartheta} \mathrm{Im}\frac{e^{3i\vartheta}(e^{(2k-4)i\vartheta} - 1)}{e^{2i\vartheta} - 1} \qquad (k \geq 2).$$

Notice that the last term is bounded in k for every fixed x. From now on, $O(1)$ will denote a quantity which depends both on x and k such that *for every fixed x it is bounded in k*. We emphasize that in Theorem 6 the limit is taken for fixed λ (which is equal to $x \cdot \sqrt{d-1}$). Thus in the proofs of this chapter we always think of x as a fixed quantity while tending to infinity with k.

Continuing our calculations, we obtain that

$$s_1(k, x) = 1 + T(k, x) + \sum_{j=1}^{k} l(j, x) = \left(1 - \frac{d-2}{d(d-1)} - \frac{(d-2)^2}{d(d-1)\sin\vartheta} \mathrm{Im}\frac{e^{3i\vartheta}}{e^{2i\vartheta} - 1}\right) k + O(1).$$

24

On the other hand, we have

$$s_1(k,x) - s_2(k,x) = \frac{d}{d-1} s_1(k,x) - \frac{d}{d-1} \sum_{j=1}^{k} l(j,x) - \frac{1}{d-1} = O(1);$$

$$s_1(k,x) - s_3(k,x) = 1 + \sum_{j=1}^{k} (d-1)^j f(2j,x) - (d-1)^j f(2j-1,x) = O(1);$$

$$s_3(k,x) - s_4(k,x) = 2 \sum_{j=1}^{k} (d-1)^j f(2j-1,x) = O(1).$$

To put it in another way, with

$$A = 1 - \frac{d-2}{d(d-1)} - \frac{(d-2)^2}{d(d-1)\sin\vartheta} \operatorname{Im} \frac{e^{3i\vartheta}}{e^{2i\vartheta}-1}$$

the expressions $s_1(k,x)$, $s_2(k,x)$, $s_3(k,x)$ and $s_4(k,x)$ are all in the form $Ak + O(1)$. If $A > 0$, then we get that

$$\log s_j(k,x) = \log A + \log k + o(1) \qquad (j=1,2,3,4),$$

which implies the statement of the lemma. Hence, in the rest of the proof, we check that $A > 0$ holds.

First notice that $1 - \frac{d-2}{d(d-1)} > \frac{(d-2)^2}{d(d-1)}$ is satisfied for all d. In addition, by using elementary trigonometric identities we obtain

$$\operatorname{Im} \frac{e^{3i\vartheta}}{\sin\vartheta(e^{2i\vartheta}-1)} = \frac{\sin\vartheta - \sin 3\vartheta}{2\sin\vartheta(1-\cos 2\vartheta)} = \frac{1 - \cos 2\vartheta - 2\cos^2\vartheta}{4\sin^2\vartheta} = \frac{\sin^2\vartheta - \cos^2\vartheta}{2\sin^2\vartheta} \le 1.$$

Putting this together, we get the positivity of A, which concludes the proof.

We have to deal with the remaining special cases. First, if $|x| = 1$, then $\sin\vartheta = 0$, but we still have $U_k(\cos\vartheta) = k+1$. This implies that $l(k,x)$ is a quadratic polynomial, and $s_1(k,x)$ is a cubic polynomial (with leading coefficient $1 - (d-2)^2/d/(d-1) > 0$). Moreover, the differences $s_1 - s_2, s_1 - s_3, s_1 - s_4$ are all of order $O(k^2)$, which implies the statement of the lemma.

The last case is when $e^{2i\vartheta} = 1$. This implies $U_k(\cos\vartheta) = 1$ for all k. That is, $q_k(x) = (d-2)/\sqrt{d(d-1)}$, and $l(k,x)$ is of the form $Bk + O(1)$ for some nonzero B. Now s_1, s_2, s_3, s_4 are all quadratic polynomials as a function of k, while their differences are linear. It follows again that the expression in the lemma goes to 0 as $k \to \infty$. $\qquad\square$

Proof of Theorem 6. As we have discussed, it is sufficient to show that (17) holds. First fix k, and recall Lemma 9.1. Notice that for every $1 \le r \le k-1$, if we take two vertices v_1, v_2 in $B_k(C)$ such that $|v_1| = |v_2| = r$, then the linear transformation Σ_k restricted to $\mathcal{E}_{v_1}$ is isomorphic to the linear transformation Σ_k restricted to $\mathcal{E}_{v_2}$ (we use again that the entries of the covariance matrix depend only on the distance of the vertices). Hence the set of eigenvalues of Σ_k corresponding to the invariant subspaces $\mathcal{E}_{v_1}$ and $\mathcal{E}_{v_2}$ are the same. Furthermore, this linear transformation is isomorphic to the linear transformation Σ'_k restricted to $\mathcal{E}'_v$, if $|v| = r$ holds. For every $1 \le r \le k-1$ we have

$$\left|\{v : |v| = r, v \in B_k(C)\}\right| = \frac{d}{2}\left|\{v : |v| = r, v \in B_k(e)\}\right|.$$

25

This means that all eigenvalues belonging to the spaces $\mathcal{E}_v$ and $\mathcal{E}'_v$ cancel out in the expression in (17) for $|v| > 0$.

Therefore only the eigenvalues corresponding to $\mathcal{G}, \mathcal{E}'_\emptyset, \mathcal{G}'_1$ and $\mathcal{G}'_2$ are left. These are calculated with multiplicities in Lemma 9.2, and hence Lemma 9.3 finishes the proof by showing that the difference goes to zero as $k \to \infty$. $\qquad\square$

10 Improved differential entropy inequality

In this chapter we use a combination of Theorem 5 and Theorem 6 to prove an improved version of Theorem 5 for the case $k = 0$. We will use the notation from Chapter 4. If S is either $B_k(C)$ or $B_k(e)$ for some $k \geq 0$ and p is in $S \setminus \partial S$, then we have for the Gaussian wave Ψ_λ with distribution μ that

$$\mathbb{D}(\mathcal{Q}(S, p, \mu)) = \mathbb{D}(B_p) + \sum_{(v,w) \in D} \mathbb{D}(A_{v,w}). \tag{18}$$

We obtain from (18) that the differential entropy $\mathbb{D}(\mathcal{Q}(S, p, \mu))$ does not depend on p for a Gaussian wave with distribution μ. Observe that changing the vertex p to p' results in a linear transformation $T_{p,p'}$ in the system $\mathcal{Q}(S, p, \mu)$. The invariance of the differential entropy shows by Lemma 13.1 that $T_{p,p'}$ has determinant 1.

Now applying Lemma 13.1 for an arbitrary smooth eigenvector process ν with eigenvalue λ we obtain that the value of $\mathcal{D}(\mathcal{Q}(S, p, \nu))$ is independent of p since the transformations $T_{p,p'}$ depend only on the quadruple p, p', λ, S and can be calculated from the eigenvector equation. For a general smooth eigenvector process ν we define $\mathbb{D}(S, \nu)$ as this unique differential entropy of $\mathbb{D}(\mathcal{Q}(S, p, \nu))$. We will need the next definition.

Definition 10.1 *A process $\{Y_v\}_{v \in V_d}$ in $I_d(X)$ is called 2-Markov if for an arbitrary edge e the distributions $\{Y_v\}_{v \in W_1}$ and $\{Y_v\}_{v \in W_2}$ are conditionally independent with respect to $\{Y_v\}_{v \in e}$ where W_1 and W_2 are the set of vertices on the two sides of e. (With this notation $V_d = W_1 \cup e \cup W_2$.)*

Note that the 2-Markov property implies that the marginal distribution $Y_C = \{Y_v\}_{v \in C}$ determines the whole process because we can build up the distribution $\{Y_v\}_{v \in V_d}$ using iterated conditionally independent couplings of Y_C along edges. More precisely, if for some connected subgraph K of T_d the distribution $\{Y_v\}_{v \in V(K)}$ is already constructed and $w \in V_d$ is a vertex such that the star $B_1(w)$ intersects K in a single edge e, then the joint distribution $\{Y_v\}_{v \in V(K) \cup B_1(w)}$ is the conditionally independent coupling of $\{Y_v\}_{v \in V(K)}$ and $\{Y_v\}_{v \in B_1(w)}$ with respect to $\{Y_v\}_{v \in e}$. The invariance of the process implies that $\{Y_v\}_{v \in B_1(w)}$ has the same distribution as Y_C. By iterating this procedure we can build up the marginal distribution on any finite connected subgraph of T_d and thus the whole process in uniquely determined.

Lemma 10.1 *We have for every $k \geq 1$ and smooth eigenvector process ν the following three inequalities.*

$$\mathbb{D}(B_k(C), \nu) \leq d\mathbb{D}(B_k(e), \nu) - (d-1)\mathbb{D}(B_{k-1}(C), \nu),$$

$$\mathbb{D}(B_k(e), \nu) \leq 2\mathbb{D}(B_{k-1}(C), \nu) - \mathbb{D}(B_{k-1}(e), \nu),$$

$$\mathbb{D}(B_k(C), \nu) - (d/2)\mathbb{D}(B_k(e), \nu) \leq \mathbb{D}(B_{k-1}(C), \nu) - (d/2)\mathbb{D}(B_{k-1}(e), \nu).$$

If ν is Gaussian, then we have equality everywhere. Furthermore if we have equality everywhere (for every k), then ν is 2-Markov.

Proof. The proof is based on the general fact (see Lemma 13.2) that for a joint distribution (X, Y, Z) we have that $\mathbb{D}(X, Z) + \mathbb{D}(Y, Z) - \mathbb{D}(Z) \geq \mathbb{D}(X, Y, Z)$ holds with equality if and only if X and Y are conditionally independent with respect to Z. To see the first inequality let us place p to the root of $B_k(C)$. We can cover $B_k(C)$ in a rotational symmetric way by d copies of $B_k(e)$ in a way that all of them contain $B_{k-1}(C)$ and they are disjoint outside of $B_{k-1}(C)$. Each copy of $B_k(e)$ covers a subset of the variables $A_{v,w}$ and B_p such that the joint differential entropy of this subset of variables is equal to $\mathbb{D}(B_k(e), \nu)$. Now Lemma 13.2 finishes the proof of the first inequality. The other two inequalities can be seen in a similar way. If we have equality everywhere for every k, then by lemma 13.2 we get that the joint distribution of ν on $B_k(C)$ is the conditionally independent coupling of d copies of $B_k(e)$ over $B_{k-1}(C)$ and the joint distribution on $B_k(e)$ is the conditionally independent coupling of 2 copies of $B_{k-1}(C)$ over $B_{k-1}(e)$. By induction the 2-markov property follows inside $B_k(C)$ for every k and thus for the whole process.

Let $\alpha(S, \nu)$ denote the difference $\mathbb{D}(\mathcal{Q}(S, p, \nu)) - \mathbb{D}(\mathcal{Q}(S, p, \mu))$ where μ is the Gaussian eigenvector process with the same eigenvalue as ν. If we apply the same change of basis to both $\mathbb{D}(\mathcal{Q}(S, p, \nu))$ and $\mathbb{D}(\mathcal{Q}(S, p, \mu))$, they change with the same additive constant by lemma 13.1 and thus $\alpha(S, \nu)$ remains unchanged. This shows the basis independence of $\alpha(S, \nu)$. In particular we have that if ν is a smooth eigenvector process then $\alpha(S, \nu) = \mathbb{D}_{\mathrm{sp}}(S, \nu) - \mathbb{D}_{\mathrm{sp}}(S, \mu)$ and so

$$\alpha(B_k(C), \nu) - (d/2)\alpha(B_k(e), \nu) =$$

$$\left(\mathbb{D}_{\mathrm{sp}}(B_k(C), \nu) - (d/2)\mathbb{D}_{\mathrm{sp}}(B_k(e), \nu)\right) - \left(\mathbb{D}_{\mathrm{sp}}(B_k(C), \mu) - (d/2)\mathbb{D}_{\mathrm{sp}}(B_k(e), \mu)\right). \quad (19)$$

Using Theorem 5 and Theorem 6 we get the following consequence.

Proposition 10.1 *If ν is a smooth typical eigenvector process then*

$$\limsup_{k \to \infty} \alpha(B_k(C), \nu) - (d/2)\alpha(B_k(e), \nu) = \limsup_{k \to \infty} \mathbb{D}_{\mathrm{sp}}(B_k(C), \nu) - (d/2)\mathbb{D}_{\mathrm{sp}}(B_k(e), \nu) \geq 0.$$

Proof. By Theorem 5 the first term in (19) is non negative and by Theorem 6 the second term converges to 0. This completes the proof.

The main theorem of this chapter is the following.

Theorem 7 *If ν is a smooth typical eigenvector process, then $\alpha(C, \nu) - (d/2)\alpha(e, \nu) \geq 0$ and equality implies that ν is 2-Markov.*

Proof. By iterating the third inequality in Lemma 10.1 we obtain that

$$\mathbb{D}(B_k(C), \nu) - (d/2)\mathbb{D}(B_k(e), \nu) \leq \mathbb{D}(B_l(C), \nu) - (d/2)\mathbb{D}(B_l(e), \nu)$$

holds for every $k \geq l \geq 0$. Furthermore if we replace ν in the above formula by the Gaussian wave μ then we get equality. This implies that

$$\alpha(B_k(C), \nu) - (d/2)\alpha(B_k(e), \nu) \leq \alpha(B_l(C), \nu) - (d/2)\alpha(B_l(e), \nu) \tag{20}$$

holds for every $k \geq l \geq 0$. By Proposition 10.1 we get the inequality of the theorem by applying (20) with $l = 0$ and $k \to \infty$. For the second statement assume that $\alpha(C, \nu) - (d/2)\alpha(e, \nu) = 0$. By (20) this is only possible if $\alpha(B_k(C), \nu) - (d/2)\alpha(B_k(e), \nu) = 0$ holds for every k and thus the inequalities of Lemma 10.1 are all equalities. This implies that ν is 2-Markov. $\square$

11 Heat equation and the proof of the main theorem

Definition 11.1 *For $\lambda \in [-2\sqrt{d-1}, 2\sqrt{d-1}]$ and $d \geq 3$ let $\mathcal{F}_{d,\lambda}$ denote the set of joint distributions $F = (X_1, X_2, \ldots, X_d, Z)$ of real valued random variables such that*

1. *$\mathbb{E}(X_i X_j) = (\lambda^2 - d)d^{-1}(d-1)^{-1}$, $\mathbb{E}(X_i Z) = \lambda/d$, $\mathbb{E}(X_i) = \mathbb{E}(Z) = 0$ and $\mathbb{E}(X_i^2) = \mathbb{E}(Z^2) = 1$ holds for $1 \leq i, j \leq d$ and $i \neq j$;*

2. *the joint distribution $(X_1, X_2, \ldots, X_d, Z)$ is symmetric under every permutation that fixes Z;*

3. *for every $1 \leq i \leq d$ the joint distribution (X_i, Z) is the same as the joint distribution (Z, X_i);*

4. *the quantities $\mathbb{D}_{\mathrm{sp}}(X_1, X_2, \ldots, X_d, Z)$ and $\mathbb{D}(X_1, Z)$ are both finite.*

We define the function $\mathcal{D} : \mathcal{F}_{d,\lambda} \to \mathbb{R}$ by $\mathcal{D}(F) = \mathbb{D}_{\mathrm{sp}}(X_1, X_2, \ldots, X_d, Z) - \frac{d}{2}\mathbb{D}(X_1, Z)$.

Notice that the covariance conditions of definition 11.1 guarantee that $\mathbb{E}((X_1 + X_2 + \cdots + X_d - \lambda Z)^2) = 0$ and thus $X_1 + X_2 + \cdots + X_d = \lambda Z$ holds with probability 1. This implies that the joint distribution F is concentrated on the 1 co-dimensional (d-dimensional) subspace $W_\lambda(C)$ in $\mathbb{R}^{d+1}$. The subspace differential entropy in definition 11.1 is measured in this subspace.

In this chapter we think of $\lambda \in [-2\sqrt{d-1}, 2\sqrt{d-1}]$ and d as fixed values and most of the times our notation will not indicate the dependence on these values even if there is such a dependence. Our goal is to solve the extremal problem of maximizing $\mathcal{D}$ inside $\mathcal{F}_{d,\lambda}$ (see Theorem 8). This will provide the last step in the proof of our main theorems (see Theorem 2 and Theorem 3) as we will explain at the end of this chapter.

It will be important that there is a unique element $F^* = (X_1^*, X_2^*, \ldots, X_d^*, Z^*)$ in $\mathcal{F}_{d,\lambda}$ such that F^* is Gaussian. (The covariances define the Gaussian system uniquely, which clearly satisfies

the symmetry conditions.) Note that F^* depends on both d and λ. Most of this chapter deals with the proof of the the next theorem which says that the entropy formula $\mathcal{D}$ in $\mathcal{F}_{d,\lambda}$ is maximized by the Gaussian distribution F^*.

Theorem 8 *For every $F \in \mathcal{F}_{d,\lambda}$ we have that $\mathcal{D}(F) \leq \mathcal{D}(F^*)$ and equality holds if and only if $F = F^*$.*

To get rid of the subspace differential entropy, we apply a change of basis to the systems in $\mathcal{F}_{d,\lambda}$. We can choose a fix linear transformation $T : \mathbb{R}^{d+1} \to \mathbb{R}^d$ (depending on d and λ) such that for every $F \in \mathcal{F}_{d,\lambda}$ the system $T(F) = (B_1, B_2, \ldots, B_d)$ satisfies $\mathbb{E}(B_i B_j) = \delta_{i,j}$ for $1 \leq i, j \leq d$. Using that $M = T(F^*)$ is Gaussian we obtain that M is the standard normal distribution on $\mathbb{R}^d$. Using that linear transformations change differential entropy with a fix constant (depending on the transformation; see Lemma 13.1), the statement of the theorem is equivalent to the fact that

$$\mathbb{D}(T(F)) - (d/2)\mathbb{D}(X_1, Z) \leq \mathbb{D}(M) - (d/2)\mathbb{D}(X_1^*, Z^*)$$

and equality holds if and only if $T(F) = M$. The proof of Theorem 8 relies on the following proposition.

Proposition 11.1 *Let $F = (X_1, X_2, \ldots, X_d, Z) \in \mathcal{F}_{d,\lambda}$ and $F_t = F + \sqrt{2t}F^*$ (using independent sum) for every $t > 0$. Then the function*

$$\Lambda_F(t) = \mathbb{D}(T(F_t)) - (d/2)\mathbb{D}(X_1 + \sqrt{2t}X_1^*, Z + \sqrt{2t}Z^*)$$

satisfies $\Lambda_F'(0) \geq 0$. If F is not Gaussian then $\Lambda_F'(t) > 0$ for some $t \geq 0$.

Note that the choice of F_t comes from the heat equation in $\mathbb{R}^d$ (see chapter 12). We first show that Proposition 11.1 implies Theorem 8. The joint distribution F_t does not satisfy the covariance conditions of definition 11.1 but it is clear that the scaled version $F_t(1 + 2t)^{-1/2}$ is in $\mathcal{F}_{d,\lambda}$. Notice that scaling does not change the differential entropy formula because the extra additive constants coming from scaling exactly cancel each other. By using the claim for $G = F_t(1 + 2t)^{-1/2}$ we obtain from $\Lambda_G'(0) \geq 0$ that $\Lambda_F'(t) \geq 0$ holds for F with every $t \geq 0$. Since $F_t(1 + 2t)^{-1/2} = F(1 + 2t)^{-1/2} + F^*(2t/(1 + 2t))^{1/2}$ converges to F^* as $t \to \infty$, we obtain that F^* maximizes $\mathcal{D}$. To see that only the Gaussian system F^* attains the maximum assume that F attains the maximum. In this case $\Lambda_F'(t) \geq 0$ is only possible if $\Lambda_F'(t) = 0$ holds for every $t \geq 0$. This implies that F is Gaussian by the second part of Proposition 11.1.

It remains to prove Proposition 11.1. We start with some notation and lemmas. Assume that the measure μ is the distribution and f is the density function of $B = T(F)$. We can choose a matrix $Q \in \mathbb{R}^{(d+1)\times d}$ such that $X_i = \sum_k Q_{i,k} B_k$ holds for $i = 1, \ldots, d$, and $Z = \sum_k Q_{d+1,k} B_k$ is also satisfied. We define $v_i = (Q_{i,1}, \ldots, Q_{i,d})$ for $i = 1, \ldots, d$ and $w = (Q_{d+1,1}, \ldots, Q_{d+1,d})$. Notice

that the covariance matrix of F is QQ^T. Hence the vectors $\{v_i\}_{i=1}^d$ and w are unit vectors such that $(v_i, v_j) = (\lambda^2 - d)d^{-1}(d-1)^{-1}$ and $(v_i, w) = \lambda/d$ for every $1 \leq i, j \leq d$. The joint distribution $((v_1, B), (v_2, B), \ldots, (v_d, B), (w, B))$ is the same as F. We have that μ is invariant with respect to every orthogonal transformation permuting the system $\{v_i\}_{i=1}^d$ and fixing w. Furthermore we have that μ projected to the space spanned by v_i and w is invariant with respect to the reflection interchanging v_i and w. We can choose two real numbers α and β such that the vectors $a_i = \alpha w + \beta v_i$ and $b_i = \alpha v_i + \beta w$ satisfy $(a_i, b_i) = 0$ and $\|a_i\|_2 = \|b_i\|_2 = 1$ for every $1 \leq i \leq d$ (in particular, we need that $(\alpha^2 + \beta^2)\lambda/d + 2\alpha\beta = 0$, and $(\alpha^2 + \beta^2) + 2\alpha\beta\lambda/d = 1$). The choice of α and β is unique up to multiplying both by -1 or switching them. Their values can be determined using elementary geometry.

Note that construction of the vector system $\{v_i\}_{i=1}^d$, w, $\{a_i\}_{i=1}^d$, $\{b_i\}_{i=1}^d$ in $\mathbb{R}^d$ is purely linear algebraic. Such system, with the scalar products given above, can be constructed for an arbitrary $|\lambda| \leq d$ however in the case of $|\lambda| \leq 2\sqrt{d-1}$ they satisfy a useful geometric property expressed in the following lemma.

Lemma 11.1 *If $|\lambda| \leq 2\sqrt{d-1}$, then there are numbers t_1, t_2 with $t_1, t_2 \geq 0$ and $t_1 + t_2 = 1$ such that for every $u \in \mathbb{R}^d$ we have*

$$\|u\|_2^2 = \sum_{i=1}^d \left(t_1 (u, a_i)^2 + t_2 (u, b_i)^2 \right).$$

Proof. The proof follows from two observations. The first one is the following. Let $\{v_i'\}_{i=1}^d$ be a system of unit vectors such that all pairwise scalar products are equal and for all $1 \leq i \leq d$ we have $(v_i', w) = c$ for some $c \leq 1$. (The system $\{a_i\}_{i=1}^d$ satisfies the conditions with $c = \alpha + \beta\lambda/d$, and $\{b_i\}_{i=1}^d$ with $c = \alpha\lambda/d + \beta$.) Then

$$\sum_{i=1}^d (u, v_i')^2 = (1 - c^2)d(d-1)^{-1}\|u - (u, w)w\|_2^2 + c^2 d(u, w)^2 \tag{21}$$

holds for every $u \in \mathbb{R}^d$. To see this, first notice that $\sum_i v_i' = cdw$, which implies that $(v_i', v_j') = (c^2 d - 1)/(d - 1)$ for $1 \leq i < j \leq d$. It follows that we can choose $\gamma \in \mathbb{R}$ such that the equality $(v_i' - \gamma w, v_j' - \gamma w) = 0$ holds for $1 \leq i < j \leq d$. On the other hand, we have $\sum_i (u - (u, w)w, v_i') = 0$. Therefore

$$\sum_{i=1}^d (u, v_i')^2 = \sum_{i=1}^d \left((u - (u, w)w, v_i')^2 + (u, w)^2 (w, v_i')^2 \right).$$

The second term is equal to the second term of (21). In the first term, we can replace v_i' with $v_i' - \gamma w$. If the latter is equal to zero, then we are done. Otherwise, since $\{v_i' - \gamma w\}_{i=1}^d$ is an orthogonal basis in $w^\perp$, we obtain

$$\sum_{i=1}^d (u, v_i')^2 = \|v_i' - \gamma w\|_2^2 \|u - (u, w)w\|_2^2 + c^2 d(u, w)^2.$$

On the other hand, for symmetry reasons, $\|v_i' - \gamma w\|_2^2$ does not depend on i. By substituting $u = v_1'$ and using $(v_1', v_j') = (c^2 d - 1)/(d - 1)$ again, we get that the value of this constant is the same as in equation (21).

The second observation says that if $|\lambda| \leq 2\sqrt{d-1}$ then there exist constants $t_1, t_2 \geq 0$ with $t_1 + t_2 = 1$ such that $t_1(a_i, w)^2 + t_2(b_i, w)^2 = 1/d$. First of all note the symmetries of the vector system imply that $(a_i, w)^2, (b_i, w)^2$ are independent from i. Elementary calculation shows that the two values $(a_i, w)^2$ and $(b_i, w)^2$ are equal to $(1 \pm \sqrt{1 - (\lambda/d)^2})/2$. This shows the existence of the constants t_1, t_2. We get the statement of the lemma by taking the convex combination of (21) applied for $\{a_i\}_{i=1}^d$ and $\{b_i\}_{i=1}^d$ with coefficients t_1 and t_2. $\qquad\square$

Now we return to the proof of Proposition 11.1. For $1 \leq i \leq d$ let f_i denote the orthogonal projection of f to the two dimensional space $V_i = \langle w, v_i \rangle_\mathbb{R}$. This means that

$$f_i(x) = \int_{z \in V_i^\perp} f(x + z)$$

for $x \in V_i$. Let $T_2 : \mathbb{R}^2 \to \mathbb{R}^2$ denote the linear transformation $T_2(x, y) = (\alpha x + \beta y, \alpha y + \beta x)$ with α and β defined above. We have that f_i (when written in the orthonormal basis a_i, b_i) is the density function of $T_2(X_i, Z)$. We can write $\Lambda_F(t)$ as

$$k + \mathbb{D}(T(F_t)) - (d/2)\mathbb{D}(T_2(X_1 + \sqrt{2t}X_1^*, Z + \sqrt{2t}Z^*)) \tag{22}$$

where the constant k comes from the change of basis T_2. Then by the de Bruijn identity (see equation (27) and Lemma 12.2) we get

$$\Lambda_F'(0) = \int_{\mathbb{R}^d} \|\nabla f\|_2^2 / f - (d/2) \int_{V_1} \|\nabla f_1\|_2^2 / f_1.$$

From Lemma 11.1 we have that

$$\|\nabla f\|_2^2 = \sum_{i=1}^d (t_1(\partial_{a_i} f)^2 + t_2(\partial_{b_i} f)^2) \tag{23}$$

holds for some $t_1, t_2 \geq 0$ such that $t_1 + t_2 = 1$. Using

$$\|\nabla f_1\|_2^2 = (\partial_{a_1} f_1)^2 + (\partial_{b_1} f_1)^2$$

and the above equations it follows that

$$\Lambda_F'(0) = \sum_{i=1}^d \int_{\mathbb{R}^d} (t_1(\partial_{a_i} f)^2 + t_2(\partial_{b_i} f)^2)/f - (d/2) \int_{V_1} ((\partial_{a_1} f_1)^2 + (\partial_{b_1} f_1)^2)/f_1.$$

By the symmetries of f we have that the terms in the above sum are all the same and thus

$$\Lambda_F'(0) = d \int_{\mathbb{R}^d} (t_1(\partial_{a_1} f)^2 + t_2(\partial_{b_1} f)^2)/f - (d/2) \int_{V_1} ((\partial_{a_1} f_1)^2 + (\partial_{b_1} f_1)^2)/f_1. \tag{24}$$

31

Lemma 11.2 *If $u \in V_1$, then*

$$\int_{\mathbb{R}^d} (\partial_u f)^2 / f \geq \int_{V_1} (\partial_u f_1)^2 / f_1.$$

Equality holds if and only if the function $g = \partial_u f(x)/f(x) = \partial_u \log f$ satisfies $g(x) = g(x + z)$ for every pair $x \in \mathbb{R}^d$ and $z \in V_1^{\perp}$.

Proof.

$$\int_{\mathbb{R}^d} (\partial_u f)^2 / f = \int_{x \in V_1} f_1(x) \int_{z \in V_1^{\perp}} (f(x+z)/f_1(x))(\partial_u f(x+z)/f(x+z))^2$$

$$\geq \int_{x \in V_1} f_1(x) \left(\int_{z \in V_1^{\perp}} (f(x+z)/f_1(x))(\partial_u f(x+z)/f(x+z)) \right)^2$$

$$= \int_{x \in V_1} f_1(x) \left(\int_{z \in V_1^{\perp}} (\partial_u f(x+z)/f_1(x)) \right)^2 = \int_{V_1} (\partial_u f_1)^2 / f_1.$$

To see the inequality in the above calculation notice that $f(x + z)/f_1(x)$ is the density function of a probability measure on $x + V_1$. We can apply the Cauchy–Schwarz inequality using this density function to get the inequality. It also shows that equality holds in the statement of the lemma if and only if $\partial_u f(x+z)/f(x+z)$ is constant almost everywhere on $x + V_1$ for almost every x. Since we work with continuous functions the almost can be omitted. $\qquad\square$

We apply Lemma 11.2 for a_1 and b_1 and (24) to obtain that

$$\Lambda'_F(0) \geq (dt_1 - d/2) \int_{V_1} (\partial_{a_1} f_1)^2 / f_1 + (dt_2 - d/2) \int_{V_1} (\partial_{b_1} f_1)^2 / f_1. \tag{25}$$

Using the symmetry of (X_1, Z) we obtain that

$$\int_{V_1} (\partial_{a_1} f_1)^2 / f_1 = \int_{V_1} (\partial_{b_1} f_1)^2 / f_1. \tag{26}$$

It follows that $\Lambda'_F(0) \geq 0$. The proof of the first part of Proposition 11.1 is now complete.

We arrived to the second part of Proposition 11.1. Assume that F satisfies $\Lambda'_F(t) = 0$ for every $t \geq 0$. Using the notation from Lemma 11.1 we have by $t_1 + t_2 = 1$ that at least one of $t_1 > 0$ and $t_2 > 0$ holds. Without loss of generality we assume that $t_1 > 0$. Let $g_t : \mathbb{R}^d \to \mathbb{R}$ denote the logarithm of the density function of $B_t = T(F_t)$. This implies by Lemma 11.2 that $\partial_{a_1} g_t$ satisfies the property that $\partial_{a_1} g_t(x) = \partial_{a_1} g_t(x + z)$ holds whenever $x \in \mathbb{R}^d$ and $z \in V_1^{\perp}$. For convenience, we will write every element $x \in \mathbb{R}^d$ as a triple $(\alpha(x), \beta(x), \gamma(x))$, where $\alpha(x) = (x, a_1)$, $\beta(x) = (x, b_1)$ and $\gamma(x)$ is the projection of x to $V_1^{\perp}$. Using this notation, we have that $\partial_{a_1} g_t(x) = h_t(\alpha(x), \beta(x))$. This means that there exists a function $\widehat{h}_t : \mathbb{R}^2 \to \mathbb{R}$ such that $\partial_{a_1} \widehat{h}_t(\alpha(x), \beta(x)) = \partial_{a_1} g_t(x)$. We have by $\partial_{a_1}(g_t - \widehat{h}_t) = 0$ that $g_t(x) - \widehat{h}_t(\alpha(x), \beta(x)) = s_t(\beta(x), \gamma(x))$ for some function s_t. We obtain that the density function of B_t can be written in the following form.

$$\exp(\widehat{h}_t(\alpha(x), \beta(x))) \cdot \exp(s_t(\beta(x), \gamma(x))).$$

In other words, this means that the random variables of $\alpha(B_t)$ and $\gamma(B_t)$ are conditionally independent with respect to $\beta(B_t)$. This implies by Lemma 12.6 that one of the following two possibilities holds: either $\alpha(B)$ is independent of $(\beta(B), \gamma(B))$ or $\gamma(B)$ is independent of $(\alpha(B), \beta(B))$. In the first case we obtain (using the terminology of Lemma 12.5) that a_1 is an independent direction for B. By symmetries of B we obtain that $\{a_i\}_{i=1}^d$ are all independent directions for B. If $t_1 < 1$ then $(a_i, a_j) \neq 0$ for every pair $1 \leq i < j \leq d$ and Lemma 12.5 finishes the proof. If $t_1 = 1$, then $\{a_i\}_{i=1}^d$ is an orthonormal basis in $\mathbb{R}^d$ and $b_1 = \sum_{i=2}^{d-1} a_i (d-1)^{-1/2}$. We have that (B, a_i) are identically distributed independent random variables and that $(B, b_1) = \sum_{i=2}^{d-1} (B, a_i)(d-1)^{-1/2}$ has the same distribution. This is only possible if this (B, a_i) is normal for every i. We obtain that B is Gaussian.

In the case when $\gamma(B)$ is independent of $(\alpha(B), \beta(B))$ we have that $\partial_u g(x) = \partial_u g(x+z)$ holds for every triple $u \in V_1, x \in \mathbb{R}^d, z \in V_1^\perp$. The symmetries of f imply that $\partial_u g(x) = \partial_u g(x+z)$ holds for every triple $u \in V_i, x \in \mathbb{R}^d, z \in V_i^\perp$. Let r_i denote the orthogonal projection of a_i to $w^\perp$ for $1 \leq i \leq d$. Note that the vector system $\{r_i\}_{i=1}^d$ is completely symmetric in the sense that the origin is the center of a regular simplex whose vertices are given by these vectors. We have for every $1 \leq i \leq d$ that $\partial_{r_i} g(x) = h_i((x, r_i), (x, w))$ for some two variable function $h_i : \mathbb{R}^2 \to \mathbb{R}$. The symmetries of f imply that h_i does not depend on i and thus $h_i = h$ for some h for every i.

The next step is to prove that $h(x, y) = xh^*(y)$ for some one variable function h^*. We have by $\sum_{i=1}^d r_i = 0$ that $\sum_{i=1}^d \partial_{r_i} g(x) = 0$ and thus $\sum_{i=1}^d h_i((x, r_i), (x, w)) = 0$ holds for every $x \in \mathbb{R}^d$. For arbitrary numbers $x_1, x_2, \ldots, x_d, y \in \mathbb{R}$ with $\sum_{i=1}^d x_i = 0$ we can choose $x \in \mathbb{R}^d$ such that $(x, r_i) = x_i$ and $(x, w) = y$. It follows that $\sum_{i=1}^d h(x_i, y) = 0$ holds for arbitrary numbers with $\sum_{i=1}^d x_i = 0$. Assume first that all x_i is 0 then we have that $dh(0, w) = 0$ and thus $h(0, w) = 0$ for every w. Then assume that $x_1 = a, x_2 = -a$ and $x_i = 0$ if $i \geq 3$. We obtain that $h(a, w) + h(-a, w) = 0$ and thus $h(-a, w) = -h(a, w)$ holds for every a and w. Finally let $x_1 = a, x_2 = b, x_3 = -a - b$ and $x_i = 0$ if $i \geq 4$. We obtain that $h(a, w) + h(b, w) = -h(-a-b, w) = h(a+b, w)$. Since h is additive and continuous in the first coordinate a well known fact implies that h is a linear function in the first coordinate and thus we obtain $h(x, y) = xh^*(y)$.

Now we have $\partial_{r_i} g(x) = (x, r_i)h^*((x, w))$ for every $1 \leq i \leq d$. It is easy to see that this implies that $g(x) = \|x - (x, w)w\|_2^2 h^*((x, w)) + c^*((x, w))$ where $c^* : \mathbb{R} \to \mathbb{R}$ is some function. We have that $\partial_w g(x) = \|x - w(x, w)\|_2^2 (h^*)'((x, w)) + (c^*)'((x, w))$. On the other hand we have that $\partial_w g(x) = \partial_w g(x+z)$ holds whenever $z \in V_i^\perp$ for every $1 \leq i \leq d$. It follows that $(h^*)' = 0$ everywhere and so $g(x) = c\|x - (x, w)w\|_2^2 + c^*((x, w))$ with some constant c. Now from $f(x) = \exp(c\|x - (x, w)w\|_2^2 + c^*((x, w)))$ we have that $B - (B, w)w$ and (B, w) are independent random variables. Furthermore $B - (B, w)w$ is a Gaussian distribution concentrated on the orthogonal space of w. This means that the pair $(B, r_1), (B, w)$ of random variables is independent and (B, r_1) is Gaussian. We know that (B, v_1) is a linear combination of (B, r_1) and

33

(B, w) (with a non-zero coefficient for (B, r_1)) and its distribution is the same as the distribution of (B, w) (here we use the symmetries of B). It follows that (B, w) is also Gaussian and thus $c^*((x, w)) = c_2(x, w)^2 + c_3$ for some constants c_2, c_3. Thus we have that B is a Gaussian joint distribution implying that F is also joint Gaussian, as it is a linear function of B. $\square$

Proof of Theorem 2 and Theorem 3 From Proposition 5.1 we have that theorem 3 implies theorem 2. Let μ be a smooth typical eigenvector process corresponding to eigenvalue λ represented by a system of random variables $\{X_v\}_{v \in V_d}$. According to the results in chapter 7 it is enough to show that μ is a Gaussian wave. We have by Lemma 5.1 that $\lambda \in [-2\sqrt{d-1}, 2\sqrt{d-1}]$. Let $F = \{X_v\}_{v \in C}$. It is clear that $F \in \mathcal{F}_{d,\lambda}$. We have by Theorem 7 that $\mathcal{D}(F) \geq \mathcal{D}(F^*)$ and thus by Theorem 8 we obtain that $F = F^*$. Again by Theorem 7 we have that μ is 2-Markov and so $\{X_v\}_{v \in V_d}$ can be obtained by iterating conditionally independent couplings of C along edges (see chapter 10). This shows the Gaussianity of the whole system $\{X_v\}_{v \in V}$.

12 Appendix A: On heated random variables

Let X be a random variable with values in $\mathbb{R}^n$ and let M be the standard normal distribution on $\mathbb{R}^n$. Let f_t denote the density function of $X + \sqrt{2t}M$ and let μ_t denote the corresponding measure on $\mathbb{R}^n$. The standard heat equation says that $\partial_t f_t = \triangle f_t$ holds for every $t > 0$. It is useful to compute the variation of the differential entropy $\mathbb{D}(f_t)$. The de Bruijn identity (see e.g. [17]) says that

$$\partial_t(\mathbb{D}(f_t)) = \partial_t \int_{\mathbb{R}^n} -f_t \log f_t = \int_{\mathbb{R}^n} -\triangle f_t(1 + \log f_t) = \int_{\mathbb{R}^n} \|\nabla f_t\|_2^2 / f_t. \quad (27)$$

However the validity of (27) relies on the fact that both $\partial_i f_t$ and $\partial_i f_t \log f_t$ vanish at infinity for every $1 \leq i \leq n$. This fact is proved in Lemma 12.2. Notice that if X has finite variance then also $X + \sqrt{2t}M$ has finite variance and if $t > 0$ then $\mathbb{D}(f_t)$ is a finite quantity (Lemma 12.3).

In general if $\sigma > 0$ then the density function f of $X + \sigma M$ is smooth, non-vanishing and analytic restricted to every line in $\mathbb{R}^n$. More precisely, if $p, q \in \mathbb{R}^n$ then the real function $\lambda \mapsto f(p + \lambda(q-p))$ extends to an entire analytic function on $\mathbb{C}$. Furthermore every partial derivative of f has this property. In the rest of this appendix we prove several other facts about heated random variables.

Lemma 12.1 *Let X be a random variable with values in $\mathbb{R}^n$ and let M be the standard normal distribution on $\mathbb{R}^n$. Let f be the density function of the independent sum $X + \sigma M$ for some $\sigma > 0$. Then for every $1 \leq i \leq n$ and $x \in \mathbb{R}^n$ we have that $|\partial_i f(x)| \leq f(x)a(1 + |\log(bf(x))|^{1/2})$ for some constants a, b depending on n and σ.*

Proof. Let Φ be the density function of σM and let μ be the distribution of X. Let $r \in \mathbb{R}^+$ the smallest positive real number such that $\partial_i \Phi(z) \leq |f(x)|$ for every z satisfying $|z| \geq r$. It can be shown that $r \leq c_2(1 + |\log(c_1 f(x))|^{1/2})$ for some constants (depending on n and σ). Let

$D = \{z : |x - z| \leq r\}$. We have that $\int_{z \in \overline{D}} \partial_i \Phi(x - z) d\mu \leq |f(x)|$. On the other hand, by $f(x) = \int_{z \in \mathbb{R}} \Phi(x - z) d\mu$ we obtain that

$$\int_{z \in D} \partial_i \Phi(x - z) d\mu \leq |f(x)| \max_{|y| \leq r} \partial_i \Phi(y) / \Phi(y).$$

Using that $\partial_i f(x) = \int_z \partial_i \Phi(x - z) d\mu$ and that $\partial_i \Phi(z)/\Phi(z) = O(z)$ the proof is complete. $\quad\square$

Lemma 12.2 *Let X be a random variable with values in $\mathbb{R}^n$ and let M be a random variable with standard normal distribution on $\mathbb{R}^n$. Let f be the density function of the independent sum $X + \sigma M$ for some $\sigma > 0$. Then for $1 \leq i \leq n$ the functions f, $\partial_i f$ and $\partial_i f \log f$ vanish at infinity.*

Proof. We start with f. For contradiction, assume that $D = \{x : f(z) \geq c\}$ is unbounded for some $c > 0$. Let Φ be the density function of σM and let us choose $r \in \mathbb{R}^+$ such that $\Phi(x) \leq c/2$ whenever $\|x\|_2 \geq r$. Let μ be the probability distribution of X. We have that if $f(x) \geq c$ and $Q_x = \{z : \|z - x\|_2 < r\}$, then $\int_{z \in Q_x} \Phi(z - x) d\mu \geq c/2$ and thus $\mu(Q_x) \geq \|\Phi\|_\infty^{-1} c/2$. From the unboundedness of D we conclude that there is an infinite set of points $\{p_i\}_{i=1}^\infty$ in D such that $\|p_i - p_j\|_2 > 2r$ holds for every pair $i \neq j$ in $\mathbb{N}$. This contradicts the fact that μ is finite. The statement for $\partial_i f$ and $\partial_i f \log f$ follows from Lemma 12.1. $\quad\square$

Lemma 12.3 *Let X be a random variable with values in $\mathbb{R}^n$ with finite covariance matrix. Let M be independent of X, with standard normal distribution on $\mathbb{R}^n$. Then, for every $\sigma > 0$, $X + \sigma M$ has finite differential entropy.*

Proof. The random variable $X + \sigma M$ has finite covariance matrix. As it is well-known, among the distributions with a given covariance matrix, Gaussian distribution maximizes differential entropy. Hence $\mathbb{D}(X + \sigma M) < \infty$. On the other hand, as in the previous lemma, let f be the density function of $X + \sigma M$. Lemma 12.2 implies that $\{t : f(t) > 1\}$ is a compact set. The continuity of f implies that $\int_{\mathbb{R}^n} f(t) \log f(t) dt < \infty$. Thus we also have $\mathbb{D}(X + \sigma M) > -\infty$. $\quad\square$

Lemma 12.4 *Let X be a random variable with values in $\mathbb{R}^n$ and let M be the standard normal distribution on $\mathbb{R}^n$. Let f be the density function of the independent sum $X + \sigma M$ for some $\sigma > 0$. Then for every $\varepsilon > 0$ there is $\varepsilon' > 0$ such that for every pair $a, b \in \mathbb{R}^n$ with $\|a - b\|_2 = r \leq \varepsilon'$ and $f(b) > c$ we have that $f(a)/f(b) \geq 1 - \varepsilon$, where $c = \sigma^{-(n-1)/2} \exp(-r^{-1}/(16\sigma^2))$.*

Proof. We start by general estimates for a pair $a, b \in \mathbb{R}^n$ with $r = \|a - b\|_2 \leq 1/4$. We have that $f(x) = \int_{y \in \mathbb{R}^n} \Phi(x - y) d\mu$ where μ is the probability distribution of X and Φ is the density function of σM. Let $D = \{z : \|z - a\|_2 \leq r^{-1/2}\}$. Let $f_1(x) = \int_{y \in \mathbb{R}^n} 1_D \Phi(x - y) d\mu$ and $f_2(x) = f(x) - f_1(x)$. We have that $f_2(x) \leq \sup_{z \in \overline{D}} \Phi(x - z)$. It follows that $|f_2(a)|, |f_2(b)| \leq (2\pi\sigma^2)^{-(n-1)/2} \Phi_0(r^{-1/2} - r)$ where Φ_0 is the density function of the one dimensional normal distribution $N(0, \sigma)$. Thus using $r^{-1/2} - r \geq r^{-1/2}/2$ and $1/\sqrt{2\pi} < 1$ we have

$$|f_2(a)|, |f_2(b)| \leq (2\pi\sigma^2)^{-(n-1)/2} \Phi_0(r^{-1/2}/2) < c^2. \tag{28}$$

To estimate $f_1(a)/f_1(b)$ we use

$$\min_{z \in D} \Phi(z - a)/\Phi(z - b) \le f_1(a)/f_1(b).$$

From

$$\Phi(z - a)/\Phi(z - b) = \exp((\|z - b\|_2^2/2 - \|z - a\|_2^2/2)/\sigma^2) = \exp(((a - z, b - a) + r^2/2)/\sigma^2)$$

it follows that

$$f_1(a)/f_1(b) \ge \exp((-r^{1/2} + r^2/2)/\sigma^2) \ge 1 - r^{1/2}/\sigma^2. \tag{29}$$

Inequality (28) implies that $f(b) = f_1(b) + f_2(b) \le f_1(b) + c^2$. Using this and $f_1(a) \le f(a)$ we obtain

$$f(a)/f(b) \ge f_1(a)/(f_1(b) + c^2) = (f_1(a)/f_1(b))(1 + c^2/f_1(b))^{-1}.$$

If $f(b) > c$, then $f_1(b) = f(b) - f_2(b) > c - c^2$ and thus by (29) we get

$$f(a)/f(b) > (f_1(a)/f_1(b))(1 + c^2/(c - c^2))^{-1} \ge (1 - r^{1/2})(1 - c).$$

The quantity c goes to 0 with r and so if r is small enough we have that $f(a)/f(b) \ge 1 - \varepsilon$. □

Let X be a random variable with values in $\mathbb{R}^d$. We say that $v \in \mathbb{R}^d$ is an independent direction for X if the $\mathbb{R}$-valued random variable (X, v) is independent from the projection of X to the $d - 1$ dimensional space $v^\perp$. Note that every direction is independent for the standard normal distribution.

Lemma 12.5 *Let X be a random variable with values in $\mathbb{R}^d$ and with $\mathbb{E}(X) = 0$. Assume that $\{v_i\}_{i=1}^d$ is a basis in $\mathbb{R}^d$ such that each v_i is an independent direction for X and furthermore for every $1 \le i \le d$ there is $1 \le j_i \le d$ such that $(v_i, v_{j_i}) \ne 0$. Then X is Gaussian.*

Proof. Let N be the standard normal distribution. It is clear that the independent sum $X + \varepsilon N$ has the same independence property as X for every $\varepsilon \ge 0$. Furthermore it is enough to prove that the heated version $X + \varepsilon N$ of X is Gaussian for every $\varepsilon > 0$. Let us fix $\varepsilon > 0$. The advantage of working with $X + \varepsilon N$ is that it has a strictly positive smooth density function f on $\mathbb{R}^d$ and so we can work with logarithms and with partial derivatives. The independence property now says that $\partial_{v_i} \log f(x)$ is equal to $h_i((x, v_i))$ for some smooth function $h_i : \mathbb{R} \to \mathbb{R}$. We obtain that

$$\partial_{v_{j_i}} \partial_{v_i} f = (v_i, v_{j_i}) h_i''((x, v_i)) \ , \ \partial_{v_i} \partial_{v_{j_i}} f = (v_i, v_{j_i}) h_{j_i}''((x, v_{j_i})).$$

and so $h_i''((x, v_i)) = h_{j_i}''((x, v_{j_i}))$. Since v_i and v_{j_i} are independent for every pair $a, b \in \mathbb{R}$ there is $x \in \mathbb{R}^d$ such that $(x, v_i) = a$ and $(x, v_{j_i}) = b$. This implies that $h_i''(a) = h_{j_i}''(b)$ holds for every a, b and so each h_i'' is a constant function for every i. Consequently h_i is linear for every i and thus (using that $\{v_i\}_{i=1}^d$ is a basis) $\partial_u \log f$ is a linear function for every $u \in \mathbb{R}^d$. It follows that $\log f$ satisfies $\partial_{u_1} \partial_{u_2} \partial_{u_3} \log f = 0$ for every $u_1, u_2, u_3 \in \mathbb{R}^d$. This means that $\log f$ is given by a quadratic d-variate polynomial Q on $\mathbb{R}^d$. We obtain that $f = c \exp(Q(x))$ holds and thus f is the density function of some Gaussian distribution.

Lemma 12.6 *Let (X, Y, Z) be a joint distribution with $X, Y \in \mathbb{R}$ and $Z \in \mathbb{R}^{d-2}$. Let (X_t, Y_t, Z_t) be the triple obtained by running the heat equation for time t with $X_0 = X, Y_0 = Y$ and $Z_0 = Z$. Assume that for every $t \geq 0$ we have that X_t and Z_t are conditionally independent with respect to Y_t. Then either (X, Y) is independent from Z or (Y, Z) is independent from X.*

Proof. We parametrize $\mathbb{R}^d$ with triples $(x, y, \underline{z})$ where the coordinates x and y are real numbers and $\underline{z}$ is a $d - 2$-dimensional vector. By Δ_z we mean the sum of the second partial derivates with respect to the coordinates belonging to $\underline{z}$. We denote by f_t, h_t, g_t and m_t the density functions of $(X_t, Y_t, Z_t), (X_t, Y_t), (Z_t, Y_t)$ and Y_t, respectively. We also introduce $s_t(\underline{z}, y) = g_t(\underline{z}, y)/m_t(y)$. Using the conditional independence and the heat equation we obtain the following equations.

$$f_t(x, y, \underline{z}) = h_t(x, y)g_t(\underline{z}, y)/m_t(y) = h_t(x, y)s_t(\underline{z}, y).$$

$$\partial_t f_t = \Delta f_t; \qquad \partial_t h_t = \Delta h_t; \qquad \partial_t g_t = \Delta g_t = \partial_{yy}g_t + \Delta_z g_t; \qquad \partial_t m_t = \Delta m_t.$$

By abusing the notation we will omit t from f_t, h_t, g_t, m_t and s_t in the following calculations. We start from the first equality, and use the other three one after the other.

$$\partial_{xx}f + \partial_{yy}f + \Delta_z f = \partial_t h \cdot s + h \cdot \partial_t s.$$

$$\partial_{xx}h \cdot s + \partial_{yy}f + \Delta_z f = \partial_{xx}h \cdot s + \partial_{yy}h \cdot s + h \cdot \partial_t s.$$

$$\partial_{yy}f + h \cdot \Delta_z g/m = \partial_{yy}h \cdot s + h \cdot \big((\partial_t g)/m - (g \cdot \partial_t m)/m^2\big).$$

$$\partial_{yy}h \cdot s + 2\partial_y h \cdot \partial_y s + h \cdot \partial_{yy}s = \partial_{yy}h \cdot s + (h \cdot \partial_{yy}g)/m - (gh \cdot \partial_{yy}m)/m^2. \qquad (30)$$

Before continuing this, we calculate the partial derivatives of s with respect to y.

$$\partial_y s = (\partial_y g)/m - (g \cdot \partial_y m)/m^2.$$

$$\partial_{yy}s = (\partial_{yy}g)/m - 2(\partial_y g \cdot \partial_y m)/m^2 - (g \cdot \partial_{yy}m)/m^2 + 2g \cdot (\partial_y m)^2/m^3.$$

Now we substitute this into equation (30).

$$2(\partial_y h \cdot \partial_y g)/m - 2g(\partial_y h \cdot \partial_y m)/m^2 + (h \cdot \partial_{yy}g)/m$$

$$- 2h(\partial_y g \cdot \partial_y m)/m^2 - (gh \cdot \partial_{yy}m)/m^2 + 2hg(\partial_y m)^2/m^3 = (h \cdot \partial_{yy}g)/m - (gh \cdot \partial_{yy}m)/m^2.$$

$$(\partial_y h \cdot \partial_y g)/m - (g \cdot \partial_y h + h \cdot \partial_y g) \cdot (\partial_y m)/m^2 + gh(\partial_y m)^2/m^3 = 0.$$

$$(\partial_y h \cdot \partial_y g)/m^2 - (g \cdot \partial_y h + h \cdot \partial_y g) \cdot (\partial_y m)/m^3 + gh(\partial_y m)^2/m^4 = 0.$$

$$\Big((\partial_y h)/m - (h \cdot \partial_y m)/m^2\Big)\Big((\partial_y g)/m - (g \cdot \partial_y m)/m^2\Big) = 0.$$

$$\big(\partial_y(h/m)\big) \cdot \big(\partial_y(g/m)\big) = 0.$$

We obtain that at least one of $\partial_y(h/m) = 0$ and $\partial_y(g/m) = 0$ holds on an open set. Assume that (without loss of generality) $\partial_y(h/m) = 0$ holds on an open set U of the domain D of h/m which is

37

$\mathbb{R}^2$. This is equivalent to $m\partial_y h - h\partial_y m = 0$ on U. Let $p \in U$ and $q \in D$ be arbitrary. Let us define the function $r : \mathbb{R} \to \mathbb{R}$ by $r(\lambda) = (m\partial_y h - h\partial_y m)(p + \lambda(q - p))$. Then r has an analytic extension to $\mathbb{C}$. In addition, $r = 0$ in a small neighborhood of 0 in $\mathbb{R}$. It follows that r is constant 0 and thus $\partial_y(h/m)$ is 0 at every $q \in D$. This implies that X_t is independent of Y_t. Similarly if $\partial_y(g/m) = 0$ holds on an open set we obtain that Z_t is independent of Y_t. The conditional independence of X_t and Z_t with respect to Y_t finishes the proof. $\qquad\square$

13 Appendix B: Differential entropy

Differential entropy is defined as follows for absolutely continuous random vectors. Some properties of the discrete entropy are preserved (e.g. it is additive if we put together independent random variables), others do not hold any more; an essential difference is that differential entropy does not have to be nonnegative.

Definition 13.1 *Let* $(X_1, X_2, \ldots, X_n)$ *be a family of random variables. Suppose that their joint distribution is absolutely continuous, and their joint density function is* f*. Then their differential entropy is defined as follows (provided that the integral exists):*

$$\mathbb{D}(X_1, X_2, \ldots, X_n) = -\int_{\mathbb{R}^n} f(t_1, \ldots, t_n) \log f(t_1, \ldots, t_n) dt_1 \ldots dt_n.$$

To see the connection between entropy in the discrete case and differential entropy, recall Theorem 9.3.1 from [17]. This says that if we divide the range of X into bins of length δ, and X^δ denotes the quantized version of X with respect to this grid, then

$$\mathbb{H}(X^\delta) + \log \delta \to \mathbb{H}(X)$$

as $\delta \to 0$, assuming that the density of X is Riemann integrable.

The following well-known lemma shows how the differential entropy is modified when we apply a linear transformation to the random vector (see e.g. corollary to Theorem 9.6.4 in [17].

Lemma 13.1 *Let* $\underline{X} = (X_1, \ldots, X_n)$ *be a family of random variables, and let* $A \in \mathbb{R}^{n \times n}$ *be an invertible matrix. Then*

$$\mathbb{D}(A\underline{X}) = \mathbb{D}(\underline{X}) + \log |\det(A)|.$$

The following lemma is equivalent to the fact that the nonnegativity of conditional mutual information holds for differential entropy as well. We include a proof for completeness.

Lemma 13.2 *Let* X, Y, Z *be random variables such that their differential entropy exist. Then we have*

$$\mathbb{D}(X, Y, Z) \le \mathbb{D}(X, Z) + \mathbb{D}(Y, Z) - \mathbb{D}(Z).$$

Proof. Let $g(x,y) = f(x,y,z)/f(z)$ on the support of Z, and 0 otherwise. Then g is a density function on $\mathbb{R}^2$. As the nonnegativity of mutual information is satisfied for differential entropy (see e.g. Corollary to Theorem 9.6.1. in [17]), we have

$$-\int g(x,y)\log g(x,y)dxdy \leq -\int g_1(x)\log g_1(x)dx - \int g_2(y)\log g_2(y)dy,$$

where g_1 and g_2 are the marginal densities of g. Multiplying both sides by $f(z)$ and integrating with respect to z we get the statement of the lemma. $\qquad\square$

14 Appendix C: Factor of i.i.d. processes

Let $f : [0,1]^{V_d} \to Y$ be a measurable function such that it is invariant under root preserving automorphisms. We can use f to construct an invariant process in the following way. First we put independent uniformly random elements from $[0,1]$ on the vertices of T_d. Then, at each vertex v, we evaluate f for this random labeling such that the root is placed to v. If f depends only on finitely many coordinates, then the corresponding process is called a block factor of i.i.d. process.

Proposition 14.1 *If $\{X_v\}_{v\in V_d}$ is a real-valued typical process and $\{Y_v\}_{v\in V_d}$ is a weak limit of factor of i.i.d. processes, then their independent sum $\{X_v + Y_v\}_{v\in V_d}$ is a typical process.*

Proof. Notice that the family of typical processes is closed with respect to the weak topology. On the other hand, every process that is a weak limit of factor i.i.d. processes is also a weak limit of block factor of i.i.d. processes [31]. Hence it is enough to prove the statement in the case when $\{Y_v\}_{v\in V_d}$ is a block factor of i.i.d. process. It is well known that block factor of i.i.d. processes can be approximated with the corresponding local algorithm computed on graphs with sufficiently large (essential) girth. Since large random d-regular graphs have large essential girth, the independent sum of the approximation of $\{X_v\}_{v\in V_d}$ and the local algorithm approximating $\{Y_v\}_{v\in V_d}$ is an approximation of the process $\{X_v + Y_v\}_{v\in V_d}$. $\qquad\square$

The next proposition was proved by Harangi and Virág in [27].

Proposition 14.2 *For $|\lambda| \leq 2\sqrt{d-1}$ the unique Gaussian wave Ψ_λ is a weak limit of factor of i.i.d processes (but not a factor of i.i.d. process itself).*

Acknowledgement.

References

[1] N. Alon, I. Benjamini, E. Lubetzky and S. Sodin, Non-backtracking random walks mix faster, Commun. Contemp. Math. **9** (2007), no. 4, 585–603.

[2] N. Anantharaman and E. Le Masson, Quantum ergodicity on large regular graphs, Duke Math. J. **164** (2015), no. 4, 723–765.

[3] Á. Backhausz and B. Szegedy, On large girth regular graphs and random processes on trees. Preprint. arXiv:1406.4420 [math.PR] (2014).

[4] Á. Backhausz, B. Szegedy and B. Virág, Ramanujan graphings and correlation decay in local algorithms, Random Structures Algorithms **47** (2015), no. 3, 424–435.

[5] Á. Backhausz and B. Virág, Spectral measures of factor of i.i.d. processes on vertex-transitive graphs. To appear in *Ann. Inst. Henri Poincaré Probab. Stat.*

[6] R. Bauerschmidt, J. Huang, A. Knowles and H.-T. Yau, Bulk eigenvalue statistics for random regular graphs. Preprint. arXiv:1505.06700 [math.PR] (2015).

[7] R. Bauerschmidt, A. Knowles and H.-T. Yau, Local semicircle law for random regular graphs. Preprint. arXiv:1503.08702 [math.PR] (2015).

[8] F. Benaych-Georges and A. Knowles, Lectures on the local semicircle law for Wigner matrices. Preprint. arXiv:1601.04055 [math.PR] (2016).

[9] A. Bloemendal, A. Knowles, H-T. Yau and J. Yin, On the principal components of sample covariance matrices, Probab. Theory Related Fields **164** (2016), no. 1-2, 459–552.

[10] B. Bollobás, *Random graphs*, second edition, Cambridge Studies in Advanced Mathematics, 73, Cambridge Univ. Press, Cambridge, 2001.

[11] B. Bollobás, O. Riordan, Sparse graphs: Metrics and random models. *Random Structures Algorithms* **39** (2011), 1–38.

[12] C. Bordenave, A new proof of Friedman's second eigenvalue theorem and its extension to random lifts. Preprint. arXiv:1502.04482 [math.CO] (2015).

[13] P. Bourgade and H.-T. Yau. The eigenvector moment flow and local quantum unique ergodicty. To appear in *Comm. Math. Phys* (2016).

[14] L. Bowen, The ergodic theory of free group actions: entropy and the f-invariant. *Groups Geom. Dyn.* **4** (2010), no. 3, 419–432.

[15] S. Brooks and E. Lindenstrauss, Non-localization of eigenfunctions on large regular graphs, Israel J. Math. **193** (2013), no. 1, 1–14.

[16] C. T. Conley, A. S. Marks, R.Tucker-Drob, Brooks's theorem for measurable colorings. Preprint. arXiv:1601.03361 [math.LO] (2016).

[17] T. M. Cover and J. A. Thomas, *Elements of information theory*, second edition, Wiley-Interscience, Hoboken, NJ, 2006.

[18] E. Csóka, B. Gerencsér, V. Harangi and B. Virág, Invariant Gaussian processes and independent sets on regular graphs of large girth, Random Structures Algorithms **47** (2015), no. 2, 284–303.

[19] I. Dumitriu, T. Johnson, S. Pal and E. Paquette, Functional limit theorems for random regular graphs, Probab. Theory Related Fields **156** (2013), no. 3-4, 921–975.

[20] I. Dumitriu and S. Pal, Sparse regular random graphs: spectral density and eigenvectors, Ann. Probab. **40** (2012), no. 5, 2197–2235.

[21] Y. Elon, Gaussian waves on the regular tree. Preprint. arXiv:0907.5065[math-ph] (2009).

[22] J. Friedman, A proof of Alon's second eigenvalue conjecture and related problems, Mem. Amer. Math. Soc. **195** (2008), no. 910, viii+100 pp.

[23] N. Friedman, D. Ornstein, On isomorphism of weak Bernoulli transformations. *Adv. Math.* **5** (1971), 365–394.

[24] D. Gaboriau and R. Lyons, A measurable-group-theoretic solution to von Neumann's problem. *Invent. Math.* **177** (2009), 533–540.

[25] D. Gamarnik and M. Sudan, Limits of local algorithms over sparse random graphs. Preprint. arXiv:1304.1831 [math.PR] (2013).

[26] L. Geisinger, Convergence of the density of states and delocalization of eigenvectors on random regular graphs. Preprint. arXiv:1305.1039 [math.PR] (2013).

[27] V. Harangi and B. Virág, Independence ratio and random eigenvectors in transitive graphs, Ann. Probab. **43** (2015), no. 5, 2810–2840.

[28] H. Hatami, L. Lovász and B. Szegedy, Limits of locally-globally convergent graph sequences, Geom. Funct. Anal. **24** (2014), no. 1, 269–296.

[29] C. Hoppen and N. Wormald, Local algorithms, regular graphs of large girth, and random regular graphs. Preprint. arXiv:1308.0266 [math.CO].

[30] J. Huang, B. Landon and H.-T. Yau, Bulk universality of sparse random matrices, J. Math. Phys. **56** (2015), no. 12, 123301, 19 pp.

[31] R. Lyons, Factors of iid on trees. To appear in *Combin. Probab. Comput.*

[32] R. Lyons and F. Nazarov, Perfect matchings as IID factors on non-amenable groups. *European J. Combin.* **32** (2011), 1115–1125.

[33] D. Puder, Expansion of random graphs: new proofs, new results, Invent. Math. **201** (2015), no. 3, 845–908.

[34] S. O'Rourke, V. Vu and K. Wang, Eigenvectors of random matrices: a survey. Preprint. arXiv:1601.03678 [math.PR] (2016).

[35] M. Rahman, Factor of iid percolation on trees. Preprint. arXiv:1410.3745 [math.PR] (2014).

[36] M. Rahman, B. Virág, Local algorithms for independent sets are half-optimal. Preprint. arXiv:1402.0485 [math.PR] (2014).

[37] T. Tao and V. Vu, Random matrices: universal properties of eigenvectors, Random Matrices Theory Appl. **1** (2012), no. 1, 1150001, 27 pp.

[38] N. C. Wormald, Models of random regular graphs, in *Surveys in combinatorics, 1999 (Canterbury)*, 239–298, London Math. Soc. Lecture Note Ser., 267, Cambridge Univ. Press, Cambridge.

ÁGNES BACKHAUSZ. Eotvos Loránd University, Budapest, Hungary and MTA Alfréd Rényi Institute of Mathematics. `agnes@math.elte.hu`

BALÁZS SZEGEDY. MTA Alfréd Rényi Institute of Mathematics, Budapest, Hungary. `szegedy.balazs@renyi.mta.hu`