The Kadison-Singer Problem for Strongly Rayleigh Measures and Applications to Asymmetric TSP

Nima Anari * Shayan Oveis Gharan †

Abstract

Marcus, Spielman, and Srivastava in their seminal work [MSS13b] resolved the Kadison-Singer conjecture by proving that for any set of finitely supported independently distributed random vectors $v_1, \dots, v_n$ which have "small" expected squared norm and are in isotropic position (in expectation), there is a positive probability that the sum $\sum v_i v_i^\intercal$ has small spectral norm. Their proof crucially employs real stability of polynomials which is the natural generalization of real-rootedness to multivariate polynomials.

Strongly Rayleigh distributions are families of probability distributions whose generating polynomials are real stable [BBL09]. As independent distributions are just special cases of strongly Rayleigh measures, it is a natural question to see if the main theorem of [MSS13b] can be extended to families of vectors assigned to the elements of a strongly Rayleigh distribution.

In this paper we answer this question affirmatively; we show that for any homogeneous strongly Rayleigh distribution where the marginal probabilities are upper bounded by ϵ_1 and any isotropic set of vectors assigned to the underlying elements whose norms are at most $\sqrt{\epsilon_2}$, there is a set in the support of the distribution such that the spectral norm of the sum of the natural quadratic forms of the vectors assigned to the elements of the set is at most $O(\epsilon_1 + \epsilon_2)$. We employ our theorem to provide a sufficient condition for the existence of spectrally thin trees. This, together with a recent work of the authors [AO15], provides an improved upper bound on the integrality gap of the natural LP relaxation of the Asymmetric Traveling Salesman Problem.

1 Introduction

Marcus, Spielman and Srivastava [MSS13b] in a breakthrough work proved the Kadison-Singer conjecture [KS59] by proving Weaver's [Wea04] conjecture KS_2 and the Akemann and Anderson's Paving conjecture [AA91]. The following is their main technical contribution.

Theorem 1.1. *If $\epsilon > 0$ and $v_1, \dots, v_m$ are independent random vectors in $\mathbb{R}^d$ with finite support where,*

$$\sum_{i=1}^{m} \mathbb{E} v_i v_i^\intercal = I,$$

such that for all i,

$$\mathbb{E} \|v_i\|^2 \le \epsilon,$$

*Computer Science Division, UC Berkeley. Email: anari@berkeley.edu.

†Department of Computer Science and Engineering, University of Washington. This work was partly done while the author was a postdoctoral Miller fellow at UC Berkeley. Email: shayan@cs.washington.edu.

then

$$\mathbb{P}\left[\left\|\sum_{i=1}^{m} v_i v_i^{\intercal}\right\| \leq (1+\sqrt{\epsilon})^2\right] > 0.$$

In this paper, we prove an extension of the above theorem to families of vectors assigned to elements of a not necessarily independent distribution.

Let $\mu : 2^{[m]} \to R_+$ be a probability distribution on the subsets of the set $[m] = \{1, 2, \ldots, m\}$. In particular, we assume that $\mu(.)$ is nonnegative and,

$$\sum_{S \subseteq [m]} \mu(S) = 1.$$

We assign a multi-affine polynomial with variables $z_1, \ldots, z_m$ to μ,

$$g_\mu(z) = \sum_{S \subseteq [m]} \mu(S) \cdot z^S,$$

where for a set $S \subseteq [m]$, $z^S = \prod_{i \in S} z_i$. The polynomial g_μ is also known as the *generating polynomial* of μ. We say μ is a *homogeneous* probability distribution if g_μ is a homogeneous polynomial.

We say that μ is a *strongly Rayleigh* distribution if g_μ is a real stable polynomial. See Subsection 3.2 for the definition of real stability. Strongly Rayleigh measures are introduced and deeply studied in the seminal work of Borcea, Brändén and Liggett [BBL09]. They are natural generalizations of product distributions and cover several interesting families of probability distributions including determinantal measures and random spanning tree distributions. We refer interested readers to [OSS11, PP14] for applications of these probability measures.

Our main theorem extends Theorem 1.1 to families of vectors assigned to the elements of a strongly Rayleigh distribution. This can be seen as a generalization because independent distributions are special classes of strongly Rayleigh measures. To state the main theorem we need another definition. The *marginal* probability of an element i with respect to a probability distribution, μ, is the probability that i is in a sample of μ,

$$\mathbb{P}_{S\sim\mu}\left[i \in S\right] = \partial_{z_i} g_\mu(z)\big|_{z_1=\ldots=z_m=1}. \tag{1}$$

Theorem 1.2 (Main)**.** *Let μ be a homogeneous strongly Rayleigh probability distributions on $[m]$ such that the marginal probability of each element is at most ϵ_1, and let $v_1, \ldots, v_m \in \mathbb{R}^d$ be vectors in an isotropic position,*

$$\sum_{i=1}^{m} v_i v_i^{\intercal} = I,$$

such that for all i, $\|v_i\|^2 \leq \epsilon_2$. Then,

$$\mathbb{P}_{S\sim\mu}\left[\left\|\sum_{i\in S} v_i v_i^{\intercal}\right\| \leq 4(\epsilon_1+\epsilon_2) + 2(\epsilon_1+\epsilon_2)^2\right] > 0.$$

The above theorem does not directly generalize Theorem 1.1, but it can be seen as a variant of Theorem 1.1 to the case where the vectors $v_1, \ldots, v_m$ are negatively dependent. We expect to see several applications of our main theorem that are not realizable by the original proof of

[MSS13b]. In the following subsections we describe our main motivation for studying the above statement, which is to design approximation algorithms for the Asymmetric Traveling Salesman Problem (ATSP).

Let us conclude this part by proving a simple application of the above theorem to prove KS_r for $r \geq 5$.

Corollary 1.3. *Given a set of vectors $v_1, \ldots, v_m \in \mathbb{R}^d$ in isotropic position,*

$$\sum_{i=1}^{m} v_i v_i^\intercal = I,$$

if for all i, $\|v_i\|^2 \leq \epsilon$, then for any r, there is an r partitioning of $[m]$ into $S_1, \ldots, S_r$ such that for any $j \leq r$,

$$\left\| \sum_{i \in S_j} v_i v_i^\intercal \right\| \leq 4(1/r + \epsilon) + 2(1/r + \epsilon)^2.$$

Proof. The proof is inspired by the lifting idea in [MSS13b]. For $i \in [m]$ and $j \in [r]$ let $w_{i,j} \in \mathbb{R}^{d \cdot r}$ be the directed sum of r vectors all of which are 0^d except the j-th one which is v_i, i.e.,

$$w_{i,1} = \begin{pmatrix} v_i \\ 0^d \\ \vdots \\ 0^d \end{pmatrix}, w_{i,2} = \begin{pmatrix} 0^d \\ v_i \\ \vdots \\ 0^d \end{pmatrix}, \text{ and so on.}$$

Let $E = \{(i,j) : i \in [m], j \in [r]\}$ and let $\mu : 2^E \to \mathbb{R}_+$ be a product distribution defined in a way that selects exactly one pair $(i,j) \in E$ for any $i \in [m]$ uniformly at random. Observe that there are m^r sets in the support of μ each of size exactly m and each has probability $1/r^m$. Therefore, μ is a homogeneous probability distribution and the marginal probability of each element of E is exactly $1/r$. In addition, since product distributions are strongly Rayleigh, μ is strongly Rayleigh. Therefore, by Theorem 1.2, there is a set S in the support of μ such that

$$\left\| \sum_{(i,j) \in S} w_{i,j} w_{i,j}^\intercal \right\| \leq \alpha,$$

for $\alpha = 4(1/r + \epsilon) + 2(1/r + \epsilon)^2$. Now, let $S_j = \{i : (i,j) \in S\}$. It follows that for any $j \in [r]$,

$$\left\| \sum_{i \in S_j} v_i v_i^\intercal \right\| \leq \alpha,$$

as desired. □

1.1 The Thin Basis Problem

In this section we use the main theorem to prove the existence of a thin basis among a given set of isotropic vectors. In the next section, we will use this theorem to prove the existence of thin trees in graphs, i.e., trees which are "sparse" in all cuts of a given graph.

Given a set of vectors $v_1, \ldots, v_m \in \mathbb{R}^d$ in the isotropic position,

$$\sum_{i=1}^m v_i v_i^\intercal = I,$$

we want to find a sufficient condition for the existence of a *thin basis*. Recall that a set $T \subset [m]$ is a basis if $|T| = d$ and all vectors indexed by T are linearly independent. We say T is α-thin if

$$\left\| \sum_{i \in T} v_i v_i^\intercal \right\| \leq \alpha.$$

An obvious necessary condition for the existence of an α-thin basis is that the set

$$V(\alpha) := \{v_i : \|v_i\|^2 \leq \alpha\},$$

contains a basis. We show that there exist universal constants $C_1, C_2 > 0$ such that the existence of C_1/α disjoint bases in $V(C_2 \cdot \alpha)$ is a sufficient condition.

Theorem 1.4. *Given a set of vectors $v_1, \ldots, v_m \in \mathbb{R}^d$ in the sub-isotropic position*

$$\sum_{i=1}^m v_i v_i^\intercal \preceq I,$$

if for all $1 \leq i \leq m$, $\|v_i\|^2 \leq \epsilon$, and the set $\{v_1, \ldots, v_m\}$ contains k disjoint bases, then there exists an $O(\epsilon + 1/k)$-thin basis $T \subseteq [m]$.

We will use Theorem 1.2 to prove the above theorem. To use Theorem 1.2 we need to define a strongly Rayleigh distribution on $[m]$ with small marginal probabilities. This is proved in the following proposition.

Proposition 1.5. *Given a set of vectors $v_1, \ldots, v_m \in \mathbb{R}^d$ that contains k disjoint bases, there is a strongly Rayleigh probability distribution $\mu : 2^{[m]} \to \mathbb{R}_+$ supported on the bases such that the marginal probability of each element is at most $O(1/k)$.*

Now, Theorem 1.4 follows simply from the above proposition. Letting μ be defined as above, we get $\epsilon_1 = \epsilon$ and $\epsilon_2 = O(1/k)$ in Theorem 1.2 which implies the existence of a basis $T \subseteq [m]$ such that

$$\left\| \sum_{i \in T} v_i v_i^\intercal \right\| \leq O(\epsilon + 1/k),$$

as desired.

In the rest of this section we prove the above proposition. In our proof μ will in fact be a homogeneous *determinantal* probability distribution. We say $\mu : 2^{[m]} \to \mathbb{R}_+$ is a determinantal probability distribution if there is a PSD matrix $M \in \mathbb{R}^{m \times m}$ such that for any set $T \subseteq [m]$,

$$\mathbb{P}_{S \sim \mu}[T \subseteq S] = \det(M_{T,T}),$$

where $M_{T,T}$ is the principal submatrix of M whose rows and columns are indexed by T. It is proved in [BBL09] that any determinantal probability distribution is a strongly Rayleigh measure, so this is sufficient for our purpose. In fact, we will find nonnegative weights $\lambda : [m] \to \mathbb{R}_+$ and for any basis T we will let

$$\mu_\lambda(T) \propto \det\left(\sum_{i \in T} \lambda_i v_i v_i^\intercal\right). \tag{2}$$

It follows by the Cauchy-Binet identity that for any λ, such a distribution is determinantal with respect to the gram matrix

$$M(i,j) = \sqrt{\lambda_i \lambda_j}\left\langle B^{-1/2} v_i, B^{-1/2} v_j\right\rangle$$

where $B = \sum_{i=1}^m \lambda_i v_i v_i^\intercal$. So, all we need to do is find $\{\lambda_i\}_{1 \le i \le m}$ such that the marginal probability of each element in μ_λ is $O(1/k)$.

For any basis $T \subset [m]$ let $\mathbf{1}_T \in \mathbb{R}^m$ be the indicator vector of the set T. Let P be the convex hull of bases' indicator vectors,

$$P := \operatorname{conv}\{\mathbf{1}_T : T \text{ is a basis}\}.$$

Recall that a point x is in the *relative interior* of P, $x \in \operatorname{relint} P$, if and only if x can be written as a convex combination of all of the vertices of P with strictly positive coefficients.

We find the weights in two steps. First, we show that for any point $x \in \operatorname{relint} P$, there exist weights $\lambda : [m] \to \mathbb{R}$ such that for any i,

$$\mathbb{P}_{S \sim \mu_\lambda}[i \in S] = x(i),$$

where $x(i)$ is the i-th coordinate of x and μ_λ is defined as in (2). Then, we show that there exists a point $x \in \operatorname{relint} P$ such that for all i, $x(i) \le O(1/k)$.

Lemma 1.6. *For any $x \in \operatorname{relint} P$ there exist $\lambda : [m] \to \mathbb{R}_+$ such that the marginal probability of each element i in μ_λ is $x(i)$.*

Proof. Let $\mu^* := \mu_{\mathbf{1}}$ be the (determinantal) distribution where $\lambda_i = 1$ for all i. The idea is to find a distribution $p(.)$ maximizing the relative entropy with respect to μ^* and preserves x as the marginal probabilities. This is analogous to the recent applications of maximum entropy distributions in approximation algorithms [AGM+10, SV14].

Consider the following entropy maximization convex program.

$$\begin{aligned} \min \quad & \sum_T p(T) \cdot \log \frac{p(T)}{\mu^*(T)} \\ \text{s.t.} \quad & \sum_{T: i \in T} p(T) = x(i) \qquad \forall i, \\ & p(T) \ge 0. \end{aligned} \tag{3}$$

Note that any feasible solution satisfies $\sum_T p(T) = 1$ so we do not need to add this as a constraint. First of all, since $x \in \operatorname{relint} P$, there exists a distribution $p(.)$ such that for all bases T, $p(T) > 0$. So, the Slater condition holds and the duality gap of the above program is zero.

Secondly, we use the Lagrange duality to characterize the optimum solution of the above convex program. For any element i let γ_i be the Lagrange dual variable of the first constraint. The Lagrangian $L(p, \gamma)$ is defined as follows:

$$L(p,\gamma) = \inf_{p \ge 0} \sum_T p(T) \cdot \log \frac{p(T)}{\mu^*(T)} - \sum_i \gamma_i \sum_{T: e \in T} (p(T) - x(i))$$

Let p^* be the optimum p, letting the gradient of the RHS equal to zero we obtain, for any bases T,

$$\log \frac{p^*(T)}{\mu^*(T)} + 1 = \sum_{i \in T} \gamma_i.$$

For all i, let $\lambda_i = \exp(\gamma_i - 1/d)$, where d is the dimension of the v_i's. Then, we get

$$\begin{aligned} p^*(T) &= \prod_{i \in T} \lambda_i \cdot \mu^*(T) \\ &= \prod_{i \in T} \lambda_i \cdot \det \left(\sum_{i \in T} v_i v_i^\intercal \right) \\ &= \det \left(\sum_{i \in T} \lambda_i v_i v_i^\intercal \right). \end{aligned}$$

Therefore $p^* \equiv \mu_\lambda$. Since the duality gap is zero, the above p^* is indeed an optimal solution of the convex program (3). Therefore, the marginal probability of every element i with respect to p^* (μ_λ) is equal to $x(i)$. □

Lemma 1.7. *If $\{v_1, \dots, v_m\}$ contains k disjoint bases, then there exists a point $x \in \operatorname{relint} P$, such that $x(i) = O(1/k)$ for all i.*

Proof. Let $T_1, \dots, T_k$ be the promised disjoint bases. Let

$$x_0 = \frac{\mathbf{1}_{T_1} + \cdots + \mathbf{1}_{T_k}}{k}.$$

The above is a convex combination of the vertices of P; so $x_0 \in P$. We now perturb x_0 by a small amount to find a point in $\operatorname{relint} P$. Let x_1 be an arbitrary point in $\operatorname{relint} P$ (such as the average of all vertices). For any $0 < \epsilon < 1$, the point $x = (1-\epsilon)x_0 + \epsilon x_1 \in \operatorname{relint} P$. If ϵ is small enough, we get $x(i) = O(1/k)$ which proves the claim. □

This completes the proof of Proposition 1.5.

1.2 Spectrally Thin Trees

For a graph $G = (V, E)$, the Laplacian of G, L_G, is defined as follows: For a vertex $i \in V$ let $\mathbf{1}_i \in \mathbb{R}^V$ be the vector that is one at i and zero everywhere else. Fix an arbitrary orientation on the edges of E and let $b_e = \mathbf{1}_i - \mathbf{1}_j$ for an edge e oriented from i to j. Then,

$$L_G = \sum_{e \in E} b_e b_e^\intercal.$$

We use $L_G^\dagger$ to denote the pseudo-inverse of L_G. Also, for a set $T \subseteq E$, we write

$$L_T = \sum_{e \in T} b_e b_e^\intercal.$$

We say a spanning tree $T \subseteq E$ is α-thin with respect to G if for any set $S \subset V$,

$$|T(S, \overline{S})| \leq \alpha \cdot |E(S, \overline{S})|,$$

where $T(S, \overline{S}), E(S, \overline{S})$ are the set of edges cross the cut $(S, \overline{S})$ in T, G respectively. We say a spanning tree T is α-spectrally thin with respect to G if

$$L_T \preceq \alpha \cdot L_G.$$

It is easy to see that spectral thinness is a generalization of the combinatorial thinness, i.e., if T is α-spectrally thin it is also α-thin.

We say a graph G is k-edge-connected if it has at least k edges in any cut. In recent works on Asymmetric TSP [AGM+10, OS11] it was shown that the existence of (combinatorially) thin trees in k-edge-connected graphs plays an important role in bounding the integrality gap of the natural linear programming relaxation of the Asymmetric TSP [AO15].

It turns out that the existence of spectrally thin trees is significantly easier to prove than combinatorially thin trees thanks to Theorem 1.1 of [MSS13b]. Given a graph $G = (V, E)$, Harvey and Olver [HO14] employ a recursive application of [MSS13b] and show that if for all edges $e \in E$, $b_e^\intercal L_G^\dagger b_e \leq \alpha$, then G has an $O(\alpha)$-spectrally thin tree. The quantity $b_e L_G^\dagger b_e$ is the effective resistance between the endpoints of e when we replace every edge of G with a resistor of resistance 1 [LP13, Ch. 2]. Unfortunately, k-edge-connectivity is a significantly weaker property than $\max_e b_e L_G^\dagger b_e \leq \alpha$ [AO15]. So, this does not resolve the thin tree problem.

The main idea of [AO15] is to slightly change the graph G in order to decrease the effective resistance of edges while maintaining the size of the cuts intact. More specifically, to add a "few" edges E' to G such that in the new graph $G' = (V, E \cup E')$, the effective resistance of every edge of E is small and the size of every cut of G' is at most twice of that cut in G. If we can prove that G' has a spectrally thin tree $T \subseteq E$ such a tree is combinatorially thin with respect to G because G, G' have the same cut structure. To show that G' has a spectrally thin tree we need to answer the following question.

Problem 1.8. *Given a graph $G = (V, E)$, suppose there is a set $F \subseteq E$ such that (V, F) is k-edge-connected, and that for all $e \in F$, $b_e^\intercal L_G^\dagger b_e \leq \alpha$. Can we say that G has a $C \cdot \max\{\alpha, 1/k\}$-spectrally thin tree for a universal constant C?*

We use Theorem 1.4 to answer the above question affirmatively. Note that the above question cannot be answered by Theorem 1.1. One can use Theorem 1.1 to show that the set F can be partitioned into two sets F_1, F_2 such that each F_i is $1/2 + O(\alpha)$-spectrally thin, but Theorem 1.1 gives no guarantee on the connectivity of F_i's. On the other hand, once we apply our main theorem to a strongly Rayleigh distribution supported on connected subgraphs of G, e.g. the spanning trees of G, we get connectivity for free.

Corollary 1.9. *Given a graph $G = (V, E)$ and a set $F \subseteq E$ such that (V, F) is k-edge-connected, if for $\epsilon > 0$ and any edge $e \in F$, $b_e^\intercal L_G^\dagger b_e \leq \epsilon$, then G has an $O(1/k + \epsilon)$ spectrally thin tree.*

Proof. Let $L_G^{\dagger/2}$ be the square root of $L_G^\dagger$. Note that since $L_G^\dagger \succeq 0$, its square root is well defined. For all $e \in F$, let

$$v_e = L_G^{\dagger/2} b_e.$$

Then, by the corollary's assumption, for each $e \in F$,

$$\|v_e\|^2 = b_e L_G^\dagger b_e \leq \epsilon,$$

and the vectors $\{v_e\}_{e\in F}$ are in sub-isotropic position,

$$\begin{aligned}\sum_{e\in F} v_e v_e^\intercal &= L_G^{\dagger/2}\left(\sum_{e\in F} b_e b_e^\intercal\right) L_G^{\dagger/2} \\ &= L_G^{\dagger/2} L_F L_G^{\dagger/2} \preceq I.\end{aligned}$$

In addition, we can show that $\{v_e\}_{e\in F}$ contains $k/2$ disjoint bases. First of all, note that each basis of the vectors $\{v_e\}_{e\in F}$ corresponds to a spanning tree of the graph (V,F). Nash-Williams [NW61] proved that any k-edge-connected graph has $k/2$ edge-disjoint spanning trees. Since (V,F) is k-edge-connected, it has $k/2$ edge-disjoint spanning trees, and equivalently, $\{v_e\}_{e\in F}$ contains $k/2$ disjoint bases.

Therefore, by Theorem 1.4, there exists a basis (i.e., a spanning tree) $T\subseteq F$ such that

$$\left\|\sum_{e\in T} v_e v_e^\intercal\right\| \le \alpha, \tag{4}$$

for $\alpha = O(\epsilon + 1/k)$. Fix an arbitrary vector $y\in\mathbb{R}^V$. We show that

$$y^\intercal L_T y \le \alpha\cdot y^\intercal L_G y, \tag{5}$$

and this completes the proof. By (4) for any $x\in\mathbb{R}^V$,

$$x^\intercal\left(\sum_{e\in T} v_e v_e^\intercal\right) x \le \alpha\cdot\|x\|^2.$$

Let $x = L_G^{1/2} y$, we get

$$y^\intercal L_G^{1/2}\left(L_G^{\dagger/2}\sum_{e\in T} b_e b_e^\intercal L_G^{\dagger/2}\right) L_G^{1/2} y \le \alpha\cdot y^\intercal L_G y.$$

The above is the same as (5) and we are done. □

The above corollary completely answers Problem 1.8 but it is not enough for our purpose in [AO15]; we need a slightly stronger statement. For a matrix $D\in\mathbb{R}^{V\times V}$ we say $D\preceq_\square L_G$, if for any set $S\subset V$,

$$\mathbf{1}_S^\intercal D\mathbf{1}_S \le \mathbf{1}_S^\intercal L_G\mathbf{1}_S,$$

where as usual $\mathbf{1}_S\in\mathbb{R}^V$ is the indicator vector of the set S. In the main theorem of [AO15] we show that for any k-edge-connected graph G (for $k = 7\log n$) there is a positive definite (PD) matrix $D\preceq_\square L_G$ and a set $F\subseteq E$ such that (V,F) is $\Omega(k)$-edge-connected and

$$\max_{e\in F} b_e^\intercal D^{-1} b_e \le \frac{\text{polylog}(k)}{k}.$$

To show that G has a combinatorially thin tree it is enough to show that there is a tree $T\subseteq E$ that is α-spectrally thin w.r.t. $L_G + D$ for $\alpha = \text{polylog}(k)/k$, i.e.,

$$L_T \preceq \frac{\text{polylog}(k)}{k}(L_G + D).$$

Such a tree is 2α-combinatorially thin w.r.t. G because $D \preceq_{\square} L_G$. Note that the above corollary does not imply $L_G + D$ has a spectrally thin tree because D is not necessarily a Laplacian matrix. Nonetheless, we can prove the existence of a spectrally thin tree with another application of Theorem 1.4.

Corollary 1.10. *Given a graph $G = (V, E)$, a PD matrix D, and $F \subseteq E$ such that (V, F) is k-edge-connected, if for any edge $e \in F$,*

$$b_e^{\intercal} D^{-1} b_e \leq \epsilon,$$

then G has a spanning tree $T \subseteq F$ such that

$$L_T \preceq O(\epsilon + 1/k) \cdot (L_G + D).$$

Proof. The proof is very similar to Corollary 4.2. For any edge $e \in F$, let $v_e = (D + L_G)^{-1/2} b_e$. Note that since D is PD, $D + L_G$ is PD and $(D + L_G)^{-1/2}$ is well defined. By the assumption,

$$\|v_e\|^2 = b_e^{\intercal}(D + L_G)^{-1} b_e \leq b_e^{\intercal} D^{-1} b_e = \epsilon,$$

where the inequality uses Lemma 3.14. In addition, the vectors are in sub-isotropic position,

$$\sum_{e \in F} v_e v_e^{\intercal} = (D + L_G)^{\dagger/2} L_F (D + L_G)^{\dagger/2} \preceq I.$$

The matrix PSD inequality uses that $L_F \preceq L_G \preceq D + L_G$. Furthermore, every basis of $\{v_e\}_{e \in E}$ is a spanning tree of G and by $\Omega(k)$-connectivity of F, there are $\Omega(k)$-edge disjoint bases. Therefore, by Theorem 1.4, there is a tree $T \subseteq F$ such that

$$\left\| \sum_{e \in T} v_e v_e^{\intercal} \right\| \leq \alpha,$$

for $\alpha = O(\epsilon + 1/k)$. Similar to Corollary 4.2 this tree satisfies

$$L_T \preceq \alpha \cdot (L_G + D),$$

and this completes the proof. □

1.3 Bounded Degree Spanning Trees

In this section we provide a simple application of Corollary 1.10. We show that k-regular k-edge-connected graphs have constant degree spanning trees.

Lemma 1.11. *Any k-edge-connected graph $G = (V, E)$ has a spanning tree $T \subseteq E$ such that for any $i \in V$,*

$$d_T(i) \leq O(1/k) d_G(i),$$

where $d_T(i), d_G(i)$ are the degree of v in T and G respectively.

In two beautiful works Goemans [Goe06] and Lau, Singh [LS15] used combinatorial and polyhedral techniques to prove a generalization of the above lemma. Next, we use Corollary 1.10 to give a *spectral* proof of the above lemma.

Our proof of the above lemma follows the high level plan of [AO15] that we discussed in the previous section. The main difference because in this we are interested in finding a spanning three which in "only thin with respect to degree cuts", we can find the best possible matrix D independent of the cut structure of graph G. In particular, we simply let $D = kI$. Then, for any edge $e \in F$, we have

$$b_e^\intercal D^{-1} b_e = b_e^\intercal \frac{I}{k} b_e = \frac{\|b_e\|^2}{k} = \frac{2}{k},$$

where we used that for any edge e oriented from i to j, $\|\mathbf{1}_e\|^2 = \|\mathbf{1}_i\|^2 + \|\mathbf{1}_j\|^2 = 2$. Since E is k-edge-connected, by Corollary 1.10, there is a spanning tree $T \subseteq E$ such that

$$L_T \preceq O(2/k + 1/k) \cdot (L_G + kI).$$

Fix a vertex $i \in V$. We show that $d_T(i) \leq O(1/k) d_G(i)$. Multiplying both sides of the above equation with $\mathbf{1}_i$ we get

$$d_T(i) = \mathbf{1}_i^\intercal L_T \mathbf{1}_i \leq O(1/k) \mathbf{1}_i^\intercal (L_G + kI) \mathbf{1}_i \tag{6}$$

We can upper bound the RHS as follows,

$$\mathbf{1}_i^\intercal (L_G + kI)\mathbf{1}_i = \mathbf{1}_i^\intercal L_G \mathbf{1}_i + k \mathbf{1}_i^\intercal I \mathbf{1}_i = d_G(i) + k \|\mathbf{1}_i\|^2 = d_G(i) + k \leq 2 d_G(i),$$

where the inequality uses that G is k-edge-connected, so $d_G(i) \geq k$. Putting the above equation together with (6) proves the lemma.

2 Proof Overview

We build on the method of interlacing polynomials of [MSS13a, MSS13b]. Recall that an interlacing family of polynomials has the property that there is always a polynomial whose largest root is at most the largest root of the sum of the polynomials in the family. First, we show that for any set of vectors assigned to the elements of a homogeneous strongly Rayleigh measure, the characteristic polynomials of natural quadratic forms associated with the samples of the distribution form an interlacing family. This implies that there is a sample of the distribution such that the largest root of its characteristic polynomial is at most the largest root of the average of the characteristic polynomials of all samples of μ. Then, we use the multivariate barrier argument of [MSS13b] to upper-bound the largest root of our expected characteristic polynomial.

Our proof has two main ingredients. The first one is the construction of a new class of expected characteristic polynomials which are the weighted average of the characteristic polynomials of the natural quadratic forms associated to the samples of the strongly Rayleigh distribution, where the weight of each polynomial is proportional to the probability of the corresponding sample set in the distribution. To show that the expected characteristic polynomial is real rooted we appeal to the theory of real stability. We show that our expected characteristic polynomial can be realized by applying $\prod_{i=1}^m (1 - \partial/\partial_{z_i}^2)$ operator to the real stable polynomial $g_\mu(z) \cdot \det(\sum_{i=1}^m z_i v_i v_i^\intercal)$, and then projecting all variables onto x.

Our second ingredient is the extension of the multivariate barrier argument. Unlike [MSS13b], here we need to prove an upper bound on the largest root of the mixed characteristic polynomial which is very close to zero. It turns out that the original idea of [BSS14] that studies the behavior of the roots of a (univariate) polynomial $p(x)$ under the operator $1 - \partial/\partial_x$ cannot establish upper bounds that are less than one. Fortunately, here we need to study the behavior of the roots of a (multivariate) polynomial $p(z)$ under the operators $1 - \partial/\partial_{z_i}^2$. The $1 - \partial/\partial_{z_i}^2$ operators allow us to impose very small shifts on the multivariate upper barrier assuming the barrier functions are sufficiently small. The intuition is that, since

$$1 - \frac{\partial}{\partial_{z_i}^2} = \left(1 - \frac{\partial}{\partial_{z_i}}\right) \cdot \left(1 + \frac{\partial}{\partial_{z_i}}\right),$$

we expect $(1 - \partial/\partial_{z_i})$ to shift the upper barrier by $1 + \Theta(\delta)$ (for some δ depending on the value of the i-th barrier function) as proved in [MSS13b] and $(1 + \partial/\partial_{z_i})$ to shift the upper barrier by $1 - \Theta(\delta)$. Therefore, applying both operators the upper barrier must be moved by no more than $\Theta(\delta)$.

3 Preliminaries

We adopt a notation similar to [MSS13b]. We write $\binom{[m]}{k}$ to denote the collection of subsets of $[m]$ with exactly k elements. We write $2^{[m]}$ to denote the family of all subsets of the set $[m]$. We write ∂_{z_i} to denote the operator that performs partial differentiation with respect to z_i. We use $\|v\|$ to denote the Euclidean 2-norm of a vector x. For a matrix M, we write $\|M\| = \max_{\|x\|=1} \|Mx\|$ to denote the operator norm of M. We use $\mathbf{1}$ to denote the all 1 vector.

3.1 Interlacing Families

We recall the definition of interlacing families of polynomials from [MSS13a], and its main consequence.

Definition 3.1. *We say that a real rooted polynomial $g(x) = \alpha_0 \prod_{i=1}^{m-1}(x - \alpha_i)$ interlaces a real rooted polynomial $f(x) = \beta_0 \prod_{i=1}^{m}(x - \beta_i)$ if*

$$\beta_1 \leq \alpha_1 \leq \beta_2 \leq \alpha_2 \leq \ldots \leq \alpha_{m-1} \leq \beta_m.$$

We say that polynomials $f_1, \ldots, f_k$ have a common interlacing if there is a polynomial g such that g interlaces all f_i. The following lemma is proved in [MSS13a].

Lemma 3.2. *Let $f_1, \ldots, f_k$ be polynomials of the same degree that are real rooted and have positive leading coefficients. Define*

$$f_\emptyset = \sum_{i=1}^{k} f_i.$$

If $f_1, \ldots, f_k$ have a common interlacing, then there is an i such that the largest root of f_i is at most the largest root of $f_\emptyset$.

Definition 3.3. *Let $\mathcal{F} \subseteq 2^{[m]}$ be nonempty. For any $S \in \mathcal{F}$, let $f_S(x)$ be a real rooted polynomial of degree d with a positive leading coefficient. For $s_1, \ldots, s_k \in \{0, 1\}$ with $k < m$, let*

$$\mathcal{F}_{s_1,\ldots,s_k} := \{S \in \mathcal{F} : i \in S \Leftrightarrow s_i = 1\}.$$

Note that $\mathcal{F} = \mathcal{F}_{\emptyset}$. *Define*

$$f_{s_1,\ldots,s_k} = \sum_{S \in \mathcal{F}_{s_1,\ldots,s_k}} f_S,$$

and

$$f_{\emptyset} = \sum_{S \in \mathcal{F}} f_S.$$

We say polynomials $\{f_S\}_{S\in\mathcal{F}}$ *form an* interlacing family *if for all* $0 \le k < m$ *and all* $s_1, \ldots, s_k \in \{0,1\}$ *the following holds: If both of* $\mathcal{F}_{s_1,\ldots,s_k,0}$ *and* $\mathcal{F}_{s_1,\ldots,s_k,1}$ *are nonempty,* $f_{s_1,\ldots,s_k,0}$ *and* $f_{s_1,\ldots,s_k,1}$ *have a common interlacing.*

The following is analogous to [MSS13b, Thm 3.4].

Theorem 3.4. *Let* $\mathcal{F} \subseteq 2^{[m]}$ *and let* $\{f_S\}_{S\in\mathcal{F}}$ *be an interlacing family of polynomials. Then, there exists* $S \in \mathcal{F}$ *such that the largest root of* $f(S)$ *is at most the largest root of* $f_{\emptyset}$.

Proof. We prove by induction. Assume that for some choice of $s_1, \ldots, s_k \in \{0,1\}$ (possibly with $k = 0$), $\mathcal{F}_{s_1,\ldots,s_k}$ is nonempty and the largest root of $f_{s_1,\ldots,s_k}$ is at most the largest root of $f_{\emptyset}$. If $\mathcal{F}_{s_1,\ldots,s_k,0} = \emptyset$, then $f_{s_1,\ldots,s_k} = f_{s_1,\ldots,s_k,1}$, so we let $s_{k+1} = 1$ and we are done. Similarly, if $\mathcal{F}_{s_1,\ldots,s_k,1} = \emptyset$, then we let $s_{k+1} = 0$ and we are done with the induction. If both of these sets are nonempty, then $f_{s_1,\ldots,s_k,0}$ and $f_{s_1,\ldots,s_k,1}$ have a common interlacing. So, by Lemma 3.2, for some choice of $s_{k+1} \in \{0,1\}$, the largest root of $f_{s_1,\ldots,s_{k+1}}$ is at most the largest root of $f_{\emptyset}$. □

We use the following lemma which appeared as Theorem 2.1 of [Ded92] to prove that a certain family of polynomials that we construct in Section 4 form an interlacing family.

Lemma 3.5. *Let* $f_1, \ldots, f_k$ *be univariate polynomials of the same degree with positive leading coefficients. Then,* $f_1, \ldots, f_k$ *have a common interlacing if and only if* $\sum_{i=1}^k \lambda_i f_i$ *is real rooted for all convex combinations* $\lambda_i \ge 0$, $\sum_{i=1}^k \lambda_i = 1$.

3.2 Stable Polynomials

Stable polynomials are natural multivariate generalizations of real-rooted univariate polynomials. For a complex number z, let $\mathrm{Im}(z)$ denote the imaginary part of z. We say a polynomial $p(z_1, \ldots, z_m) \in \mathbb{C}[z_1, \ldots, z_m]$ is *stable* if whenever $\mathrm{Im}(z_i) > 0$ for all $1 \le i \le m$, $p(z_1, \ldots, z_m) \ne 0$. We say $p(.)$ is real stable, if it is stable and all of its coefficients are real. It is easy to see that any univariate polynomial is real stable if and only if it is real rooted.

One of the most interesting classes of real stable polynomials is the class of determinant polynomials as observed by Borcea and Brändén [BB08].

Theorem 3.6. *For any set of positive semidefinite matrices* $A_1, \ldots, A_m$*, the following polynomial is real stable:*

$$\det\Big(\sum_{i=1}^m z_i A_i\Big).$$

Perhaps the most important property of stable polynomials is that they are closed under several elementary operations like multiplication, differentiation, and substitution. We will use these operations to generate new stable polynomials from the determinant polynomial. The following is proved in [MSS13b].

Lemma 3.7. *If $p \in \mathbb{R}[z_1, \dots, z_m]$ is real stable, then so are the polynomials $(1 - \partial_{z_1})p(z_1, \dots, z_m)$ and $(1 + \partial_{z_1})p(z_1, \dots, z_m)$.*

The following corollary simply follows from the above lemma.

Corollary 3.8. *If $p \in \mathbb{R}[z_1, \dots, z_m]$ is real stable, then so is*

$$(1 - \partial_{z_1}^2)p(z_1, \dots, z_m).$$

Proof. First, observe that

$$(1 - \partial_{z_1}^2)p(z_1, \dots, z_m) = (1 - \partial_{z_1})(1 + \partial_{z_1})p(z_1, \dots, z_m).$$

So, the conclusion follows from two applications of Lemma 3.7. □

The following closure properties are elementary.

Lemma 3.9. *If $p \in \mathbb{R}[z_1, \dots, z_m]$ is real stable, then so is $p(\lambda \cdot z_1, \dots, \lambda_m \cdot z_m)$ for real-valued $\lambda_1, \dots, \lambda_m > 0$.*

Proof. Say $(z_1, \dots, z_m) \in \mathbb{C}^m$ is a root of $p(\lambda \cdot z_1, \dots, \lambda_m \cdot z_m)$. Then $(\lambda_1 \cdot z_1, \dots, \lambda_m \cdot z_m)$ is a root of $p(z_1, \dots, z_m)$. Since p is real stable, there is an i such that $\mathrm{Im}(\lambda_i \cdot z_i) \le 0$. But, since $\lambda_i > 0$, we get $\mathrm{Im}(z_i) \le 0$, as desired. □

Lemma 3.10. *If $p \in \mathbb{R}[z_1, \dots, z_m]$ is real stable, then so is $p(z_1 + x, \dots, z_m + x)$ for a new variable x.*

Proof. Say $(z_1, \dots, z_m, x) \in \mathbb{C}^m$ is a root of $p(z_1 + x, \dots, z_m + x)$. Then $(z_1 + x, \dots, z_m + x)$ is a root of $p(z_1, \dots, z_m)$. Since p is real stable, there is an i such that $\mathrm{Im}(z_i + x) \le 0$. But, then either $\mathrm{Im}(x) \le 0$ or $\mathrm{Im}(z_i) \le 0$, as desired. □

3.3 Facts from Linear Algebra

For a Hermitian matrix $M \in \mathbb{C}^{d \times d}$, we write the characteristic polynomial of M in terms of a variable x as

$$\chi[M](x) = \det(xI - M).$$

We also write the characteristic polynomial in terms of the square of x as

$$\chi[M](x^2) = \det(x^2 I - M).$$

For $1 \le k \le n$, we write $\sigma_k(M)$ to denote the sum of all principal $k \times k$ minors of M, in particular,

$$\chi[M](x) = \sum_{k=0}^{d} x^{d-k}(-1)^k \sigma_k(M).$$

The following lemma follows from the Cauchy-Binet identity. See [MSS13b] for the proof.

Lemma 3.11. *For vectors $v_1, \ldots, v_m \in \mathbb{R}^d$ and scalars $z_1, \ldots, z_m$,*

$$\det\left(xI + \sum_{i=1}^{m} z_i v_i v_i^{\intercal}\right) = \sum_{k=0}^{d} x^{d-k} \sum_{S \subseteq \binom{[m]}{k}} z^S \sigma_k\big(\sum_{i \in S} v_i v_i^{\intercal}\big).$$

In particular, for $z_1 = \ldots = z_m = -1$,

$$\det\left(xI - \sum_{i=1}^{m} v_i v_i^{\intercal}\right) = \sum_{k=0}^{d} x^{d-k} (-1)^k \sum_{S \subseteq \binom{[m]}{k}} \sigma_k\big(\sum_{i \in S} v_i v_i^{\intercal}\big).$$

The following is Jacboi's formula for the derivative of the determinant of a matrix.

Theorem 3.12. *For an invertible matrix A which is a differentiable function of t,*

$$\partial_t \det(A) = \det(A) \cdot \mathrm{Tr}(A^{-1} \partial_t A).$$

Lemma 3.13. *For an invertible matrix A which is a differentiable function of t,*

$$\frac{\partial A^{-1}}{\partial t} = -A^{-1}(\partial_t A) A^{-1}.$$

Proof. Differentiating both sides of the identity $A^{-1}A = I$ with respect to t, we get

$$A^{-1}\frac{\partial A}{\partial t} + \frac{\partial A^{-1}}{\partial t} A = 0.$$

Rearranging the terms and multiplying with A^{-1} gives the lemma's conclusion. □

The following two standard facts about trace will be used throughout the paper. First, for $A \in \mathbb{R}^{k \times n}$ and $B \in \mathbb{R}^{n \times k}$,

$$\mathrm{Tr}(AB) = \mathrm{Tr}(BA).$$

Secondly, for positive semidefinite matrices A, B of the same dimension,

$$\mathrm{Tr}(AB) \geq 0.$$

Also, we use the fact that for any positive semidefinite matrix A and any Hermitian matrix B, BAB is positive semidefinite.

Lemma 3.14. *If $A, B \in \mathbb{R}^{n \times n}$ are PD matrices and $A \preceq B$, then $B^{-1} \preceq A^{-1}$.*

Proof. Since $A \preceq B$,

$$B^{-1/2} A B^{-1/2} \preceq B^{-1/2} B B^{-1/2} = I.$$

So,

$$B^{1/2} A^{-1} B^{1/2} = (B^{-1/2} A B^{-1/2})^{-1} \succeq I$$

Multiplying both sides of the above by $B^{-1/2}$, we get

$$A^{-1} = B^{-1/2} B^{1/2} A^{-1} B^{1/2} B^{-1/2} \succeq B^{-1/2} I B^{-1/2} = B^{-1}.$$

□

4 The Mixed Characteristic Polynomial

For a probability distribution μ, let d_μ be the degree of the polynomial g_μ.

Theorem 4.1. *For $v_1, \ldots, v_m \in \mathbb{R}^d$ and a homogeneous probability distribution $\mu : [m] \to \mathbb{R}_+$,*

$$x^{d_\mu - d} \mathop{\mathbb{E}}_{S\sim\mu} \chi\left[\sum_{i\in S} 2v_iv_i^\intercal\right](x^2) = \prod_{i=1}^m \left(1 - \partial_{z_i}^2\right)\left(g_\mu(x\mathbf{1}+z)\cdot \det\left(xI + \sum_{i=1}^m z_iv_iv_i^\intercal\right)\right)\Bigg|_{z_1=\ldots=z_m=0}. \tag{7}$$

We call the polynomial $\mathbb{E}_{S\sim\mu}\chi[\sum_{i\in S} 2v_iv_i^\intercal](x^2)$ the *mixed characteristic polynomial* and we denote it by $\mu[v_1,\ldots,v_m](x)$.

Proof. For $S \subseteq [m]$, let $z^{2S} = \prod_{i\in S} z_i^2$. By Lemma 3.11, the coefficient of z^{2S} in

$$g_\mu(x\mathbf{1}+z)\cdot\det(xI + \sum_{i=1}^m z_iv_iv_i^\intercal)$$

is equal to

$$\left(\prod_{i\in S}\partial_{z_i}^2\right)\left(g_\mu(x\mathbf{1}+z)\cdot\det\left(xI+\sum_{i=1}^m z_iv_iv_i^\intercal\right)\right)\Bigg|_{z_1=\ldots=z_m=0}.$$

Each of the two polynomials $g_\mu(x\mathbf{1}+z)$ and $\det(xI + \sum_{i=1}^m z_iv_iv_i^\intercal)$ is multi-linear in $z_1,\ldots,z_m$. Therefore, for $k = |S|$, the above is equal to

$$2^k \cdot \left(\prod_{i\in S}\partial_{z_i}\right) g_\mu(x\mathbf{1}+z)\Bigg|_{z_1=\ldots=z_m=0} \cdot \left(\prod_{i\in S}\partial_{z_i}\right)\det\left(xI+\sum_{i=1}^m z_iv_iv_i^\intercal\right)\Bigg|_{z_1=\ldots=z_m=0}. \tag{8}$$

Since g_μ is a homogeneous polynomial of degree d_μ, the first term in the above is equal to

$$x^{d_\mu-k}\mathbb{P}_{T\sim\mu}\left[S\subseteq T\right].$$

And, by Lemma 3.11, the second term of (8) is equal to

$$x^{d-k}\sigma_k\left(\sum_{i\in S} v_iv_i^\intercal\right).$$

Applying the above identities for all $S\subseteq[m]$,

$$\begin{aligned}
&\prod_{i=1}^m \left(1-\partial_{z_i}^2\right)\left(g_\mu(x\mathbf{1}+z)\cdot\det\left(xI+\sum_{i=1}^m z_iv_iv_i^\intercal\right)\right)\Bigg|_{z_1=\ldots=z_m=0}\\
&\qquad= \sum_{k=0}^m (-1)^k \sum_{S\subseteq\binom{[m]}{k}} \left(\prod_{i\in S}\partial_{z_i}^2\right)\left(g_\mu(x\mathbf{1}+z)\cdot\det\left(xI+\sum_{i=1}^m z_iv_iv_i^\intercal\right)\right)\Bigg|_{z_1=\ldots=z_m=0}\\
&\qquad= \sum_{k=0}^d (-1)^k 2^k x^{d_\mu+d-2k} \sum_{S\in\binom{[m]}{k}} \mathbb{P}_{T\sim\mu}\left[S\subseteq T\right]\cdot\sigma_k\left(\sum_{i\in S} v_iv_i^\intercal\right)\\
&\qquad= x^{d_\mu-d}\mathop{\mathbb{E}}_{S\sim\mu}\chi\left[\sum_{i\in S}2v_iv_i^\intercal\right](x^2).
\end{aligned}$$

The last identity uses Lemma 3.11. $\square$

Corollary 4.2. *If μ is a strongly Rayleigh probability distribution, then the mixed characteristic polynomial is real-rooted.*

Proof. First, by Theorem 3.6,

$$\det\left(xI + \sum_{i=1}^{m} z_i v_i v_i^{\intercal}\right)$$

is real stable. Since μ is strongly Rayleigh, $g_\mu(z)$ is real stable. So, by Lemma 3.10, $g_\mu(x\mathbf{1}+z)$ is real stable. The product of two real stable polynomials is also real stable, so

$$g_\mu(x\mathbf{1}+z)\cdot\det\left(xI + \sum_{i=1}^{m} z_i v_i v_i^{\intercal}\right)$$

is real stable. Corollary 3.8 implies that

$$\prod_{i=1}^{m}\left(1-\partial_{z_i}^2\right)\left(g_\mu(x\mathbf{1}+z)\cdot\det\left(xI + \sum_{i=1}^{m} z_i v_i v_i^{\intercal}\right)\right)$$

is real stable as well. Wagner [Wag11, Lemma 2.4(d)] tells us that real stability is preserved under setting variables to real numbers, so

$$\prod_{i=1}^{m}\left(1-\partial_{z_i}^2\right)\left(g_\mu(x\mathbf{1}+z)\cdot\det\left(xI + \sum_{i=1}^{m} z_i v_i v_i^{\intercal}\right)\right)\Bigg|_{z_1=\ldots=z_m=0}$$

is a univariate real-rooted polynomial. The mixed characteristic polynomial is equal to the above polynomial up to a term $x^{d_\mu - d}$. So, the mixed characteristic polynomial is also real rooted. □

Now, we use the real-rootedness of the mixed characteristic polynomial to show that the characteristic polynomials of the set of vectors assigned to any set S with nonzero probability in μ form an interlacing family. For a homogeneous strongly Rayleigh measure μ, let

$$\mathcal{F} = \{S : \mu(S) > 0\},$$

and for $s_1,\ldots,s_k \in \{0,1\}$ let $\mathcal{F}_{s_1,\ldots,s_k}$ be as defined in Definition 3.3. For any $S\in\mathcal{F}$, let

$$q_S(x) = \mu(S)\cdot\chi\left[\sum_{i\in S} 2v_i v_i^{\intercal}\right](x^2).$$

Theorem 4.3. *The polynomials $\{q_S\}_{S\in\mathcal{F}}$ form an interlacing family.*

Proof. For $1\le k\le m$ and $s_1,\ldots,s_k\in\{0,1\}$, let $\mu_{s_1,\ldots,s_k}$ be μ conditioned on the sets $S\in\mathcal{F}_{s_1,\ldots,s_k}$, i.e., μ conditioned on $i\in S$ for all $i\le k$ where $s_i=1$ and $i\notin S$ for all $i\le k$ where $s_i=0$. We inductively write the generating polynomial of $\mu_{s_1,\ldots,s_k}$ in terms of g_μ. Say we have written $g_{\mu_{s_1,\ldots,s_k}}$ in terms of g_μ. Then, we can write,

$$g_{\mu_{s_1,\ldots,s_k,1}}(z) = \frac{z_{k+1}\cdot\partial_{z_{k+1}} g_{\mu_{s_1,\ldots,s_k}}(z)}{\partial_{z_{k+1}} g_{\mu_{s_1,\ldots,s_k}}(z)\big|_{z_i=1}}, \tag{9}$$

$$g_{\mu_{s_1,\ldots,s_k,0}}(z) = \frac{g_{\mu_{s_1,\ldots,s_k}}(z)\big|_{z_{k+1}=0}}{g_{\mu_{s_1,\ldots,s_k}}(z)\big|_{z_{k+1}=0, z_i=1 \text{ for } i\neq k+1}}. \tag{10}$$

Note that the denominators of both equations are just normalizing constants. The above polynomials are well defined if the normalizing constants are nonzero, i.e., if the set $\mathcal{F}_{s_1,\ldots,s_k,s_{k+1}}$ is nonempty. Since the real stable polynomials are closed under differentiation and substitution, for any $1 \leq k \leq m$, and $s_1, \ldots, s_k \in \{0, 1\}$, if $g_{\mu_{s_1,\ldots,s_k}}$ is well defined, it is real stable, so $\mu_{s_1,\ldots,s_k}$ is a strongly Rayleigh distribution.

Now, for $s_1, \ldots, s_k \in \{0, 1\}$, let

$$q_{s_1,\ldots,s_k}(x) = \sum_{S \in \mathcal{F}_{s_1,\ldots,s_k}} q_S(x).$$

Since $\mu_{s_1,\ldots,s_k}$ is strongly Rayleigh, by Corollary 4.2, $q_{s_1,\ldots,s_k}(x)$ is real rooted.

By Lemma 3.5, to prove the theorem it is enough to show that if $\mathcal{F}_{s_1,\ldots,s_k,0}$ and $\mathcal{F}_{s_1,\ldots,s_k,1}$ are nonempty, then for any $0 < \lambda < 1$,

$$\lambda \cdot q_{s_1,\ldots,s_k,1}(x) + (1 - \lambda) \cdot q_{s_1,\ldots,s_k,0}(x)$$

is real rooted. Equivalently, by Corollary 4.2, it is enough to show that for any $0 < \lambda < 1$,

$$\lambda \cdot g_{\mu_{s_1,\ldots,s_k,1}}(z) + (1 - \lambda) \cdot g_{\mu_{s_1,\ldots,s_k,0}}(z) \tag{11}$$

is real stable. Let us write,

$$\begin{aligned} g_{\mu_{s_1,\ldots,s_k}}(z) &= z_{k+1} \cdot \partial_{z_{k+1}} g_{\mu_{s_1,\ldots,s_k}}(z) + g_{\mu_{s_1,\ldots,s_k}}(z)\big|_{z_{k+1}=0} \\ &= \alpha \cdot g_{\mu_{s_1,\ldots,s_k,1}}(z) + \beta \cdot g_{\mu_{s_1,\ldots,s_k,0}}(z), \end{aligned}$$

for some $\alpha, \beta > 0$. The second identity follows by (9) and (10). Let $\lambda_{k+1} > 0$ such that

$$\frac{\lambda_{k+1} \cdot \alpha}{\lambda} = \frac{\beta}{1 - \lambda}. \tag{12}$$

Since $g_{\mu_{s_1,\ldots,s_k}}$ is real stable, by Lemma 3.9

$$g_{\mu_{s_1,\ldots,s_k}}(z_1, \ldots, z_k, \lambda_{k+1} \cdot z_{k+1}, z_{k+2}, \ldots, z_m)$$

is real stable. But, by (12) the above polynomial is just a multiple of (11). So, (11) is real stable. □

5 An Extension of [MSS13b] Multivariate Barrier Argument

In this section we upper-bound the roots of the mixed characteristic polynomial in terms of the marginal probabilities of elements of $[m]$ in μ and the maximum of the squared norm of vectors $v_1, \ldots, v_m$.

Theorem 5.1. *Given vectors $v_1, \ldots, v_m \in \mathbb{R}^d$, and a homogeneous strongly Rayleigh probability distribution $\mu : [m] \to \mathbb{R}_+$, such that the marginal probability of each element $i \in [m]$ is at most ϵ_1, $\sum_{i=1}^m v_i v_i^\intercal = I$ and $\|v_i\|^2 \leq \epsilon_2$, the largest root of $\mu[v_1, \ldots, v_m](x)$ is at most $4(2\epsilon + \epsilon^2)$, where $\epsilon = \epsilon_1 + \epsilon_2$,*

First, similar to [MSS13b] we derive a slightly different expression.

Lemma 5.2. *For any probability distribution μ and vectors $v_1, \dots, v_m \in \mathbb{R}^d$ such that $\sum_{i=1}^m v_i v_i^\intercal = I$,*

$$x^{d_\mu - d} \mu[v_1, \dots, v_m](x) = \prod_{i=1}^m \left(1 - \partial_{y_i}^2\right) \left(g_\mu(y) \cdot \det\left(\sum_{i=1}^m y_i v_i v_i^\intercal \right) \right) \Bigg|_{y_1 = \dots = y_m = x}.$$

Proof. This is because for any differentiable function f, $\partial_{y_i} f(y_i)|_{y_i = z_i + x} = \partial_{z_i} f(z_i + x)$. □

Let

$$Q(y_1, \dots, y_m) = \prod_{i=1}^m \left(1 - \partial_{y_i}^2\right) \left(g_\mu(y) \cdot \det\left(\sum_{i=1}^m y_i v_i v_i^\intercal \right) \right).$$

Then, by the above lemma, the maximum root of $Q(x, \dots, x)$ is the same as the maximum root of $\mu[v_1, \dots, v_m](x)$. In the rest of this section we upper-bound the maximum root of $Q(x, \dots, x)$.

It directly follows from the proof of Theorem 5.1 in [MSS13b] that the maximum root of $Q(x, \dots, x)$ is at most $(1 + \sqrt{\epsilon})^2$. But, in our setting, any upper-bound that is more than 1 obviously holds, as for any $S \subseteq [m]$,

$$\left\| \sum_{i=1}^m v_i v_i^\intercal \right\| \leq 1.$$

The main difficulty that we are facing is to prove an upper-bound of $O(\epsilon)$ on the maximum root of $Q(x, \dots, x)$.

We use an extension of the multivariate barrier argument of [MSS13b] to upper-bound the maximum root of Q. We manage to prove a significantly smaller upper-bound because we apply $1 - \partial_{y_i}^2$ operators as opposed to the $1 - \partial_{y_i}$ operators used in [MSS13b]. This allows us to impose significantly smaller shifts on the barrier upper-bound in our inductive argument.

Definition 5.3. *For a multivariate polynomial $p(z_1, \dots, z_m)$, we say $z \in \mathbb{R}^m$ is above all roots of p if for all $t \in \mathbb{R}_+^m$,*

$$p(z + t) > 0.$$

We use $\mathbf{Ab}_p$ to denote the set of points which are above all roots of p.

We use the same barrier function defined in [MSS13b].

Definition 5.4. *For a real stable polynomial p, and $z \in \mathbf{Ab}_p$, the barrier function of p in direction i at z is*

$$\Phi_p^i(z) := \frac{\partial_{z_i} p(z)}{p(z)} = \partial_{z_i} \log p(z).$$

To analyze the rate of change of the barrier function with respect to the $1 - \partial_{z_i}^2$ operator, we need to work with the second derivative of p as well. We define,

$$\Psi_p^i(z) := \frac{\partial_{z_i}^2 p(z)}{p(z)}.$$

Equivalently, for a univariate restriction $q_{z,i}(t) = p(z_1, \ldots, z_{i-1}, t, z_{i+1}, \ldots, z_m)$*, with real roots* $\lambda_1, \ldots, \lambda_r$ *we can write,*

$$\Phi_p^i(z) = \frac{q'_{z,i}(z_i)}{q_{z,i}(z_i)} = \sum_{j=1}^{r} \frac{1}{z_i - \lambda_j},$$
$$\Psi_p^i(z) = \frac{q''_{z,i}(z_i)}{q_{z,i}(z_i)} = \sum_{1 \le j < k \le r} \frac{2}{(z_i - \lambda_j)(z_i - \lambda_k)}.$$

The following lemma is immediate from the above definition.

Lemma 5.5. *If* p *is real stable and* $z \in \mathbf{Ab}_p$*, then for all* $i \le m$*,*

$$\Psi_p^i(z) \le \Phi_p^i(z)^2.$$

Proof. Since $z \in \mathbf{Ab}_p$, $z_i > \lambda_j$ for all $1 \le j \le r$, so,

$$\Phi_p^i(z)^2 - \Psi_p^i(z) = \left(\sum_{j=1}^{r} \frac{1}{z_i - \lambda_j} \right)^2 - \sum_{1 \le j < k \le r} \frac{2}{(z_i - \lambda_j)(z_i - \lambda_k)} = \sum_{j=1}^{r} \frac{1}{(z_i - \lambda_j)^2} > 0.$$

□

The following monotonicity and convexity properties of the barrier functions are proved in [MSS13b].

Lemma 5.6. *Suppose* $p(.)$ *is a real stable polynomial and* $z \in \mathbf{Ab}_p$*. Then, for all* $i, j \le m$ *and* $\delta \ge 0$*,*

$$\Phi_p^i(z + \delta \mathbf{1}_j) \le \Phi_p^i(z) \text{ and,} \qquad \textit{(monotonicity)} \tag{13}$$
$$\Phi_p^i(z + \delta \mathbf{1}_j) \le \Phi_p^i(z) + \delta \cdot \partial_{z_j} \Phi_p^i(z + \delta \mathbf{1}_j) \qquad \textit{(convexity).} \tag{14}$$

Recall that the purpose of the barrier functions Φ_p^i is to allow us to reason about the relationship between $\mathbf{Ab}_p$ and $\mathbf{Ab}_{p - \partial_{z_i}^2 p}$; the monotonicity property and Lemma 5.5 imply the following lemma.

Lemma 5.7. *If* p *is real stable and* $z \in \mathbf{Ab}_p$ *is such that* $\Phi_p^i(z) < 1$*, then* $z \in \mathbf{Ab}_{p - \partial_{z_i}^2 p}$*.*

Proof. Fix a nonnegative vector t. Since Φ is nonincreasing in each coordinate,

$$\Phi_p^i(z + t) \le \Phi_p^i(z) < 1.$$

Since $z + t \in \mathbf{Ab}_p$, by Lemma 5.5,

$$\Psi_p^i(z + t) \le \Phi_p^i(z + t)^2 < 1.$$

Therefore,

$$\partial_{z_i}^2 p(z + t) < p(z + t) \Rightarrow (1 - \partial_{z_i}^2) p(z + t) > 0,$$

as desired. □

We use an inductive argument similar to [MSS13b]. We argue that when we apply each operator $(1-\partial^2_{z_j})$, the barrier functions, $\Phi^i_p(z)$, do not increase by shifting the upper bound along the direction $\mathbf{1}_j$. As we would like to prove a significantly smaller upper bound on the maximum root of the mixed characteristic polynomial, we may only shift along direction $\mathbf{1}_j$ by a small amount. In the following lemma we show that when we apply the $(1-\partial^2_{z_j})$ operator we only need to shift the upper bound proportional to $\Phi^j_p(z)$ along the direction $\mathbf{1}_j$.

Lemma 5.8. *Suppose that $p(z_1,\ldots,z_m)$ is real stable and $z \in \mathbf{Ab}_p$. If for $\delta > 0$,*

$$\frac{2}{\delta}\Phi^j_p(z) + \Phi^j_p(z)^2 \le 1,$$

then, for all i,

$$\Phi^i_{p-\partial^2_{z_j}p}(z+\delta\cdot\mathbf{1}_j) \le \Phi^i_p(z).$$

To prove the above lemma we first need to prove a technical lemma to upper-bound $\frac{\partial_{z_i}\Psi^j_p(z)}{\partial_{z_i}\Phi^j_p(z)}$. We use the following characterization of the bivariate real stable polynomials proved by Lewis, Parrilo, and Ramana [LPR05]. The following form is stated in [BB10, Cor 6.7].

Lemma 5.9. *If $p(z_1,z_2)$ is a bivariate real stable polynomial of degree d, then there exist $d\times d$ positive semidefinite matrices A, B and a Hermitian matrix C such that*

$$p(z_1,z_2) = \pm\det(z_1A + z_2B + C).$$

Lemma 5.10. *Suppose that p is real stable and $z \in \mathbf{Ab}_p$, then for all $i,j \le m$,*

$$\frac{\partial_{z_i}\Psi^j_p(z)}{\partial_{z_i}\Phi^j_p(z)} \le 2\Phi^j_p(z).$$

Proof. If $i = j$, then we consider the univariate restriction $q_{z,i}(z_i) = \prod_{k=1}^r (z_i - \lambda_k)$. Then,

$$\frac{\partial_{z_i}\sum_{1\le k<\ell\le r}\frac{2}{(z_i-\lambda_k)(z_i-\lambda_\ell)}}{\partial_{z_i}\sum_{k=1}^r\frac{1}{(z_i-\lambda_k)}} = \frac{\sum_{k\neq\ell}\frac{-2}{(z_i-\lambda_k)^2(z_i-\lambda_\ell)}}{\sum_{k=1}^r\frac{-1}{(z_i-\lambda_k)^2}} \le \sum_{\ell=1}^r\frac{2}{(z_i-\lambda_\ell)} = 2\Phi^j_p(z).$$

The inequality uses the assumption that $z \in \mathbf{Ab}_p$.

If $i \neq j$, we fix all variables other than z_i, z_j and we consider the bivariate restriction

$$q_{z,ij}(z_i,z_j) = p(z_1,\ldots,z_m).$$

By Lemma 5.9, there are Hermitian positive semidefinite matrices B_i, B_j, and a Hermitian matrix C such that

$$q_{z,ij}(z_i,z_j) = \pm\det(z_iB_i + z_jB_j + C).$$

Let $M = z_iB_i + z_jB_j + C$. Marcus, Spielman, and Srivastava [MSS13b, Lem 5.7] observed that the sign is always positive, that $B_i + B_j$ is positive definite. In addition, M is positive definite since $B_i + B_j$ is positive definite and $z \in \mathbf{Ab}_p$.

By Theorem 3.12, the barrier function in direction j can be expressed as

$$\Phi^j_p(z) = \frac{\partial_{z_j}\det(M)}{\det(M)} = \frac{\det(M)\operatorname{Tr}(M^{-1}B_j)}{\det(M)} = \operatorname{Tr}(M^{-1}B_j). \tag{15}$$

By another application of Theorem 3.12,

$$
\begin{aligned}
\Psi_p^j(z) = \frac{\partial_{z_j}^2 \det(M)}{\det(M)} &= \frac{\partial_{z_j}(\det(M)\operatorname{Tr}(M^{-1}B_j))}{\det(M)} \\
&= \frac{\det(M)\operatorname{Tr}(M^{-1}B_j)^2}{\det(M)} + \frac{\det(M)\operatorname{Tr}((\partial_{z_j}M^{-1})B_j)}{\det(M)} \\
&= \operatorname{Tr}(M^{-1}B_j)^2 + \operatorname{Tr}(-M^{-1}B_jM^{-1}B_j) \\
&= \operatorname{Tr}(M^{-1}B_j)^2 - \operatorname{Tr}((M^{-1}B_j)^2).
\end{aligned}
$$

The second to last identity uses Lemma 3.13. Next, we calculate $\partial_{z_i}\Phi_p^j$ and $\partial_{z_i}\Psi_p^j$. First, by another application of Lemma 3.13,

$$\partial_{z_i} M^{-1}B_j = -M^{-1}B_iM^{-1}B_j =: L.$$

Therefore,

$$\partial_{z_i}\Phi_p^j(z) = \partial_{z_i}\operatorname{Tr}(M^{-1}B_j) = \operatorname{Tr}(L),$$

and

$$
\begin{aligned}
\partial_{z_i}\Psi_p^j(z) &= \partial_{z_i}\operatorname{Tr}(M^{-1}B_j)^2 - \partial_{z_i}\operatorname{Tr}((M^{-1}B_j)^2) \\
&= 2\operatorname{Tr}(M^{-1}B_j)\operatorname{Tr}(L) - \operatorname{Tr}\left(L(M^{-1}B_j) + (M^{-1}B_j)L\right) \\
&= 2\operatorname{Tr}(M^{-1}B_j)\operatorname{Tr}(L) - 2\operatorname{Tr}(LM^{-1}B_j).
\end{aligned}
$$

Putting above equations together we get

$$
\begin{aligned}
\frac{\partial_{z_i}\Psi_p^j(z)}{\partial_{z_i}\Phi_p^j(z)} &= 2\frac{\operatorname{Tr}(M^{-1}B_j)\operatorname{Tr}(L) - \operatorname{Tr}(LM^{-1}B_j)}{\operatorname{Tr}(L)} \\
&= 2\operatorname{Tr}(M^{-1}B_j) - 2\frac{\operatorname{Tr}(LM^{-1}B_j)}{\operatorname{Tr}(L)} \\
&= 2\Phi_p^j(z) - 2\frac{\operatorname{Tr}(LM^{-1}B_j)}{\operatorname{Tr}(L)}
\end{aligned}
$$

where we used (15).

To prove the lemma it is enough to show that $\frac{\operatorname{Tr}(LM^{-1}B_j)}{\operatorname{Tr}(L)} \geq 0$. We show that both the numerator and the denominator are nonpositive. First,

$$\operatorname{Tr}(L) = -\operatorname{Tr}(M^{-1}B_iM^{-1}B_j) \leq 0$$

where we used that $M^{-1}B_iM^{-1}$ and B_j are positive semidefinite and the fact that the trace of the product of positive semidefinite matrices is nonnegative. Secondly,

$$\operatorname{Tr}(LM^{-1}B_j) = \operatorname{Tr}(-M^{-1}B_iM^{-1}B_jM^{-1}B_j) = -\operatorname{Tr}(B_iM^{-1}B_jM^{-1}B_jM^{-1}) \leq 0,$$

where we again used that $M^{-1}B_jM^{-1}B_jM^{-1}$ and B_i are positive semidefinite and the trace of the product of two positive semidefinite matrices is nonnegative. □

Proof of Lemma 5.8. We write ∂_i instead of ∂_{z_i} for the ease of notation. First, we write $\Phi^i_{p-\partial_j^2 p}$ in terms of Φ^i_p and Ψ^j_p and $\partial_i \Psi^j_p$.

$$\begin{aligned}\Phi^i_{p-\partial_j^2 p} &= \frac{\partial_i(p-\partial_j^2 p)}{p-\partial_j^2 p} \\ &= \frac{\partial_i((1-\Psi^j_p)p)}{(1-\Psi^j_p)p} \\ &= \frac{(1-\Psi^j_p)(\partial_i p)}{(1-\Psi^j_p)p} + \frac{(\partial_i(1-\Psi^j_p))p}{(1-\Psi^j_p)p} \\ &= \Phi^i_p - \frac{\partial_i \Psi^j_p}{1-\Psi^j_p}.\end{aligned}$$

We would like to show that $\Phi^i_{p-\partial_j^2 p}(z+\delta\mathbf{1}_j) \le \Phi^i_p(z)$. Equivalently, it is enough to show that

$$-\frac{\partial_i \Psi^j_p(z+\delta\mathbf{1}_j)}{1-\Psi^j_p(z+\delta\mathbf{1}_j)} \le \Phi^i_p(z) - \Phi^i_p(z+\delta\mathbf{1}_j).$$

By (14) of Lemma 5.6, it is enough to show that

$$-\frac{\partial_i \Psi^j_p(z+\delta\mathbf{1}_j)}{1-\Psi^j_p(z+\delta\mathbf{1}_j)} \le \delta \cdot (-\partial_j \Phi^i_p(z+\delta\mathbf{1}_j)).$$

By (13) of Lemma 5.6, $\delta \cdot (-\partial_j \Phi^i_p(z+\delta\mathbf{1}_j)) > 0$ so we may divide both sides of the above inequality by this term and obtain

$$\frac{-\partial_i \Psi^j_p(z+\delta\mathbf{1}_j)}{-\delta \cdot \partial_i \Phi^j_p(z+\delta\mathbf{1}_j)} \cdot \frac{1}{1-\Psi^j_p(z+\delta\mathbf{1}_j)} \le 1,$$

where we also used $\partial_j \Phi^i_p = \partial_i \Phi^j_p$. By Lemma 5.10, $\frac{\partial_i \Psi^j_p}{\partial_i \Phi^j_p} \le 2\Phi^j_p$. So, we can write,

$$\frac{2}{\delta}\Phi^j_p(z+\delta\mathbf{1}_j) \cdot \frac{1}{1-\Psi^j_p(z+\delta\mathbf{1}_j)} \le 1.$$

By Lemma 5.5 and (13) of Lemma 5.6,

$$\begin{aligned}&\Phi^j_p(z+\delta\mathbf{1}_j) \le \Phi^j_p(z), \\ &\Psi^j_p(z+\delta\mathbf{1}_j) \le \Phi^j_p(z+\delta\mathbf{1}_j)^2 \le \Phi^j_p(z)^2.\end{aligned}$$

So, it is enough to show that

$$\frac{2}{\delta}\Phi^j_p(z) \cdot \frac{1}{1-\Phi^j_p(z)^2} \le 1$$

Using $\Phi^j_p(z) < 1$ we may multiply both sides with $1-\Phi^j_p(z)$ and we obtain,

$$\frac{2}{\delta}\Phi^j_p(z) + \Phi^j_p(z)^2 \le 1,$$

as desired. □

Now, we are read to prove Theorem 5.1.

Proof of Theorem 5.1. Let

$$p(y_1, \ldots, y_m) = g_\mu(y) \cdot \det\left(\sum_{i=1}^{m} y_i v_i v_i^\intercal\right).$$

Set $\epsilon = \epsilon_1 + \epsilon_2$ and

$$\delta = t = \sqrt{2\epsilon + \epsilon^2}.$$

For any $z \in \mathbb{R}^m$ with positive coordinates, $g_\mu(z) > 0$, and additionally

$$\det\left(\sum_{i=1}^{m} z_i v_i v_i^\intercal\right) > 0.$$

Therefore, for every $t > 0$, $t\mathbf{1} \in \mathbf{Ab}_p$.

Now, by Theorem 3.12,

$$\begin{aligned}\Phi_p^i(y) &= \frac{(\partial_i g_\mu(y)) \cdot \det(\sum_{i=1}^m y_i v_i v_i^\intercal)}{g_\mu(y) \cdot \det(\sum_{i=1}^m y_i v_i v_i^\intercal)} + \frac{g_\mu(y) \cdot (\partial_i \det(\sum_{i=1}^m y_i v_i v_i^\intercal))}{g_\mu(y) \cdot \det(\sum_{i=1}^m y_i v_i v_i^\intercal)} \\ &= \frac{\partial_i g_\mu(y)}{g_\mu(y)} + \mathrm{Tr}\left(\left(\sum_{i=1}^{m} y_i v_i v_i^\intercal\right)^{-1} v_i v_i^\intercal\right)\end{aligned}$$

Therefore, since g_μ is homogeneous,

$$\begin{aligned}\Phi_p^i(t\mathbf{1}) &= \frac{1}{t} \cdot \frac{\partial_i g_\mu(\mathbf{1})}{g_\mu(\mathbf{1})} + \frac{\|v_i\|^2}{t} \\ &= \frac{\mathbb{P}_{S\sim\mu}[i \in S]}{t} + \frac{\|v_i\|^2}{t} \leq \frac{\epsilon_1}{t} + \frac{\epsilon_2}{t} = \frac{\epsilon}{t}.\end{aligned}$$

The second identity uses (1). Let $\phi = \epsilon/t$. Using $t = \delta$, it follows that

$$\frac{2}{\delta}\phi + \phi^2 = \frac{2\epsilon}{t^2} + \frac{\epsilon^2}{t^2} = 1.$$

For $k \in [m]$ define

$$p_k(y_1, \ldots, y_m) = \prod_{i=1}^{k} \left(1 - \partial_{y_i}^2\right) \left(g_\mu(y) \cdot \det\left(\sum_{i=1}^{m} y_i v_i v_i^\intercal\right)\right),$$

and note that $p_m = Q$. Let x^0 be the all-t vector and x^k be the vector that is $t + \delta$ in the first k coordinates and t in the rest. By inductively applying Lemma 5.7 and Lemma 5.8 for any $k \in [m]$, x^k is above all roots of p_k and for all i,

$$\Phi_{p_k}^i(x_k) \leq \phi \Rightarrow \frac{2}{\delta}\Phi_{p_k}^i(x_i) + \Phi_{p_k}^i(x_i)^2 \leq 1.$$

Therefore, the largest root of $\mu[v_1, \ldots, v_m](x)$ is at most

$$t + \delta = 2\sqrt{2\epsilon + \epsilon^2}.$$

□

Proof of Theorem 1.2. Let $\epsilon = \epsilon_1 + \epsilon_2$ as always. Theorem 5.1 implies that the largest root of the mixed characteristic polynomial, $\mu[v_1, \ldots, v_m](x)$, is at most $2\sqrt{2\epsilon + \epsilon^2}$. Theorem 4.3 tells us that the polynomials $\{q_S\}_{S:\mu(S)>0}$ form an interlacing family. So, by Theorem 3.4 there is a set $S \subseteq [m]$ with $\mu(S) > 0$ such that the largest root of

$$\det\left(x^2 I - \sum_{i\in S} 2 v_i v_i^\intercal\right)$$

is at most $2\sqrt{2\epsilon + \epsilon^2}$. This implies that the largest root of

$$\det\left(xI - \sum_{i\in S} 2 v_i v_i^\intercal\right)$$

is at most $(2\sqrt{2\epsilon + \epsilon^2})^2$. Therefore,

$$\left\|\sum_{i\in S} v_i v_i^\intercal\right\| = \frac{1}{2}\left\|\sum_{i\in S} 2 v_i v_i^\intercal\right\| \leq \frac{1}{2}(2\sqrt{2\epsilon + \epsilon^2})^2 = 4\epsilon + 2\epsilon^2.$$

□

6 Discussion

Similar to [MSS13b] our main theorem is not algorithmic, i.e., we are not aware of any polynomial time algorithm that for a given homogeneous strongly Rayleigh distribution with small marginal probabilities and for a set of vectors assigned to the underlying elements with small norm finds a sample of the distribution with spectral norm bounded away from 1. Such an algorithm can lead to improved approximation algorithms for the Asymmetric Traveling Salesman Problem.

Although our main theorem can be seen as a generalization of [MSS13b] the bound that we prove on the maximum root of the mixed characteristic polynomial is incomparable to the bound of Theorem 1.1. In Corollary 1.3 we used our main theorem to prove Weaver's KS_r conjecture [Wea04] for $r > 4$. It is an interesting question to see if the dependency on ϵ in our multivariate barrier can be improved, and if one can reprove KS_2 using our machinery.

Acknowledgement

We would like to thank Adam Marcus, Dan Spielman, and Nikhil Srivastava for stimulating discussions regarding the main obstacles in generalizing their proof of the Kadison-Singer problem.

References

[AA91] Charles A. Akemann and Joel Anderson. *Lyapunov theorems for operator algebras*, volume 94. Memoirs of the American Mathematical Society, 1991. 1

[AGM+10] Arash Asadpour, Michel X. Goemans, Aleksander Madry, Shayan Oveis Gharan, and Amin Saberi. An $O(\log n/\log\log n)$-approximation Algorithm for the Asymmetric Traveling Salesman Problem. In *SODA*, pages 379–389, 2010. 5, 7

[AO15] Nima Anari and Shayan Oveis Gharan. Effective-Resistance-Reducing Flows and Asymmetric TSP. In *FOCS*, pages 20–39, 2015. 1, 7, 8, 10

[BB08] Julius Borcea and Petter Brändén. Applications of stable polynomials to mixed determinants: Johnson's conjectures, unimodality, and symmetrized Fischer products. *Duke Math. Journal*, 143(2):205–223, 2008. 12

[BB10] J. Borcea and P. Brädén. Multivariate Pólya-Schur classification problems in the Weyl algebra. *Proceedings of the London Mathematical Society*, 101(3):73–104, 2010. 20

[BBL09] Julius Borcea, Petter Branden, and Thomas M. Liggett. Negative dependence and the geometry of polynomials. *Journal of American Mathematical Society*, 22:521–567, 2009. 1, 2, 4

[BSS14] Joshua D. Batson, Daniel A. Spielman, and Nikhil Srivastava. Twice-Ramanujan Sparsifiers. *SIAM Review*, 56(2):315–334, 2014. 11

[Ded92] Jean Pierre Dedieu. Obreschkoff's theorem revisited: what convex sets are contained in the set of hyperbolic polynomials? *Journal of Pure and Applied Algebra*, 81(3):269–278, 1992. 12

[Goe06] Michel X. Goemans. Minimum Bounded Degree Spanning Trees. In *FOCS*, pages 273–282, 2006. 10

[HO14] Nicholas J. A. Harvey and Neil Olver. Pipage Rounding, Pessimistic Estimators and Matrix Concentration. In *SODA*, pages 926–945, 2014. 7

[KS59] Richard V. Kadison and I. M. Singer. Extensions of Pure States. *American Journal of Mathematics*, 81(2):383–400, 1959. 1

[LP13] Russell Lyons and Yuval Peres. *Probability on Trees and Networks*. Cambridge University Press, 2013. To appear. 7

[LPR05] A. Lewis, P. Parrilo, and M. Ramana. The Lax conjecture is true. *Proceedings of the American Mathematical Society*, 133(9), 2005. 20

[LS15] Lap Chi Lau and Mohit Singh. Approximating Minimum Bounded Degree Spanning Trees to within One of Optimal. *J. ACM*, 62(1):1–19, 2015. 10

[MSS13a] Adam Marcus, Daniel A. Spielman, and Nikhil Srivastava. Interlacing Families I: Bipartite Ramanujan Graphs of All Degrees. In *FOCS*, pages 529–537, 2013. 10, 11

[MSS13b] Adam Marcus, Daniel A Spielman, and Nikhil Srivastava. Interlacing Families II: Mixed Characteristic Polynomials and the Kadison-Singer Problem. 2013. 1, 3, 7, 10, 11, 12, 13, 17, 18, 19, 20, 24

[NW61] C. St. J. A. Nash-Williams. Edge disjoint spanning trees of finite graphs. *J. London Math. Soc.*, 36:445–45, 1961. 8

[OS11] Shayan Oveis Gharan and Amin Saberi. The Asymmetric Traveling Salesman Problem on Graphs with Bounded Genus. In *SODA*, pages 967–975, 2011. 7

[OSS11] Shayan Oveis Gharan, Amin Saberi, and Mohit Singh. A Randomized Rounding Approach to the Traveling Salesman Problem. In *FOCS*, pages 550–559, 2011. 2

[PP14] Robin Pemantle and Yuval Peres. Concentration of Lipschitz Functionals of Determinantal and Other Strong Rayleigh Measures. *Combinatorics, Probability and Computing*, 23:140–160, 1 2014. 2

[SV14] Mohit Singh and Nisheeth K. Vishnoi. Entropy, optimization and counting. In *STOC*, pages 50–59, 2014. 5

[Wag11] David G. Wagner. Multivariate stable polynomials: theory and applications. *Bulletin of the American Mathematical Society*, 48(1):53–53, January 2011. 16

[Wea04] Nik Weaver. The Kadison-Singer problem in discrepancy theory. *Discrete Mathematics*, 278(1-3):227–239, 2004. 1, 24