Zeros of Polynomials and their Applications to Theory: A Primer

Nisheeth K. Vishnoi*

October 22, 2013

Abstract

Problems in many different areas of mathematics reduce to questions about the zeros of complex univariate and multivariate polynomials. Recently, several significant and seemingly unrelated results relevant to theoretical computer science have benefited from taking this route: they rely on showing, at some level, that a certain univariate or multivariate polynomial has no zeros in a region. This is achieved by inductively constructing the relevant polynomial via a sequence of operations which *preserve* the property of not having roots in the required region. The goal of this article is to present this viewpoint and to convey why the study of zeros is a natural, powerful, and versatile tool. It is meant to be a gentle introduction for the essential techniques underlying these results, is largely self-contained and aimed at a broad theory audience.

Contents

*Microsoft Research, Bangalore, India. Email: `nisheeth.vishnoi@gmail.com`

1 Introduction

Consider the following results relevant in theoretical computer science:

1. The permanent of an $n \times n$ stochastic matrix is at least $n!/n^n$, [6, 7, 20, 11]. (This has been used to show that every k-regular graph on n vertices has a traveling salesman tour of length at most $(1 + O(1/\sqrt{\log k}))n$, [23].)

2. The polynomial time approximation algorithm for the Traveling Salesman Problem on undirected, unweighted graphs with approximation ratio $3/2 - \varepsilon$, for some constant $\varepsilon > 0$, [10, 16].

3. The seminal result by Lee and Yang [25] in statistical physics that shows the lack of phase transition in the Ising model, and the mean magnetization of the Ising model and the average size of a matching in the monomer-dimer model are both #P-hard to compute, [21].

4. For every d, there is an infinite sequence of d-regular bipartite Ramanujan graphs, whose adjacency matrices have all nontrivial eigenvalues bounded by $2\sqrt{d-1}$, [14].

5. Every transitive graph with m edges and n vertices can be partitioned into $O(m/n)$ edge-disjoint subgraphs of size $O(n)$, each of which approximates the cuts of the original graph up to a constant factor. This is a special case of a spectral discrepancy theorem about partitioning sets of vectors in $\mathbb{C}^n$, which also resolves the Kadison-Singer problem in operator theory, [15].

While the above problems and results seem unrelated, their solutions share a common thread: they all rely on showing, at some level, that a certain univariate or multivariate polynomial has no zeros in a region of $\mathbb{C}^n$ (e.g., the upper complex half-plane, or the unit disk). This is achieved by inductively constructing the relevant polynomial via a sequence of operations which *preserve* the property of not having roots in the required region.

For instance, when the coefficients of the polynomial are real and the region of no zeros is the upper complex-half plane, the polynomial is called *real stable* and this property is preserved under operations such as multiplication, taking derivatives and specialization to real values. While there are extensive and difficult characterizations of real stable polynomials (see [2, 3, 17, 24]), the above properties of real stable polynomials are rather simple to prove and, surprisingly, are sufficient for the applications listed above. Moreover, when the polynomials are of combinatorial origin, these operations have clear algebraic and combinatorial interpretations. Thus, there is a robust way to encode many kinds of combinatorial objects as polynomials, and to draw useful conclusions from their analytic properties. More generally, this serves as evidence against the stereotype that the roots of polynomials are brittle and ill-behaved (which is the case under unnatural operations such as perturbing the coefficients), and therefore difficult to exploit.

Roughly, these principles are evident in the applications listed above as follows: In [11], closure of real stability under taking derivatives allows one to lower bound the value of the permanent of a doubly stochastic matrix. In [10], using the fact that the polynomials corresponding to the max-entropy probability distributions on spanning trees are real-stable, robust and novel negative correlation and anti-concentration properties of them are established. The result in [21] relies on

a stability result (w.r.t the unit disk) for derivatives of the partition function of the Ising model and extends the famous Lee-Yang theorem [25, 12]. In [14], real stability allows one to relate the behavior of one polynomial to the behavior of a sum of polynomials leading to a new existence argument. Lastly, in [15], this theory allows the authors to control the evolution of roots of a polynomial under the application of differential operators.

One may argue that some of the applications above have alternative proofs that do not require this machinery. However, the fact remains that understanding the zeros of the relevant polynomials is important, and, in certain cases, has led to major progress in problems of interest. Moreover, with dramatic progress in the mathematics of this area, such techniques have recently reached a certain maturity which makes them ripe for applications. Thus, we feel that there is need to communicate the essential techniques underlying these results, in a largely self-contained manner, to a broad theory audience, and that is the goal of this article. For more in depth exposition of techniques, the reader is referred to the extensive surveys of [17, 24].

2 Basics

We are primarily concerned with univariate and multivariate polynomials $f(z_1,\ldots,z_n) \in \mathbb{R}[z_1,\ldots,z_n]$. On occasion we may run into polynomials with complex coefficients. Of interest will be zeros of such a polynomial which is always a subset of $\mathbb{C}^n$. For a number $z \in \mathbb{C}$, its real part is denoted by $\Re(z)$ and its imaginary part by $\Im(z)$. Let $\mathscr{H} \stackrel{\text{def}}{=} \{z \in \mathbb{C} : \Im(z) > 0\}$ denote the upper-half complex plane.

2.1 Stability

A polynomial $f(z_1,\ldots,z_n)$ is said to be *stable* with respect to (w.r.t.) a region $\Omega \subseteq \mathbb{C}^n$ if no root of f lies in Ω. Of particular interest is the region

$$\mathscr{H}^n = \{(z_1,\ldots,z_n) \in \mathbb{C}^n : \ \forall i, \ \Im(z_i) > 0\}$$

and polynomials with no root in this region will be referred to as $\mathscr{H}$-*stable* or simply *stable*. To emphasize the fact that the coefficients of f are all real numbers, we often call such polynomials *real stable*. When f is a univariate polynomial, real stability amounts to saying that all the roots of f are real, or f is *real-rooted*. This is because of the following simple lemma which states that the complex roots of a univariate polynomial with real coefficients appear as pairs and, hence, if there is a complex root, there would be one with a positive imaginary part.

Lemma 2.1. *For $f \in \mathbb{R}[z]$, if for $a,b \in \mathbb{R}$ $f(a+\iota b) = 0$, then $f(a-\iota b) = 0$.*

Proof. $f(a+\iota b) = \sum_i a_i(a+\iota b)^i = 0 = \overline{\sum_i a_i(a+\iota b)^i} = \sum_i a_i\overline{(a+\iota b)^i} = \sum_i a_i\overline{(a+\iota b)}^i = \sum_i a_i(a-\iota b)^i = f(a-\iota b)$. $\qquad\square$

The benefits of being real-rooted.   In the univariate case, if the polynomial has coefficients that are non-negative, then all its roots have to be non-positive. Thus, if $f(z) = \sum_{i=0}^d a_i z^i$ is real-rooted

with $a_i \geq 0$ for all i, it can be written as $a_d \prod_{i=1}^{d}(z + \alpha_i)$ where $\alpha_i \geq 0$. For a moment, assume that $\alpha_i > 0$ for all i. Thus, at any positive $z = t$, using the AM-GM inequality we obtain that

$$a_d \prod_{i=1}^{d}(t + \alpha_i) \leq a_d \left(t + \frac{\sum_i \alpha_i}{d} \right)^d.$$

Since $\prod_i \alpha_i = \frac{a_0}{a_d}$,

$$a_d \prod_{i=1}^{d}(t + \alpha_i) = a_0 \prod_{i=1}^{d}\left(1 + \frac{t}{\alpha_i} \right) \leq a_0 \left(1 + \frac{t}{d}\sum_i \frac{1}{\alpha_i} \right)^d = a_0 \left(1 + \frac{t f'(0)}{a_0 d} \right)^d.$$

Optimizing for t then shows that

$$f'(0) \geq \left(\frac{d-1}{d} \right)^{d-1} \inf_{t>0} \frac{f(t)}{t}.$$

Now, if some $\alpha_i = 0$, since $f'(0) = a_1 \geq 0$, then $f(0) = 0$, and the above holds trivially. Note that it was possible for us to apply the AM-GM inequality because all roots of the polynomial were negative. Here, we used the fact that f has non-negative coefficients and is real-rooted. This fact, summarized as the lemma below, was used by Gurvits [11] to lower bound the number of perfect matchings in a k-regular bipartite graph with n vertices on each side by $(k/e)^n$, see Section 3.

Lemma 2.2. *Let* $f(z) = \sum_{i=0}^{d} a_i z^i$ *with* $a_i \geq 0$ *for all* i, *then* $f'(0) \geq \left(\frac{d-1}{d} \right)^{d-1} \inf_{t>0} \frac{f(t)}{t}$.

As another example of real-rootedness we derive an interesting property for probability distributions whose generating functions are real-rooted. For a probability distribution $a_0, a_1, \ldots, a_d$ over $\{0, 1, \ldots, d\}$, its generating function is defined to be the degree d polynomial $g(z) \overset{\text{def}}{=} \sum_{i=0}^{d} a_i z^i$. Suppose $p(z)$ is real-rooted. What does this say about the probability distribution itself? Start by observing that if $g(z)$ is real-rooted, then all its roots have to be non-positive as $a_i \geq 0$ for all i. Thus, $g(z) = a_d \prod_{i=1}^{d}(z + \alpha_i)$ for non-negative α_is. Let $p_i \overset{\text{def}}{=} \frac{1}{1+\alpha_i}$ so that $\alpha_i = \frac{1-p_i}{p_i}$. Since $\alpha_i \geq 0$, $0 < p_i \leq 1$ for all i. Thus,

$$g(z) = a_d \prod_{i=1}^{d}\left(z + \frac{1-p_i}{p_i} \right) = a_d \sum_{S \subseteq [d]} z^{|S|} \prod_{i \notin S} \frac{1-p_i}{p_i} = \frac{a_d}{\prod_{i=1}^{d} p_i} \sum_{S \subseteq [d]} z^{|S|} \prod_{i \notin S}(1-p_i) \prod_{i \in S} p_i.$$

Note that, since, $\sum_{i=1}^{d} a_i = 1$,

$$g(1) = a_d \prod_{i=1}^{d}(1 + \alpha_i) = a_d \prod_{i=1}^{d} \frac{1}{p_i} = 1.$$

Hence,

$$g(z) = \sum_{S \subseteq [d]} z^{|S|} \prod_{i \notin S}(1 - p_i) \prod_{i \in S} p_i = \sum_{k=0}^{d} z^k \sum_{S \subseteq [d]:|S|=k} \prod_{i \notin S}(1-p_i) \prod_{i \in S} p_i.$$

4

Consider a sequence of independent random variables $Y_1, \ldots, Y_d$ such that each Y_i is 1 with probability p_i and 0 with probability $1 - p_i$. Further, let $X \overset{\text{def}}{=} \sum_{i=0}^{d} Y_i$ denote the number of 1s obtained if we sample from each Y_i. Then, $\Pr[X = k] = \sum_{S \subseteq [d]:|S|=k} \prod_{i \notin S}(1 - p_i) \prod_{i \in S} p_i$. Thus,

$$g(z) = \sum_{k=0}^{d} z^k \Pr[X = k].$$

In other words, if the generating function of a probability distribution is real-rooted, then the distribution corresponds to a sum of independent Bernoulli random variables. Thus, for a real-rooted polynomial with non-negative coefficients, its coefficients are unimodal; this is the content of Newton's theorem 2.4.

When is a polynomial real-rooted? Given a polynomial, it is not obvious by looking at its coefficients if it is real-rooted or not. So how would we ever *know* for a polynomial if it is real-rooted or not? How robust is real-rootedness? For instance, if f is real-rooted, so are the polynomials $f(cz)$ for a real number c, and $z^d \cdot f(1/z)$. A bit more non-trivially, so is the derivative of f: $f'(z)$. To see this, recall from calculus that between any two roots of f there is exactly one root of f'. Thus, if all the d roots of f are real, then so are all the $d - 1$ roots of f'. Another way of stating this result is that, if f is real-rooted, then the roots of f and f' alternate, or f and f' *interlace*, see Section 5 for more on interlacing. This latter fact is a manifestation of the more general Gauss-Lucas theorem which states that the convex hull of the set of roots of a real (or complex) polynomial f contains the set of roots of f'. To see a proof of this, write $f(z) = a_d \prod_i (z - \alpha_i)$. Thus,

$$\frac{f'(z)}{f(z)} = \sum_i \frac{1}{z - \alpha_i}.$$

Thus, if $\beta \in \mathbb{C}$ is such that $f'(\beta) = 0$ and $f(\beta) \neq 0$, then $\sum_i \frac{1}{\beta - \alpha_i} = 0$. This implies that $\sum_i \frac{\bar{\beta} - \bar{\alpha}_i}{|\beta - \alpha_i|^2} = 0$. Thus, separating $\bar{\beta}$ out and conjugating, we obtain $\sum_i p_i \alpha_i$ where $p_i \overset{\text{def}}{=} \frac{\frac{1}{|\beta - \alpha_i|^2}}{\sum_j \frac{1}{|\beta - \alpha_i|^2}}$. The Gauss-Lucas theorem follows by noticing that $p_i \geq 0$ for all i and $\sum_i p_i = 1$.

Theorem 2.3 (Gauss-Lucas). *Let $f \in \mathbb{C}[z]$, then all the roots of $f'(z)$ can be written as a convex combination of the roots of $f(z)$.*

As a simple but useful application of this theorem, we prove what are called Newton's identities.

Theorem 2.4 (Newton). *If $f(z) = \sum_{i=0}^{d} a_i z^i$ is real stable with $a_i \geq 0$, then the sequence $\{a_i\}_{i=0}^{d}$ is ultra log-concave, i.e, for all $1 \leq i \leq d - 1$, $\frac{a_{i-1}}{\binom{d}{i-1}} \frac{a_{i+1}}{\binom{d}{i+1}} \leq \left(\frac{a_i}{\binom{d}{i}} \right)^2$.*

Proof. If $f(z)$ is real-rooted, then by Theorem 2.3, so is the polynomial $f_1(z) \overset{\text{def}}{=} \frac{d^{i-1}}{dz^{i-1}} f(z)$. This kills off any coefficients up to $i - 2$. Trivially, the polynomial $f_2(z) \overset{\text{def}}{=} z^{d-i+1} f_1(1/z)$ is real-rooted. Finally, the polynomial $f_3(z) \overset{\text{def}}{=} \frac{d^{d-i+1}}{dz^{d-i+1}} f_2(z)$ is real-rooted, again by Theorem 2.3. But this differentiation kills off any terms after $i + 1$. Hence, observe that $f_3(z) = \frac{d!}{2} \left(\frac{a_{i-1}}{\binom{d}{i-1}} z^2 + \frac{2a_i}{\binom{d}{i}} z + \frac{a_{i+1}}{\binom{d}{i+1}} \right)$. However, a quadratic is real-rooted if and only if its discriminant is non-negative. This gives us the conclusion of the theorem. $\qquad\square$

2.2 Multivariate Polynomials

Recall that $f(z_1, \ldots, z_n)$ is said to be real stable if $f \in \mathbb{R}[z_1, \ldots, z_n]$ and no root of it lies in $\mathscr{H}^n$. It seems harder to show that a multivariate polynomial is real stable. The first lemma reduces the problem of checking real stability of a multivariate polynomial to checking real-rootedness of a set of univariate polynomials, and turns out to be quite effective.

Lemma 2.5. *A multivariate polynomial $f(z_1, \ldots, z_n) \in \mathbb{R}[z_1, \ldots, z_n]$ is stable if and only if for all $v \in \mathbb{R}^n$ and all $u \in \mathbb{R}^n_{>0}$, the univariate polynomial $f(v + tu)$ is real-rooted.*

Proof. Suppose that $f(v + tu)$ is real-rooted for all $v \in \mathbb{R}^n$ and all $u \in \mathbb{R}^n_{>0}$, but f is not real stable. The latter implies that there is an $a = (a_1, \ldots, a_n) \in \mathscr{H}^n$ such that $f(a) = 0$. Let $v \stackrel{\text{def}}{=} \mathfrak{R}(a)$ and $u \stackrel{\text{def}}{=} \mathfrak{I}(a)$. Since $a \in \mathscr{H}^n$, $u_i > 0$ for all i. But then $f(a) = f(v + \iota u) = 0$ and, hence, ι is a root of $f(v + tu)$ which contradicts the real-rootedness of $f(v + tu)$.

For the other direction, suppose that there are $v \in \mathbb{R}^n$ and $u \in \mathbb{R}^n_{>0}$ and a $t = t_1 + \iota t_2$ such that $f(v + tu) = 0$. Since complex roots of a univariate polynomial appear in conjugates (Lemma 2.1), we may assume that $t_2 > 0$. Thus, the imaginary part of each component of $v + tu$ is strictly positive contradicting the fact that f is real stable. $\qquad \square$

Using the lemma above, several multivariate polynomials can be shown to be real stable. Perhaps the simplest such polynomial (which can be seen to be real stable without appealing to the lemma above) is $\sum_i a_i z_i$ when $a_i \geq 0$ for all i. Since a root of a polynomial that is a product of two polynomials is a root of at least one of those two polynomials, the polynomial $\prod_j \sum_i a_{ij} z_i$ is also real stable. A bit more non-trivially, the following important class of polynomials arising from determinants can be shown to be real stable.

Lemma 2.6. *Let $A_1, \ldots, A_n \in \mathbb{R}^{m \times m}$ be positive definite matrices[1] and B be a symmetric $m \times m$ real matrix. Then the polynomial $f(z_1, \ldots, z_n) \stackrel{\text{def}}{=} \det(z_1 A_1 + \cdots + z_n A_n + B)$ is real stable.*

Proof. By Lemma 2.5, it is sufficient to prove that for all $v \in \mathbb{R}^n$ and $u \in \mathbb{R}^n_{>0}$, $f(v + tu)$ is real-rooted. This is the same as showing that

$$\det \left(B + \sum_{i=1}^n v_i A_i + t \sum_{i=1}^n u_i A_i \right)$$

is real-rooted. Since $A_i \succ 0$ and $u_i > 0$ for all i, $\sum_{i=1}^n u_i A_i \succ 0$. Thus, letting $M \stackrel{\text{def}}{=} \sum_{i=1}^n u_i A_i$, we need to show that

$$\det \left(M^{-1/2} \left(B + \sum_{i=1}^n v_i A_i \right) M^{-1/2} + tI \right).$$

This latter is true because $M^{-1/2} \left(B + \sum_{i=1}^n v_i A_i \right) M^{-1/2}$ is symmetric and every real-symmetric has all real eigenvalues. To see this, if A is a real symmetric matrix and λ is an eigenvalue with an eigenvector v, then $Av = \lambda v$. Conjugating both sides we obtain that $v^\star A^\top = \overline{\lambda} v^\star$, where $v^\star$ is the conjugate transpose of v. Hence, $v^\star A v = \overline{\lambda} v^\star v$, since A is symmetric. Thus, $\lambda |v|^2 = \overline{\lambda} |v|^2$ which implies that $\lambda = \overline{\lambda}$. Thus, $\lambda \in \mathbb{R}$. $\qquad \square$

---

[1] Semi-positive definite and positive definite matrices over reals are symmetric.

The above lemma can be proved in the setting when A_is are positive semi-definite (PSD) as opposed to being positive-definite. This is quite useful for applications. However, extending Lemma 2.6 requires the following theorem from complex analysis whose proof is beyond the scope of the current article.

Theorem 2.7 (Hurwitz). *Let $\{f_k\}_{k\geq 0}$ be a sequence of Ω-stable polynomials over $z_1,\ldots,z_n$ for a connected and open set Ω that uniformly converge to a polynomial f over compact subsets of Ω. Then f is Ω-stable.*

To use this theorem for a matrix A_i which is just guaranteed to be PSD one approximates each A_i by a sequence of matrices $A_i + \frac{1}{2^k}I$ which are positive definite and converge to A_i as k goes to infinity. One can ask if all real stable polynomials arise from such determinants. This is the content of the Lax Conjecture and the interested reader is referred to [13].

2.3 Closure Properties

What makes the stability theory particularly powerful is that many of the *closure* properties discussed in Section 2.1 hold in the multivariate setting as well. Thus, we can start with real stable polynomials and prove stability for new ones. For us, the key closure properties will be closure under *inversion*, *specialization* and *differentiation*.

Inversion. If $f(z_1,\ldots,z_n)$ is real stable with the degree of z_i in f being d_i, then the polynomial $f(1/z_1,\ldots,1/z_n)\prod_{i=1}^n z_i^{d_i}$ is also real stable. Suppose $(a_1,\ldots,a_n) \in \mathscr{H}^n$ be such that $f(1/a_1,\ldots,1/a_n) = 0$. Since the coefficients of f are real, if $\overline{f(1/a_1,\ldots,1/a_n)} = f(\overline{1/a_1},\ldots,\overline{1/a_n}) = 0$. Since, if the imaginary part of a_i is positive that of $1/a_i$ is negative, the imaginary part of $\overline{1/a_i}$ is positive for each i. This contradicts the real stability of f and establishes our first closure result.

Specialization. It is easy to see that if $f(z_1,\ldots,z_n)$ is a stable polynomial, then $f(a,z_2,\ldots,z_n)$ is also stable if $\Im(a) > 0$. However, if $f(z_1,\ldots,z_n)$ is real, $f(a,z_2,\ldots,z_n)$ may have complex coefficients and, thus, may not be real stable. The following lemma, which relies on Hurwitz's theorem (Theorem 2.7), shows that if $a \in \mathbb{R}$ then $f(a,z_2,\ldots,z_n)$ is real stable.

Lemma 2.8. *If $f(z_1,\ldots,z_n)$ is real stable, then for all $a \in \mathbb{R}$ (the closure of $\mathscr{H}$), $f(a,z_2,\ldots,z_n)$ is also real stable.*

Proof. If $a \in \mathscr{H}$, then the proof follows from the discussion above. Thus, it is sufficient to prove this lemma for $a \in \mathbb{R}$. We will only sketch a proof here. Suppose, for sake of contradiction, that $f(a,a_2,\ldots,a_n) = 0$ with some a_j such that $\Im(a_j) > 0$. It follows from the definition that $f_k \stackrel{\text{def}}{=} f(a+\iota 2^{-k},z_2,\ldots,z_n)$ is stable for any $k > 0$. The lemma now follows from Hurwitz's theorem since $\lim_{k\to\infty} f_k = f(a,z_2,\ldots,z_n)$ is stable and, being real, is real stable. □

Differentiation. The next crucial closure property is closure of real stability under taking partial derivatives. Following is some basic notation for partial derivatives of multivariate polynomials. Let $\partial_i \stackrel{\text{def}}{=} \partial/\partial z_i$. For a tuple $\alpha : [n] \mapsto \mathbb{Z}_{\geq 0}$, let $\partial^\alpha \stackrel{\text{def}}{=} \prod_{i=1}^n \partial_i^{\alpha_i}$.

Lemma 2.9. *Let f be real stable. Then, $\partial_1 f$ is also real stable.*

Since the choice of the first variable is arbitrary, for any $\alpha : [n] \mapsto \mathbb{Z}_{\geq 0}$, $\partial^\alpha f$ is real stable. Thus, the real stability of $\partial^\alpha f$ follows by an inductive application of this lemma.

Proof. Assume on the contrary that $\partial_1 f$ is not real stable and let $a = (a_1, a_2, \ldots, a_n) \in \mathcal{H}^n$ such that $\partial_1 f(a_1, a_2, \ldots, a_n) = 0$. Let $g(z) \stackrel{\text{def}}{=} f(z, a_2, \ldots, a_n)$. If $g \equiv 0$, then $f(a_1, a_2, \ldots, a_n) = 0$, contradicting the stability of f. Hence, $g \not\equiv 0$. Since f is real stable, so is $g(z)$ by Lemma 2.8. By the Gauss-Lucas theorem, the roots of $g'(z)$ are in the convex hull of the roots of $g(z)$ and, hence, $g'(z)$ is real stable. Since $g'(z) = \partial_1 f(z, a_2, \ldots, a_n)$, $g'(a_1) = \partial_1 f(a_1, a_2, \ldots, a_n) = 0$ by assumption. Thus, $g'(a_1) = 0$ for a_1 such that $\Im(a_1) > 0$, contradicting the stability of g'. $\qquad\square$

3 Lower Bounding the Permanent

As a simple but powerful application of the closure properties we show how, starting with simple polynomials, we can argue about non-trivial (and computationally intractable) objects such as the permanent of a matrix. For a matrix $A = (a_{ij})_{i \in [n], j \in [n]}$ with real entries, its permanent is defined to be

$$\text{per}(A) \stackrel{\text{def}}{=} \sum_{\sigma \in S_n} \prod_{i=1}^{n} a_{i\sigma(i)}.$$

Consider the polynomial $f_A(z_1, \ldots, z_n) \stackrel{\text{def}}{=} \prod_{i=1}^{n} \sum_{j=1}^{n} a_{ij} z_j$, and note that f_A is clearly real stable. Moreover, it follows from a repeated application of Lemma 2.8 and Lemma 2.9 that, for any $1 \leq i < n$, the polynomial

$$g_i(z_1, \ldots, z_i) \stackrel{\text{def}}{=} \partial^{(i+1,\ldots,n)} f_A(z_1, \ldots, z_i, 0, \ldots, 0)$$

is real stable. Note that $g_0 = \partial^{(1,\ldots,n)} f_A(0, \ldots, 0) = \text{per}(A)$. If all entries of A are nonnegative, then it follows from Lemma 2.2 and Lemma 2.8 that, for any fixed positive $b_1, \ldots, b_{i-1}$,

$$g_{i-1}(b_1, \ldots, b_{i-1}) = \partial_i g_i(b_1, \ldots, b_{i-1}, 0) \geq \left(\frac{d_i - 1}{d_i}\right)^{d_i - 1} \frac{g_i(b_1, \ldots, b_i)}{b_i},$$

where d_i is the degree of the polynomial $g_i(b_1, \ldots, b_{i-1}, z_i)$. Fixing $s_1, s_2, \ldots, s_{i-1}$, let s_i be defined to be

$$\arg \inf_{t > 0} \frac{g_i(s_1, \ldots, s_{i-1}, t)}{t}.$$

Thus, applying the above inequality for $i = 0$ to $n - 1$ and letting $d \stackrel{\text{def}}{=} \max_{i=1}^{n} d_i$, we obtain that $\text{per}(A) = g_0$, which is at least

$$\left(\frac{d-1}{d}\right)^{d-1} \frac{g_1(s_1)}{s_1} \geq \cdots \geq \left(\frac{d-1}{d}\right)^{(d-1)n} \frac{g_n(s_1, \ldots, s_n)}{\prod_{i=1}^{n} s_i} = \left(\frac{d-1}{d}\right)^{(d-1)n} \frac{f_A(s_1, \ldots, s_n)}{\prod_{i=1}^{n} s_i}.$$

Since $\frac{f_A(s_1, \ldots, s_n)}{\prod_{i=1}^{n} s_i} \geq \inf_{b_1 > 0, \ldots, b_n > 0} \frac{f_A(b_1, \ldots, b_n)}{\prod_{i=1}^{n} b_i}$, we need to calculate $\inf_{b_1 > 0, \ldots, b_n > 0} \frac{f_A(b_1, \ldots, b_n)}{\prod_{i=1}^{n} b_i}$. It turns out that when A is a doubly stochastic matrix, then this quantity can be lower bounded by 1. Recall

that a matrix A is said to be doubly stochastic matrix, i.e., $a_{ij} \geq 0$ and, $\sum_{i=1}^n a_{ij} = 1$ for all j and $\sum_{j=1}^n a_{ij} = 1$ for all i. To see the claim, observe that for any positive $b_1, \ldots, b_n$,

$$f_A(b_1, \ldots, b_n) = \prod_{i=1}^n \sum_{j=1}^n a_{ij} b_j \overset{\text{AM}-\text{GM}; \sum_j a_{ij}=1}{\geq} \prod_{i=1}^n \prod_{j=1}^n b_j^{a_{ij}} = \prod_{j=1}^n \prod_{i=1}^n b_j^{a_{ij}} \overset{\sum_i a_{ij}=1}{=} \prod_{j=1}^n b_j.$$

Thus, when A is doubly stochastic, $\inf_{b_1>0,\ldots,b_n>0} \frac{f_A(b_1,\ldots,b_n)}{\prod_{i=1}^n b_i} \geq 1$. Noting that $\left(\frac{d-1}{d}\right)^{d-1} \geq \frac{1}{e}$, we have proved the van der Waerden conjecture.

Theorem 3.1. *For a $n \times n$ doubly stochastic matrix A, $\mathrm{per}(A) \geq \left(\frac{1}{e}\right)^n$.*

As a corollary, let $G = (V, W, E)$ be a k-regular bipartite graph with $|V| = |W| = n$. Let A be the matrix with rows indexed by V, columns by W, and $a_{ij} = 1$ if $(i, j) \in E$. Then, $\frac{1}{k}A$ is doubly stochastic and, hence, $\mathrm{per}(A) \geq \left(\frac{k}{e}\right)^n$. Note that $\mathrm{per}(A)$ counts exactly the number of perfect matchings in G.

4 Probability Measures and Real Stability

In this section we study probability distributions over $\{0,1\}^n$ by looking at their generating function. For a distribution μ, the generating function is the multivariate affine polynomial

$$g_\mu \overset{\text{def}}{=} \sum_{S \subseteq [n]} \mu(S) \prod_{i \in S} z_i = \sum_{S \subseteq [n]} \mu(S) z^S.$$

If g_μ is real stable, then one can derive a host of properties of μ by appealing to the closure properties enjoyed by real stable polynomials. In this case, μ is said to be *strongly Rayleigh*.

Definition 4.1. *A measure μ over $\{0,1\}^n$ is said to be strongly Rayleigh if its generating function $\sum_{S \subseteq [n]} \mu(S) z^S$ is real stable.*

Strongly Rayleigh measures satisfy the strongest forms of negative dependence, a consequence of which is the concentration of measure for a sum of random variables drawn from a such a measure. As a starting point, we prove the pairwise negative correlation property of strongly Rayleigh measures.

Definition 4.2. *A measure μ is said to be pairwise negatively correlated if*

$$\sum_{S: S \supseteq \{i\}} \mu(S) \sum_{T: T \supseteq \{j\}} \mu(T) \geq \sum_{S: S \supseteq \{i,j\}} \mu(S)$$

for all $i \neq j$. In terms of polynomials, this condition is equivalent to

$$\partial_i g_\mu(1, 1, \ldots, 1) \partial_j g_\mu(1, 1, \ldots, 1) \geq g_\mu(1, 1, \ldots, 1) \partial^{(i,j)} g_\mu(1, 1, \ldots, 1). \tag{1}$$

In fact, for strongly Rayleigh measures, one can show something stronger: the condition (1) holds for all $(a_1, \ldots, a_n) \in \mathbb{R}^n$ rather than just the vector $(1, \ldots, 1)$. This property, in fact, implies the strong Rayleigh measures but we just prove the forward direction.

Lemma 4.3. *If $f \in \mathbb{R}[z_1, \ldots, z_n]$ is affine and stable, then*

$$\partial_i f(a_1, \ldots, a_n) \partial_j f(a_1, \ldots, a_n) \geq f(a_1, \ldots, a_n) \partial^{(i,j)} f(a_1, \ldots, a_n)$$

for all $(a_1, \ldots, a_n) \in \mathbb{R}^n$.

Thus, the lemma implies that strongly Rayleigh measures are pairwise negatively correlated.

Proof. Fix $(a_1, \ldots, a_n) \in \mathbb{R}^n$ and let $g(s,t) \overset{\text{def}}{=} f(a_1, \ldots, a_i + s, \ldots, a_j + t, \ldots, a_n)$. It follows from Lemma 2.8 along with Theorem 2.7 that $g(s,t)$ is stable. On the other hand, since f is affine,

$$g(s,t) = f(a_1, \ldots, a_n) + s \partial_i f(a_1, \ldots, a_n) + t \partial_j f(a_1, \ldots, a_n) + st \partial^{(i,j)} f(a_1, \ldots, a_n).$$

Since $g(s,t)$ is stable, for any $s = s_1 + \iota s_2$ such that $g(s, \iota) = 0$, $s_2 \leq 0$. If $g(s, \iota) = 0$, then (dropping $(a_1, \ldots, a_n)$ for the easy of reading),

$$f + s_1 \partial_i f - s_2 \partial^{(i,j)} f = 0$$

$$s_2 \partial_i f + \partial_j f + s_1 \partial^{(i,j)} f = 0.$$

Multiplying the first equation by $-\partial^{(i,j)} f$ and the second by $\partial_i f$ and adding them, we obtain

$$-f \partial^{(i,j)} f + s_2 (\partial^{(i,j)} f)^2 + s_2 (\partial_i f)^2 + \partial_i f \partial_j f = 0.$$

Since, $s_2 \leq 0$, this implies that $f \partial^{(i,j)} f \leq \partial_i f \partial_j f$ completing the proof. $\qquad\square$

4.1 Closure Properties

We now show how certain measures derived from a strongly Rayleigh measure μ remain strongly Rayleigh. Such measures include measures obtained by conditioning a strongly Rayleigh measure μ on certain types of events, e.g., $(\mu | X_i = \sigma_1)$ or $(\mu | \sum_{i=1}^n X_i = k)$. This is a consequence of the closure properties of the associated real stable polynomials. Thus, these conditioned measures, being strongly Rayleigh, also end up satisfying strong negative dependence properties. We emphasize that, in most cases, deriving such properties about the measure without appealing to the closure properties of real stability is not known. For an extensive discussion on the properties derivable for strongly Rayleigh measures, the reader is referred to [4].

Fixations. Let μ be a strongly Rayleigh measure on $\{0, 1\}^n$, let $i_1, i_2, \ldots, i_k$ be some indices in $[n]$ and let $\sigma_1, \ldots, \sigma_k \in \{0, 1\}$. Then the measure $(\mu | X_{i_1} = \sigma_1, \ldots, X_{i_k} = \sigma_k)$ is strongly Rayleigh. The proof follows from the repeated application of the observations that, if we define $\mu' \overset{\text{def}}{=} (\mu | X_1 = 1)$ and $\mu'' \overset{\text{def}}{=} (\mu | X_1 = 0)$, then $g_{\mu'} = \partial_1 g_\mu$ and $g_{\mu''} = g_\mu(0, z_2, \ldots, z_n)$. Here we have used Lemma 2.9 and Lemma 2.8 respectively.

Ultra-Log-Concavity. For a probability measure μ on $\{0,1\}^n$, consider the probability measure on $\{0,1,\ldots,n\}$ defined as:

$$p_k \stackrel{\text{def}}{=} \Pr_{(a_1\ldots,a_n)\leftarrow\mu}\left[\sum_{i=1}^{n} a_i = k\right] = \sum_{S\subseteq[n]:|S|=k} \mu(S).$$

Thus, if $g(z) \stackrel{\text{def}}{=} g_\mu(z,z,\ldots,z)$, then $g(z) = \sum_{i=0}^{n} p_i z^i$. If μ is strongly Rayleigh then g_μ is real stable and, hence, g is a real-rooted polynomial with non-negative coefficients. Thus, it follows from Theorem 2.4 that for all $1 \leq i \leq n-1$, $\frac{p_{i-1}}{\binom{n}{i-1}}\frac{p_{i+1}}{\binom{n}{i+1}} \leq \left(\frac{p_i}{\binom{n}{i}}\right)^2$. This is the same as saying that $\{p_k\}_{k\in\{0,1,\ldots\}}$ is *ultra-log-concave*.

Ultra-log-concavity of a distribution implies that it is *unimodal*, i.e., the distribution has only one peak. This property can then be used to show a concentration result of the following kind, which is easily seen to be false in general and is useful in applications. If X is a random variable that takes values in the set $\{0,1,\ldots,n\}$, and $\mathbb{E}[X] \in [1.5, 2.5]$, then $\Pr[X = 2] = \Omega(1)$, see [10]. (Here the constants are arbitrary.)

Conditioning. As mentioned before that if μ is strongly Rayleigh over $\{0,1\}^n$, then it has many nice properties, e.g., it is negatively associated. What about the measure obtained by conditioning μ on events such as $(\mu|\sum_{i=1}^{n} X_i = k)$? We show that such a measure is also strongly Rayleigh and, hence, has equally nice properties. To prove this, consider the polynomial $f(z_1,\ldots,z_n,y)$ which is the homogenization of $g_\mu(z_1,\ldots,z_n)$ and defined as

$$f(z_1,\ldots,z_n,y) \stackrel{\text{def}}{=} \sum_{S\subseteq[n]} \mu(S)z^S y^{n-|S|} = \sum_{S:|S|=k} \mu(S)z^S y^{n-k} = \sum_{k=0}^{n} g_{\mu,k}(z_1,\ldots,z_n)y^{n-k}.$$

Thus, to show that $(\mu|\sum_{i=1}^{n} X_i = k)$ is strongly Rayleigh, we need to show that $g_{\mu,k}$ is real stable. To see this we construct a sequence of polynomials: let $f_1 \stackrel{\text{def}}{=} \frac{d^{n-k} f}{dy^{n-k}}$. $f_2 \stackrel{\text{def}}{=} (\prod_{i=1}^{n} z_i) y^k f_1(1/z_1,\ldots,1/z_n,1/y)$. Thus, observe that $g_{\mu,k} = \frac{d^k f_2}{dy^k}$. Since g_μ is real stable, so is f. The real stability of f_1 follows from that of f and the Gauss-Lucas theorem. The real stability of f_2 follows from closure under inversion. Finally, the real stability of $g_{\mu,k}$ is obtained by again appealing to the Gauss-Lucas theorem.

This proof readily extends to showing that measures $(\mu|p \leq \sum_{i=1}^{n} X_i \leq q)$ remain strongly Rayleigh if $|p - q| \leq 1$.

Negative Association. Perhaps the most important property, from the point of view of applications, satisfied by strongly Rayleigh measures is that they are *negatively associated*. A probability measure μ is said to be negatively associated if for any monotonically non-decreasing functions g_1, g_2 on $\{0,1\}^m$, on disjoint sets of variables

$$\mathbb{E}_\mu[g_1(X_1,\ldots,X_n)]\mathbb{E}_\mu[g_2(X_1,\ldots,X_n)] \geq \mathbb{E}_\mu[g_1(X_1,\ldots,X_n)g_2(X_1,\ldots,X_n)].$$

Negative association is important as it implies Chernoff bounds: if $(X_1,\ldots,X_n)$ is drawn from a negatively associated measure μ, then $\sum_{i=1}^{n} X_i$ is concentrated around its mean. While we do not

11

prove that strongly Rayleigh measures are negatively associated in this article, following a proof of [8], it can be shown to be a consequence of Lemma 4.3 and the closure properties of strongly Rayleigh measures mentioned earlier.

In fact, for strongly Rayleigh measures, an even stronger form of Chernoff bound was shown in [18] which is not known to be true for negatively associated random variables: if $f : \{0,1\}^n \mapsto \mathbb{R}$ be a function with Lipschitz constant 1, and if $(X_1, \ldots, X_n)$ is drawn from a strongly Rayleigh measure μ, then $f(X_1, \ldots, X_n)$ is concentrated around $\mathbb{E}[f(X_1, \ldots, X_n)]$.

4.2 Determinantal Measures

But how do we show that for a measure μ, p_μ is real stable? One important class of measures for which we can prove real stability of the corresponding generating functions are *determinantal measures*. For a matrix $n \times n$ matrix A and a set $S \subseteq [n]$, let $A_{S,S}$ denote the matrix obtained from A by picking rows and columns corresponding to S.

Definition 4.4. *A measure μ on $\{0,1\}^n$ is said to be determinantal if there exists an $n \times n$ matrix A such that $\forall S \subseteq [n]$,*

$$\sum_{T:T \supseteq S} \mu(T) = \det\left(A_{S,S}\right).$$

Lemma 4.5. *If μ is determinantal w.r.t. the matrix A which satisfies $0 \preceq A \preceq I$, then p_μ is real stable.*

Proof. We show this result when $0 \prec A \prec I$ and the lemma follows from an application of Hurwitz's theorem (Theorem 2.7). Let Z be the diagonal matrix where the (i,i)th entry is z_i. We claim that $p_\mu = \det(I - A + AZ)$. Assuming this claim, the stability of p_μ follows from Lemma 2.6 and Hurwitz's theorem. since $p_\mu = \det(A)\det(A^{-1} - I + Z)$. To prove that $p_\mu(z_1, \ldots, z_n) = \det(I - A + AZ)$, we need to show that $\sum_{T:T \supseteq S} \mu(S) = \det(A_{S,S})$. Let Z_S denote the diagonal matrix derived from Z by setting all variables not in S to 1. Then the coefficient of $\prod_{i \in S} z_i$ in $\det(I - A + AZ_S)$ is exactly $\sum_{T:T \supseteq S} \mu(S)$. The matrix $AZ_S - A$ has zeros in every entry in any column corresponding to $\overline{S}$. Hence, $I + AZ_S - A$ is identity on the rows and columns indexed by $\overline{S}$ and 0 on the rows corresponding to S and columns corresponding to $\overline{S}$. Thus, $\det(I - A + AZ_S) = \det(I_{S,S} + A_{S,S}(Z_{S,S} - I_{S,S}))$. The coefficient of $\prod_{i \in S} z_i$ in this is exactly $\det(A_{S,S})$. Since all matrices involved are positive $(A^{-1} - I \succ 0)$, all determinants are positive and, hence, the determinants are positive. $\square$

The Spanning Tree Measure. Let $G = (V,E)$ be a graph with n vertices, m edges, and edge weights $\lambda : E \mapsto \mathbb{R}_{\geq 0}$. Consider the measure on subsets of edges of E defined by $\lambda(T) \overset{\text{def}}{=} \prod_{e \in T} \lambda(e)$. We are interested in the case when T is a spanning tree in G. Let $\mathscr{T}$ be the set of all spanning trees in G. Consider the probability measure

$$\mu(T) \overset{\text{def}}{=} \frac{\lambda(T)}{\sum_{T \in \mathscr{T}} \lambda(T)}.$$

This measure can be shown to be determinantal and we end this section by describing the matrix whose sub-determinants give rise to this measure. Let Λ be an $m \times m$ diagonal matrix with $\Lambda_{(e,e)} = \lambda(e)$. Orient the edges of G arbitrarily such that every edge has a head and a tail. For this fixed

orientation, let B be an $m \times n$ matrix such that $B_{e,v} = 1$ if v is the head of e, -1 if v is the tail of e and 0 otherwise. The combinatorial Laplacian of G is then defined to be $L \stackrel{\text{def}}{=} B^{\top} \Lambda B$ and is independent of the orientation. Further, let L^{+} denote the inverse of L restricted to its image. Define the matrix $\Pi \stackrel{\text{def}}{=} BL^{+}B^{\top}$. A generalization of the matrix-tree theorem proves that the measure μ defined above is determinantal w.r.t the matrix $\Lambda \Pi$. (Note that Π is a projection matrix and all its eigenvalues are either 0 or 1.)

5 Real Stability and Interlacing

In this section we show how real stability can be used to show that a bound on the largest root of a sum of a certain family of polynomials implies the same bound on the largest root of one polynomial in the family. This novel technique is central to the proof of the existence of Ramanujan graphs of all degrees [14], and the resolution of the Kadison-Singer problem [15].

More formally, the family of polynomials we consider is $\{f_{\sigma}(z)\}_{\sigma \in \{-1,1\}^n}$ where each polynomial in the family is of the same degree and has a positive leading coefficient. For $p = (p_1, \ldots, p_n) \in [0,1]^n$, we define a random polynomial $f_{(X_1,\ldots,X_n)}$ where X_i is an independent Bernoulli random variable which is 1 with probability p_i and -1 with probability $1 - p_i$. Assume, the seemingly strong hypothesis, that this family polynomials satisfies, for every $p \in [0,1]^n$ the polynomial $\mathbb{E}_{(p_1,\ldots,p_n)}[f_{(X_1,\ldots,X_n)}]$ is real-rooted. Such a family is shown to have the property that if the largest root of $\sum_{\sigma} f_{\sigma}$ is bounded by ρ, then there is a σ such that the largest root of f_{σ} is also bounded by ρ. This is captured in the following theorem.

Theorem 5.1. *[14] Suppose $\{f_{\sigma}(z)\}_{\sigma \in \{-1,1\}^n}$ is a family of real-rooted polynomials with positive leading coefficients where all have the same degree. Then, there is a σ such that the largest root of $f_{\sigma}(z)$ is at most the largest root of $\sum_{\sigma \in \{-1,1\}^n} f_{\sigma}(z)$.*

While we do not go into the the proof of the hypothesis for any specific family in this article, we mention that the real-rootedness of $\mathbb{E}_{(p_1,\ldots,p_n)}[f_{(X_1,\ldots,X_n)}]$ is shown by constructing a suitable starting multivariate polynomial that is real stable (using Lemma 2.6) and then applying a carefully chosen sequence of closure properties such as the ones presented in Section 2.2. We start the proof of Theorem 5.1 for a family which satisfies the above hypothesis by observing the following implication of the hypothesis.

Lemma 5.2. *Under the hypothesis, for any fixing $\sigma_1, \ldots, \sigma_k$ any convex combination of*

$$\sum_{\sigma_{k+1},\ldots,\sigma_n} f_{\sigma_1,\ldots,\sigma_k,1,\sigma_{k+1},\ldots\sigma_n} \quad \text{and} \quad \sum_{\sigma_{k+1},\ldots,\sigma_n} f_{\sigma_1,\ldots,\sigma_k,-1,\sigma_{k+1},\ldots\sigma_n}$$

are real-rooted.

Proof. For a parameter $\lambda \in [0,1]$, set $p_{k+1} \stackrel{\text{def}}{=} \lambda$ and $p_{k+2} = \cdots = p_n = 1/2$ and $p_i \stackrel{\text{def}}{=} \frac{1+\sigma_i}{2}$ for $1 \leq i \leq k$. It follows that $\mathbb{E}_{(\sigma_1,\ldots,\sigma_n) \leftarrow \mu_p}[f_{\sigma}(z)] = \lambda \sum_{\sigma_{k+1},\ldots,\sigma_n} f_{\sigma_1,\sigma_2,\ldots,\sigma_k,1} + (1-\lambda) \sum_{\sigma_{k+1},\ldots,\sigma_n} f_{\sigma_1,\sigma_2,\ldots,\sigma_k,-1}$, which is real-rooted by the hypothesis. $\square$

The conclusion of the above lemma is interesting because if any convex combination of two univariate polynomials with leading positive coefficients is real-rooted, then they have a *common*

interlacing. Two real-rooted polynomials $f(z)$ and $g(z)$ of the same degree (d) are said to interlace if their roots alternate. More formally, if $a_1 \leq \cdots \leq a_d$ are roots of f and $b_1 \leq \cdots \leq b_d$ are the roots of g, then

$$a_1 \leq b_1 \leq a_2 \leq b_2 \leq \cdots \leq a_d \leq b_d \text{ or } b_1 \leq a_1 \leq b_2 \leq a_2 \leq \cdots \leq b_d \leq a_d.$$

Further, if there is a polynomial which interlaces with both $f(z)$ and $g(z)$, we say that they have a common interlacing. The following lemma can be proved by showing that, if one looks at the intervals corresponding to the successive roots of each polynomial and order them from left to right, the corresponding intervals have non-empty intersection. This is a consequence of the fact that two interlacing polynomials with positive leading coefficients cannot differ in the number of roots they have in any interval of the form $[a,\infty)$ by more than 1. We omit the elementary but somewhat tedious proof, see [9, 5].

Lemma 5.3. *Let $f(z)$ and $g(z)$ be two real polynomials of the same degree with positive leading coefficients such that every convex combination of them is real-rooted. Then $f(z)$ and $g(z)$ have a common interlacing.*

Finally, we need the following lemma which shows that if two polynomials have a common interlacing, then largest root of one of the polynomials is at most the largest root of the sum of the two polynomials.

Lemma 5.4. *Let $f(z)$ and $g(z)$ be two real polynomials of the same degree that have a common interlacing and positive largest coefficients. Then, the largest root of both $f(z)$ and $g(z)$ cannot be larger than the largest root of $f(z) + g(z)$.*

Proof. Let $h(z)$ be polynomial which interlaces with both $f(z)$ and $g(z)$, let a_d be the largest root of $f(z)$, b_d be the largest root of $g(z)$ and c_d be the largest root of $h(z)$. Then $c_d \leq a_d$ and $c_d \leq b_d$. Since f,g have positive leading coefficients they both go to infinity as z goes to infinity. Thus, $f(z) + g(z) > 0$ for all $z \geq \max\{a_d, b_d\}$. Moreover, since the second largest roots of $f(z)$ and $g(z)$ are both at most c_d, $f(c_d) \leq 0$ and $g(c_d) \leq 0$. Thus, $f(z) + g(z) \leq 0$ for all $z \in [c_d, \min\{a_d, b_d\}]$. Thus, at least on of a_d or b_d is at most the largest root of $f(z) + b(z)$. □

The proof of Theorem 5.1 now follows easily by iteratively applying Lemma 5.3 and Lemma 5.4. We just show the first step.

Proof. (of Theorem 5.1) Assume that the family $\{f_\sigma(z)\}_{\sigma \in \{-1,1\}^n}$ satisfies the hypothesis of Theorem 5.1. Start by writing

$$\sum_{\sigma_1,\ldots,\sigma_n} f_{\sigma_1,\ldots,\sigma_n} = \sum_{\sigma_2,\ldots,\sigma_n} f_{1,\sigma_2,\ldots,\sigma_n} + \sum_{\sigma_2,\ldots,\sigma_n} f_{-1,\sigma_2,\ldots,\sigma_{n-1}}.$$

Combining Lemma 5.2, Lemma 5.3 and Lemma 5.4, we obtain that the largest root of at least one of the two polynomials on the r.h.s. of the equality above is no larger than the largest root of the polynomial $\sum_{\sigma_1,\ldots,\sigma_n} f_{\sigma_1,\ldots,\sigma_n}$. The theorem follows from inducting on the polynomial which satisfies this property. □

6 Stability w.r.t. the Unit Disk: The Lee-Yang Theorem

In this section we move away from real stability and present a result due to Lee and Yang [12] that may be considered one of the founding results in multivariate stability theory: the roots of the partition function of the Ising model on any connected graph lie on the unit circle (for any fixed *temperature* between 0 and 1). Lee-Yang used this theorem to show that the Ising model indeed does not show a phase transition with respect to the external field at any non-zero value of the field, see [22].

The Ising Model.   We begin by looking at the partition functions of the Ising model: given a graph $G = (V, E)$ with $|V| = n$, the Ising model is a probability distribution over spin assignments $\sigma : V \to \{-1, 1\}$ to the vertices. The *temperature* of the system is modeled by a parameter $\beta \in (0, 1]$ (the temperature is $-\log \beta$), and the preference for particular spins (the magnetic field) through an *activity* parameter λ in the following way: the probability of a spin assignment (or configuration) σ is proportional to its weight $w(\sigma)$ which is defined to be $w(\sigma) \stackrel{\text{def}}{=} \beta^{d(\sigma)} \lambda^{m(\sigma)}$, where $d(\sigma)$ is the number of edges $e = \{u, v\}$ such that $\sigma(u) \neq \sigma(v)$ and $m(\sigma)$ is the number of vertices which are assigned a positive spin by σ. The partition function of the Ising model is then defined as $Z_{G,\beta}(\lambda) \stackrel{\text{def}}{=} \sum_\sigma w(\sigma)$. More generally, we can consider a setting where λ is different for each vertex. We denote the corresponding partition function by $Z_{G,\beta}(\lambda_1, \ldots, \lambda_n)$. The Lee-Yang circle theorem can now be stated as follows:

Theorem 6.1 (Lee-Yang). *Let G be an undirected graph and suppose $\beta \in (0, 1)$. Then the zeros of $Z_{G,\beta}(\lambda)$ all lie on the unit circle.*

The proof given here is due to Asano [1], who simplified the original proof by Lee and Yang [12] while proving a version of the Lee-Yang theorem for the quantum Ising model. As is usual when proving stability results, it helps to consider the more general case of the multivariate partition function. We prove the following generalization of the Lee-Yang theorem.

Theorem 6.2 (Multivariate-Lee-Yang). *Let G be an undirected graph and suppose $\beta \in (0, 1)$. Suppose $(\lambda_i)_{i \in V}$ are complex numbers then $Z_{G,\beta}(\lambda_1, \ldots, \lambda_n) \neq 0$ if $|\lambda_i| > 1$ for all i.*

Setting $\lambda_i = \lambda$ for each i, we obtain that, if $\beta \in (0, 1)$, then the zeros of $Z_{G,\beta}(\lambda)$ all lie on the unit circle, Theorem 6.1.

Proof. The main idea used in the proof is often called the *Asano contraction* (see for example, Ruelle's article [19], whose presentation we loosely follow). Consider a graph H on n vertices, and identify two vertices v_1 and v_2 in G that are not connected by an edge. Let λ_1, λ_2 be variables representing the vertex activities at v_1 and v_2. The vertex activities of the other vertices in H are denoted by $\lambda_3, \lambda_4, \ldots, \lambda_n$.

We say that a graph G satisfies the *Lee-Yang property* if the partition functions of all the induced subgraphs of G satisfy the conclusion of the Theorem 6.2 (that is, whenever all the vertex activities have magnitude greater than 1, the partition function is non-zero). The reason for considering the partition functions of all induced subgraphs is technical and will become clear later in the proof.

Now suppose that a graph H satisfies the Lee-Yang property. The Asano contraction lemma states then that the graph H' obtained by contracting the vertex v_1 and v_2 into a single vertex v also obeys the Lee-Yang property. To see this, we first ignore the issue of the induced subgraphs and consider the partition functions of H and H' themselves. We can write the partition function of H as

$$A\lambda_1\lambda_2 + B\lambda_1 + C\lambda_2 + D, \tag{2}$$

where A, B, C, D are polynomials in the other vertex activities $\lambda_3, \ldots, \lambda_n$. The first, and crucial, observation is that the partition function of the new graph H' is then simply

$$A\lambda + D, \tag{3}$$

which follows from a consideration of the definition of the Ising model partition function.

Now, consider any fixing of values of $\lambda_3, \ldots, \lambda_n$ such that they are all greater than 1 in magnitude. We would be done if we can show that this implies that the expression in (3) is non-zero for $|\lambda| > 1$. Now, since the expression in eq. (2) satisfies the Lee-Yang property, we see by substituting these values into (2) and setting $\lambda_1 = \lambda_2 = z$ that the quadratic equation $Az^2 + (B+C)z + D = 0$ has no solution with $|z| > 1$. In particular, this implies that the product of its zeros, D/A, must have magnitude at most 1. But this implies that for the expression in (3) to be zero, we must have $|\lambda| = |D/A| \leq 1$, which is what we wanted to prove.[2] A similar argument applies to the partition functions of the induced subgraphs of H'; we omit the details.

We now proceed to the proof of Theorem 6.2 for a given graph G. We first consider the case of a graph with just one vertex; in this case, the partition function is simply $1 + \lambda_1$, and hence the Lee-Yang property is trivially satisfied. We now consider the first important case, that of a single edge. In this case, the partition function is $1 + \beta(\lambda_1 + \lambda_2) + \lambda_1\lambda_2$. Now suppose $|\lambda_1| > 1$. For the partition function to be 0, we must have

$$|\lambda_2| = \left| \frac{1 + \beta\lambda_1}{\beta + \lambda_1} \right| < 1, \tag{4}$$

where the last equation uses the fact that for $\beta < 1$, the above fraction is a Möbius transformation that takes the exterior of the unit disk to its interior. These two observations show that a single edge has the Lee-Yang property.

We now notice that if two graphs G_1 and G_2 on disjoint sets of vertices have the Lee-Yang property, then so does the graph $G' = G_1 \cup G_2$. This follows simply because the partition function of G' is the product of the partition functions of G_1 and G_2 (and a similar argument applies to their subgraphs). Now, to prove that a given graph G with m edges has the Lee-Yang property, we start with a disjoint union of m edges, each corresponding to one of the edges of G; note that this "graph" has the Lee-Yang property. At each stage, we take the end-points of two edges which are supposed to be incident to the same vertex in G, and merge these end-points. Note that this process culminates with the graph G. By the Asano contraction lemma, the resulting graph after each of these steps continues to have the Lee-Yang property, and hence G must also have the Lee-Yang property. $\qquad\square$

---

[2]We implicitly assumed in the above argument that for our fixed values of $\lambda_3, \ldots, \lambda_n$, we have $A \neq 0$. This technicality can be avoided by noting that A can be seen as the partition function of the Ising model on the induced subgraph of H obtained by removing v_1 and v_2, and hence satisfies the Lee-Yang property. We omit the details in this article. This is the reason why we needed the stronger hypothesis that involves all sub-graphs.

References

[1] Taro Asano. Lee-Yang theorem and the Griffiths inequality for the anisotropic Heisenberg ferromagnet. *Physical Review Letters*, 24(25):1409–1411, June 1970.

[2] Julius Borcea and Petter Brändén. The Lee-Yang and Pólya-Schur programs. I. linear operators preserving stability. *Inventiones mathematicae*, 177(3):541–569, 2009.

[3] Julius Borcea and Petter Brändén. The Lee-Yang and Pólya-Schur programs. II. theory of stable polynomials and applications. *Communications on Pure and Applied Mathematics*, 62(12):1595–1631, 2009.

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

[5] Maria Chudnovsky and Paul Seymour. The roots of the independence polynomial of a clawfree graph. *Journal of Combinatorial Theory, Series B*, 97(3):350–357, May 2007.

[6] G. P. Erorychev. Proof of the van der Waerden conjecture for permanents. *Siberian Mathematical Journal*, 22:854–859, 1981.

[7] D. I. Falikman. Proof of the van der Waerden conjecture regarding the permanent of a doubly stochastic matrix. *Mathematical Notes*, 29:475–479, 1981.

[8] Tomás Feder and Milena Mihail. Balanced matroids. In *Proceedings of the twenty-fourth annual ACM symposium on Theory of computing*, pages 26–38. ACM, 1992.

[9] Steve Fisk. Polynomials, roots, and interlacing. *arXiv preprint math/0612833*, 2006.

[10] Shayan Oveis Gharan, Amin Saberi, and Mohit Singh. A randomized rounding approach to the traveling salesman problem. In *FOCS*, pages 550–559, 2011.

[11] Leonid Gurvits. Hyperbolic polynomials approach to van der Waerden/Schrijver-valiant like conjectures: sharper bounds, simpler proofs and algorithmic applications. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 417–426. ACM, 2006.

[12] T. D. Lee and C. N. Yang. Statistical theory of equations of state and phase transitions. II. Lattice gas and Ising model. *Physical Review*, 87(3):410–419, August 1952.

[13] Adrian Lewis, Pablo Parrilo, and Motakuri Ramana. The lax conjecture is true. *Proceedings of the American Mathematical Society*, 133(9):2495–2499, 2005.

[14] Adam Marcus, Daniel A Spielman, and Nikhil Srivastava. Interlacing families I: Bipartite Ramanujan graphs of all degrees. *arXiv preprint arXiv:1304.4132*, 2013.

[15] Adam Marcus, Daniel A Spielman, and Nikhil Srivastava. Interlacing families II: Mixed characteristic polynomials and the Kadison-Singer problem. *arXiv preprint arXiv:1306.3969*, 2013.

[16] Tobias Mömke and Ola Svensson. Approximating graphic TSP by matchings. In *FOCS*, pages 560–569, 2011.

[17] Robin Pemantle. Hyperbolicity and stable polynomials in combinatorics and probability. *arXiv preprint arXiv:1210.3231*, 2012.

[18] Robin Pemantle and Yuval Peres. Concentration of lipschitz functionals of determinantal and other strong rayleigh measures. *arXiv preprint arXiv:1108.0687*, 2011.

[19] David Ruelle. Is our mathematics natural? The case of equilibrium statistical mechanics. *Bulletin (New Series) of the American Mathematical Society*, 19(1):259–268, July 1988.

[20] Alexander Schrijver. Counting 1-factors in regular bipartite graphs. *Journal of Combinatorial Theory, Series B*, 72(1):122–135, 1998.

[21] Alistair Sinclair and Piyush Srivastava. Lee-Yang theorems and the complexity of computing averages. In *Proceedings of the 45th annual ACM symposium on Symposium on theory of computing*, pages 625–634. ACM, 2013.

[22] Piyush Srivastava. The Lee-Yang theory of phase transitions, 2013.

[23] Nisheeth K. Vishnoi. A permanent approach to the traveling salesman problem. *In FOCS*, 2012.

[24] David Wagner. Multivariate stable polynomials: theory and applications. *Bulletin of the American Mathematical Society*, 48(1):53–84, 2011.

[25] C. N. Yang and T. D. Lee. Statistical theory of equations of state and phase transitions. I. Theory of condensation. *Physical Review*, 87(3):404–409, August 1952.