

Hyperbolicity and stable polynomials in combinatorics and probability

Robin Pemantle^{1,2}

ABSTRACT: These lectures survey the theory of hyperbolic and stable polynomials, from their origins in the theory of linear PDE's to their present uses in combinatorics and probability theory.

Keywords: amoeba, cone, dual cone, Gårding-hyperbolicity, generating function, half-plane property, homogeneous polynomial, Laguerre–Pólya class, multiplier sequence, multi-affine, multivariate stability, negative dependence, negative association, Newton's inequalities, Rayleigh property, real roots, semi-continuity, stochastic covering, stochastic domination, total positivity.

Subject classification: Primary: 26C10, 62H20; secondary: 30C15, 05A15.

¹Supported in part by National Science Foundation grant # DMS 0905937

²University of Pennsylvania, Department of Mathematics, 209 S. 33rd Street, Philadelphia, PA 19104 USA, pemantle@math.upenn.edu

Contents

1	Introduction	1
2	Origins, definitions and properties	4
2.1	Relation to the propagation of wave-like equations	4
2.2	Homogeneous hyperbolic polynomials	7
2.3	Cones of hyperbolicity for homogeneous polynomials	10
3	Semi-continuity and Morse deformations	14
3.1	Localization	14
3.2	Amoeba boundaries	15
3.3	Morse deformations	17
3.4	Asymptotics of Taylor coefficients	19
4	Stability theory in one variable	23
4.1	Stability over general regions	23
4.2	Real roots and Newton's inequalities	26
4.3	The Laguerre–Pólya class	31
5	Multivariate stability	34
5.1	Equivalences	36
5.2	Operations preserving stability	38
5.3	More closure properties	41
6	Negative dependence	41
6.1	A brief history of negative dependence	41
6.2	Search for a theory	43

6.3	The grail is found: application of stability theory to joint laws of binary random variables	47
6.4	Proof of Theorem 6.9	48
7	Further applications of stability: determinants, permanents and moments	56
7.1	The van der Waerden conjecture	56
7.2	The Monotone Column Permanent Conjecture	60
7.3	The BMV conjecture	61

1 Introduction

These lectures concern the development and uses of two properties, **hyperbolicity** and **stability**. Hyperbolicity has been used chiefly in geometric and analytic contexts, concerning wave-like partial differential equations, lacunas, and generalized Fourier transforms. Stability has its origins in control theory but has surfaced more recently in combinatorics and probability, where its algebraic properties, such as closure under various operations, are paramount. We begin with the definitions.

Definition 1.1 (hyperbolicity). *A homogeneous complex d -variable polynomial $p(z_1, \dots, z_d)$ is said to be hyperbolic in direction $\mathbf{x} \in \mathbb{R}^d$ if and only if*

$$p(\mathbf{x} + i\mathbf{y}) \neq 0 \text{ for all } \mathbf{y} \in \mathbb{R}^d. \quad (1.1)$$

A non-homogeneous d -variable polynomial q with leading homogeneous part p is said to be hyperbolic if and only if $p(\mathbf{x}) \neq 0$ and there is some $t_0 > 0$ such that

$$q(it\mathbf{x} + \mathbf{y}) \neq 0 \text{ for all } \mathbf{y} \in \mathbb{R}^d \text{ and real } t < t_0. \quad (1.2)$$

A polynomial is said to be hyperbolic if it is hyperbolic in direction ξ for some $\xi \in \mathbb{R}^d$.

Definition 1.2 (stability). *The complex polynomial q is said to be **stable** if and only if*

$$\Im\{z_j\} > 0 \text{ for all } j = 1, \dots, d \text{ implies } q(z_1, \dots, z_d) \neq 0. \quad (1.3)$$

*A **real stable** polynomial is one that is both real and stable.*

The following relation holds between these two definitions.

Proposition 1.3. *A real homogeneous polynomial p is stable if and only if it is hyperbolic in direction ξ for every ξ in the positive orthant.*

Hyperbolicity is a real geometric property, that is, it is invariant under the general linear group. It is not invariant under complex linear transformations, rather it distinguishes the real subspace of \mathbb{C}^d . Stability is invariant under yet fewer maps than is hyperbolicity, due to the fact that it distinguishes positive from negative. Stability is invariant under coordinatewise transformations of the upper half-plane, such as inversion $z \mapsto -1/z$ and dilation $z \mapsto \lambda z$, from which one can get a surprising amount of mileage.

The notion of hyperbolicity for polynomials has been around since its introduction by L. Gårding sixty years ago [Går51].³ The notion of stability for real or complex polynomials in several variables goes back at least to multivariate versions of Hurwitz stability in the 1980's, but appears to have become established as tools for combinatorialists only in the last five years [COSW04, Brä07, BBL09, Gur08]. These two related notions have now been applied in several seemingly unrelated fields.

³Gårding credits many of the ideas to I. G. Petrovsky in his seminal paper [Pet45].

Gårding's original purpose was to find the right condition to guarantee analytic stability of a PDE as it evolves in time from an original condition in space; the notion of stability here is that small perturbations of the initial conditions should produce small perturbations of the solution at time t . That Gårding found the right definition is undeniable because his result (Theorem 2.1 below) is an equivalence. His proof uses properties of hyperbolic functions and their cones that are derived directly from the definitions. Specifically, in order to construct solutions to boundary value problems via the **Riesz kernel**, an inverse Fourier transform is computed on a linear space $\{\mathbf{x} + i\mathbf{y} : \mathbf{y} \in \mathbb{R}^d\}$; hyperbolicity in the direction \mathbf{x} is used to ensure that this space does not intersect the set $\{q = 0\}$ where q^{-a} is singular. Gårding developed some properties of hyperbolic polynomials in this paper and later devoted a separate paper to the further development of some inequalities [Går59].

Twenty years later, with Atiyah and Bott [ABG70], this work was extended considerably. They were able to compute inverse Fourier transforms more explicitly for a number of homogeneous hyperbolic functions. To do this, they developed semi-continuity properties of hyperbolic functions. These properties enable the construction of certain vector fields and deformations, which in turn are used to deform the plane $\{\mathbf{x} + i\mathbf{y} : \mathbf{y} \in \mathbb{R}^d\}$ into a cone on which the inverse Fourier transforms are directly integrable. Direct integrability provides useful estimates and also enables explicit computation via topological dimension reduction theorems of Leray and Petrovsky. My introduction to this exceptional body of work came in the early 2000's when I needed to apply these deformations to multivariate analytic combinatorics; refining the relatively crude techniques in [PW02, PW04] led to [BP11].

In control theory, a continuous-time (respectively discrete-time) system whose transfer function is rational will be stable if its poles all lie in the open left half-plane (respectively the open unit disk). Accordingly, we say a univariate polynomial is **Hurwitz stable** (respectively **Schur stable**) if its zeros are all in the open left half-plane (respectively the open unit disk). Multivariate generalizations of Hurwitz and Schur stability have arisen in a variety of problems. The development of these generalizations splits into two veins. The review paper [Sok05] surveys a number of “hard” results on zero-free regions, meaning results that give good estimates on zero-free regions that depend on specific parameters of the models or graphs from which the polynomial is formed. Most relevant to the applications in Sections 6 and 7 are “soft” results, which give simple domains free of zeros (such as products of half planes), valid for all graphs or for wide classes of graphs.

The development of this vein of multivariate stable function theory took place mainly in the area of statistical physics and combinatorial extensions thereof. Multivariate Schur stability arises in the celebrated Lee–Yang Theorem for ferromagnetic Ising models [YL52, LY52], where it implies the absence of phase transitions at nonzero magnetic field. Likewise, multivariate Hurwitz stability arises in electrical circuit theory [FB85, COSW04, Sok11] and matroid generalizations thereof [COSW04, Brä07, WW09], and has applications to combinatorial enumeration [Wag06, Wag08]. It also arises in generalizations of the Lee–Yang Theorem [LS81] and in the Heilmann-Lieb Theorem for matching polynomials (also known as monomer-dimer models) [HL72, COSW04]. It should be noted that

Hurwitz stability [Hur96], differs from the notion in Definition 1.2 in that the zero-free region is the right half-plane rather than the upper half-plane. For general complex polynomials the two notions are equivalent under a linear change of variables, but for real polynomials the two notions are very different. Except for brief portions of Sections 4 and 5, our concern will be with stability as defined in Definition 1.2

Multivariate stability behaves nicely under certain transformations. The closure properties were investigated further by Borcea and Brändén [BB09a]. These closure properties turn out to have powerful implications for systems of negatively dependent random variables, as worked out in 2009 by Borcea, Brändén and Liggett [BBL09]. My second brush with this subject was when I needed to apply some of their results to determinantal point processes [PP11].

These two encounters with hyperbolic/stable polynomials in seemingly dissimilar areas of mathematics prompted my search for a unified understanding. As pointed out at the end of the introduction of [BBL09], such a viewpoint was taken 25 years ago by Gian-Carlo Rota.

“The one contribution of mine that I hope will be remembered has consisted in just pointing out that all sorts of problems of combinatorics can be viewed as problems of the locations of zeros of certain polynomials...”

to which we would add “and also problems in differential equations, number theory, probability, and perhaps many more areas.”

Presently, the greatest interest in the subject of hyperbolic/stable polynomials seems to be its potential for unlocking some combinatorial conjectures concerning determinants such as the Bessis–Moussa–Villani conjecture, Lieb’s “permanent on top” conjecture and extensions of the van der Waerden conjecture, proved in 1979, but still an active subject (see, e.g., [Gur06]). The distances between areas of mathematics in which hyperbolicity plays a role are such that it was difficult to find a primary subject classification for these lectures. It also poses a unique challenge for me, as I am an expert in none of these, save for applications to analytic combinatorics. I therefore apologize in advance for any historical inaccuracies perpetrated here, or for idiosyncratic viewpoints arising from my personal history with the subject. Here follows an attempt to lay out its two main pillars, hyperbolicity and stability, and to follow their progress up to the present day.

Part I: Hyperbolicity

2 Origins, definitions and properties

2.1 Relation to the propagation of wave-like equations

Gårding's objective was to prove stability results for wave-like partial differential equations. Endow the space of smooth complex valued functions of d real variables, denoted $C^\infty(\mathbb{R}^d)$, with the topology of uniform convergence on compact sets of all the partial derivatives. For any polynomial q in d variables, let $D[q] = q(\partial/\partial\mathbf{x})$ denote the linear differential operator obtained by formally substituting $(\partial/\partial x_i)$ for x_i in q . If we consider \mathbb{R}^d as spacetime with the positive time direction given by a vector ξ , and denote by H_ξ the hyperplane orthogonal to ξ , then the boundary value equation

$$D[q](f) = 0, \quad f = g \text{ on } H_\xi \tag{2.1}$$

on the halfspace $\{\mathbf{x} : \mathbf{x} \cdot \xi \geq 0\}$ may be interpreted as the evolution of g under the equation $D[q](f) = 0$ with initial condition g . One would expect $m - 1$ further initial conditions, such as $(\partial^j/\partial x_\xi^j)f = g_j$ for $1 \leq j \leq m - 1$, where $(\partial/\partial x_\xi)$ denotes a derivative in direction ξ and m is the degree of q .

We consider (2.1) to be **stable under perturbations** (in time direction ξ) if for all sequences $\{f_n\}$ of functions in $C^\infty(\mathbb{R}^d)$ satisfying $D[q](f_n) = 0$, convergence to zero of the restrictions of f_n to H_ξ implies convergence to zero of f_n on the whole space. Intuitively, arbitrary small change in the initial conditions cannot lead to macroscopic changes in the evolution at a later time. The first of Gårding's results can be stated as follows.

Theorem 2.1 ([Går51, Theorem III]). *The differential equation (2.1) is stable in direction ξ if and only if q is hyperbolic in direction ξ .*

It should be noted that the definition of hyperbolicity in direction \mathbf{x} in Gårding's 1951 paper is that $q(t\mathbf{x} + i\mathbf{y}) \neq 0$ when $\mathbf{y} \in \mathbb{R}^d$ and $t > t_0$. This does not appear to me to be equivalent to Definition 1.1, which appears in all later work including [ABG70]. The two definitions specialize to the same thing when q is homogeneous. Although the primary focus of these notes is on later

uses outside of PDE's, I will give indications of the proof. Not only does this satisfy curiosity and a historical sense, but it promotes the goal of using some of the geometric understanding when dealing with contemporary, more algebraic problems.

To keep things simple, assume q is homogeneous. Examples will be discussed shortly, but for now, a good mental picture is to keep in mind the example $d = 3$ and $q(x, y, z) = z^2 - x^2 - y^2$. Even though \mathbb{R}^d and its dual are isomorphic, I find it useful to classify a vector as belonging to the dual space $(\mathbb{R}^d)^*$ rather than \mathbb{R}^d if it plays the role of a linear functional on the original space. For example the frequency ξ of a wave $\mathbf{x} \mapsto \exp(i\xi \cdot \mathbf{x})$ is thought of as living in $(\mathbb{R}^d)^*$.

FORWARD DIRECTION: For any $\xi \in (\mathbb{R}^d)^*$, let f_ξ denote the function $\mathbf{x} \mapsto e^{i\xi \cdot \mathbf{x}}$ mapping \mathbb{R}^d to \mathbb{C} . Of course $(\mathbb{R}^d)^*$ is isomorphic to \mathbb{R}^d but we call it $(\mathbb{R}^d)^*$ to remind ourselves that ξ runs over frequency space. When \mathbf{r} is real, the function $f_{\mathbf{r}}$ is a sinusoidal wave and is bounded, as in Figure 1. Let us assume without loss of generality that $\xi = (0, \dots, 0, 1)$. If $q(\xi) = 0$ then f_ξ is a solution to

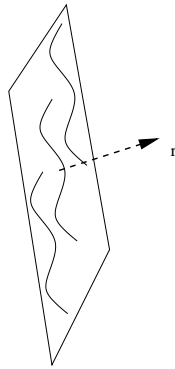


Figure 1: $f_{\mathbf{r}}$ when \mathbf{r} is real

$D[q](f) = 0$. If q is not hyperbolic in direction ξ then there is some (r_1, \dots, r_{d-1}) such that the d solutions to $q(r_1, \dots, r_{d-1}, t) = 0$ include at least one conjugate pair of values that are not real. Denoting by r_d the one with the negative imaginary part, we see that $f_{\mathbf{r}}(\mathbf{x})$ grows exponentially as x_d increases. By homogeneity, $f_{\lambda \mathbf{r}}$ grows at λ times the exponential rate. Choose c_λ so that $c_\lambda f_{\lambda \mathbf{r}}(0, \dots, 0, 1) = 1$. Then $f_{n\mathbf{r}} \rightarrow 0$ on the hyperplane H_ξ in the topology of uniform convergence (due to periodicity, no compact set restriction is required), while $f_{n\mathbf{r}} \equiv 1$ at $(0, \dots, 0, 1)$, proving the contrapositive, namely that lack of hyperbolicity implies lack of stability. \square

BACKWARD DIRECTION, HANDWAVING ARGUMENT: For fixed (r_1, \dots, r_{d-1}) , the set $\{f_{\mathbf{r}} : \mathbf{r} \in (\mathbb{R}^d)^*\}$ of solutions to $D[q]f = 0$ spans a vector space $V(r_1, \dots, r_{d-1})$ of dimension k . Since we are waving our hands, we have assumed **strong** hyperbolicity, namely that the d values of r_d are distinct. Any integral $\int c(r_1, \dots, r_{d-1}) dr_1 \cdots dr_{d-1}$ is also a solution, where $c(\cdot)$ is a section of $V(\cdot)$. The handwaving part is that this gives all solutions of $D[q]f = 0$. Assuming this, we write a generic solution f as such an integral. Each f_ξ evolves unitarily, so more handwaving along the lines of a

Parseval relation shows that if f and its first k derivatives are small at time zero then the time t value is small as well. \square

Removing all the handwaving takes considerable work, beyond our scope here. The key is the construction of the Riesz kernel, which is interesting enough to merit a brief digression. Suppose that q is m -homogeneous and hyperbolic in direction \mathbf{x} and let α is any complex number. Define the Riesz kernel Q_α by

$$Q_\alpha(\mathbf{y}) := (2\pi)^{-d} \int_{\mathbf{z} \in \mathbf{x} + i\mathbb{R}^d} q(\mathbf{z})^{-\alpha} \exp(\mathbf{y} \cdot \mathbf{z}) d\mathbf{z}. \quad (2.2)$$

We have not yet defined cones of hyperbolicity, but let us nonetheless try to visualize why the Riesz kernel is well defined and supported on the dual cone. To explain the terminology, the convex dual of a cone $K \subseteq \mathbb{R}^d$ is the cone $K^* \subseteq (\mathbb{R}^d)^*$ consisting of vectors \mathbf{y} such that $\mathbf{w} \cdot \mathbf{y} \geq 0$ for all $\mathbf{w} \in K$. Duality maps fat cones to skinny cones and vice versa (see Figure 2). The connected

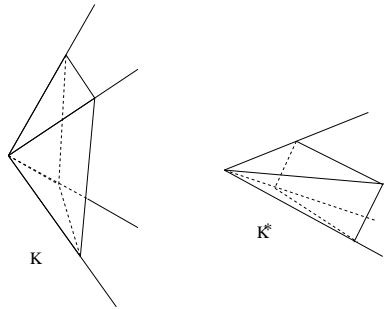


Figure 2: The cone K and its dual K^* .

components of the open set $\{\mathbf{w} \in \mathbb{R}^d : q(\mathbf{w}) \neq 0\}$ are cones. If convex they have duals. Let C be the component containing \mathbf{x} . We will see later that hyperbolicity guarantees this is one of the convex ones. Hyperbolicity in direction \mathbf{x} guarantees that q is nonvanishing on the domain of integration $\mathbf{x} + i\mathbb{R}^d$. This is illustrated in Figure 3, where the dashed lines signify that the contour of integration is varying in imaginary directions only and therefore does not hit the surface $\{q = 0\}$. The magnitude of the exponential term is constant, whence the integral is convergent at infinity

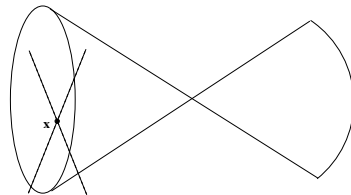


Figure 3: Chain of integration in the definition of Q_α

when $m \operatorname{Re} \{\alpha\} > 1 + d$. Because the chain of integration avoids $\{q = 0\}$, integrability is assured

as is a well defined branch of $q^{-\alpha}$. It follows easily from Cauchy's Theorem that the Riesz kernel is independent of the choice of \mathbf{x} within the component C . Furthermore, if $\mathbf{y} \in (\mathbb{R}^d)^*$ is a vector for which some $\mathbf{x} \in C$ has $\mathbf{x} \cdot \mathbf{y} < 0$, then it is clear that the integral defining $Q_\alpha(\mathbf{y})$ can be deformed so as to be arbitrarily small, hence is equal to zero. In other words, Q_α is supported on the set of \mathbf{y} for which there is no such \mathbf{x} , which is precisely the dual cone C^* . \square

To complete the sketchy proof of the backward direction, convolving with the Riesz kernel gives an operator I_q inverted by $D[q]$. This construction is stable under perturbations. If $D[q]f = 0$, one shows that f is equal to $(I - I_q D[q])\tilde{f}$, where \tilde{f} is defined stably in terms of f and its first $d - 1$ normal derivatives on H_ξ . Small changes in the initial conditions thus give rise to small changes in \tilde{f} , which give rise to small changes in f , proving the theorem.

The Riesz kernels with parameters α and $\alpha - 1$ are related by a differential identity as long as $\alpha \neq 1$. When $\alpha = 1$, the Riesz kernel is called the **fundamental solution** by reason of a result of Gårding. This is explained more fully in Theorem 2.8 below, but the short statement, found for instance in [Gül97, Theorem 2.2], is that the solution of

$$D[q]f(\mathbf{x}) = \delta(\mathbf{x})$$

exists, is unique, and is supported on the dual cone, C^* .

2.2 Homogeneous hyperbolic polynomials

Many of the properties of hyperbolic polynomials are easier to state, prove and understand in the homogeneous case. Throughout this section, therefore, we deal only with homogeneous polynomials. We will use p rather than q as a visual cue when speaking about homogeneous polynomials. Hyperbolicity is easily seen to be equivalent to the equivalent to the following “real root” property.

Proposition 2.2. *The homogeneous polynomial p is hyperbolic in direction \mathbf{x} if and only if for any $\mathbf{y} \in \mathbb{R}^d$ the univariate polynomial $t \mapsto p(\mathbf{y} + t\mathbf{x})$ has only real roots.*

PROOF: Because p is homogeneous, when $\lambda \neq 0$, we have $p(\lambda\mathbf{z}) = 0$ if and only if $p(\mathbf{z}) = 0$. With $\lambda = is$ for some real nonzero value of s we see that $p(\mathbf{x} + i\mathbf{y}) \neq 0$ for all $\mathbf{y} \in \mathbb{R}^d$ (the definition of hyperbolicity) is equivalent to $p(\mathbf{y} + is\mathbf{x}) \neq 0$ for all $\mathbf{y} \in \mathbb{R}^d$ and all nonzero real s . This is equivalent to $p(\mathbf{y} + t\mathbf{x} + is\mathbf{x}) \neq 0$ for all $\mathbf{y} \in \mathbb{R}^d$ and real s, t with $s \neq 0$. Writing $z = t + is$, this is equivalent to $p(\mathbf{y} + z\mathbf{x}) \neq 0$ when z is not real. \square

Denote by $t_k(\mathbf{x}, \mathbf{y})$ the roots of $t \mapsto p(\mathbf{y} + t\mathbf{x})$. Then

$$p(\mathbf{y} + t\mathbf{x}) = p(\mathbf{x}) \prod_{k=1}^m [t - t_k(\mathbf{x}, \mathbf{y})]. \quad (2.3)$$

By Proposition 2.2, when p is hyperbolic in direction ξ , all values $t_k(\mathbf{x}, \mathbf{y})$ are real, hence, setting $t = 0$, we see that $p/p(\mathbf{x})$ is a real polynomial. We see that little generality is lost in restricting our discussion of homogeneous polynomials to those with real coefficients.

The two nontrivial examples of hyperbolic polynomials given in Gårding's original paper, namely Lorentzian quadratics and the determinant, provide much of the intuition as to the meaning of hyperbolicity. We now discuss these. A preliminary observation is that, trivially, all real homogeneous polynomials of degree one are hyperbolic in all directions in which they do not vanish. Next, observe that hyperbolicity of the homogeneous polynomial p in direction ξ is preserved when p and ξ are transformed by the same invertible real linear map. This allows us to classify all nondegenerate quadratics: it suffices to consider all polynomials of the form $\sum_{j=1}^d \pm x_j^2$, hyperbolicity evidently being determined by signature. A necessary and sufficient condition for hyperbolicity is Lorentzian signature, that is precisely one sign different from the others.

Example 2.3 (Lorentzian quadratics). *Let $p(\mathbf{x}) := x_1^2 - \sum_{j=2}^d x_j^2$. Real vectors ξ may be classified as time-like, light-like or space-like according to whether $p(\xi)$ is respectively positive, zero or negative. The time-like vectors form the open convex cone $\{\mathbf{x} : (x_2/x_1)^2 + \cdots + (x_d/x_1)^2 < 1\}$. Fix any time-like vector ξ . If $\eta = \lambda\xi$ then the line $t \mapsto \eta + t\xi$ contains a doubled root of p at the origin. For any other real η , the line $\eta + t\xi$ intersects the hyperplane H_ξ orthogonal to ξ at some point other than the origin and p takes a negative value there. On the other hand, the quadratic $p(\eta + t\xi)$ has positive leading term, so goes to $+\infty$ at $t = \pm\infty$. Hence the line intersects the zero set of p twice. The degree of p is two, hence for any η , the polynomial $t \mapsto p(\eta + t\xi)$ has all real roots. Hence p is hyperbolic. Replacing p by $-p$ does not affect hyperbolicity. Hence homogeneous quadratics with signature 1 or $d - 1$ are hyperbolic in timelike directions.*

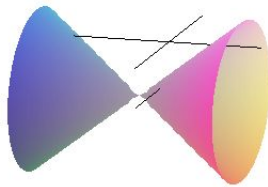


Figure 4: This Lorentzian quadratic is hyperbolic in the x -direction but not the y -direction.

For the converse, elliptic quadratics have lines in every direction with no real roots, hence are obviously not hyperbolic. If $d > 4$ and there are at least two positive and two negative directions, the any line on which p has a minimum may be translated in a positive direction not on the line so that the minimum becomes positive, and similarly any line on which p has a maximum may be translated so that the maximum becomes negative. We conclude there are no directions of hyperbolicity.

The other classical example is as follows.

Example 2.4 (Determinants of Hermitian matrices). *An $n \times n$ matrix M is Hermitian if $M_{ij} = \overline{M_{ji}}$. The space of Hermitian matrices of size n is parametrized by n^2 real parameters, these being the $\binom{n}{2}$ arbitrary complex numbers $\{M_{ij} : 1 \leq i < j \leq n\}$ and the n real numbers $\{M_{ii} : 1 \leq i \leq n\}$. Let p be the homogeneous polynomial on \mathbb{R}^{n^2} defined by the determinant as a function of these n^2 real numbers. Hyperbolicity of p in the direction of any positive definite matrix A is equivalent to the well known fact that the zeros of $M + zA$ are all real if M is Hermitian and A is positive definite. To see this, let $z = a + bi$ with $b > 0$ and write*

$$\text{Det}(M + zA) = \text{Det}(bA)\text{Det}\left[(bA)^{-1/2}(aA + M)(bA)^{-1/2} + iI\right],$$

where $(bA)^{1/2}$ is the Hermitian positive definite square root of bA . If this quantity is equal to zero then $-i$ is an eigenvalue of $(bA)^{-1/2}(aA + M)(bA)^{-1/2}$ which is impossible because this matrix is Hermitian.

There are several ways to construct new hyperbolic functions from old ones. Note that the first does not require homogeneity.

Proposition 2.5 (products and polarization).

- (i) Let q_1 and q_2 be hyperbolic with respect to \mathbf{x} . Then q_1q_2 is also hyperbolic with respect to \mathbf{x} .
- (ii) Let p be homogeneous of degree m and hyperbolic with respect to \mathbf{x} . Let p_0, \dots, p_m be the coefficients of $p(\mathbf{y} + t\mathbf{x})$ as a polynomial in t :

$$p(\mathbf{y} + t\mathbf{x}) = \sum_{k=0}^m p_k(\mathbf{y})t^k. \tag{2.4}$$

Then p_k is hyperbolic with respect to \mathbf{x} for all $0 \leq k \leq m$.

PROOF: The fact for products is immediate from the definition. For polarization, we first observe that hyperbolicity of p in direction \mathbf{x} implies hyperbolicity of the directional derivative $D[\mathbf{x}][p]$ in direction \mathbf{x} . To see this, observe that $f(t)$ having only real roots implies $f'(t)$ has only real roots. Setting $f(t) = p(\mathbf{y} + t\mathbf{x})$, we see that $f'(t)$, namely $t \mapsto D_{\mathbf{x}}(p)(\mathbf{y} + t\mathbf{x})$, has only real roots for any $\mathbf{y} \in \mathbb{R}^d$, which is hyperbolicity of $D_{\mathbf{x}}(p)$ in direction \mathbf{x} .

Now consider the expansion (2.4) of $p(\mathbf{y} + t\mathbf{x})$ into homogeneous parts. The parts are given by

$$p_k(\mathbf{y}) = \frac{1}{k!} \left(\frac{d}{dt} \right)^k \Big|_{t=0} p(\mathbf{y} + t\mathbf{x}) = (D_{\mathbf{x}})^k (p)(\mathbf{y}).$$

Hyperbolicity in direction \mathbf{x} is stable under $D_{\mathbf{x}}$, which shows that p_k is hyperbolic in direction \mathbf{x} . \square

2.3 Cones of hyperbolicity for homogeneous polynomials

Hyperbolic polynomials have associated with them certain convex cones. In the homogeneous case the definition is simple and is contained in the following proposition. This was proved first in [Går51] and reproduced many times; the proof below follows [Gül97]. Let p be a (complex) homogeneous hyperbolic polynomial, hyperbolic in direction ξ . Dividing by a real multiple of $p(\xi)$ we may assume, cf. (2.3), that p is real and $p(\xi) = 1$.

Proposition 2.6. *Let $K(p, \xi)$ denote the connected component of the set $\mathbb{R}^d \setminus \{\mathbf{x} : p(\mathbf{x}) = 0\}$ that contains ξ .*

- (i) *p is hyperbolic in direction \mathbf{x} for every $\mathbf{x} \in K(p, \xi)$.*
- (ii) *The set $K(p, \xi)$ is an open convex cone; we call this a **cone of hyperbolicity** for p .*
- (iii) *$K(p, \xi)$ is equal to the set K of vectors \mathbf{x} for which all roots of $t \mapsto p(\mathbf{x} + t\xi)$ are real and negative, which (by hyperbolicity of p in direction ξ) is the same as the set of vectors \mathbf{x} for which no root of this polynomial is real and nonnegative.*

PROOF: By continuity of the roots of a polynomial with respect to its coefficients, the set K is open. Also $\xi \in K$ because $p(\xi + t\xi) = (1+t)^m p(\xi)$ where m is the degree of the homogeneous polynomial p . If \mathbf{x} is in the closure \overline{K} then, again by continuity of the roots, $p(\mathbf{x} + t\xi) \neq 0$ for $t > 0$; in particular, $p(\mathbf{x}) \neq 0$ implies $\mathbf{x} \in K$, which means that K is closed in $\mathbb{R}^d \setminus \{p = 0\}$. Also K is connected: by definition of K , if $\mathbf{x} \in K$ then $\mathbf{x} + t\xi \in K$ for $t > 0$, and by homogeneity $s\mathbf{x} + t\xi \in K$ for $s, t > 0$, and sending s to zero proves that K is star-convex at ξ (hence connected). Being connected and both closed and open in $\mathbb{R}^d \setminus \{p = 0\}$, K is a component of $\mathbb{R}^d \setminus \{p = 0\}$ and is thus equal to $K(p, \xi)$.

Next we check that for $\mathbf{v} \in K$, \mathbf{x} real, and s and t complex,

$$p(\mathbf{x} + t\mathbf{v} + s\xi) \neq 0 \text{ whenever } \Im(s) \leq 0, \Im(t) \leq 0, \Im(s) + \Im(t) < 0. \quad (2.5)$$

First, suppose $\Im(s) < 0$. If $u \geq 1$ then the polynomial

$$t \mapsto p \left[\frac{\mathbf{x}}{u} + t\mathbf{v} + \left(\frac{s}{u} + i \left(\frac{1}{u} - 1 \right) \right) \xi \right] \quad (2.6)$$

is nonvanishing for real t because this is p evaluated at the sum of a real vector and a complex multiple of ξ with nonzero imaginary part. As $u \rightarrow \infty$, the number of roots of (2.6) in the lower half-plane remains constant. The limit at $u = \infty$ is $t \mapsto p(t\mathbf{v} - i\xi)$. This has roots in the upper half-plane because its roots are $-i$ divided by the roots of $t \mapsto p(\mathbf{v} + t\xi)$, the latter of which we have seen to be negative real. We conclude that for $u = 1$ there are no roots in the lower half-plane, in other words,

$$p(\mathbf{x} + t\mathbf{v} + s\xi) \neq 0 \text{ whenever } \Im(s) < 0, \Im(t) \leq 0. \quad (2.7)$$

To see that this also holds for $\Im(s) = 0$ and $\Im(t) < 0$, note that K is open, so $\mathbf{v} - \epsilon\xi \in K$ for some $\epsilon > 0$, whence

$$p(\mathbf{x} + t\mathbf{v} + s\xi) = p(\mathbf{x} + t(\mathbf{v} - \epsilon\xi) + (s + \epsilon t)\xi)$$

which is nonzero by (2.7). This completes the verification of (2.5).

Setting $s = 0$ in (2.5) shows that any root of $t \mapsto p(\mathbf{x} + t\mathbf{v})$ satisfies $\Im(t) \geq 0$. Because p is real, complex conjugation may be applied to the entire argument, showing that also $\Im(t) \leq 0$, and hence that all roots of $t \mapsto p(\mathbf{x} + t\mathbf{v})$ are real. The vector \mathbf{v} was chosen arbitrarily in K . We conclude that p is hyperbolic with respect to every $\mathbf{v} \in K$. It follows that K is star convex with respect to \mathbf{v} as well, and since $\mathbf{v} \in K$ is arbitrary, that K is convex. \square

Because the cones of hyperbolicity are characterized as components of the nonzero set in real space, it follows that the cones of hyperbolicity of p and q are pairwise intersections of the cones of hyperbolicity of p with the cones of hyperbolicity of q .

Example 2.7 (coordinate planes). *Each coordinate function z_j is homogeneous of degree 1, therefore hyperbolic in every direction not contained in the plane $\{z_j = 0\}$. The cones of hyperbolicity are the two half spaces bounded by this plane. It follows that the product $\prod_{j=1}^d z_j$ is hyperbolic in every direction in which no coordinate vanishes, and that the cones of hyperbolicity are the orthants. This is also obvious from Figure 5.*

Returning briefly to the origins of hyperbolicity theory in the properties of wave-like partial differential equations, the following result was proved in [ABG70, Theorem 4.1]. It says roughly that cones of hyperbolicity are propagation cones for solutions to wave-like equations. Given any convex cone $K \subseteq \mathbb{R}^d$ over the origin, let K^* denote the dual cone of vectors $\mathbf{v}^* \in (\mathbb{R}^d)^*$ such that $\langle \mathbf{v}^*, \mathbf{x} \rangle \leq 0$ for all $\mathbf{x} \in K$. Recall the linear partial differential operator $D[p]$ defined by substituting $\partial/\partial x_i$ for x_i in p . Let $E = E(p, \mathbf{x})$ be a solution to

$$D[p]E = \delta_{\mathbf{x}} \quad (2.8)$$

where $\delta_{\mathbf{x}}$ is a delta function at \mathbf{x} .

Theorem 2.8. *Suppose the homogeneous polynomial p is hyperbolic in direction ξ . Then there exists a unique distribution E solving (2.8) and its support is the cone $K(p, \xi)^*$ dual to the cone of*

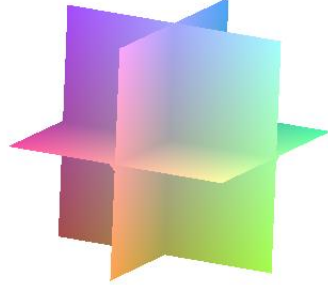


Figure 5: The cones of hyperbolicity for $\prod_{j=1}^d z_j$ are the orthants

hyperbolicity of p containing ξ . In fact E is given as in inverse Fourier transform:

$$E(p, \xi) = (2\pi)^{-d} \int_{\mathbb{R}^d} p_-(\mathbf{y})^{-1} e^{i(\mathbf{x}, \mathbf{y})} d\mathbf{y}$$

where $p_-(y)^{-1} := \lim_{t \downarrow 0} p(\mathbf{y} - it\nu)^{-1}$ as distributions and ν is any element of $K(p, \xi)$. □

Although there is no time here to explore the properties of hyperbolic polynomials as barrier functions, I will mention this example briefly, as it has proved to be of some importance to convex programming.

Example 2.9 (self-concordant barrier functions). *Let $Q \subseteq \mathbb{R}^d$ be an open convex set. A function $F : Q \rightarrow \mathbb{R}$ is a **logarithmically homogeneous self-concordant barrier function** for Q if it is smooth, convex, and satisfies several properties:*

$$\begin{aligned} F(\mathbf{x}) &\rightarrow \infty \text{ as } \mathbf{x} \rightarrow \partial Q \\ |D^3 F(\mathbf{x})[u, u, u]| &\leq 2 (D^2 F(\mathbf{x})[u, u])^{3/2} \\ \|DF(\mathbf{x})[u]\|^2 &\leq \theta D^2 F(\mathbf{x})[u, u] \\ F(t\mathbf{x}) &= F(\mathbf{x}) - \theta \log t. \end{aligned}$$

The interior point method for convex programming problems (see, e.g. [NN93]) is based on finding such a function for a given region, Q . There is a universal construction but it is not always useful for computations. Also, properties beyond those satisfied by the universal construction are required for long term stability of the interior point method. Such properties will depend on the region Q . When Q is the cone of hyperbolicity for a homogeneous polynomial p , it turns out that the function $F(\mathbf{x}) := -\log p(\mathbf{x})$ is a logarithmically homogeneous self-concordant barrier function with a number of other useful properties. These are detailed in [Gül97].

Hyperbolicity without homogeneity

Delving into the relation between hyperbolicity in the homogeneous and non-homogeneous cases, (definitions both given in Definition 1.1, we begin by looking at the homogeneous parts of a polynomial, f . Sorting the terms of f by total degree, the highest degree part will be denoted $LT(f)$ for “leading term(s)”. The lowest degree is called the **localization** and will be discussed further in the next section. If f is hyperbolic, then both the leading term and the localization are hyperbolic as well (respectively [ABG70, Lemma 3.20] and [ABG70, Lemma 3.42]). By definition, the cones of hyperbolicity of a non-homogeneous polynomial f are the cones of hyperbolicity of $LT(f)$. When f is hyperbolic, these cones still characterize the directions of stable propagation of PDE’s and supports of solutions to these.

Because hyperbolicity is easier to understand in the homogeneous case, Gårding looked for a converse to these and came up with a criterion for homogeneous polynomials called strong hyperbolicity. This is most naturally stated using the equivalent definition of hyperbolicity from Proposition 2.2: the roots of $p(\mathbf{y} + t\mathbf{x})$ must not only be real but also distinct (except when \mathbf{y} is a multiple of \mathbf{x} , when all the zeros coincide perforce). We then have:

Theorem 2.10. *If A is strongly hyperbolic then f is hyperbolic for any f such that $LT(f) = A$.*

This is proved as (3.9) of [ABG70] and in fact is already in [Går51] after the statement and proof of Lemma 2.5. Necessary and sufficient conditions for f to be hyperbolic are given in [Hör83, Theorem 12.4.6]; these were conjectured by Gårding and proved first by Svensson [Sve68].

3 Semi-continuity and Morse deformations

3.1 Localization

Given a function f analytic on a neighborhood of the origin, its order of vanishing is the least total degree of a nonvanishing term in its Taylor series. The sum of all such terms is called the homogeneous part of f at the origin and denoted $\text{hom}(f)$. For any \mathbf{x} we let

$$\text{hom}(f, \mathbf{x}) := \text{hom}(f(\mathbf{x} + \cdot))$$

denote the homogeneous part of f at \mathbf{x} . When $\mathbf{x} \neq 0$, Atiyah *et al.* [ABG70] define this by taking the $t = 0$ term of $t^m f(t^{-1}\mathbf{x} + \mathbf{y})$ as a function of \mathbf{y} , where $m = \deg(f)$, and they refer to this as the **localization** of f at \mathbf{x} .

Proposition 3.1. *Let p be any hyperbolic homogeneous polynomial, and let m be its degree. Fix \mathbf{x} with $p(\mathbf{x}) = 0$ and let $\tilde{p} := \text{hom}(p, \mathbf{x})$ denote the leading homogeneous part of p at \mathbf{x} . If p is hyperbolic in direction \mathbf{u} then \tilde{p} is also hyperbolic in direction \mathbf{u} . Consequently, if B is any cone of hyperbolicity for p then there is some cone of hyperbolicity for \tilde{p} containing B .*

PROOF: This follows from the conclusion (3.45) of [ABG70, Lemma 3.42]. Because the development there is long and complicated, we give here a short, self-contained proof, provided by J. Borcea [BP11, Proposition 2.8]. If Q is a polynomial whose degree at zero is k , we may recover its leading homogeneous part $\text{hom}(Q)$ by

$$\text{hom}(Q)(\mathbf{y}) = \lim_{\lambda \rightarrow \infty} \lambda^k Q(\lambda^{-1}\mathbf{y}).$$

The limit is uniform as \mathbf{y} varies over compact sets. Indeed, monomials of degree k are invariant under the scaling on the right-hand side, while monomials of degree $k + j$ scale by λ^{-j} , uniformly over compact sets.

Apply this with $Q(\cdot) = p(\mathbf{x} + \cdot)$ and $\mathbf{y} + t\mathbf{u}$ in place of \mathbf{y} to see that for fixed \mathbf{x}, \mathbf{y} and \mathbf{u} ,

$$\tilde{p}(\mathbf{y} + t\mathbf{u}) = \lim_{\lambda \rightarrow \infty} \lambda^k p(\mathbf{x} + \lambda^{-1}(\mathbf{y} + t\mathbf{u}))$$

uniformly as t varies over compact sub-intervals of \mathbb{R} . Because p is hyperbolic in direction \mathbf{u} , for any fixed λ , all the zeros of this polynomial in t are real. Hurwitz' theorem on the continuity of zeros [Con78, Corollary 2.6] says that a limit, uniformly on bounded intervals, of polynomials having all real zeros will either have all real zeros or vanish identically. The limit $\tilde{p}(\mathbf{y} + t\mathbf{u})$ has degree $k \geq 1$; it does not vanish identically and therefore it has all real zeros. This shows \tilde{p} to be hyperbolic in direction \mathbf{u} . \square

Definition 3.2 (family of cones in the homogeneous case). *Let p be a hyperbolic homogeneous polynomial and let B be a cone of hyperbolicity for p . If $p(\mathbf{x}) = 0$, define*

$$\mathbf{K}^{p,B}(\mathbf{x})$$

to be the cone of hyperbolicity of $\text{hom}(p, \mathbf{x})$ containing B , whose existence we have just proved. If $p(\mathbf{x}) \neq 0$ we define $\mathbf{K}^{p,B}(\mathbf{x})$ to be all of \mathbb{R}^d .

Example 3.3 (strongly hyperbolic functions). *As mentioned in passing in Section 2, a homogeneous polynomial p is said to be strongly hyperbolic if the roots of $p(\mathbf{y} + t\mathbf{x})$, in addition to being real, are distinct. Equivalently, the projective variety $\mathcal{V} := \{\mathbf{x} : q(\mathbf{x}) = 0\}$ is smooth. In this case, for each nonzero \mathbf{x} , the localization $\text{hom}(p, \mathbf{x})$ has degree one. The cones of hyperbolicity $K^{p,B}(\mathbf{x})$ are halfspaces whose common tangent hyperplane is tangent to B at \mathbf{x} ; the convexity of B implies that the tangent hyperplane is a support hyperplane to B at \mathbf{x} , and we see that indeed $K^{p,B}(\mathbf{x})$ contains B .*

Suppose $\mathbf{x}_n \rightarrow \mathbf{x}$. It is not in general true that $\text{hom}(f, \mathbf{x}_n) \rightarrow \text{hom}(f, \mathbf{x})$. However, it is true that

$$\text{hom}(f, \mathbf{x}_n) = \text{hom}(\text{hom}(f, \mathbf{x}), \mathbf{x}_n)$$

for n sufficiently large. This implies that if p is homogeneous and hyperbolic with cone of hyperbolicity B , then $K^{p,B}(\mathbf{x})$ is semi-continuous in \mathbf{x} :

$$K^{p,B}(\mathbf{x}) \subseteq \liminf K^{p,B}(\mathbf{x}_n) \text{ as } \mathbf{x}_n \rightarrow \mathbf{x}. \quad (3.1)$$

This is proved in [ABG70, Lemma 5.9]. We will want a version of this valid for polynomials that are not necessarily homogeneous. If q is hyperbolic but not homogeneous, the cone $K^{q,B}(\mathbf{x})$ has not been defined. To do this, at least for some points \mathbf{x} , we need the notion of the amoeba of q .

3.2 Amoeba boundaries

Definition 3.4 (amoeba). *The **amoeba** of any polynomial q is defined to be the image of the zero set $\mathcal{V} := \{\mathbf{z} : q = 0\}$ of q under the coordinatewise log modulus map $\text{ReLog}(\mathbf{z}) := (\log |z_1|, \dots, \log |z_d|)$. We denote this image by $\text{amoeba}(q)$.*

The connected components of $\mathbb{R}^d \setminus \text{amoeba}(q)$ are convex sets and are in one to one correspondence with the Laurent series expansions of F , with each expansion converging on precisely the set $\{\exp(\mathbf{x} + i\mathbf{y}) : \mathbf{y} \in B\}$ for some component B of $\mathbb{R}^d \setminus \text{amoeba}(q)$. This is well known and is presented, for instance, in Chapter 6 of [GKZ94]. For those who have not seen an amoeba before, one of the simplest nontrivial amoebas is shown in Figure 6.

The following result, proved in [BP11, Proposition 2.12], defines a family of cones via localizations on the boundary of an amoeba.

Proposition 3.5. *Let q be any polynomial and let B be a component of the complement of $\text{amoeba}(q)$. Denote the boundary of B by ∂B . Fix $\mathbf{x} \in \partial B$ and let $f = q \circ \exp$ so that f vanishes at some point $\mathbf{x} + i\mathbf{y}$. Let $f_{\mathbf{y}} := \text{hom}(f, \mathbf{x} + i\mathbf{y})$. Then each $f_{\mathbf{y}}$ is hyperbolic (meaning hyperbolic in at least one*

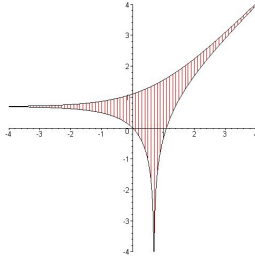


Figure 6: The amoeba of the polynomial $2 - x - y$

direction) and one of its cones of hyperbolicity contains B . We denote this cone by $K^{q,B}(\mathbf{y})$. (The point \mathbf{x} is considered fixed and is suppressed from the notation.) \square

One may extend semi-continuity for localizations beyond the homogeneous case, to the families in Proposition 3.5.

Theorem 3.6 ([BP11, Corollary 2.15]). *If B is a component of $\text{amoeba}(q)^c$ and $\mathbf{x} \in \partial B$, then the family of cones $K^{q,B}(\mathbf{y})$ defined in Proposition 3.5 satisfies*

$$K^{q,B}(\mathbf{y}) \subseteq \liminf K^{q,B}(\mathbf{y}_n) \text{ whenever } \mathbf{y}_n \rightarrow \mathbf{y}. \quad (3.2)$$

\square

The fact that localizations of *any* polynomial to points on the amoeba boundary are hyperbolic allows us to give numerous examples for hyperbolic polynomials beyond the classic ones: planes, quadrics, and the determinant function. For example, the homogeneous polynomial variety shown in Figure 7 is a localization of the famous so-called fortress generating function denominator.



Figure 7: Another hyperbolic homogeneous polynomial

It is instructive to see what role hyperbolicity plays in Theorem 3.6 by considering the counterexample $f(x, y, z) = xy + z^3$ as in Figure 8. As \mathbf{y} varies over a neighborhood U of the origin, is it possible to choose open convex cones $\{K(\mathbf{y}) : \mathbf{y} \in U\}$ over \mathbf{y} in such a way that $K(\mathbf{y})$ varies semi-continuously with \mathbf{y} and each $K(\mathbf{y})$ is, locally, a subset of $\{\mathbf{y} : f(\mathbf{y}) \neq 0\}$? The points $\mathbf{y} = (0, y, 0)$

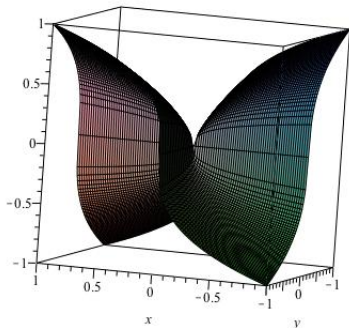


Figure 8: The sheared cubic $xy + z^3$

are forced to choose whether $K(\mathbf{y})$ contains points with positive x components or negative x components. One of these violates semi-continuity with cones $K(0, y, \epsilon)$ as $\epsilon \downarrow 0$ while the other violates semi-continuity with cones $K(0, y, -\epsilon)$ as $-\epsilon \uparrow 0$. The key to avoiding this in Theorem 3.6 is that $f_{\mathbf{y}}$ must be hyperbolic, and this cannot happen with hyperbolic functions.

3.3 Morse deformations

Suppose that $K(\mathbf{y})$ is any family of convex cones in \mathbb{R}^d varying semi-continuously with \mathbf{y} in the sense of (3.2). Let \mathbf{r} be any vector such that each $K(\mathbf{y})$ contains a vector \mathbf{v} with $\mathbf{r} \cdot \mathbf{v} > 0$.

Proposition 3.7. *Over any compact set S a section $\{\mathbf{v}(\mathbf{y}) : \mathbf{y} \in S\}$ may be chosen continuously with $\mathbf{v}(\mathbf{y}) \cdot \mathbf{r} > 0$ for all $\mathbf{y} \in S$.*

PROOF: Given \mathbf{y} , we may choose \mathbf{v} in the interior of $K(\mathbf{y})$ with $\mathbf{v} \cdot \mathbf{r} > 0$, whence by semi-continuity, $\mathbf{v} \in K(\mathbf{y}')$ for all \mathbf{y}' in some neighborhood of \mathbf{y} which we denote $\mathcal{N}(\mathbf{y}, \mathbf{v})$. We may cover the compact set S with finitely many of these neighborhoods, $\{\mathcal{N}(\mathbf{y}^{(k)}, \mathbf{v}^{(k)}) : 1 \leq k \leq m\}$. Let $\{\psi^{(k)} : 1 \leq k \leq m\}$ be a partition of unity subordinate to this cover and for any $\mathbf{y} \in S$ define

$$\mathbf{v}(\mathbf{y}) := \sum_{k=1}^m \psi^{(k)}(\mathbf{y}) \mathbf{v}^{(k)}.$$

It is clear that $\mathbf{r} \cdot \mathbf{v}(\mathbf{y}) > 0$ for all $\mathbf{y} \in S$. For each $\mathbf{y} \in S$, if $\psi^{(k)}(\mathbf{y}) > 0$ then $\mathbf{v}^{(k)} \in K(\mathbf{y})$. By convexity of $K(\mathbf{y})$, it follows that $\mathbf{v}(\mathbf{y}) \in K(\mathbf{y})$. \square

Two applications of this are as follows. Suppose p is a homogeneous hyperbolic function and let B be a cone of hyperbolicity for p . Due to homogeneity, $K^{p,B}(\lambda\mathbf{y}) = K^{p,B}(\mathbf{y})$, whence the family may be described as $\{K^{p,B}(\hat{\mathbf{y}}) : \hat{\mathbf{y}} \in S^{d-1}\}$. Suppose \mathbf{r} is such that each $K(\mathbf{y})$ contains a \mathbf{v} with $\mathbf{r} \cdot \mathbf{v} > 0$. Because S^{d-1} is compact, an application of Proposition 3.7 yields a continuous section $\{\mathbf{v}(\mathbf{y}) : \mathbf{y} \in S^{d-1}\}$ with $\mathbf{v}(\mathbf{y}) \in K^{p,B}(\mathbf{y})$ and $\mathbf{r} \cdot \mathbf{v}(\mathbf{y}) > 0$. Extending to all of $\mathbb{R}^d \setminus \{\mathbf{0}\}$ by $\mathbf{v}(\lambda\mathbf{y}) := \lambda\mathbf{v}(\mathbf{y})$, we arrive at:

Corollary 3.8. *There is a vector field $\mathbf{v}(\cdot)$ on \mathbb{R}^d which is 1-homogeneous, vanishes only at $\mathbf{0}$, is a section of $K^{p,B}(\cdot)$, and for some $\epsilon > 0$ satisfies $\mathbf{r} \cdot \mathbf{v}(\mathbf{y}) \geq \epsilon|\mathbf{y}|$. \square*

Example 3.9. *If the homogeneous polynomial q is strongly hyperbolic then its zero set is the cone over a smooth projective hypersurface and each $K(\mathbf{y})$ is a halfspace (see Example 3.3). The condition that every $K(\mathbf{y})$ contain a vector \mathbf{v} with $\mathbf{r} \cdot \mathbf{v} > 0$ is the same as requiring that \mathbf{r} not be the outward normal to the bounding hyperplane of $K(\mathbf{y})$. In other words, \mathbf{r} may not be on the boundary of the dual cone to B .*

The second application is when B is a component of $\text{amoeba}(q)^c$. In this case the family $\{f_{\mathbf{y}}\}$ is periodic with period 2π in each coordinate. Accordingly, the cones $K(\mathbf{y}) = K^{q,B}(\mathbf{y})$ from Proposition 3.5 are indexed by a compact set, namely the torus $\mathbb{R}^d/(2\pi\mathbb{Z})^d$. Applying Proposition 3.7 then yields:

Corollary 3.10. *If each cone $K^{q,B}(\mathbf{y})$ in Proposition 3.5 contains a vector \mathbf{v} with $\mathbf{r} \cdot \mathbf{v} > 0$ then there is a continuously varying section $\{\mathbf{v}(\mathbf{y}) : \mathbf{y} \in \mathbb{R}^d\}$ with $\mathbf{v}(\mathbf{y}) \in K^{q,B}(\mathbf{y})$ and $\mathbf{r} \cdot \mathbf{v}(\mathbf{y}) > 0$ for all \mathbf{y} . \square*

With these vector fields in hand we have enough to carry out the programs of [ABG70] and [BP11]. I will briefly describe the former, then go into a little more detail on the latter. Let q be a homogeneous polynomial, strongly hyperbolic in the direction \mathbf{x} . We wish to compute its inverse Fourier transform, which we have called the Riesz kernel, Q_α . We recall that its value at \mathbf{y} is given by

$$Q_\alpha(\mathbf{y}) := (2\pi)^{-d} \int_{\mathbf{z} \in \mathbf{x} + i\mathbb{R}^d} q(\mathbf{z})^{-\alpha} \exp(\mathbf{y} \cdot \mathbf{z}) d\mathbf{z}. \quad (2.2)$$

We set $\mathbf{r} = -\mathbf{y}$ and assume \mathbf{r} not to be on the boundary of B^* , the dual cone of the cone B of hyperbolicity of q that contains \mathbf{x} . As we have seen in Example 3.9, this guarantees the existence of the 1-homogeneous vector field $\mathbf{v}(\cdot)$. The sets $K(\mathbf{y})$ are convex and each contains both \mathbf{x} and $\mathbf{v}(\mathbf{y})$. Therefore, each contains the line segment joining \mathbf{x} to $\mathbf{v}(\mathbf{y})$. Define a homotopy by

$$H(\mathbf{u}, t) := i\mathbf{u} + (1-t)\mathbf{x} + t\mathbf{v}(\mathbf{u}).$$

This deforms the domain of integration $\mathbf{x} + i\mathbb{R}^d$ in (2.2) to the cone $\mathcal{C} := \{i\mathbf{u} + \mathbf{v}(\mathbf{u}) : \mathbf{u} \in \mathbb{R}^d\}$ while avoiding the set \mathcal{V} where q vanishes. Cauchy's theorem implies that deforming the contour does not change the integral and therefore that

$$Q_\alpha(\mathbf{y}) = \int_{\mathcal{C}} q(\mathbf{z})^{-\alpha} \exp(\mathbf{y} \cdot \mathbf{z}) d\mathbf{z}.$$

Omitting pages of detail and skipping to the punch line, this representation allows us to factor the integral. Integrating radially reduces the integral to an integral over the **leray cycle**, which is a $(d-1)$ -dimensional homology class in the complement of $\{q = 0\}$ in $\mathbb{C}\mathbb{P}^{d-1}$. The dimension reduction allows a number of explicit computations which are carried out in [ABG70, Section 7].

3.4 Asymptotics of Taylor coefficients

Let $F = P/Q = \sum_{\mathbf{r}} a_{\mathbf{r}} \mathbf{Z}^{\mathbf{r}}$ be a Laurent series expansion converging on the component B of $\text{amoeba}(Q)^c$. The coefficients $\{a_{\mathbf{r}}\}$ of F may be recovered via Cauchy's formula

$$a_{\mathbf{r}} = (2\pi i)^{-d} \int_T \mathbf{Z}^{-\mathbf{r}} F(\mathbf{Z}) \frac{d\mathbf{Z}}{\mathbf{Z}} \quad (3.3)$$

where T is the torus $T(\mathbf{x}) = \{\exp(\mathbf{x} + i\mathbf{y}) : \mathbf{y} \in \mathbb{R}^d / (2\pi\mathbb{Z})^d\}$ for some $\mathbf{x} \in B$ and $d\mathbf{Z}/\mathbf{Z} := dz_1 \wedge \cdots \wedge dz_d / (z_1 \cdots z_d)$. The motivation for the evaluation or asymptotic estimation of the coefficients $\{a_{\mathbf{r}}\}$ comes from analytic combinatorics, where the primary object of study is the array $\{a_{\mathbf{r}}\}$, which counts something of interest. One constructs the generating function $F(\mathbf{Z}) := \sum_{\mathbf{r}} a_{\mathbf{r}} \mathbf{Z}^{\mathbf{r}}$ and hopes to identify a closed form representation of F . If $\{a_{\mathbf{r}}\}$ satisfy a recursion, then (depending on boundary values) this will usually succeed. A number of examples of combinatorial interest are surveyed in [PW08]. In most of the cases surveyed there, the generating function F is rational. In [RW08], based on a result of [Saf00], it is shown how to embed any algebraic function of d variables as a diagonal of a rational function of $(d+1)$ variables. Thus the evaluation of Taylor coefficients of rational functions solves the enumeration problem for any array of numbers whose generating function is rational or algebraic.

Exact evaluation of the Cauchy integral (3.3) is not easy but some methods are known for evaluating it asymptotically. In the case where the pole variety $\mathcal{V} := \{\mathbf{z} : Q(\mathbf{z}) = 0\}$ is smooth, a formula was given in [PW02]; a coordinate free version is given in [BBBP08]. Normal self-intersection in \mathcal{V} can also be dealt with [PW04]. A number of examples from combinatorics and statistical physics have generating functions whose pole variety has a singularity with nontrivial monodromy. In the remainder of this section I will explain how hyperbolicity and the resulting deformations allow explicit asymptotic evaluation of some of these generating functions.

So as to keep the conversation more concrete, we consider as a running example the generating function for the probability of a Northgoing diamond in a uniform random tiling of the Aztec Diamond. This example is taken from [BP11, Section 4]. We have

$$F(X, Y, Z) = \frac{Z/2}{(1 - (X + X^{-1} + Y + Y^{-1})Z + Z^2)(1 - YZ)} = \sum_{r,s,t} a_{r,s,t} X^r Y^s Z^t$$

where a_{rst} is the probability that the domino covering the square (r, s) in the order t Aztec Diamond is oriented in a Northgoing direction. Figure 3.4 illustrates the Aztec Diamond shape of order 4 and the macroscopic features of a random tiling by dominoes of a larger Aztec Diamond (order 47).

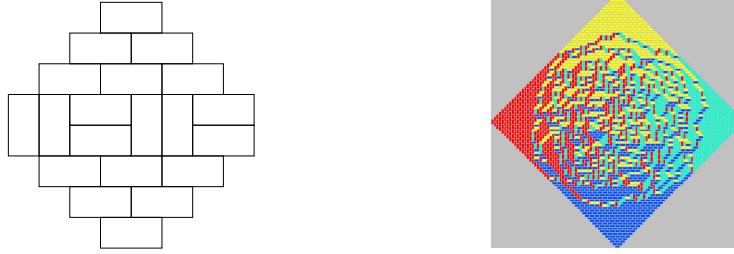


Figure 9: An Aztec diamond and a random tiling of another Aztec diamond

Observe that r and s can be both positive and negative, while t is always positive and at least $|r| + |s|$. The denominator $Q(\mathbf{Z})$ is a Laurent polynomial whose zero set in $(\mathbb{C}^*)^d$ has two isolated singularities at $\pm(1, 1, 1)$. Near each of these, the zero sets of the two factors in the denominator look like a cone and a plane respectively: letting $Q = JH$, where J is the quadratic and $H = 1 - YZ$ is the log-linear term, we have

$$\text{hom}(Q, (1, 1, 1)) = (Z^2 - \frac{1}{2}(U^2 + V^2))(Y - Z).$$

The zero set of this homogeneous polynomial is shown in Figure 10.

Let T' denote the flat torus $\mathbf{x} + \mathbb{R}^d / (2\pi\mathbb{Z})^d$ where \mathbf{x} is any element of B , the component of $\text{amoeba}(Q)^c$ over which the series converges. Changing variables via $\mathbf{Z} = \exp(\mathbf{z}) = (e^{z_1}, e^{z_2}, e^{z_3})$ and $d\mathbf{z} = d\mathbf{Z}/\mathbf{Z}$, then writing $\mathbf{z} = \mathbf{x} + i\mathbf{y}$ and $f(\cdot) = F \circ \exp(\mathbf{x} + i\cdot)$, yields

$$a_{\mathbf{r}} = (2\pi)^{-d} e^{-\mathbf{r} \cdot \mathbf{x}} \int_{T'} \exp(-i\mathbf{r} \cdot \mathbf{y}) f(\mathbf{y}) d\mathbf{y}. \quad (3.4)$$

Up to this point the expression for $a_{\mathbf{r}}$ is exact. One can show that approximating f by $\text{hom}(f)$ and T' by \mathbb{R}^d does not change the leading asymptotic term. This is a somewhat lengthy verification and it is here that the conical deformation of Corollary 3.10 is required. The upshot is that $a_{\mathbf{r}} = E(\mathbf{r})$ where E is the inverse Fourier transform of $\text{hom}(f)$. For quadratics, the inverse Fourier transform is the dual quadratic. For a product of a quadratic and a log-linear function, an explicit computation yields [BP11, Theorem 3.9]

$$C \arctan \left(\frac{\sqrt{A^*(\mathbf{r}, \mathbf{r})} \sqrt{-A^*(\ell, \ell)}}{A^*(\mathbf{r}, \ell)} \right)$$

where A^* is the quadratic form dual to $\text{hom}(f)$ and ℓ is the the log-linear factor viewed as an element of $(\mathbb{R}^d)^*$ in the logarithmic space. For example, when F is the Aztec generating function one obtains the following result, a pictorial version of which is given in Figure 11.

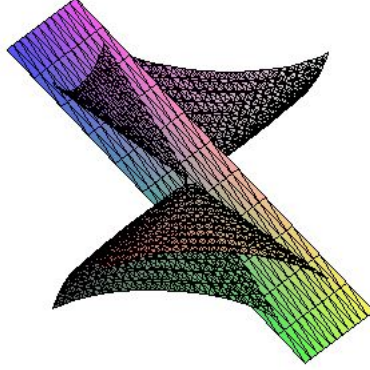


Figure 10: The homogenized zero variety near $(1, 1, 1)$

Theorem 3.11 ([BP11, Theorem 3.7]). *The Northgoing placement probabilities $\{a_{rst}\}$ for the Aztec Diamond are asymptotically given by*

$$a_{rst} \sim \frac{1}{\pi} \arctan \left(\frac{\sqrt{t^2 - 2r^2 - 2s^2}}{t - 2s} \right)$$

when (r, s, t) is in the cone $\{(r, s, t) : t^2 > 2r^2 + 2s^2\}$ and $r + s + t$ is odd.

Remark. The way the generating function is indexed, $a_{rst} = 0$ when $r + s + t$ is even. Asymptotics in the corners of the diamond outside the inscribed circle are given by $a_{rst} \rightarrow 1$ in one corner and $a_{rst} \rightarrow 0$ in the other three, where convergence occurs at an exponential rate as $t \rightarrow \infty$.

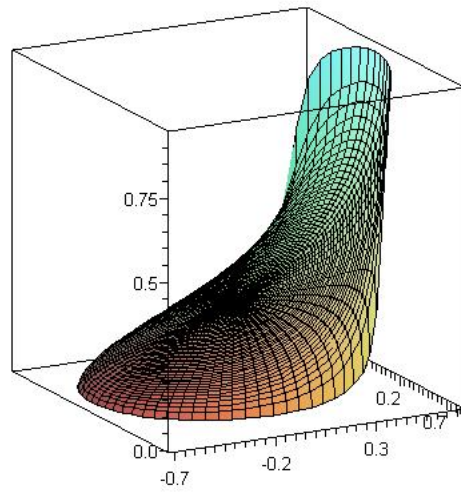


Figure 11: Scaling limit of $a_{r,s,t}$

Part II: Stability

We now turn to the notion of stability. As we will see in Sections 4.3 and 5, the most useful properties of stability are closure properties: stability is preserved by a wealth of operations that are natural from an algebraic or probabilistic point of view. Stability is defined for general multivariate complex polynomials. Some of these properties or their proofs simplify when restricted to certain subclasses, such as polynomials whose coefficients are real or positive, or **multi-affine** polynomials, whose degrees in each variable never exceed 1. Even in these cases, however, some properties are proved only by going through the more general setting of complex coefficients.

4 Stability theory in one variable

As usual, the univariate theory is older and simpler. As with hyperbolicity, the origins of stability theory are in differential equations and control theory. After discussing this, we turn to combinatorial uses of the univariate theory. Stable generating polynomials produce coefficient sequences satisfying Newton’s inequalities, implying, among other things, log-concavity. We end the section on univariate stability with a discussion of stability preserving operations and the so-called Laguerre–Pólya class.

4.1 Stability over general regions

In this section we will begin by thinking more generally of polynomials that avoid roots in some region Ω . We will call these Ω -stable. Throughout the remainder of this paper we will let

$$\mathcal{H} := \{z : \Im\{z\} > 0\}.$$

denote the open upper half-plane. I will always use the term “stable” to mean \mathcal{H} -stable, as in Definition 1.2, and will use “Hurwitz stable” or “ Ω -stable” for other notions of stability. Recalling G.-C. Rota’s philosophical observation, one might keep in mind that the Riemann Hypothesis is equivalent to Ω -stability where $\Omega = \{z : \Re\{z\} \neq 1/2\}$. This may seem like a strained connection, but in fact some of the literature we review in Section 7 is explicitly motivated by the desire to understand the zeta function.

Statistical physicists have a different motive for understanding regions free of zeros. To explain this, we examine some graph theoretic models, paraphrasing the exposition in [Sok01]. Let $G = (V, E)$ be a finite graph and let q be a positive integer. Define a polynomial in variables $\{x_e : e \in E\}$

$$Z(q, \{x_e : e \in E\}) := \sum_{\sigma} \prod_{e=\overline{vw} \in E} (1 + x_e \delta_{\sigma(v), \sigma(w)})$$

where σ ranges over q -colorings of the vertices, that is, all maps from V to $\{1, \dots, q\}$. This is elementarily seen to be equivalent to the alternative definition

$$Z(q, \{x_e : e \in E\}) := \sum_{E' \subseteq E} q^{k(E')} \prod_{e \in E'} x_e$$

where $k(E')$ denotes the number of connected components in the subgraph (V, E') . The last formulation makes it obvious that Z is a polynomial in q as well as in $\{x_e\}$. When $x_e = -1$ for all e this specializes to the chromatic polynomial $\chi(q)$; more generally, taking $x_e = x$ for all e gives the bivariate **Tutte polynomial**.

The polynomial Z is also the **partition function** for the q -state Potts model. Statistical physicists are interested in **phase transitions** where the behavior of the system depends non-analytically on its parameters. Being a polynomial, Z is of course analytic in all its parameters. Typically though, one is interested in infinite-volume limits such as the free energy

$$f(q, x) := \lim_{n \rightarrow \infty} n^{-1} \log Z_n(q, x)$$

where Z_n is taken on a graph G_n in a family of graphs whose increasing limit is \mathbb{Z}^d . The function f may fail to be analytic at (q, x) if the zeros of $Z_n(q, x)$ have a limit point at (q, x) . Therefore, there is physical significance in keeping the zeros of the polynomial Z_n out of specified regions. Such results are often known as **Lee-Yang** theorems, after [YL52, LY52] (see also [LS81]).

One such theorem was proved by D. Wagner in 2000. The **reliability polynomial** R_G of the graph $G = (V, E)$ is the probability that the graph is connected when each edge is kept or deleted independently with probability x , which is a polynomial in the parameter x . Brown and Colbourn conjectured [BC92] that all zeros of $R_G(x)$ are in the closed disk $\{z : |z - 1| \leq 1\}$. In other words, R_G is \mathcal{D}^c -stable where $\mathcal{D} := \{z : |z - 1| \leq 1\}$. This was proved for the class of series-parallel graphs in [Wag00, Theorem 0.2]. A simpler proof was found by Sokal [Sok01, Section 4.1, Remark 3], who actually proved the stronger multivariate stability result. On the other hand, Royle and Sokal [RS01] showed that the original univariate conjecture is false for general graphs; they also showed that multivariate \mathcal{D}^c -stability of R_G holds *if and only if* G is series-parallel. Sokal [Sok11] has recently conjectured that R_G is (univariate) \mathcal{D}^c -stable for the complete graphs, K_n .

Hurwitz stability

The association of zero-free regions with the term “stability” originated in ODE’s and control theory. Its use is attributed to Hurwitz. In fact, Ω -stability with $\Omega = \{z : \Re\{z\} \geq 0\}$ is sometimes called

“Hurwitz stability”. In ODE’s, it is easy to see the physical significance of Hurwitz stability. Let M be a matrix and consider the linear system $\mathbf{y}' = M\mathbf{y}$. Hurwitz stability is equivalent to all eigenvalues of M having negative real parts, which is equivalent to all homogeneous solutions decaying, and hence good control over the system $\mathbf{y}' - M\mathbf{y} = \mathbf{v}(t)$. One may also consider discrete-time analogues, such as the system $\mathbf{y}^{(n+1)} - Q\mathbf{y}^{(n)} = \mathbf{v}^{(n)}$ where Q would correspond to the exponential of M in the previous system. Now decay of homogeneous solutions is equivalent to Ω -stability when Ω is the complement of the open unit disk; polynomials whose zeros are all in the open unit disk are said to be **Schur-stable**.

For a less trivial example we turn to control theory. Following [Hen91, Section 10.3], we consider systems which turn an input signal $f(t)$ into an output $u(t)$. Many systems, and in particular those built from networks of impedances, are not only linear but also time homogeneous, that is, the map $f \mapsto u$ commutes with time translation. In terms of Laplace transforms, this means that the system acts multiplicatively, meaning that if \mathcal{L} denotes the Laplace transform, then such systems obey the law

$$\mathcal{L}u = g \cdot \mathcal{L}f \tag{4.1}$$

for some function $g : [0, \infty) \rightarrow \mathbb{R}$ that is called the **transfer function**. For example an L-C-R circuit, with an input voltage function $f(t)$ between the inductor and the resistor producing an output $u(t)$ of current from the inductor to the capacitor, satisfies (4.1) with $g(s) = (R + Ls + C^{-1}/s)^{-1}$. In impedance networks the transfer function is in fact always rational, $g(s) = p(s)/q(s)$ for some polynomials p and q . Generally, a system is said to be stable if bounded input produces bounded output.

To check whether the system with transfer function g is stable, it suffices to test it on inputs of the form $f(t) = e^{i\omega t}$, the Laplace transform of which is $(\mathcal{L}f)(s) = 1/(s - i\omega)$. Let the poles of the rational function g be a_1, \dots, a_r . The $\mathcal{L}u = g \cdot \mathcal{L}f$ has poles $a_1, \dots, a_r, i\omega$. We may write $\mathcal{L}u$ in a partial fraction expansion resulting in $\sum_{j=0}^r p_j/(s - a_j)$ for some polynomials $\{p_j\}$; here we have set $a_0 := i\omega$. Inverting the Laplace transform gives a sum $\sum_{j=1}^r P_j(t)e^{a_j t}$ for some collection $\{P_j\}$ of polynomials. This is bounded if and only if each a_j is either in the open left half-plane or is imaginary and a simple zero. But when g has an imaginary zero a_j , then setting $a_0 = a_j$ (that is, taking the input to be $e^{ia_j t}$) produces a doubled root and an unbounded output. Hence, bounded inputs produce bounded outputs if and only if all poles of g lie strictly in the left half-plane.

We conclude that stable behavior of the system corresponds to Hurwitz stability of the denominator of the transfer function g . For example, in the L-C-R circuit above, the denominator of g is a quadratic with positive coefficients. The real parts of the roots are always negative, whence such a system always behaves stably.

Differential equations of order d may be transformed into first order systems in d variables by the well known trick of representing the first $d - 1$ derivatives as new variables. It is not surprising, therefore, that Hurwitz stability also arises in the stability analysis of order- d linear differential

equation with constant coefficients. Given d initial conditions and an inhomogeneous term, we may write such an equation as

$$f^{(d)} + a_{d-1}f^{(d-1)} + \cdots + a_0f = H \quad (4.2)$$

with initial conditions $f^{(j)}(0) = b_j$ for $0 \leq j \leq d-1$. Assuming H to grow at most exponentially, a Laplace transform exists near the origin. We may take the Laplace transform of both sides of (4.2). Using linearity and the rule $\mathcal{L}f'(s) = s\mathcal{L}f(s) - f(0^+)$ and denoting $g := \mathcal{L}f$ gives inductively

$$\mathcal{L}(f^{(d)})(s) = s^d g(s) - s^{d-1}f(0^+) - s^{d-2}f'(0^+) - \cdots - f^{(d-1)}(0^+).$$

Plugging this and the boundary conditions into (4.2) yields the following equation for g :

$$\begin{aligned} & s^d g(s) - s^{d-1}b_0 - s^{d-2}b_1 - \cdots - b_{d-1} \\ & + a_{d-1} [s^{d-1}g(s) - s^{d-2}b_0 - \cdots - b_{d-2}] \\ & \quad \vdots \\ & + a_1 [s^1 g(s) - b_0] \\ & + a_0 g(s) = \mathcal{L}H(s). \end{aligned}$$

Letting $p(s) := s^d + a_{d-1}s^{d-1} + \cdots + a_0$ denote the characteristic polynomial of the equation (4.2), and $p_k(s)$ denote the shifted polynomial $s^k + a_{d-1}s^{k-1} + \cdots + a_{d-k}$, we may rewrite the equation for g as

$$p \cdot g = \mathcal{L}H + \sum_{j=0}^{d-1} b_j p_{d-1-j}$$

and hence

$$g(s) = \frac{\mathcal{L}H(s) + \sum_{j=0}^{d-1} b_j p_{d-1-j}(s)}{p(s)}. \quad (4.3)$$

Again, poles of g with positive real part produce unbounded output, as do purely imaginary poles $i\omega$ when the driving term is taken to be $e^{i\omega t}$, whose Laplace transform $\mathcal{L}H$ has a pole at $i\omega$. Thus, again, Hurwitz stability is equivalent to bounded output on bounded input.

4.2 Real roots and Newton's inequalities

I will now return, permanently, to upper half-plane stability, which will be called, simply, "stability". In one variable, when the coefficients of f are real, the zeros come in conjugate pairs. Stability, therefore, is equivalent to having only real zeros. When the coefficients are nonnegative, a strictly positive zero is impossible, whence stability is further equivalent to having all zeros on the negative half line.

Combinatorialists have long sought to exploit the properties of stable generating functions. The universal problem in combinatorial enumeration is to count a family of structures indexed by one or

more positive integer parameters. In the case of one parameter, say n , the sequence $\{a_n\}$ of counts and the corresponding generating function $f(z) := \sum_n a_n z^n$ are of fundamental interest. In the setting of probability generating functions, the coefficients a_n are nonnegative and sum to one. In this case, if f has only real roots, then the distribution which gives probability a_n to the value n is representable as the sum of independent random variables each taking the value 0 or 1 (Bernoulli random variables). There may or may not be a natural interpretation for these Bernoulli variables; see [HKPV09, Example 23] for an example in which there can be no natural interpretation.

Often it is intuitively plausible that the sequence $\{a_n\}$ is **unimodal**, meaning that for some k , we have $a_0 \leq \dots \leq a_k \geq a_{k+1} \geq a_{k+2} \geq \dots$. The enumeration literature is littered with examples in which unimodality is conjectured (see for example the survey [Sta89] and the follow-up to this [Bre94]), but a proof is often elusive for the reason that there is no obvious theoretical framework within which to prove unimodality. There are, however some stronger properties for which natural avenues of proof exist.

Definition 4.1 (log-concavity). *A finite or infinite sequence $\{a_k\}$ of nonnegative numbers is said to have no internal zeros if the indices of the nonzero terms form an interval $[r, s]$. The sequence is said to be **log-concave** if it has no internal zeros and if*

$$a_{k-1}a_{k+1} \leq a_k^2$$

for $k \in [r+1, s-1]$. The sequence $\{a_r, \dots, a_{r+k}\}$ with no internal zeros is said to be **ultra log-concave** if

$$\left(\frac{a_j}{\binom{k}{j}}\right)^2 \geq \frac{a_{j+1}}{\binom{k}{j+1}} \frac{a_{j-1}}{\binom{k}{j-1}}.$$

It is immediate that ultra log-concavity implies log-concavity which implies unimodality. While ultra log-concavity appears to be the least natural of these properties, it was shown three centuries ago by Newton to follow from stability of the generating function.

Theorem 4.2 (Newton's inequalities). *Suppose $f(z) := \sum_{k=0}^n a_k z^k$ is a real stable polynomial. Then for $1 \leq k \leq n-1$,*

$$\left(\frac{a_k}{\binom{n}{k}}\right)^2 \geq \frac{a_{k+1}}{\binom{n}{k+1}} \frac{a_{k-1}}{\binom{n}{k-1}}. \quad (4.4)$$

If furthermore the coefficients a_k are all nonnegative then the roots of f are all nonpositive and the sequence is ultra log-concave.

PROOF: By Rolle's Theorem, if a univariate polynomial f has only real zeros then so does its derivative. This observation will be useful many times below. We use it now to deduce that

$$Q(z) := \left(\frac{d}{dz}\right)^{k-1} f(z)$$

has only real zeros. Reversing a sequence of coefficients via $R(z) := z^{n-k+1}Q(1/z)$ also preserves the property of having all real roots. By Rolle's Theorem again, $S(z) := (d/dz)^{n-k+1}R(z)$ has all real roots. But $S(z)$ is the trinomial

$$\frac{n!}{2} \left(\frac{a_{k-1}}{\binom{n}{k-1}} z^2 + 2 \frac{a_k}{\binom{n}{k}} z + \frac{a_{k+1}}{\binom{n}{k+1}} \right),$$

and the theorem follows from the discriminant test for quadratics. \square

A curious result in a similar vein was proved by Gurvits [Gur08, Lemma 3.2]; the proof, which is a few lines of calculus, is omitted.

Proposition 4.3. *Let f be a probability generating polynomial of degree d and let $C := \inf f(t)/t$. If f is stable then*

$$a_1 = f'(0) \geq \left(\frac{d-1}{d} \right)^{d-1} C.$$

Equality holds if and only if $f = (1 - q + qt)^d$ generates a binomial distribution. The value $[(d-1)/d]^{d-1}$ increases to e^{-1} as $d \rightarrow \infty$; for infinite series, the bound $a_1 \geq e^{-1}C$ holds with equality if and only if $f(t) = e^{\lambda(t-1)}$ is the generating function for a Poisson distribution. \square

There are a number of techniques that can be used to establish stability (also known, in the univariate case, as the real root property). In some cases one has a sequence $\{f_n\}$ where f_n has degree n and the roots of f_{n-1} interlace the roots of f_n , meaning that each interval between roots of f_n contains a root of f_{n-1} . If this is true, then often it is possible to prove it by induction. The following example from [Sta89] illustrates this.

Example 4.4 (Hermite polynomials). *The Hermite polynomials are a sequence of polynomials defined by*

$$H_n(z) := \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{(-1)^k n! (2z)^{n-2k}}{k!(n-2k)!}.$$

They are orthogonal with respect to the Gaussian measure $e^{-x^2} dx$. They satisfy the recursion

$$H_n(z) = -e^{z^2} \frac{d}{dz} \left(e^{-z^2} H_{n-1}(z) \right).$$

Assume for induction that $H_{n-1}(z)$ has $n-1$ real zeros. It is clear from this that H_n has $n-2$ zeros interlacing the zeros of H_{n-1} . To see that H_n also has a zero less than all the zeros of H_{n-1} , observe that $e^{-z^2} H_{n-1}$ tends to zero as $z \rightarrow -\infty$, therefore $e^{-z^2} H_{n-1}$ has an extreme value to the left of its leftmost zero, which is a zero of H_n . Similarly, H_n has a zero greater than the greatest zero of H_{n-1} , and the induction is established.

There is a more methodical way that often works to prove interlacing by induction. If f and g are real polynomials of respective degrees n and $n-1$ having all real roots and the zeros of g

interlace the zeros of f , it may be seen elementarily that $f + \lambda g$ has n real zeros for any real λ . A simple application of this idea is to the sequence $\{g_n\}$ defined by the three-term recurrence

$$g_{n+1}(x) = axg_n(x) + bg_{n-1}(x),$$

which include the Chebyshev polynomials and Laguerre polynomials (a and b depending on n in the latter case). The same idea is the basis for a theorem independently proved by Heilmann and Lieb [HL72], by Gruber and Kunz [GK71] and, in part, by Nijenhuis [Nij76]. A matching on a finite weighted graph or multi-graph is a subset M of the edge set such that any two edges in M are disjoint (share no vertex). Fix a graph $G = (V, E)$ and let $\{W(e) : e \in E\}$ be a set of nonnegative weights. Let a_j denote the number of matchings with j edges, counted by weight, where the weight of a matching is the product of the weights of the edges in the matching; by convention we take $a_0 = -1$. In the simplest case, $W(e) \equiv 1$ and a_j simply counts matchings of size j , size being the number of edges.

If we are allowed to set $W(e) = 0$ for some edges, then the complete graph is universal. We therefore assume that $G = K_n$ for some n .

Theorem 4.5 (matchings). *The number of weighted matchings of K_n enumerated by size is ultra log-concave.*

PROOF: We show stability of a certain generating function, the most convenient being

$$Q_n(z) := \sum_{j=0}^{\lfloor n/2 \rfloor} (-1)^j a_j z^{n-2j}. \quad (4.5)$$

This is monic, of degree n , and is odd or even depending on n . If we show Q_n is stable and has n distinct roots, then letting $Q_n(z) = R_n(z^2)$ in the even case and $Q_n(z) = zR_n(z^2)$ in the odd case, it follows that $R_n(-z)$ has $\lfloor n/2 \rfloor$ negative real roots, hence is stable; it is also the generating function for matchings by size, hence the proof will be complete.

Taking limits at the end, we may assume without loss of generality that the weights $W(e)$ are all strictly positive. To see that Q_n is stable, let Q_n^S denote the generating function defined by (4.5) on the graph $K_n \setminus S$. We use the recursion

$$Q_n(z) = zQ_n^u(z) - \sum_{v \neq u} W(u, v) Q_n^{u, v}(z) \quad (4.6)$$

which is obvious from the definition. The induction hypothesis is that Q_n^u is stable with distinct roots, which are interlaced by the roots of $Q_n^{u, v}$ for any $v \neq u$.

To verify the induction, we consider the sign of Q_n at the $n - 1$ zeros of Q_n^u in decreasing order. By the interlacing property, the sign of each $Q_n^{u, v}$ alternates, starting out positive because $Q_n^{u, v}(+\infty) = +\infty$. Therefore, $zQ_n^u - \sum_v W(u, v) Q_n^{u, v}$ alternates, starting out negative. This implies

the existence of $n - 2$ roots of Q_n interlaced by the roots of Q_n^u . But also there is one root of Q_n to the right of every root of Q_n^u because Q_n is negative at the rightmost root of Q_n^u and positive at $+\infty$. Similarly there is one root of Q_n to the left of every root of Q_n^u . This completes the inductive proof. \square

A second method by which one can establish univariate stability is via closure properties of the class of univariate stable polynomials. This topic will be expanded in the next section, but for now we mention a result from [Bre89], whose proof we will omit. Let $(x)_i$ denote the falling product $x(x - 1) \cdots (x - i + 1)$ and let $(x)^i$ denote the rising product $x(x + 1) \cdots (x + i - 1)$.

Theorem 4.6 ([Bre89, Theorems 2.4.2–2.4.3]). *Let $f(x) := \sum_{k=0}^n a_k x^k$ be real stable with non-negative coefficients. Then the polynomials $\sum_{k=0}^n a_k (x)_k$ and $\sum_{k=0}^n a_k (x)^k$ are real stable as well.* \square

In the following example, a function from the set $\{1, \dots, n\}$ to itself is represented as a directed graph with an edge from j to $f(j)$ for each j .

Example 4.7 (functions enumerated by components). *Let $b(n, k)$ be the number of functions from the set $\{1, \dots, n\}$ to itself whose directed graph has precisely k components. It is elementary, cf. [GJ83, Example 3.3.28], that*

$$b(n, k) = \sum_{i=1}^n \binom{n-1}{i-1} n^{n-i} c(i, k)$$

where $c(i, k)$ is a signless Stirling number of the first kind. Multiplying by x^k and summing over k gives

$$f_n(x) := \sum_{k=1}^n b(n, k) x^k = \sum_{i=1}^n \binom{n-1}{i-1} n^{n-i} (x)^i.$$

The corresponding sum replacing $(x)^i$ by x^i has the short closed form

$$\sum_{i=1}^n \binom{n-1}{i-1} n^{n-i} x^i = x(x+n)^{n-1}.$$

Evidently this has only real zeros, so applying the conclusion of Theorem 4.6, we see that f_n is stable as well.

A third method is an equivalent characterization for ultra log-concave that is due to Edrei [Edr53]. This relies on the notion of **total positivity**, a theme developed at length by Karlin [Kar68] and for which a number of combinatorial applications are given in [Bre95].

Definition 4.8 (Total positivity; Pólya frequency sequence).

- (i) *An infinite matrix is said to be totally positive (TP) if all of its minors are nonnegative.*

(ii) A sequence $\{a_n\}$ is said to be a Pólya frequency sequence (PF-sequence) if the matrix $A_{nk} := a_{n-k}$ is totally positive. Here we take $a_i := 0$ if $i < 0$ or the sequence $\{a_n\}$ is finite and has length less than i .

A proof of the following equivalence may be found in in [Kar68, Theorem 5.3].

Theorem 4.9 (Edrei's equivalence theorem). *The nonnegative sequence (a_0, \dots, a_d) is a PF-sequence if and only if the generating polynomial $\sum_{k=0}^d a_k z^k$ is stable.* \square

Brenti [Bre94, Section 3] points out that only the 2×2 minors are required for log-concavity, hence unimodality. Nevertheless, nonnegativity of all minors has a combinatorial interpretation which is given in [Bre95, Theorem 3.5]. This interpretation is somewhat abstract, but in special cases the interpretation can be more concrete.

Example 4.10 (r -derangements). *A permutation of $\{1, \dots, n\}$ is said to be an r -derangement if all its cycles have length at least r . Thus a 1-derangement is any permutation, a 2-derangement is a classical derangement, and so forth. Let $b(n, r)$ count the number of r -derangements of $\{1, \dots, n\}$. Brenti [Bre95, Corollary 5.9] shows that $f_n := \sum_{r=1}^n b(n, r) z^r$ is stable. To do so, he shows that a related sequence $\{c_r(n)\}$ is a PF-sequence, hence stable, then applies Theorem 4.6 to recover the result with $b(n, r)$ in place of $c_r(n)$.*

We end this section with one more example of a stable generating polynomial, this one taken from [Sta89]. Recall from Example 2.4 that if A is a real symmetric matrix and B is real positive semi-definite then $f(z) := \text{Det}(A + zB)$ is a real stable polynomial.

Example 4.11 (spanning forests enumerated by component). *Let G be a finite graph with possibly multiple edges and let*

$$a_k(G) := \sum_F \gamma(F)$$

where the sum is over all spanning forests F of G such that F has precisely k components, where $\gamma(F)$ is the product of the cardinalities (number of vertices) of the k components. The factor of γ changes the enumeration from forests to rooted forests. Following [Sta89, Proposition 4], let us see that the polynomial $f_G := \sum_k a_k(G) z^k$ is stable. Let A be the matrix whose rows and columns are indexed by the vertices of G with entry $A_{ij} := \deg(u)$ if $u = v$ and otherwise equal to $-N(i, j)$ where $N(i, j)$ is the number of edges between i and j . It is known that $f_G = \text{Det}(A + zI)$. It follows that f_G is stable. Therefore, the number of rooted forests enumerated by components is ultra log-concave, hence unimodal. Stanley attributes the identity $f_G = \text{Det}(A + zI)$ to Kelmans.

4.3 The Laguerre–Pólya class

Suppose we transform a polynomial $f(z) = \sum_{k=1}^n a_k z^k$ by multiplying each coefficient a_k by a specified constant λ_k . We may ask which sequences $\lambda := \{\lambda_k\}$ are **multiplier sequences**, meaning

that the resulting operator T_λ preserves the class of real stable polynomials. A classical theorem due to Pólya and Schur [PS14] gives a complete and beautiful characterization of multiplier sequences.

Theorem 4.12 (Pólya-Schur 1914). *Let $\lambda = \{\lambda_n : n \geq 0\}$ be a sequence of real numbers and let T_λ denote the linear operator defined by*

$$T_\lambda \left(\sum_{k=0}^n a_k z^k \right) = \sum_{k=0}^n \lambda_k a_k z^k .$$

Denote by Φ the formal power series

$$\Phi(z) := \sum_{k=0}^{\infty} \frac{\lambda_k}{k!} z^k .$$

Then the following are equivalent.

- (i) λ is a multiplier sequence;
- (ii) Φ is an entire function and is the limit, uniformly on compact sets, of the polynomials with all zeros real and of the same sign;
- (iii) Φ is entire and either $\Phi(z)$ or $\Phi(-z)$ has a representation

$$C z^n e^{\alpha_0 z} \prod_{k=1}^{\infty} (1 + \alpha_k z)$$

where n is a nonnegative integer, C is real, and α_k are real, nonnegative and summable.

- (iv) For all nonnegative integers n , the polynomial $T_\lambda[(1+z)^n]$ is real stable with all roots of the same sign.

We will not prove the general Pólya-Schur theorem here, proving only some special cases later as we need them. Some examples and remarks will clarify its meaning and possible uses. Condition (iv) may be thought of as saying that the polynomials $(1+z)^n$ are universal test cases for stability preserving: a multiplier sequences preserving stability of these will preserve stability of all real stable polynomials.

Example 4.13 (dilation). *If b is a nonnegative integer, setting $\lambda_k := b^k$ produces the operator $T_\lambda f(z) = f(bz)$. Clearly this preserves stability. The representation in (iii) is obvious because $\Phi(z) = e^{bz}$.*

Example 4.14 (factorials). *If $n \geq 1$ is an integer then the sequence $\lambda_k := (n)_k := n!/(n-k)!$, defined to vanish when $k > n$, produces the exponential generating function $\Phi(z) = (1+z)^n$. By criterion (iii) this is a multiplier sequence. Dividing by the constant $n!$ we see that $\lambda_k = 1/(n-k)!$ for $k \leq n$ and zero for $k > n$ defines a multiplier sequence. For real polynomials of degree n , stability of f is equivalent to stability of the inversion $z^n f(1/z)$, hence $\lambda_k = 1/k!$ defines a multiplier sequence on polynomials of degree n . This is true for every n , whence $\{1/k!\}_{k=0}^{\infty}$ is a multiplier sequence. This result is due to Laguerre.*

Example 4.15 (coefficientwise multiplication). Applying the previous example to a polynomial $g(z) = \sum_{k=0}^n a_k z^k$ with negative real roots shows that $\sum_{k=0}^n (a_k/k!)z^k$ also has negative real roots. Setting $\lambda_k := a_k$ for $k \leq n$ and zero for $k > n$ we see that its exponential generating function $\Phi(z) = \sum_{k=0}^n a_k z^k/k!$ is of the form in (iii) with $\alpha_0 = 0$. We conclude that $\{a_k\}$ is a multiplier sequence. In other words, if we multiply term by term the coefficient sequences of two real rooted polynomials, at least one of which has nonnegative coefficients, we get another real rooted coefficient sequence. Thus the real root property for polynomials with nonnegative coefficients is closed under **Hadamard products**.

Multiplier sequences are related to the so-called **Laguerre–Pólya** class. An entire function is said to be in the Laguerre–Pólya class, denoted $\mathcal{L}\text{-}\mathcal{P}$, if it is the limit, uniformly on compact subsets of \mathbb{C} , of polynomials with only real zeros; the terminology goes back at least to [Sch47]. It is said to belong to the subclass $\mathcal{L}\text{-}\mathcal{P}I$ if it is the limit of polynomials with only real zeros all of the same sign. Thus Theorem 4.12 asserts that generating functions of multiplier sequences are precisely the functions of class $\mathcal{L}\text{-}\mathcal{P}I$.

A related notion to that of a multiplier sequence is the notion of a **complex zero decreasing sequence** (CZDS). Let $Z_c(f)$ denote the number of non-real zeros of f . Say that a finite or infinite sequence $\{a_k\}$ is a CZDS if for any real polynomial $f(z) = \sum_{k=0}^n b_k z^k$,

$$Z_c\left(\sum_{k=0}^n a_k b_k z^k\right) \leq Z_c(f). \quad (4.7)$$

In particular, in order for $\{a_k\}$ to be a CZDS, a value of zero on the right of (4.7) (no non-real zeros) implies no non-real zeros on the left, so any CZDS is a multiplier sequence. The converse, however, is not true.

To see that there is any nontrivial CZDS, we observe that the operator $z(d/dz)$ is represented by the sequence $a_k := k$ and can never increase the number of non-real zeros. Although not every multiplier sequence is a CZDS, each multiplier sequence leads to a CZDS via the following result going back to Laguerre.

Proposition 4.16. *Let $\Phi \in \mathcal{L}\text{-}\mathcal{P}$ have zeros only in $(-\infty, 0]$. Then the sequence $\{\Phi(k) : k \geq 0\}$ is a CZDS. \square*

Related to the notions of multiplier sequences and CZDS is the notion of a multiplier sequence for the property of being nonnegative on real inputs. Say that $\{c_k\}$ is a Λ -sequence if multiplication of coefficients term by term preserves the property of being everywhere nonnegative:

$$\sum_{k=0}^n a_k x^k > 0 \text{ for all } x \in \mathbb{R} \implies \sum_{k=0}^n c_k a_k x^k > 0 \text{ for all } x \in \mathbb{R}.$$

Note that negative values are allowed for the numbers c_k . The following classical relationship is proved in [Wid41].

Proposition 4.17. *If $\{a_k\}$ is a CZDS then $\{a_k^{-1}\}$ is a Λ -sequence. □*

Connections abound between these notions: stability, multiplier sequences, the Laguerre–Pólya class, Λ -sequences, etc., which we will not have time to survey here; the reader is referred to [CC95] for an introduction. Λ -sequences, for example, are neatly characterized by a determinant condition and also by the so-called Hamburger moment problem: $\{c_n\}$ is a Λ -sequence if and only if $c_n = \int_{-\infty}^{\infty} t^n d\mu(t)$ for some measure μ on \mathbb{R} not supported on finitely many points. At the root of much of the present interest about these properties is their connection to the Riemann hypothesis. An account is given in [CV90, Cso03]. Finally, a small related literature on nonlinear transformations that preserve real stable polynomials may be of interest. It was conjectured independently by S. Fisk and R. Stanley that if the real polynomial $\sum_k a_k z^k$ has all real roots then so does the polynomial whose z^k coefficient is $a_k^2 - a_{k-1}a_{k+1}$. Some progress was made by McNamara and Sagan in [MS10]; a proof of the conjecture and more was recently found by Brändén [Brä11].

5 Multivariate stability

Let \mathcal{H} denote the open upper half-plane; thus stability is equivalent to having no zeros in the region \mathcal{H}^d . We remark on some connections to other notions of stability. For homogeneous polynomials, the zero set is invariant under multiplication by $e^{i\theta}$ in each coordinate, therefore all notions of half-plane stability coincide. Any **circular** region (the interior or exterior of a circle) is transformed into \mathcal{H} by a Möbius transformation, which allows the theory of Ω -stable functions to be mapped to the ordinary theory of stable functions via a bi-rational change of variables whenever Ω is the product of circular regions. This mapping acts nicely with respect to some aspects such as closure properties but not as nicely with respect to coefficient sequences. Thus, for example, it may be shown that **Schur-stability**, where $\Omega = \mathcal{D}^d$ and \mathcal{D} is the open unit disk, has an unexpected closure property: if f and g are multi-affine Schur-stable polynomials then the **Hadamard product** $f \bullet g$ is Schur-stable as well; here the Hadamard product of $\sum_S a(S)\mathbf{z}^S$ and $\sum_S b(S)\mathbf{z}^S$ is defined to be $\sum_S a(S)b(S)\mathbf{z}^S$. Our principal interest is in \mathcal{H} -stable polynomials and their coefficients. We will not henceforth consider notions of stability other than upper half-plane stability.

A number of the closure properties of the class of stable polynomials extend immediately from the univariate to the multivariate setting or are otherwise elementary. In particular, if $f \in \mathbb{C}[z_1, \dots, z_d]$ is stable then so are the polynomials obtained by the following operations:

- (a) dilation: replacing f by $f(b_1 z_1, \dots, b_d z_d)$, where $\{b_j\}$ are nonnegative constants;
- (b) permuting the names of the variables;
- (c) specialization: setting z_j equal to a constant in \mathcal{H} ;

- (d) diagonalization: setting z_j equal to z_i ;
- (e) inversion: replacing f by $f(-1/z_1, z_2, \dots, z_d)$;
- (f) differentiation: replacing f by $\partial f / \partial z_j$;
- (g) limits: $f_n \rightarrow f$ uniformly on compact sets and f_n stable implies f is stable;

see for example [Wag11, Lemma 2.4]; here, and after, in results such as the this one, we will take “stable” to include the zero polynomial in order not to have to make exceptions.

We also recall that Proposition 1.3 equates stability of a real homogeneous polynomial to hyperbolicity in all directions in the positive orthant. Although the definition of hyperbolicity is more complicated for non-homogeneous polynomials, the corresponding fact characterization of real stable polynomials is not.

Proposition 5.1. *A (non identically zero) real d -variable polynomial f is stable if and only if for all $\mathbf{v} \in \mathbb{R}^d$ and $\mathbf{u} \in \mathbb{R}_+^d$, the polynomial $t \mapsto f(\mathbf{v} + t\mathbf{u})$ has only real zeros.*

PROOF: If f has a zero (z_1, \dots, z_d) in \mathcal{H}^d , then writing $z_j = v_j + iu_j$, we see that $\mathbf{u} \in \mathbb{R}_+^d$ and i is a zero of $f(\mathbf{v} + t\mathbf{u})$. Conversely, if $\mathbf{u} \in \mathbb{R}_+^d$ and $\mathbf{v} \in \mathbb{R}^d$ such that $f(\mathbf{v} + t\mathbf{u})$ has a root with nonvanishing imaginary part, we may conjugate if necessary to find a root $t = a + bi$ with $b > 0$; then $\mathbf{v} + t\mathbf{u}$ is a zero of f in \mathcal{H}^d . \square

There is one geometric feature of functions of two or more complex variables, absent from the univariate theory, that is worth noting because it strengthens closure property (c). Let f be any function in d variables, vanishing at $\mathbf{b} = (b_1, \dots, b_d)$ and analytic in a neighborhood of \mathbf{b} . Suppose that $f(\cdot, b_2, \dots, b_d)$ is not the zero function. Then there is a continuous multi-valued function ϕ on a neighborhood U of (b_2, \dots, b_d) in \mathbb{C}^{d-1} such that $\phi(\mathbf{z}) \rightarrow b_1$ as $\mathbf{z} \rightarrow (b_2, \dots, b_d)$ and such that $f(\phi(\mathbf{z}), \mathbf{z}) = 0$. This perturbation property follows from the fact that at least one multivariate Puiseux series solution exists for z_1 as a function of z_2, \dots, z_d which converges to b_1 . This leads to the following lemma.

Lemma 5.2. *Let $B_j, 1 \leq j \leq d$ be open subsets of \mathbb{C} . Suppose a polynomial f is nonvanishing on $\prod_{j=1}^d B_j$. Then f is nonvanishing on $\prod_{j=1}^d \overline{B_j}$ except possibly in two cases: f can vanish on the product of boundaries $\prod_{j=1}^d \partial B_j$, and f can vanish at a point (b_1, \dots, b_d) with coordinates $\{b_j : j \in S\}$ in the interior of B_j for some set $S \subseteq \{1, \dots, d\}$ provided that f vanishes identically for all values of those coordinates, that is, $f(b'_1, \dots, b'_d) = 0$ whenever $b'_j = b_j$ for all $j \notin S$.*

In particular, taking $B_j = \mathcal{H}$ for all j , we see that stability implies no zeros in $\overline{\mathcal{H}}^d$ except for real zeros or degenerate cases.

A further consequence of this lemma is that (c) can be strengthened to include setting z_j equal to a real constant.

PROOF: Let f vanish at $\mathbf{b} = (b_1, \dots, b_d)$ where $b_j \in \partial B_j$ for $j \leq r$ and $b_j \in B_j$ for $j > r$. We need to show that the hypotheses are contradicted if $k > r$ is a coordinate such that f does not vanish identically when the remaining coordinates are fixed. This follows if we perturb each coordinates $b_j, j \neq k$ so as to lie in the open region B_j , by sufficiently small amounts so that the perturbation solution for z_k lies in B_k . \square

The above facts are generally quite elementary. Before we get to the fun stuff, there is some serious overhead in deriving further closure properties. One route to developing properties of multivariate stable functions is via the notion of **proper position**. This is in some sense a multivariate version of interlacing of roots. The development here specifically avoids this because I do not find it intuitive and because we can do everything we need without it. For a development incorporating the notion of proper position, see [Wag11].

5.1 Equivalences

Recall from Proposition 1.3 the relation between stability and hyperbolicity: a real homogeneous polynomial is stable if and only if it has a cone of hyperbolicity containing the (strictly) positive orthant. A number of equivalent formulations of stability will prove useful. If $f \in \mathbb{C}[z_1, \dots, z_d]$ is any polynomial, let m denote the maximum total degree of any monomial in f and define the homogenization f_H of f to be the unique degree- m homogeneous polynomial in the variables z_1, \dots, z_{d+1} such that $f_H(z_1, \dots, z_d, 1) = f(z_1, \dots, z_d)$. We may write explicitly $f_H(z_1, \dots, z_{d+1}) = z_{d+1}^m f(z_1/z_{d+1}, \dots, z_d/z_{d+1})$. The following proposition is proved in [BBL09, Section 4].

Proposition 5.3 (homogenization). *The real polynomial $f \in \mathbb{R}[z_1, \dots, z_d]$ is stable if and only if the homogeneous polynomial f_H is hyperbolic in all directions \mathbf{v} such that $v_j > 0$ for $1 \leq j \leq d$ and $v_{d+1} = 0$.*

If f also has nonnegative coefficients then the four properties are all equivalent:

- (i) f is stable;
- (ii) f_H is stable;
- (iii) f_H is hyperbolic with respect to some vector in the nonnegative orthant;
- (iv) f_H is hyperbolic with respect to every vector in the positive orthant.

PROOF: For one direction, assume f_H to be hyperbolic in every direction \mathbf{y} with y_j positive when $j \leq d$ and $y_{d+1} = 0$. For any $\mathbf{x} \in \mathbb{R}^d$ and $\mathbf{y} \in \mathbb{R}_+^d$, the polynomial

$$t \mapsto f(\mathbf{x} + t\mathbf{y}) = f_H((\mathbf{x}, 1) + t(\mathbf{y}, 0))$$

is not identically zero because $\lim_{t \rightarrow \infty} t^{-m} f(\mathbf{x} + t\mathbf{y})$ is equal to $f_H(\mathbf{y}, 0) \neq 0$. It has only real zeros, by hyperbolicity of f_H in direction $(\mathbf{y}, 0)$, hence i is not a root, hence $f(\mathbf{x} + i\mathbf{y}) \neq 0$; but $\mathbf{x} \in \mathbb{R}^d$ and $\mathbf{y} \in \mathbb{R}_+^d$ are arbitrary, so this is stability of f .

For the other direction, suppose f_H fails to be hyperbolic in a direction \mathbf{v} for which $v_j > 0, 1 \leq j \leq d$, and $v_{d+1} = 0$. Then $f_H(\mathbf{x} + t\mathbf{v})$ has a non-real zero $t = u$ for some $\mathbf{x} \in \mathbb{R}^{d+1}$. The complex conjugate of u is also a zero so we may assume without loss of generality that the imaginary part of u is positive. The vector $\mathbf{y} := \mathbf{x} + u\mathbf{v}$ lies in $\mathcal{H}^d \times \mathbb{R}$. There are three cases. First, if $y_{d+1} > 0$ then dividing through by this gives a zero of f in \mathcal{H}^d , implying f is not stable. Second, if $y_{d+1} < 0$, dividing yields a zero in $-\mathcal{H}^d$ and conjugating, we see again that f is not stable. Finally, if $y_{d+1} = 0$, we may use perturbation to reduce to the case where $y_{d+1} \neq 0$.

For the equivalence of the four properties in the case of nonnegative coefficients, recall that the cone of hyperbolicity of f_H containing some vector \mathbf{v} is the connected component in \mathbb{R}^{d+1} that contains \mathbf{v} of the set where f_H is nonzero. By nonnegativity of the coefficients, if this cone contains any vector in the closed nonnegative orthant then it contains all vectors in the positive orthant. This establishes the equivalence of (iii) and (iv). Equivalence between (iii) and (ii) is Proposition 1.3. Finally, we note that (ii) implies (i) by setting $z_{d+1} = 1$ (as is allowed by the last consequence of Lemma 5.2), while (i) implies (iii) by the first part of this proposition. \square

Remark. The perturbation argument also proves that stability of f implies stability of $\text{hom}(f)$, where $\text{hom}(f)(z_1, \dots, z_d) := f_H(z_1, \dots, z_d, 0)$ is the leading homogeneous part of f ; see, e.g., [COSW04, Proposition 2.2].

The special case of multi-affine polynomials

Suppose f is multi-affine, that is, no variable appears in any monomial with power two or higher. The following equivalence is taken by some to be the definition of stability in the multi-affine case. Its significance will become evident in the next section.

Theorem 5.4 ([Brä07, Theorem 5.6]). *Let f be real and multi-affine. Then f is stable if and only if the inequality*

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) \frac{\partial f}{\partial x_j}(\mathbf{x}) \geq f(\mathbf{x}) \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) \quad (5.1)$$

holds for all real \mathbf{x} and distinct $i, j \leq n$.

The proof given here is considerably simpler than the published proof, which is a “proper position” argument. It comes from the same source (P. Brändén, personal communication) and begins with the following lemma.

Lemma 5.5. *Let $Q, R \in \mathbb{C}[z_1, \dots, z_n]$ and define $P = Q + z_{n+1}R \in \mathbb{C}[z_1, \dots, z_{n+1}]$. Let $\Omega \subseteq \mathbb{C}^n$ be any connected set and let D_1, D_2 be closed subsets of \mathbb{C} with disjoint interiors, such that $D_1 \cup D_2 = \mathbb{C}$ and $J := D_1 \cap D_2$ a simple curve separating D_1 and D_2 .*

If P has no roots in $\Omega \times J$ and R has no roots in Ω then P either has no roots in $\Omega \times D_1$ or has no roots in $\Omega \times D_2$.

PROOF: P having no roots in $\Omega \times J$ is equivalent to $-Q(\mathbf{z})/R(\mathbf{z}) \notin J$ for all $\mathbf{z} \in \Omega$. Hence $-Q/R$ maps Ω either to the interior of D_1 or the interior of D_2 , hence to D_2^c or D_1^c , and the result follows.

□

PROOF OF THEOREM: For the forward implication, suppose f is stable. Let e_i denote the i^{th} standard basis vector and define $G(s, t) := f(\mathbf{x} + te_i + se_j)$. For any $\mathbf{x} \in \mathbb{R}^d$, the function G is stable in the variables (s, t) . To see this, note that this is equivalent to f having no zeros in $\overline{\mathcal{H}}^d$ with coordinates i and j in \mathcal{H} ; this is ruled out by Lemma 5.2. By multi-affinity, we may express G as $a + bs + ct + dst$ where a, b, c, d are given by partial derivatives:

$$G(s, t) = f(\mathbf{x}) + \frac{\partial f}{\partial z_i}(\mathbf{x})s + \frac{\partial f}{\partial z_j}(\mathbf{x})t + \frac{\partial^2 f}{\partial z_i \partial z_j}(\mathbf{x})st.$$

For bivariate multi-affine polynomials $a + bs + ct + dst$, stability is equivalent to $ad \leq bc$, and the forward direction follows.

For the reverse direction, use induction on n . When $n = 1$ the result is true because all nonzero real polynomials $a + bx$ are stable. Assume the reverse result for n and suppose $P = Q + z_{n+1}R$ satisfies (5.1). If Q or R is identically zero then we are done by induction, so assume not. For any real x , it is clear that $P(z_1, \dots, z_n, x)$ satisfies (5.1), hence, by induction, the polynomials Q, R and $Q + xR$ are all stable or identically zero. If $Q + \alpha R \equiv 0$ for some real α then $P = (z_{n+1} - \alpha)R$ which is stable. We may assume, therefore, that there is no α for which $Q + \alpha R$ is identically zero. It follows that P has no zeros in $H^n \times \mathbb{R}$ where H is the open upper half-plane. By Lemma 5.5 this implies that either P or $P(z_1, \dots, z_n, -z_{n+1})$ is stable. In the latter case, by the forward direction, equation (5.1) holds with the signs reversed for $j = n+1$ and $i \leq n$, and hence with equality for such i and j . This works out to $(\partial R / \partial x_i)Q = (\partial Q / \partial x_i)R$ for all $i \leq n$, that is, Q and R are multiples of each other, which again implies that P is stable. □

5.2 Operations preserving stability

In a brilliant series of papers [BB09a, BB09b, BB10], Borcea and Brändén finished off the problem of extending the Pólya-Schur Theorem to multivariate stability preserving maps. This work not only generalizes from one variable to several, but also from multiplier sequences to arbitrary \mathbb{C} -linear maps. I will present here two of their results, one for multiplier sequences and one giving universal test functions for general linear maps, proving only those parts that will be of use later.

As before, if $\boldsymbol{\lambda} = \{\lambda_{\mathbf{r}} : \mathbf{r} \in \mathbb{Z}_+^d\}$ is an array of real numbers, we denote by $T_{\boldsymbol{\lambda}}$ the operator for

which

$$T_{\lambda} \left(\sum_{\mathbf{r}} a_{\mathbf{r}} \mathbf{z}^{\mathbf{r}} \right) = \sum_{\mathbf{r}} \lambda_{\mathbf{r}} a_{\mathbf{r}} \mathbf{z}^{\mathbf{r}}.$$

Also generalizing from before, we say that λ is a **multivariate multiplier sequence** if T_{λ} preserves the class of real stable polynomials. The following theorem is proved in [BB10, Theorem 1.8].

Theorem 5.6 (multivariate multiplier sequences). *The array λ is a d -variate multiplier sequence if and only if there are d univariate multiplier sequences $\lambda^{(1)}, \dots, \lambda^{(d)}$ such that*

$$\lambda_{\mathbf{r}} = \lambda_{r_1}^{(1)} \cdots \lambda_{r_d}^{(d)}$$

and satisfying a further sign condition: either every $\lambda_{\mathbf{r}}$ is nonnegative, or every $\lambda_{\mathbf{r}}$ is nonpositive, or the same holds for $(-1)^{|\mathbf{r}|} \lambda$.

PROOF IN ONE DIRECTION: The easy direction, and all we will need below, is to see that the product of univariate nonnegative multiplier sequences is a multiplier sequence (also that nonnegative multiplier sequences preserve complex stability, not just real stability). Let $\lambda_{\mathbf{r}} = \prod_{j=1}^d \lambda_{r_j}^{(j)}$ where each sequence $\lambda^{(j)}$ is a univariate multiplier sequence. First assume that $\lambda^{(j)}$ is identically 1 for $j \geq 2$. Let f be a real stable polynomial of d variables. Fixing z_2, \dots, z_d in the upper half-plane, the polynomial $f_1 := f(\cdot, z_2, \dots, z_d)$ is univariate stable, and by the stability preserving assumption on $\lambda^{(1)}$, we see that $T_{\lambda_1}(f_1)$ is univariate stable. This polynomial having no zeros in the upper half-plane is equivalent to $T_{\lambda} f$ having no zeros with z_1 in the upper half-plane and the specified values of z_2, \dots, z_d . Because these values were arbitrary values in the upper half-plane, this finishes the proof in this special case. For the general case, write T_{λ} as the composition of operators of this form in each coordinate. \square

One consequence of this is that the multi-affine part of a stable polynomial is stable.

Corollary 5.7. *Let f be a stable polynomial in d variables and let f_{ma} denote the multi-affine part of f , that is, the sum of the square-free monomials in f . Then f_{ma} is stable.*

PROOF: The sequence $a_0 = 1, a_1 = 1, a_n = 0$ for $n \geq 2$ is a multiplier sequence. By Theorem 5.6, the multivariate sequence defined by $\lambda_{\mathbf{r}} = 1$ if $r_j = 0$ or 1 for all $j \leq d$ and $\lambda_{\mathbf{r}} = 0$ otherwise is a multiplier sequence. This multiplier sequence extracts the multi-affine part. \square

A complete characterization of linear operators preserving stability of complex polynomials was proved in [BB09a, Theorem 2.3]. We will discuss here only the version of this result concerning operators preserving stability of multi-affine functions. We need to define what Borcea and Brändén call the **algebraic symbol** G_T of a linear operator T . Their definition requires a degree bound on each variable. We consider here only the multi-affine case; if you compare to the original text, you should set the degree bound vector κ to $(1, \dots, 1)$. Let $\mathbb{C}_{\text{ma}}[z_1, \dots, z_d]$ denote the space of complex

multiaffine polynomials in d variables. For a linear operator $T : \mathbb{C}_{\text{ma}}[z_1, \dots, z_d] \rightarrow \mathbb{C}[z_1, \dots, z_d]$ define its algebraic symbol to be the element $G_T \in \mathbb{C}[z_1, \dots, z_d, w_1, \dots, w_d]$ satisfying

$$G_T(z_1, \dots, z_d, w_1, \dots, w_d) = \sum_{S \subseteq \{1, \dots, d\}} \prod_{j \notin S} w_j T \left(\prod_{j \in S} z_j \right). \quad (5.2)$$

In other words, G applies T to $\prod_{j=1}^d (z_j + w_j)$ by treating $\{w_j\}$ as constants and applying T to the resulting monomials in $\{z_j\}$.

Theorem 5.8 ([BB09a, Lemma 3.2]). *Let $T : \mathbb{C}_{\text{ma}}[z_1, \dots, z_d] \rightarrow \mathbb{C}[z_1, \dots, z_d]$ be any \mathbb{C} -linear operator. Then T preserves stability if and only one of the two following conditions hold:*

- (a) *There is a stable polynomial P and a linear functional $\alpha : \mathbb{C}_{\text{ma}}[z_1, \dots, z_d] \rightarrow \mathbb{C}$ such that $T(f) = \alpha(f)P$;*
- (b) *The polynomial $G_T(z_1, \dots, z_d, w_1, \dots, w_d)$ is stable as a complex polynomial in $2d$ variables.*

Remark. This is a very powerful theorem and it is worth asking what part of it one needs to understand in order to understand the Borcea–Brändén–Liggett theory of negatively dependent random variables surveyed in Section 6. The answer is that we only use one direction (sufficiency) and this is only to prove Proposition 6.17, which could be proved elementarily as it concerns only bivariate functions. Nevertheless, because this would be messy and unenlightening, we will prove the sufficiency direction of Theorem 5.8 here, which we will then use to prove Proposition 6.17. The proof relies on a lemma proved thirty years ago by Lieb and Sokal [LS81], which we will not prove here (the proof is half a page in [Wag11]).

Lemma 5.9 ([LS81, Lemma 2.3]). *Let $P(\mathbf{z}) + wQ(\mathbf{z})$ be stable in $\mathbb{C}[z_1, \dots, z_d, w]$. If the degree of the variable z_j is at most 1 then the polynomial*

$$P(\mathbf{z}) - \frac{\partial Q(\mathbf{z})}{\partial z_j}$$

is either identically zero or is stable. □

PROOF OF SUFFICIENCY: If T satisfies (a) the result is immediate, so assume T satisfies (b). Because $w \mapsto -w^{-1}$ preserves \mathcal{H} , we see that $G_T(\mathbf{z}, \mathbf{w})$ is stable if and only if $w_1 \cdots w_d G_T(\mathbf{z}, -\mathbf{w}^{-1})$ is stable; here $-\mathbf{w}^{-1}$ denotes $(-w_1^{-1}, \dots, -w_d^{-1})$. From this, we see that if $f \in \mathbb{C}[v_1, \dots, v_d]$ is stable and multi-affine, then

$$(w_1 \cdots w_d) \cdot G_T(\mathbf{z}, -\mathbf{w}^{-1}) \cdot f(\mathbf{v}) = \sum_{S \subseteq \{1, \dots, n\}} T[\mathbf{z}^S] (-\mathbf{w})^S f(\mathbf{v}) \quad (5.3)$$

is stable in $\mathbb{C}[\mathbf{z}, \mathbf{w}, \mathbf{v}]$. We now apply the Lieb-Sokal lemma d times: the j^{th} time we take $w = w_j$ and $z_j = v_j$. Each time, the Lieb Sokal lemma replaces w_j in (5.3) by $-(\partial/\partial v_j)$. The final result is that

$$\sum_{S \subseteq \{1, \dots, n\}} T[\mathbf{z}^S] f^{(S)}(\mathbf{v}) \quad (5.4)$$

is stable in $\mathbb{C}[\mathbf{z}, \mathbf{v}]$. Specializing each v_j to zero preserves stability. But this specialization yields $\sum_S f^{(S)}(0)T[z^S]$, which is another name for $Tf(\mathbf{z})$. Thus stability of f implies stability of Tf , proving the theorem. \square

5.3 More closure properties

The multivariate Pólya-Schur theorems of Borcea and Brändén are powerful but there are yet more closure properties that are useful in probabilistic settings because they have specific probabilistic meanings. We state these now but defer the proofs to the next section.

Let $n_j, 1 \leq j \leq d$ be the maximum degree of the variable z_j in a given polynomial $f \in \mathbb{C}[z_1, \dots, z_d]$. One natural way to create a multi-affine polynomials out of f is to replace each occurrence of each power z_j^r by $\binom{n_j}{r}^{-1} e_r(z_{j,1}, \dots, z_{j,n_j})$, where e_r is the r^{th} elementary symmetric function. Here $\{z_{j,s} : 1 \leq j \leq d, 1 \leq s \leq n_j\}$ are a collection of **clone** variables. Formally, define the **polarization** f_p of f to be the unique polynomial symmetric separately in each set of clone variables $z_{j,1}, \dots, z_{j,n_j}$ such that substituting the value z_j for all the clones $z_{j,s}$ yields the original function f . In Section 6.4 we will show that stability is closed under polarization.

One final useful closure result is the following. The proof will be given after a probabilistic interpretation is given in Section 6.4.3.

Theorem 5.10 (partial symmetrization). *Let f be a d -variate complex polynomial. Fix $1 \leq i < j \leq d$ and define τf to be f with the roles of z_i and z_j swapped:*

$$\tau f(z_1, \dots, z_d) = f(z_1, \dots, z_{i-1}, z_j, z_{i+1}, \dots, z_{j-1}, z_i, z_{j+1}, \dots, z_d).$$

Fix $\theta \in [0, 1]$ and define $f_{\theta;i,j} := (1 - \theta)f + \theta\tau(f)$. If f is stable, so is $f_{\theta;i,j}$.

6 Negative dependence

6.1 A brief history of negative dependence

When random variables in a collection influence each other in a consistent direction, this can be very useful for obtaining one-sided bounds on moments and probabilities of natural events. For example, when all pairwise correlations are negative, the variance of the sum is bounded above by the sum of the variances. For this reason, the literature is sprinkled with definitions of various properties of positive and negative dependence, that hold in examples of interest and which imply useful consequences. Some of these apply to random variables taking values in $\mathbb{R}, \mathbb{R}^+, \mathbb{Z}$ or \mathbb{Z}^+ . To keep matters simple and coherent, we will discuss only the case where the variables are **binary**

valued, that is, take values in the set $\{0, 1\}$. The joint law of d binary variables is a probability measure on $\mathcal{B}_d := \{0, 1\}^d$.

The simplest and one of the weakest conditions is pairwise positive (respectively negative) correlation: $\mathbb{E}X_i X_j \geq (\mathbb{E}X_i)(\mathbb{E}X_j)$ (respectively \leq), which may also be written as $\mathbb{P}(X_i = X_j = 1) \geq \mathbb{P}(X_i = 1)\mathbb{P}(X_j = 1)$ (respectively \leq). As we have just seen, this condition is strong enough to imply one-sided bounds on the variance of $\sum_j X_j$. It is not strong enough to give bounds on the moments. For this, one requires at least positive or negative **cylinder dependence**. Positive cylinder dependence is said to hold if

$$\mathbb{E} \prod_{j \in S} X_j \geq \prod_{j \in S} \mathbb{E}X_j \tag{6.1}$$

for all subsets $S \subseteq \{1, \dots, n\}$. Negative cylinder dependence is the reverse inequality. Sometimes this is strengthened to require as well that

$$\mathbb{E} \prod_{j \in S} (1 - X_j) \geq \prod_{j \in S} (1 - \mathbb{E}X_j). \tag{6.2}$$

The equation (6.1) immediately implies that the moments of $Z := \sum_j X_j$ are at least what they would be for independent Bernoullis with the same marginals, while negative cylinder dependence implies the moments are at most what they would be in the independent case. Expanding the moment generating function $\phi(\lambda) := \mathbb{E}e^{\lambda Z}$ in powers of Z , and then into monomials in the variables X_j shows that positive (respectively negative) cylinder dependence implies that for $\lambda \geq 0$, the moment generating function $\phi(\lambda)$ is at least (respectively at most) what it would be for independent Bernoullis of the same marginal; including the inequality for complementary cylinders (6.2) extends the inequality for moment generating functions to $\lambda < 0$. The bounds on moment generating functions transfer to concentration inequalities (see, e.g., [PS97, Theorem3.4]).

A yet stronger property is association. Say that the collection of random variables $\{X_1, \dots, X_d\}$ is **positively associated** if

$$\mathbb{E}f(X_1, \dots, X_d)g(X_1, \dots, X_d) \geq \mathbb{E}f(X_1, \dots, X_d)\mathbb{E}g(X_1, \dots, X_d)$$

whenever $f, g : \mathcal{B}_d \rightarrow \mathbb{R}$ are nondecreasing functions (with respect to the coordinatewise partial order on \mathcal{B}_d). Negative association cannot be defined by simply reversing the inequality because any function f is always nonnegatively correlated with itself. Therefore, we amend the definition to say that $\{X_1, \dots, X_d\}$ are **negatively associated** if $\mathbb{E}fg \leq \mathbb{E}f \mathbb{E}g$ whenever $f, g : \mathcal{B}_d \rightarrow \mathbb{R}$ are nondecreasing and there is a set $S \subseteq \{1, \dots, n\}$ such that f depends only on $\{X_j : j \in S\}$ while g depends only on $\{X_j : j \notin S\}$.

One of the most useful results on positive association is a result of Fortuin, Kasteleyn and Ginibre [FKG71]. Say that y covers z in the Boolean lattice \mathcal{B}_d if $y > z$ in the coordinatewise partial order but there is no element strictly between y and z ; we denote this relation by

$$y \succ z,$$

If x and y both cover z and w covers x and y , we call this configuration a face of the lattice \mathcal{B}_d . Say that the probability measure μ satisfies the **positive lattice condition** if whenever x, y, z, w form a face of \mathcal{B}_d (with z and w at the bottom and top), $\mu(x)\mu(y) \leq \mu(z)\mu(w)$. It is immediate to see that this property for faces implies $\mu(x \wedge y)\mu(x \vee y) \geq \mu(x)\mu(y)$ for every pair x, y , hence we also call in (multiplicative) submodularity. What is not so obvious is the theorem now known as the FKG Theorem.

Theorem 6.1 (FKG). *If μ satisfies the positive lattice condition then μ is positively associated.* \square

The power of this theorem is that its hypotheses are easy to check while its conclusions are quite strong. In many statistical mechanical systems, ratios of probabilities over a face of \mathcal{B}_d are easy to compute even if the absolute probabilities are not. The FKG theorem immediately implies positive association for ferro-magnetic Ising models, the random cluster model with $q \geq 1$, and a number of other models from physics, many of which are surveyed in [Lig85].

For negative dependence the situation is not as nice. The negative lattice condition (reverse the inequality on each face) does not imply negative association, or seemingly anything of value. Unlike the positive lattice condition it is not closed under the most natural of operations, namely integrating out one variable (probabilistically this means ignoring this variable and viewing μ as a measure on \mathcal{B}_{n-1}). As a result, it has been difficult to identify negatively dependent laws and to establish properties such as negative association.

6.2 Search for a theory

For some time now, a more satisfying theory of negative dependence has been sought. The goal was to find a property of laws μ on \mathcal{B}_d which would be:

- checkable for a handful of known or conjectured **examples**;
- shown to imply **consequences** such as negative association;
- closed under natural **operations**.

This goal was popularized in the article [Pem00] which posed several challenges and stated many conjectures but contained few concrete results. **Spoiler alert:** for binary valued random variables, the property is that the probability generating function is stable; such a probability distribution is called *strong Rayleigh*; see Theorem 6.9 below. In the forthcoming lists of examples, consequences and closure properties, only in two cases are they not known to hold for the class of strong Rayleigh measures: Example 6.4 (the random cluster measure) is conjectured to be Rayleigh but known not to be strong Rayleigh, and the proposed closure property (d) from Section 6.2.3 (restriction to an interval) is known to fail.

One open question remaining from [Pem00] is whether the property known there as h-NLC⁺ implies negative association. To define h-NLC⁺, let us first define an **external field**. Let μ be a probability measure on \mathcal{B}_n and let $\lambda_j > 0$ be positive real numbers, for $1 \leq j \leq d$. Define a measure μ' by

$$\mu'(x_1, \dots, x_d) = \frac{\prod_{j=1}^d \lambda_j^{x_j}}{Z} \mu(x_1, \dots, x_d)$$

where $Z := \sum_{\mathbf{x}} \lambda^{\mathbf{x}} \mu(\mathbf{x})$ is the normalizing constant. We call μ' the perturbation of μ by the external field λ . The negative lattice condition is not closed under integrating out a variable; we say that the hereditary negative lattice condition plus external fields holds (h-NLC⁺) if the negative lattice condition holds for all measures obtained from μ by imposing an external field and integrating out some of the variables. This was later shown to be equivalent to Wagner's **Rayleigh condition** [Wag08], which will discuss further in Section 6.3. We now survey the proposed examples, consequences and closure properties for the desired class of negatively dependent laws.

6.2.1 Proposed examples

Example 6.2 (conditioned Bernoullis). *Let μ be the law of independent Bernoulli variables $\{X_1, \dots, X_d\}$ with means p_1, \dots, p_d , let k be an integer between 1 and d and let $\mu' := (\mu | \sum_{j=1}^d X_j = k)$ be the conditional law of $\{X_j\}$ given that they sum to k . This should be negatively dependent. Note that this is equal to the uniform measure on k -subsets of $\{1, \dots, d\}$ under the external field $\lambda_j = p_j/(1-p_j)$.*

Example 6.3 (spanning trees). *Let $G = (V, E)$ be a finite connected simple graph and let $\{\lambda_e : e \in E\}$ be nonnegative weights. Let W be the set of subsets $T \subseteq E$ such that the resulting (V, T) is connected and has no cycles. Each $T \in W$ will have cardinality exactly $|V| - 1$. Such subgraphs are called **spanning trees** of G . Let $w(T) := \prod_{e \in T} \lambda_e$ denote the weight of T under the multiplicative weighting scheme determined by λ . We define the weighted spanning measure μ (depending on G and λ) by*

$$\mu(T) = \frac{w(T)}{\sum_{T \in W} w(T)}.$$

This was proven in [Pem91] to have negative correlations and in [FM92] to have negative association, and we would hope it to be included under any reasonable definition of “negatively dependent”.

Example 6.4 (random cluster measure). *The random cluster model on a graph $G = (V, E)$ with parameters $q > 0$ and $\{\lambda_e : e \in E\}$ is the probability measure μ on $\{0, 1\}^E$ obtained by normalizing the weights*

$$w(T) = \prod_e \lambda_e^{T(e)} q^{N(T)}$$

where $T \subseteq E$ and $N(T)$ is the number of connected components of the subgraph (V, T) of G . When $q = 2$ this is equivalent to the Ising model for ferromagnetism and for integers $q \geq 2$ it is equivalent to the Potts model, also from statistical physics (see, e.g. [Wu88]). When $q \geq 1$, the FKG Theorem immediately shows μ to be positively associated. When $0 < q < 1$, the measure is conjectured but not

known to be negatively associated. We would hope that the right definition of negative dependence would settle this conjecture.

Example 6.5 (exclusion measures). The **exclusion process** is a continuous time Markov chain on \mathcal{B}_d as follows. The state $\mathbf{x} \in \mathcal{B}_d$ is interpreted as a collection of particles at the sites $j \in \{1, \dots, d\}$ for which $x_j = 1$; the sites j for which $x_j = 0$ are considered vacant. Let $\lambda_{\{i,j\}}$ be nonnegative real numbers, and independently at rate $\lambda_{\{i,j\}}$, let the values x_i and x_j swap. This is interpreted as a particle at one site i or j jumping to the other site. If both sites or neither site is occupied, then nothing happens. The exclusion process is a special case of the more general **exchange process**, where the labels of the sites are arbitrary, and in particular might be distinct rather than being drawn from the two element set $\{0, 1\}$. Let μ_t be the law of the exclusion process at time t , starting from some deterministic state η at time 0. It was conjectured that μ_t is always negatively associated. We would like the definition of negatively dependent measures to include μ .

Example 6.6 (determinantal measures with Hermitian kernels). Let K be a $d \times d$ Hermitian matrix with spectrum in $[0, 1]$. For $S \subseteq \{1, \dots, d\}$, let $w(S)$ denote the determinant of the submatrix K_S of K when restricted to rows and columns in the set S . There is a unique probability measure μ on subsets of $\{1, \dots, d\}$ such that

$$\mu\{T : T \supseteq S\} = |K_S|$$

for each S . This is called the **determinantal** measure with kernel K . Such measures were proved in [Lyo03] to be negatively associated. We would hope these measures satisfy our definition of negative dependence.

6.2.2 Proposed consequences

It was proposed in [Pem00] that the right definition of negative dependence for binary variables would imply the following properties, with definitions immediately to follow.

- (a) negative association
- (b) stochastically increasing levels
- (c) stochastic covering property
- (d) log-concave rank sequence

Recall that a probability measure μ on a lattice W is said to **stochastically dominate** a law ν , written $\mu \succeq \nu$, if $\mu(A) \geq \nu(A)$ for every upwardly closed set A (the set A is upwardly closed if $x \in A$ and $y > x$ implies $y \in A$). An equivalent condition is stated in terms of coupling: $\mu \succeq \nu$ if and only if there is a measure Q on $W \times W$ whose first marginal is μ , whose second marginal is ν , and which is supported on $\{(x, y) \in W^2 : x \geq y\}$. In other words, we may simultaneously sample

from μ and ν in such a way that the sample from μ is always greater than or equal to the sample from ν .

Given a measure μ on \mathcal{B}_d and an event $G \subseteq \mathcal{B}_d$ with nonzero measure, denote by $(\mu|G)$ the measure $\mu(\cdot)/\mu(\cdot \cap G)$ obtained by conditioning on G . The conditional measure $(\mu|X_j = 1)$ may be identified with a measure on \mathcal{B}_{d-1} by ignoring the j^{th} coordinate.

Property 6.1 (stochastically increasing levels). *Say that a measure μ on \mathcal{B}_d has stochastically increasing levels if $(\mu|\sum_j x_j = k) \preceq (\mu|\sum_j x_j = k + 1)$ for each k such that $\mu(\sum_j x_j = k)$ and $\mu(\sum_j x_j = k + 1)$ are both nonzero.*

The property of having stochastically increasing levels is not implied by negative association, as was hoped in [Pem00], but is a desired consequence of the right definition of negative dependence.

Property 6.2 (stochastic covering). *Say that a measure ν on a lattice W stochastically covers the measure μ , and denote this by $\nu \triangleright \mu$, if there is a coupling measure Q on W^2 with marginals ν and μ , such that Q is supported on the set $\{(x, y) \in W^2 : x = y \text{ or } x > y\}$. In other words, ν stochastically dominates μ but “only by one”, in the sense that you can sample from μ by sampling from ν and then flipping at most one one to a zero.*

Negative association of a measure μ implies that $(\mu|X_j = 0) \succeq (\mu|X_j = 1)$ when viewed as a measure on \mathcal{B}_{d-1} . The **stochastic covering property** is defined in [PP11] to hold when $(\mu|X_j = 0) \triangleright (\mu|X_j = 1)$, viewed as measures on \mathcal{B}_{d-1} . This property was wrongly conjectured in [Pem00] to follow from negative association. We would hope for it to follow from the right definition of negative dependence.

Property 6.3 ((ultra) log-concave rank sequence). *The **rank sequence** for the measure μ on \mathcal{B}_d is the sequence $(a_k := \mu\{\sum_j X_j = k\})_{0 \leq k \leq d}$. It was wrongly conjectured to follow from negative association that the rank sequence is log-concave. The right definition of negative dependence turns out to imply not only log-concavity but **ultra log-concavity**, namely that for $0 \leq k \leq d$, equation (4.4) holds, which I repeat here for convenience.*

$$\frac{a_k^2}{\binom{d}{k}^2} \geq \frac{a_{k-1}}{\binom{d}{k-1}} \frac{a_{k+1}}{\binom{d}{k+1}}. \quad (6.3)$$

6.2.3 Proposed closure properties

In addition to negative association, it was thought that the right negative dependence property would imply the following closure properties.

- (a) closure under products, projections (integrating out a variable), and limits;

- (b) closure under external fields;
- (c) conditioning on the event $\{\sum_j X_j = k\}$;
- (d) more generally, conditioning on the event $k_1 \leq \sum_j X_j \leq k_2$;
- (e) total symmetrization: replacing $\mu(\mathbf{x})$ by $(1/n!) \sum_{\pi \in S_d} \mu(\pi \mathbf{x})$ where π acts by permuting the coordinates;
- (f) more generally, evolution via the symmetric exclusion process for any finite time.

It turns out that (c) is too strong, but that the class of strong Rayleigh measures is indeed closed under **rank rescaling**, (see the upcoming definition) when the rank sequence \mathbf{b} is ultra log-concave.

Definition 6.7 (rank rescaling). *Let μ be a measure on \mathcal{B}_d and let $\mathbf{b} := (b_0, \dots, b_d)$ be a vector of nonnegative real numbers. For $\mathbf{x} = (x_1, \dots, x_d) \in \mathcal{B}_d$ let $N(\mathbf{x}) := \sum_{j=1}^d x_j$ denote the sum of the coordinates. For nontriviality assume that $Z := \sum_{i=0}^d b_i \mu\{N = i\} > 0$. The rank rescaling of μ by \mathbf{b} is the measure $\mu_{\mathbf{b}}$ defined by*

$$\mu_{\mathbf{b}}(\mathbf{x}) = \frac{N(\mathbf{x})}{Z} \mu(\mathbf{x}).$$

6.3 The grail is found: application of stability theory to joint laws of binary random variables

It is now acknowledged that the “correct” definition of negative dependence for binary random variables is the following definition due to [BBL09].

Definition 6.8 (strong Rayleigh). *Say that the probability measure μ on \mathcal{B}_d is strongly Rayleigh (also “strong Rayleigh”) if its generating polynomial $f \in \mathbb{C}[z_1, \dots, z_d]$ defined by*

$$f(\mathbf{z}) := \sum_{\omega \in \mathcal{B}_d} \mu(\omega) \mathbf{z}^\omega$$

is a real stable polynomial.

As a generating polynomial, f is automatically multi-affine, with real nonnegative coefficients. Recalling Theorem 5.4, we see that strong Rayleigh is equivalent to the inequality (5.1) on the mixed partial derivatives of f holding for all real \mathbf{x} . This is sometimes taken as the definition of strong Rayleigh.

The origin of the name comes from Wagner’s Rayleigh property. This is the mixed partial inequality (5.1) but only required to hold for nonnegative real arguments. As was mentioned before, this is equivalent to h-NLC⁺, which is not hard to believe because varying \mathbf{x} over \mathbb{R}_+^d corresponds to varying the external field arbitrarily. Strangely, when this is required to hold for negative values

of the external field as well, the property becomes more natural and robust. Because the Rayleigh property is equivalent to h-NLC^+ and is obviously a consequence of the strong Rayleigh property, we see immediately that strong Rayleigh implies pairwise negative correlations. This is, of course, just the tip of the iceberg.

Theorem 6.9 (properties of strong Rayleigh measures).

(i) *Strong Rayleigh measures are negatively associated.*

(ii) *Strong Rayleigh measures have stochastically increasing levels, satisfy the stochastic covering property 6.2, and have ultra log-concave rank sequences as in (6.3). The class of strong Rayleigh measures is also closed under rank rescaling by sequences (b_0, \dots, b_d) when the nonzero values b_i are the coefficient sequence of a stable univariate polynomial $\sum_{i=0}^d b_i z^i$. In particular, this implies closure of the class of strong Rayleigh measures under conditioning on $\{\sum_j X_j = k\}$ or $\{k \leq \sum_j X_j \leq k+1\}$.*

(iii) *The class of strong Rayleigh measures is closed under products, projections, external fields, total symmetrization and the symmetric exclusion dynamics.*

As an example of the utility of this theorem, here is one of the two applications that brought the theory to my attention.

Example 6.10 (concentration inequalities for Lipschitz functions). *Let μ be a measure on \mathcal{B}_d and let $f : \mathcal{B}_d \rightarrow \mathbb{R}$ be Lipschitz with constant 1. When μ is a product measure, the martingale $W_k := \mathbb{E}(f | \omega_1, \dots, \omega_k)$ has bounded increments $|W_k - W_{k-1}| \leq 1$. It follows from Azuma's inequality (see, e.g. [AS08]) that f satisfies the Gaussian concentration inequality*

$$\mathbb{P}(f - \mathbb{E}f \geq a) \leq e^{-a^2/(2n)}. \quad (6.4)$$

In fact independence is not needed. The stochastic covering property was defined in [PP11] precisely to imply $|W_k - W_{k-1}| \leq 1$. It follows that (6.4) holds for any strong Rayleigh measure.

6.4 Proof of Theorem 6.9

6.4.1 Proof of part (i) : negative association

“Strong Rayleigh implies negative association” was the result of Borcea–Brändén–Liggett that pointed the way to application of stable polynomial theory to the probabilistic setting. We follow their proof, which begins by paving the way for a reduction to the case of homogeneous measures.

Lemma 6.11 (homogenization). *Let $\{X_1, \dots, X_d\}$ be a collection of random variables whose law is strong Rayleigh and define $X_{d+1} := d - \sum_{j=1}^d X_j$. Then the collection $\{X_1, \dots, X_{d+1}\}$ is strong Rayleigh.*

PROOF: The generating function for the collection $\{X_1, \dots, X_{d+1}\}$ is just the homogenization of the generating function for $\{X_1, \dots, X_d\}$. Therefore, what we need to check is that the homogenization of a stable generating function is stable. This is not true for all polynomials⁴ but for probability generating functions, or more generally any stable polynomial with nonnegative coefficients, this follows from (i) \Rightarrow (ii) of Proposition 5.3. \square

Next we recall the definition of polarization via clone variables from Section 5.3: it is the unique polynomial symmetric separately in each set of clone variables $\{z_{j,1}, \dots, z_{j,n_j}\}$ such that substituting in z_j for all the clones $z_{j,s}$ yields the original function f . The following result does not rely on f having nonnegative, or even real coefficients.

Theorem 6.12 (polarization). *The complex polynomial f is stable if and only if its polarization f_p is stable.*

To prove this we require the Grace–Walsh–Szegő theorem. This result may be derived from the theory of stable polynomials, as shown in [BB09b] and streamlined in [Wag11, Section 4]. We will be content here to quote this century old result; for a proof, see [COSW04, Theorem 2.12] or Section 4 of [Wag11].

Lemma 6.13 (GraceWalsh-Szegő Theorem). *Let $f \in \mathbb{C}[z_1, \dots, z_d]$ be symmetric and multi-affine and let \mathcal{C} be a convex circular region containing the points ζ_1, \dots, ζ_d , for instance the upper half-plane \mathcal{H} . Then there exists at least one point $\zeta \in \mathcal{C}$ such that $f(\zeta_1, \dots, \zeta_d) = f(\zeta, \dots, \zeta)$.* \square

PROOF OF THEOREM 6.12: One direction is elementary: if f_p is stable then the diagonalization property allows us to set $z_{j,s} = z_j$ for all j and we deduce stability of f . For the nontrivial direction, suppose f_p is not stable and let $\mathbf{a} := \{a_{j,s}\}$ be numbers in the upper half-plane such that $f_p(\mathbf{a}) = 0$. Fixing the numbers $\{a_{j,s} : j > 1\}$, the Grace-Walsh-Szegő Theorem implies the existence of a number a_1 in the upper half-plane such that $f(a_1, \dots, a_1, \mathbf{a}') = 0$, where \mathbf{a}' are the numbers $a_{j,s}$ for $j > 1$. Iterating, we arrive at numbers $a_2, \dots, a_n \in \mathcal{H}$ for which $f(a_1, \dots, a_n) = 0$, showing that f is not stable. \square

A measure μ on \mathcal{B}_d is called **Projected Homogeneous Rayleigh** (PHR) if there is a measure ν on \mathcal{B}_m for some $m \geq d$ such that ν is a homogeneous measure with the Rayleigh property and the projection of ν to the law of the first d variables gives μ . If, furthermore, the law ν is strong Rayleigh, we say that μ is PHSR.

Corollary 6.14. *Strong Rayleigh measures are PHSR.*

PROOF: First extend the law μ of the binary variables $\{X_1, \dots, X_d\}$ by homogenization to a law μ_H on variables $\{X_1, \dots, X_{d+1}\}$, the last of which takes integer values. By Lemma 6.11 this is

⁴Indeed $f(x) = x - 1$ is a trivial example of a stable polynomial whose homogenization $f_H(x, y) = x - y$ fails to be stable.

still in the class of strong Rayleigh measures. Next replace the integer variable X_{d+1} by clones $\{X_{d+1,1}, \dots, X_{d+1,d}\}$ by polarization; the resulting measure ν on \mathcal{B}_{2d} is strong Rayleigh by Theorem 6.12 and its projection onto the first d coordinates is μ . \square

PROOF OF NEGATIVE ASSOCIATION: To prove negative association, we use an argument due to Feder and Mihail [FM92]. This argument shows that homogeneous strong Rayleigh measures (thus by Corollary 6.14, all strong Rayleigh measures) are negatively associated. All that is needed in the Feder–Mihail lemma is for this class to be closed under conditioning on $\{X_j = x\}$ (which is obvious) and have pairwise negative correlations, which we have seen in (5.1). In short, negative association for strong Rayleigh measures is reduced to the following lemma.

Lemma 6.15 (Feder–Mihail). *Let \mathcal{S} be a class of homogeneous measures on finite Boolean algebras (of differing sizes) which is closed under conditioning and each of which has pairwise negative correlations. Then all measures in \mathcal{S} are negatively associated.*

Assuming this lemma, the proof of negative association finishes as follows. The class of homogeneous strong Rayleigh measures is closed under conditioning on the value of any variable. We have noted that strong Rayleigh implies negative pairwise correlations, so we conclude from the Feder–Mihail lemma that homogeneous strong Rayleigh measures are negatively associated. Any strong Rayleigh measure is a projection of a homogeneous strong Rayleigh measure, hence inherits the negative association property. It remains only to prove the lemma.

PROOF OF FEDER–MIHAIL LEMMA: Any increasing function on \mathcal{B}_d is a positive linear combination of increasing indicator functions, that is, indicator functions $\mathbf{1}_A$ of upwardly closed events A . It therefore suffices to prove negative correlation for upwardly closed events A and B depending on disjoint sets of variables. (Note: throughout the proof, the phrase “negatively correlated” includes the case of zero correlation.)

We first prove this in the special case where A is the event $\{X_j = 1\}$. We use induction on the number of variables, m . When $m = 2$, the only nontrivial case is $\mathbf{1}_A = X_1$ and $\mathbf{1}_B = X_2$, which are negatively correlated by hypothesis. Now assume for induction that for all measures $\nu \in \mathcal{S}$ on Boolean lattices \mathcal{B}_d for $d \leq m$, for all j , and for all upwardly closed events B not depending on X_j , the functions X_j and $\mathbf{1}_B$ are negatively correlated. Let μ be a law on \mathcal{B}_{m+1} , and let $j \leq m + 1$ and $B \subseteq \mathcal{B}_{m+1}$ not depending on X_j be given. We will use the inequality

$$pa + (1 - p)b \leq p'c + (1 - p')d$$

holding when $p, p', a, b, c, d \in [0, 1]$ with $p \leq p', a \leq c, b \leq d$ and $a \geq b$. To use this inequality, fix $i \leq m + 1$ to be determined later. Let $p := \mu(X_i = 1 \mid X_j = 1)$ and $p' := \mu(X_i = 1 \mid X_j = 0)$, the inequality $p \leq p'$ holds because it is equivalent to pairwise negative correlation of X_i and X_j . We

let

$$\begin{aligned}
a &:= \mu(B | X_i = 1, X_j = 1); \\
b &:= \mu(B | X_i = 0, X_j = 1); \\
c &:= \mu(B | X_i = 1, X_j = 0); \\
d &:= \mu(B | X_i = 0, X_j = 0).
\end{aligned}$$

The inequalities $a \leq c$ and $b \leq d$ follow because the conditional measures $(\mu | X_j = 1)$, $(\mu | X_i = 1)$ and $(\mu | X_i = 0)$ are all subject to the induction hypothesis. Finally, to ensure that $a \geq b$, we now choose i judiciously. Equivalent conditions for the inequality $a \geq b$ are

$$\begin{aligned}
\frac{\mu(B, X_i = 1, X_j = 1)}{\mu(B^c, X_i = 1, X_j = 1)} &\geq \frac{\mu(B, X_i = 0, X_j = 1)}{\mu(B^c, X_i = 0, X_j = 1)} \\
&\Downarrow \\
\frac{\mu(B, X_i = 1, X_j = 1)}{\mu(B, X_i = 0, X_j = 1)} &\geq \frac{\mu(B^c, X_i = 1, X_j = 1)}{\mu(B^c, X_i = 0, X_j = 1)} \\
&\Downarrow \\
\mathbb{E}(X_i | B, X_j = 1) &\geq \mathbb{E}(X_i | B^c, X_j = 1).
\end{aligned}$$

Homogeneity of the measure μ implies that the sum over i of the left and right-hand sides of the last inequality are both equal to the deterministic value $\sum_{i=1}^{m+1} X_i$. Therefore the inequality must hold for at least one $i \leq m + 1$. We conclude that

$$\mu(B | X_j = 1) = pa + (1 - p)b \leq p'c + (1 - p')d = \mu(B | X_j = 0).$$

Having established that single variables are negatively correlated with upwardly closed events for all measures in \mathcal{S} , we consider the general case where A and B are upwardly closed events depending on disjoint sets of variables. Again we use induction, assuming the result for all measures on fewer variables and also already possessing the result in the special case. We run the argument with $\mathbf{1}_A$ in place of X_j , thus $p := \mu(X_i = 1 | A)$, $p' := \mu(X_i = 1 | A^c)$, and so forth. The inequality $p \leq p'$ is a consequence of the special case. The inequalities $a \leq c$ and $b \leq d$ come from the induction hypothesis applied to the conditional measures $(\mu | X_i = 1)$ and $(\mu | X_i = 0)$ respectively. Again, constancy of $\sum_{i=1}^{m+1} X_i$ forces the inequality $\mathbb{E}(X_i | B, A) \geq \mathbb{E}(X_i | B^c, A)$ to hold for some i , finishing the proof. \square

6.4.2 Proof of part (ii) : stochastic inequalities

We begin with the stochastic covering property, which follows directly from Corollary 6.14 and the following lemma, first proved in [PP11].

Lemma 6.16 ([PP11, Proposition 2.2]). *If a measure is projected homogeneous strong Rayleigh then it has the stochastic covering property.*

PROOF: Let μ be a projection of a homogeneous strong Rayleigh measure ν . Without loss of generality, we assume that the projection is onto the last d out of m coordinates. Let ν_0 denote the probability measure on \mathcal{B}_{m-1} which is the ν -law of (X_1, \dots, X_{m-1}) conditioned on $X_m = 0$; similarly, let ν_1 denote the probability measure on \mathcal{B}_{m-1} which is the ν -law of (X_1, \dots, X_{m-1}) conditioned on $X_m = 1$. Negative association of ν implies that ν_0 stochastically dominates ν_1 : let $B \subseteq \mathcal{B}_m$ be any event not depending on the m^{th} coordinate then B , that is $B = B_* \times \{0, 1\}$ for some $B_* \in \mathcal{B}_{m-1}$; then negative correlation of B and X_m implies that $\nu_0(B_*) \geq \nu_1(B_*)$. The equivalent coupling formulation of stochastic domination (see the beginning of Subsection 6.2.2) shows there is a random variable (X, Y) such that X has law ν_0 , Y has law ν_1 and $X \geq Y$. By homogeneity of ν , we see in fact that X always covers Y . Projecting back to \mathcal{B}_d , this yields a random pair (\bar{X}, \bar{Y}) such that \bar{X} has law $(\mu | X_m = 0)$, \bar{Y} has law $(\mu | X_m = 1)$, and \bar{X} is always equal to or covering \bar{Y} . \square

PROOF OF ULTRA LOG-CONCAVITY OF THE RANK SEQUENCE: Next we examine the rank function. Let μ be a strong Rayleigh measure on \mathcal{B}_d . The rank sequence $a_k := \mu(N = k)$ has generating function $g(z) := \sum_{k=0}^d a_k z^k = f_\mu(x, \dots, x)$. This is a stable polynomial because it is a diagonalization of the stable polynomial f_μ , recalling property (d) at the beginning of Section 5. Newton's inequalities (Theorem 4.2) for univariate real stable polynomials then yield ultra log-concavity of the coefficients of the rank sequence. \square

PROOF OF RESCALING BY A STABLE COEFFICIENT SEQUENCE: Let b_0, \dots, b_d be a sequence of non-negative real numbers whose nonzero elements are an interval b_r, b_{r+1}, \dots, b_s and are the coefficients of a stable polynomial. We have seen in Example 4.15 that the sequence $\{b_k\}$ is a multiplier sequence. It follows from Theorem 5.6 that if $f(\mathbf{z}) = \sum_{\mathbf{r}} a_{\mathbf{r}} \mathbf{z}^{\mathbf{r}}$ is a real stable function of $d+1$ variables then $g(z) := \sum_{\mathbf{r}} a_{\mathbf{r}} b_{r_{d+1}} \mathbf{z}^{\mathbf{r}}$ is also stable.

Now let μ be a strong Rayleigh measure on \mathcal{B}_d with probability generating function f_μ . By Lemma 6.11, the measure ν whose generating function is the homogenization $f_\nu = (f_\mu)_H$ is also stable (recall that ν is the law of $(X_1, \dots, X_d, d - \sum_{j=1}^d X_j)$ when (X_1, \dots, X_d) has law μ). With $\{b_k : 0 \leq k \leq d\}$ any ultra log-concave sequence as above, we have seen that $g(\mathbf{z}) := \sum_{\mathbf{r}} a_{\mathbf{r}} b_{r_{d+1}} \mathbf{z}^{\mathbf{r}}$ is stable. Setting z_{d+1} equal to 1, we see that $g(\mathbf{z})/g(1, \dots, 1)$ is a real stable probability generating function, whence the measure described at the end of part (ii) of Theorem 6.9 is strong Rayleigh. \square

PROOF OF STOCHASTICALLY INCREASING LEVELS: If the interval of support $[r, s]$ of the ultra log-concave sequence \mathbf{b} satisfies $s - r = 0$ or 1 , then it is automatically ultra log-concave. Taking $s = r$, we see that if the law μ of (X_1, \dots, X_d) is strong Rayleigh then so is $(\mu | N = k)$ as long as $\mu(N = k) > 0$; here again $N := \sum_{j=1}^d X_j$. Taking $s = r + 1$ we see that $(\mu | k \leq N \leq k + 1)$ is strong Rayleigh, provided that the event conditioned on is not null.

Fix k such that $\mu(N = k)$ and $\mu(N = k + 1)$ are both nonzero, and let $\mu_k := (\mu | N = k)$ and $\mu_{k+1} := (\mu | N = k + 1)$. The measure $(\mu | k \leq N \leq k + 1)$ is strong Rayleigh. Homogenizing by adding the check bit $x_{d+1} := k + 1 - \sum_{j=1}^d x_j$ gives a measure ν on \mathcal{B}_{d+1} that is also strong Rayleigh.

Let $A \subseteq \mathcal{B}_d$ be any upwardly closed set and let $A_* := A \times \{0, 1\}$. By negative association, the event A_* is negatively correlated with X_{d+1} . Thus $\nu(A_* | X_{d+1} = 1) \leq \nu(A_* | X_{d+1} = 0)$, which is equivalent to $\mu_k(A) \leq \mu_{k+1}(A)$. Because A was an arbitrary upwardly closed event, this proves that $\mu_k \preceq \mu_{k+1}$. \square

6.4.3 Proof of part (iii) : exclusion and other closure properties

Most of the closure properties in part (iii) of Theorem 6.9 are easy, following from the closure properties of the class of stable polynomials listed at the beginning of Section 5. The generating function for the product measure $f_{\mu \times \nu}$ is the product of functions $f_\mu \cdot f_\nu$, whence the strong Rayleigh property for products of strong Rayleigh measures follows from stability of the product of stable polynomials. Projection corresponds to setting some variables equal to 1, whence closure under projections follows from specialization to real values (Lemma 5.2). External fields correspond to replacing f by $cf(a_1 z_1, \dots, a_d z_d)$ for positive real values of the a_j , which follows from the closure of stable polynomials under dilations. To prove closure of the class of strong Rayleigh measures under total symmetrization, there are two avenues. One is to deduce this from the more difficult result for partial symmetrization (see below), observing that total symmetrization may be achieved by repeated partial symmetrization. As pointed out in [BBL09, Remark 4.5], a more direct argument is as follows. Let μ be a strong Rayleigh measure on \mathcal{B}_d . The sum $N := \sum_{j=1}^d X_j$ has generating function $f_N(z) := f_\mu(z, \dots, z)$, which is stable because it is a diagonalization of f_μ . The generating function of the total symmetrization $\bar{\mu}$ is the polarization of $f_N(z)$, hence is stable by Theorem 6.12.

The key to proving closure of the class of strong Rayleigh measures under exclusion dynamics is to prove that a single step of **partial symmetrization** preserves the strong Rayleigh property. Let μ be a probability measure on \mathcal{B}_d , let $1 \leq i < j \leq d$ be indices and fix $0 < \theta < 1$. The θ -partial symmetrization of μ with respect to indices i and j is defined to be the measure $\mu_{\theta; i, j}$ whose verbal description is as follows.

To sample from $\mu_{\theta; i, j}$, first sample from μ , then flip an independent θ -coin to decide whether to transpose the i and j coordinates.

At the level of generating functions, this is Theorem 5.10, which we now recall and prove. The generating polynomial $f_{\mu, \theta; i, j}$ is the one described in Theorem 5.10 by

$$f_{\mu, \theta; i, j} = (1 - \theta)f_\mu + \theta\tau(f_\mu)$$

where τ operates by switching arguments i and j . The theorem states that if f is stable then so is $(1 - \theta)f + \theta\tau(f)$, which when applied to $f = f_\mu$ yields closure of strong Rayleigh measures under partial symmetrization. We now prove the theorem.

PROOF OF THEOREM 5.10: Let μ be a strong Rayleigh measure. Assume without loss of generality that $i = 1$ and $j = 2$. We need to show that $f_{\mu,\theta,1,2}$ is stable. For every choice of complex numbers $a_3, a_4, \dots, a_d \in \mathcal{H}$, we know that $f_\mu(\cdot, \cdot, a_3, \dots, a_d)$ is multi-affine and stable and we need to show that $f_{\mu,\theta,1,2}(\cdot, \cdot, a_3, \dots, a_d)$ is stable. Therefore it suffices to show:

Proposition 6.17. *For all multi-affine stable functions $f : \mathbb{C}^2 \rightarrow \mathbb{C}$, the function $(1 - \theta)f(x, y) + \theta f(y, x)$ is stable.*

Remark. Because of the restriction to the multi-affine case, we know that $f(x, y) = a + bx + cy + dxy$, so it cannot be too hard to prove this! We observe that if we needed this only for real stable functions, it would follow immediately from the characterization that f is stable if and only if $ad \leq bc$, which follows from the mixed partial derivative criterion (5.1). Indeed, the proof of Theorem 5.10 appearing in [BBL09, Section 4.4] begins with this and extends via the multivariate Obreschkoff theorem. To keep things self-contained, we will instead derive this from the portion of Theorem 5.8 that we have proved; another self-contained proof is given in [Lig09, Theorem 7]; and exhaustive description of this four parameter family is yet another way to finish this.

PROOF OF PROPOSITION 6.17: The operator T on $\mathbb{C}_{\text{ma}}[x, y]$ corresponding to partial symmetrization is defined by its action on the four basis elements $1, x, y$ and xy :

$$T(1) = 1; T(x) = (1 - \theta)x + \theta y; T(y) = (1 - \theta)y + \theta x; T(xy) = xy.$$

Plugging into the definition (5.2) of G_T gives

$$G_T(x, y, u, v) = xy + [(1 - \theta)x + \theta y]u + [(1 - \theta)y + \theta x]v + uv.$$

We need to show this is stable. It is multi-affine and has real coefficients, so we may apply the mixed partial derivative test (5.1). There are six unordered pairs of variables. In each case we easily verify that the left side of (5.1) minus the right side is equal to a nonnegative multiple of a square when $0 \leq \theta \leq 1$. For example, for the pair (x, u) , we evaluate $(\partial^2 f / \partial x \partial u) f - (\partial f / \partial x)(\partial f / \partial u)$ to obtain $\theta(y + v)^2$, while for the pair (x, y) we obtain $(\partial^2 f / \partial x \partial y) f - (\partial f / \partial x)(\partial f / \partial y) = \theta(1 - \theta)(u - v)^2$; up to the symmetries of f , these are the only two cases, therefore the proposition is proved. \square

The final step, from partial symmetrization to exclusion dynamics, is a small one. The operators $T_{\theta;i,j}$ defined by $\mu \mapsto \mu_{\theta;i,j}$ generated a semigroup S of operators on the space of probability measures on \mathcal{B}_d . For any set $\{\lambda_{\{i,j\}}\}$ of swap rates and any $t \geq 0$, the operator mapping μ to the time t law of the exclusion process with swap rates λ started from the measure μ is in the closure of S . This is more or less self-evident; for a more formal proof, note that when only one swap rate is nonzero, the exclusion evolution operator is already equal to an operator $T_{\theta;i,j}$, after which one can use the Trotter product formula (see [BBL09, Proposition 5.1]) to write the general exclusion evolution operator as a limit of products of those with only one nonvanishing swap rate. This completes the proof of Theorem 6.9. \square

To finish the discussion of probabilistic applications, we discuss examples. The theorem explicitly addresses exclusion measures (Example 6.5). Conditioned Bernoullis (Example 6.2) are strong Rayleigh due stability of $1 - p + pz$ for $0 \leq p \leq 1$, closure under products, (yielding all measures with independent coordinates), and closure under conditioning on $N = k$. Random cluster measures with $q < 1$ are conjectured to be Rayleigh and known not to be strong Rayleigh and spanning trees are a special case of determinantal measures, so among the examples 6.2–6.6 it remains only to show that determinantal measures are strong Rayleigh. We will require a result which is the analogue of Example 2.4 but for stability instead of hyperbolicity. The exposition is taken from [BBL09, Proposition 3.2] though the result itself has been known for a long time.

Proposition 6.18. *Let A_1, \dots, A_d be (complex) positive semi-definite $n \times n$ matrices and let B be a Hermitian matrix, also $n \times n$.*

1. *The polynomial*

$$(z_1, \dots, z_d) \mapsto f(z_1, \dots, z_d) := \text{Det}(z_1 A_1 + \dots + z_d A_d + B)$$

is either identically zero or it is real stable.

2. *If B is also positive semi-definite then the f has all nonnegative coefficients.*

3. *It follows that if $Z := \text{diag}(z_1, \dots, z_d)$ and B is any positive semi-definite $d \times d$ matrix, then $\text{Det}(B + Z)$ is a multi-affine real stable polynomial with all nonnegative coefficients, hence equal to $c f_\mu$ for some strong Rayleigh measure μ .*

PROOF: Stability is closed under limits, so it suffices to prove the proposition in the case where the matrices A_1, \dots, A_d are positive definite. The polynomial f is real on real inputs because the determinant of a complex Hermitian matrix is real. Pick a vector $\mathbf{u} \in \mathbb{R}_+^d$ and $\mathbf{v} \in \mathbb{R}^d$. Define a function \mathbf{z} by $\mathbf{z}(t) := \mathbf{v} + t\mathbf{u}$. The matrix $P := \sum_{j=1}^d v_j A_j$ is positive definite, hence invertible with a positive definite square root, call it Q . We may write

$$f(\mathbf{z}(t)) = \text{Det}(P) \text{Det}(tI + QHQ^*)$$

where $H := B + \sum_{j=1}^d u_j A_j$ is Hermitian. Thus $f(\mathbf{z}(t))$ is a polynomial in t that is a constant multiple of the characteristic polynomial of a Hermitian matrix, hence has all real zeros. Because $\mathbf{v} \in \mathbb{R}^d$ and $\mathbf{u} \in \mathbb{R}_+^d$ are arbitrary, the criterion for stability in Proposition 5.1 is satisfied, proving the first statement of the theorem.

Expanding the determinant, we find that the coefficients of f are products of principal minors of the matrices A_1, \dots, A_d and B . When these are all positive semi-definite, the principal minors are positive, proving the second statement. The last statement follows immediately from setting A_i to the matrix with a 1 in the (i, i) -entry and zeros elsewhere, noting that this matrix is positive semi-definite. \square

PROOF THAT DETERMINANTAL MEASURES ARE STRONG RAYLEIGH: Assume first that the Hermitian kernel K is invertible. It is easily seen that the generating polynomial f_μ is given by

$$f_\mu(z_1, \dots, z_d) = \text{Det}(I - K + KZ) = \text{Det}(K) \text{Det}(K^{-1} - I + Z),$$

where $Z := \text{diag}(z_1, \dots, z_d)$. Because K is also a contraction, $K^{-1} - I$ is positive semi-definite. The constant $\text{Det}(K)$ is positive, so it follows from Example 2.4 that f_μ is stable. The general case may be obtained by taking limits because positive definite kernels are dense in the space of positive semi-definite kernels. \square

7 Further applications of stability: determinants, permanents and moments

From the outset, determinants have been prominent in the theory of hyperbolic polynomials. Already in [Går51, Example 2], the determinant function on the space of Hermitian matrices was given as an example of a hyperbolic polynomial, with the nonnegative definite matrices being a cone of hyperbolicity (see Example 2.4 above). Related to the determinant but more enigmatic is the **permanent**. The definition of the permanent,

$$\text{Per}(A) := \sum_{\sigma \in S_n} \prod_{i=1}^n a_{i, \sigma(i)},$$

differs from that of the determinant only in that there is no alternating sign factor $(-1)^{\text{parity}(\sigma)}$ in the summand. This makes the permanent much less tractable than the determinant. For example, it is $\#P$ -hard to compute the permanent of a zero-one matrix, while determinants may be evaluated in polynomial time. In this section we discuss two results on permanents and one on determinants. The first of these is a lower bound on the permanent, conjectured by van der Waerden in 1926, proved independently by Egorychev and Falikman in 1981, and re-proved in a simpler and more general way by Gurvits in 2009 using the theory of stable functions. The second of these results, the Monotone Column Permanent Conjecture asserts the stability of a certain polynomial obtained as a permanent. It was conjectured in 1999 and proved in 2009. The third result, the so-called BMV conjecture, is only tangentially related to stability theory, but has garnered enough attention to mandate its inclusion here. It was conjectured in 1975. A proof posted recently to the arXiv is believed to be correct.

7.1 The van der Waerden conjecture

One prolific area of research in understanding the permanent has been to identify extremal cases for various families of matrices. Some motivation for understanding permanents comes from graph

theory. If A is the incidence matrix of a bipartite graph G , then $\text{Per}(A)$ is the number of perfect matchings of G . This interpretation has led to an emphasis on extrema over zero-one matrices. For example, one might consider matrices of zeros and ones with prescribed row sums. An upper bound, conjectured by Minc and proved by Brégman [Bré73] is as follows.

Theorem 7.1 (Brégman’s Theorem). *Let A be a nonnegative $n \times n$ matrix of zeros and ones and let r_1, \dots, r_n denote the row sums of A . Then*

$$\text{Per}(A) \leq \prod_{j=1}^n (r_j!)^{1/r_j}.$$

□

The lower bound is of course zero because there could be a column of zeros. If we prescribe the column sums as well as the row sums, can we find a nontrivial lower bound? Removing the restriction to zero-one matrices, this question was posed by van der Waerden in 1926 for **doubly stochastic** matrices. A matrix $A = (a_{ij})_{1 \leq i, j \leq n}$ is said to be stochastic if it has nonnegative entries and all row sums $\sum_j a_{ij}$ are equal to 1. The terminology comes from the fact that these matrices are precisely the transition kernels for Markov chains. The matrix A is said to be doubly stochastic if all column sums $\sum_i a_{ij}$ are equal to 1 as well. One might not see at first that the permanent of a doubly stochastic matrix must be nonzero, but this follows from the well known fact that doubly stochastic matrices are positive linear combinations of permutation matrices (and such matrices have permanent equal to 1).

Intuition may suggest that the minimum occurs when all entries are equal to $1/n$. Indeed, van der Waerden conjectured in 1926 that if A is doubly stochastic then

$$\text{Per}(A) \geq \frac{n!}{n^n} \tag{7.1}$$

with equality if and only if $a_{ij} = 1/n$ for all i, j . This theorem was proved thirty years ago, independently by Egorychev [Ego81] and Falikman [Fal81]. Recently, Gurvits [Gur08] gave a different proof using stability. Not only does this represent a considerable simplification, but the result is general enough to imply several other well known results. Gurvits begins by identifying the permanent as a coefficient of a polynomial, as follows. If $A = (a_{ij})$ is any $n \times n$ matrix, we may define a homogeneous polynomial p by

$$p(\mathbf{x}) = p_A(\mathbf{x}) := \prod_{i=1}^n \sum_{j=1}^n a_{ij} x_j.$$

The permanent of A is then the $(1, \dots, 1)$ coefficient of p . Thus,

$$\text{Per}(A) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} p(0, \dots, 0). \tag{7.2}$$

When A is stochastic, each factor in the product evaluates to 1 at $(1, \dots, 1)$. Taking the derivative with respect to x_j and evaluating at $(1, \dots, 1)$ gives the j^{th} column sum, hence if A is doubly stochastic then

$$\frac{\partial}{\partial x_j} p_A(1, \dots, 1) = 1, \quad 1 \leq j \leq n. \quad (7.3)$$

Let \mathcal{C}_n be the class of homogeneous polynomials of degree n in n variables with all coefficients nonnegative. Following Gurvits, we extend the definition of the term ‘‘doubly stochastic’’ from matrices to \mathcal{C}_n by saying that p is doubly stochastic if (7.3) holds.

To prove the van der Waerden result, we need to derive $\text{Per}(A) \geq n!/n^n$ from (7.3). This will not be true for every $p \in \mathcal{C}_n$ satisfying (7.3) but Gurvits’ idea was that it should hold for all stable $p \in \mathcal{C}_n$. Because stability is closed under product and each linear polynomial with positive coefficients is stable, we see immediately that p_A is always stable. The other ingredient in Gurvits’ proof is to strengthen the induction, replacing the lower bound of $n!/n^n$ by $(n!/n^n)\text{Cap}(p)$, where $\text{Cap}(p)$ is a constant that evaluates to 1 when $p = p_A$ for a doubly stochastic matrix A .

Definition 7.2. *Let \mathcal{C}_n be the class of homogeneous polynomials of degree n in n variables with all coefficients nonnegative. For $p \in \mathcal{C}_n$, define the capacity of p by*

$$\text{Cap}(p) := \inf \frac{p(x_1, \dots, x_n)}{\prod_{i=1}^n x_i}$$

where the infimum is over nonnegative values of the variables x_i .

Proposition 7.3 ([Gur08, Fact 2.2]). *If p is doubly stochastic then $\text{Cap}(p) = 1$.*

PROOF: We may view $p \in \mathcal{C}_n$ as a generating function for a probability distribution μ supported on $\Delta_n := \{\omega \in \mathbb{Z}_+^n : |\omega| = n\}$. Letting \mathbb{E} denote expectation with respect to μ we see that (7.3) is equivalent to $\mathbb{E}\omega_j = 1$ for all j .

For $\mathbf{x} \in \mathbb{R}_+^n$, we may evaluate $p(\mathbf{x}) = \log \mathbb{E} \prod_j x_j^{\omega_j}$. Applying Jensen’s inequality to the concave function $\log(x)$ gives

$$\begin{aligned} \log p(x_1, \dots, x_n) &= \log \mathbb{E} \prod_{j=1}^n x_j^{\omega_j} \\ &\geq \mathbb{E} \log \prod_{j=1}^n x_j^{\omega_j} \\ &= \mathbb{E} \sum_j \omega_j \log x_j. \end{aligned}$$

Because $\mathbb{E}\omega_j = 1$ for all j , this last quantity is just $\log \prod_j x_j$. Thus, on the positive orthant, $p(\mathbf{x}) \geq \prod_{j=1}^n x_j$ showing that $\text{Cap}(p) \geq 1$. Setting $x_j = 1$ for all j shows that $\text{Cap}(p) \leq 1$. \square

Recalling the lower bound on $f'(0)$ for univariate stable polynomials with nonnegative coefficients from Proposition 4.3 it is not hard to envision some kind of induction. The engine of Gurvits’ proof, used in the induction step, is the following inequality.

Lemma 7.4 ([Gur08, Theorem 4.10]). *Let $p \in \mathcal{C}_n$ be stable and define*

$$q(x_1, \dots, x_{n-1}) := \frac{\partial}{\partial x_n} p(x_1, \dots, x_{n-1}, 0).$$

Then

$$\text{Cap}(q) \geq \text{Cap}(p) \cdot \left(\frac{d-1}{d} \right)^{d-1} \quad (7.4)$$

where d is the maximum degree of x_n in p .

PROOF: Letting x_1, \dots, x_{n-1} range over positive numbers whose product is 1, we need to show that

$$q(x_1, \dots, x_{n-1}) \geq \text{Cap}(p)G(d)$$

where $d \leq n$ is the maximum degree of x_n in p and $G(m) := [(m-1)/m]^{m-1}$. Fix x_1, \dots, x_{n-1} . The specialization property implies that the univariate polynomial $R(t) := p(x_1, \dots, x_{n-1}, t)$ is stable. By definition of capacity, $\text{Cap}(p) \leq \inf_{t>0} R(t)/t$. The degree of R is equal to d , whence it follows from Proposition 4.3 that

$$q(x_1, \dots, x_{n-1}) = R'(0) \geq \text{Cap}(p)G(d).$$

□

We may now give Gurvits' result implying the van der Waerden conjecture.

Theorem 7.5. *Let p be stable with nonnegative coefficients. Then*

$$\frac{\partial^n}{\partial x_1 \cdots \partial x_n} p(0, \dots, 0) \geq \text{Cap}(p) \frac{n!}{n^n}.$$

PROOF: Let $q_n = p$ and in general define

$$q_i(x_1, \dots, x_i) := \frac{\partial^{n-i}}{\partial x_{i+1} \cdots \partial x_n} p(x_1, \dots, x_i, 0, \dots, 0).$$

Stability is closed under differentiation and specializing to real values, hence by induction on $n-i$, each q_i is stable. Also, $q_i \in \mathcal{C}_i$ for each i . Applying Lemma 7.4 with q_i in place of p and i in place of n shows that

$$\text{Cap}(q_{i+1}) \geq \text{Cap}(q_i)G(i)$$

because i is an upper bound for the degree of x_i in q_i . Inductively, we see that

$$\text{Cap}(q_1) \geq \text{Cap}(p) \prod_{i=2}^n G(i). \quad (7.5)$$

The right-hand side may be identified as

$$\prod_{i=2}^n G(i) = n!/n^n. \quad (7.6)$$

To identify the left-hand side, observe that $q_1(x_1)$ is homogeneous of degree 1 in x_1 , that is, $q_1 = cx_1$. Thus

$$\text{Cap}(q_1) = c = q_1'(0) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} p(0, \dots, 0). \quad (7.7)$$

Plugging (7.7) and (7.6) into (7.5) proves the theorem. \square

Another result that succumbs to this method is the Schrijver–Valiant conjecture, proved in 1998. This concerns an integer version of doubly stochastic matrices. Let $\Lambda(k, n)$ denote the collection of $n \times n$ matrices whose entries are nonnegative integers and whose rows and columns all sum to k . Define

$$\lambda(k, n) = k^{-n} \min\{\text{Per}(A) : A \in \Lambda(k, n)\}$$

and

$$\theta(k) = \lim_{n \rightarrow \infty} \lambda(k, n)^{1/n}.$$

In 1980 it was proved [SV80] that $\theta(k) \leq G(k)$ and equality was conjectured. The proof by Schrijver [Sch98] was difficult. This result is a corollary of Theorem 7.5. For details on this, as well as a separate application of Theorem 7.5 to prove a lower bound on the mixed discriminant, see [Gur08].

7.2 The Monotone Column Permanent Conjecture

Say that the $n \times n$ matrix A is a **monotone column matrix** if its entries are real and weakly decreasing down each column, that is, $a_{i,j} \geq a_{i+1,j}$ for $1 \leq i \leq n-1$ and $1 \leq j \leq n$. Let J_n denote the $n \times n$ matrix of all ones. It was conjectured in [HOW99] that whenever A is a monotone column matrix, the univariate polynomial $\text{Per}(zJ_n + A)$ has only real roots. A proof was given there for the case where A is a zero-one matrix.

Although there was strong intuition already present, the proof had to wait for the development of multivariate stable function theory. In 2009, Brändén, Haglund, Visontai and Wagner [BHVW09] were able to prove this conjecture by proving something stronger, namely multivariate stability.

Theorem 7.6 (Multivariate Monotone Column Permanent Theorem (MMCPT)). *Let Z_n be the diagonal matrix whose entries are the n indeterminates z_1, \dots, z_n . Let A be an $n \times n$ monotone column matrix. Then $\text{Per}(J_n Z_n + A)$ is a stable polynomial in the variables z_1, \dots, z_n . Specializing to $z_j \equiv z$ for all j preserves stability, hence the original conjecture follows.*

I will give only a brief sketch of the proof, which relies on two results proved by Borcea and Brändén. One of these is the criterion, Theorem 5.8 for a linear operator on multi-affine polynomials to be stability preserving. They use the sufficiency direction, namely that stability of the $(2n)$ -variable “algebraic symbol” $G_T(\mathbf{z}, \mathbf{w})$ implies that T is a stability preserving operator on $\mathbb{C}_{\text{ma}}[z_1, \dots, z_d]$; this is the direction for which a proof was included above. The other lemma they require is the following result.

Lemma 7.7 ([BHVW09, Proposition 2.5]). *Let V be a real vector space and let $\phi : V^n \rightarrow \mathbb{R}$ be a multilinear form. Let $e_1, \dots, e_n, v_2, \dots, v_n$ be fixed vectors in V and suppose that the polynomial*

$$\phi(e_1, v_2 + z_2 e_2, \dots, v_n + z_n e_n)$$

is not identically zero in $\mathbb{R}[z_2, \dots, z_n]$. Then the set of all $v_1 \in V$ for which the polynomial

$$\phi(v_1 + z_1 e_1, \dots, v_n + z_n e_n)$$

is stable is either empty or a convex cone over the origin containing $\pm e_1$. □

By means of this lemma, the MMCPT may be reduced to the special case where the entries of A are all zero or one. I will not reproduce this reduction here, but the main idea is an induction. If we have proved the result in the case where k columns are real monotone and $n - k$ are zero-one monotone, then applying the lemma to one of the $n - k$ columns shows that stability is achieved over a convex cone containing all zero-one monotone columns, and such a cone necessarily contains all real monotone columns.

After a change of variables to $y_j = 1 + 1/z_j$ and the introduction of new variables x_1, \dots, x_n , stability of $\text{Per}(z_j + a_{ij})$ will follow if we show stability of $\text{Per}(a_{ij}y_j + (1 - a_{ij}x_i))$ for monotone zero-one matrices. Denote this last permanent as $\text{Per}(B(A))$. Expanding the permanent on the last row gives a differential recurrence relation which may be written in form $f_j = T f_{j-1}$ where f_j is the permanent of the upper $j \times j$ submatrix of $B(A)$. Here T is an operator of the form $k + z_j \sum_{i=1}^{j-1} \partial/\partial z_j$. The sufficiency criterion reduces the task to checking stability of the algebraic symbol, G_T , which may be accomplished by a simple computation.

7.3 The BMV conjecture

In 1975, Bessis, Moussa and Villani [BMV75] formulated what is now known as the BMV conjecture.

Conjecture 7.8 (BMV). *Let A and B be Hermitian $n \times n$ matrices with B positive semi-definite. Then the function*

$$\lambda \mapsto \text{Tr} \exp(A - \lambda B)$$

is the Laplace transform of a positive measure on $[0, \infty)$; here, $\text{Tr}(\cdot)$ denotes the trace.

This seems vaguely tied to a number of the themes we have been discussing, but not necessarily related in a direct way to the theory of stable functions. In fact, both the conclusion and the hypotheses will reveal more upon further scrutiny. Beginning with the obvious, we recall that the linear cone of positive semi-definite matrices is a cone of hyperbolicity for the determinant function on the space of Hermitian matrices. Thus if we take the determinant instead of the trace of the exponential, the resulting function $\lambda \mapsto \text{Det}(A - \lambda B)$ is stable.

Recall from Edrei’s Equivalence Theorem (Theorem 4.9) that stability of the polynomial $f := \sum_{j=0}^k a_j z^j$ is equivalent to a sequence of inequalities, which can be summed up by saying that $\{a_k\}$ is a Pólya frequency sequence. Similarly, a function ϕ on $[0, \infty)$ is a Laplace transform of a positive measure on \mathbb{R}^+ if and only if ϕ is a **completely monotone** function, meaning that the derivatives of ϕ do not change sign on $[0, \infty)$ and alternate in sign: for each nonnegative integer n and positive real t ,

$$(-1)^n \frac{d^n}{dt^n} \phi(t) \geq 0.$$

This result is the Bernstein–Widder Theorem [Ber28]. Evidently the conclusion is a property whose definition has features in common with stability, a property we know to be true for $\lambda \mapsto \text{Det}(A - \lambda B)$.

In 2004, Lieb and Seiringer [LS04] found an equivalent formulation of the BMV conjecture that looks even more similar to stability theory.

Theorem 7.9. *The BMV conjecture is equivalent to the polynomials*

$$\text{Tr}(A + \lambda B)^n$$

having nonnegative coefficients for all integers $n \geq 1$ and all pairs of matrices A, B that are both positive semi-definite.

The exponential is gone, the coefficient on λ is positive, only integer powers are involved, and what is needed is a countable sequence of inequalities: all coefficients in a sequence of polynomials must be nonnegative. Another equivalence brings the BMV conjecture directly into the realm of stable function theory. Let $p(x, y)$ be any stable bivariate polynomial with nonnegative coefficients and nonzero constant term. Let a_{mn} be the Taylor coefficients of $(\partial p / \partial x) / p$. The BMV conjecture is equivalent to the proposition that $(-1)^{m+n} a_{mn} \geq 0$ for all m, n .

In July, 2011, a proof of the BMV conjecture by H. Stahl [Stal1] was posted on arXiv. At the time of writing, the refereed publication has not appeared, but word on the street is that the proof should be correct. The proof is sufficiently complicated that understanding its specialization to the 3×3 case is considered a good project, perhaps for a dissertation. There is a sense that there may be a simpler proof out there, and that the “book proof” of the BMV-Stahl Theorem is likely to involve stable polynomials.

Acknowledgements

I owe a huge debt to Yuliy Baryshnikov for helping me to understand the material in Part I and to Petter Brändén for helping me to understand the material in Part II. Without their help, this survey could not have been written. Thanks are also due to Mirkó Visontai for his careful reading and comments, to Alan Sokal for crucial help and corrections on the development of multivariate stability, and to David Wagner for helpful discussions.

References

- [ABG70] M. Atiyah, R. Bott, and L. Gårding. Lacunas for hyperbolic differential operators with constant coefficients, I. *Acta Mathematica*, 124:109–189, 1970.
- [AS08] N. Alon and J. Spencer. *The probabilistic method*. John Wiley & Sons, Hoboken, New Jersey, third edition, 2008.
- [BB09a] J. Borcea and P. Brändén. The Lee-Yang and Pólya-Schur programs, I: Linear operators preserving stability. *Invent. Math.*, 177:541–569, 2009.
- [BB09b] J. Borcea and P. Brändén. The Lee-Yang and Pólya-Schur programs, II: Theory of stable polynomials and applications. *Comm. Pure Appl. Math.*, 62:1595–1631, 2009.
- [BB10] J. Borcea and P. Brändén. Multivariate Pólya-Schur classification problems in the Weyl algebra. *Proc. London Math. Soc.*, 101:73–104, 2010.
- [BBBP08] Y. Baryshnikov, W. Brady, A. Bressler, and R. Pemantle. Two-dimensional quantum random walk. *arXiv*, <http://front.math.ucdavis.edu/0810.5495>: 34 pages, 2008.
- [BBL09] J. Borcea, P. Brändén, and T. Liggett. Negative dependence and the geometry of polynomials. *J. AMS*, 22:521–567, 2009.
- [BC92] J. Brown and C. Colbourn. Roots of the reliability polynomial. *SIAM J. Disc. Math.*, 5:571–585, 1992.
- [Ber28] S. Bernstein. Sur les fonctions absolument monotones. *Acta Math.*, 52:1–66, 1928.
- [BHVW09] P. Brändén, J. Haglund, M. Visontai, and D. Wagner. Proof of the monotone column permanent conjecture. In *FPSAC 2009*, volume AK, pages 443–454, Nancy, 2009. Assoc. Discrete Math. Theor. Comput. Sci.
- [BMV75] D. Bessis, P. Moussa, and M. Villani. Monotonic converging variational approximations to the functional integrals in quantum statistical mechanics. *J. Math. Phys.*, 16:2318–2325, 1975.
- [BP11] Y. Baryshnikov and R. Pemantle. Asymptotics of multivariate sequences, part iii: quadratic points. *Adv. Math.*, 228:3127–3206, 2011.
- [Brä07] P. Brändén. Polynomials with the half-plane property and matroid theory. *Adv. Math.*, 216:302–320, 2007.
- [Brä11] P. Brändén. Iterated sequences and the geometry of zeros. *J. Reine Angew. Math.*, 658:115–131, 2011.
- [Bré73] L. Brégman. Some properties of nonnegative matrices and their permanents. *Sov. Math. Dokl.*, 14:945–949, 1973.

- [Bre89] F. Brenti. Unimodal, log-concave and Pólya frequency sequences in combinatorics. *Memoirs of the AMS*, 413:106+viii, 1989.
- [Bre94] F. Brenti. Log-concave and unimodal sequences in algebra, combinatorics and geometry: an update. In *Jerusalem Combinatorics '93*, volume 178, pages 71–89, Providence, 1994. AMS.
- [Bre95] F. Brenti. Combinatorics and total positivity. *J. Comb. Theory, ser. A*, 71:175–218, 1995.
- [CC95] T. Craven and G. Csordas. Complex zero decreasing sequences. *Meth. Appl. Anal.*, 2:420–441, 1995.
- [Con78] John B. Conway. *Functions of One Complex Variable*. Springer-Verlag, New York, second edition, 1978.
- [COSW04] Y.-B. Choe, J. Oxley, A. Sokal, and D Wagner. Homogeneous multivariate polynomials with the half-plane property. *Adv. Appl. Math.*, 32:88–187, 2004.
- [Cso03] G. Csordas. Complex zero decreasing sequences and the Riemann hypothesis, II. In *ISSAC 2001*, volume 10, pages 121–134, Dordrecht, 2003. Kluwer Academic Publishers.
- [CV90] G. Csordas and R. Varga. Necessary and sufficient conditions and the Riemann hypothesis. *Adv. Appl. Math.*, 11:328–357, 1990.
- [Edr53] A. Edrei. Proof of a conjecture on Schoenberg on the generating function of a totally positive sequence. *Canadian J. Math.*, 5:86–94, 1953.
- [Ego81] G. Egorychev. The solution of van der Waerden’s problem for permanents. *Adv. Math.*, 42:299–305, 1981.
- [Fal81] D. Falikman. Proof of the van der Waerden’s conjecture on the permanent of a doubly stochastic matrix. *Mat. Zametki*, 29:931–938, 1981. (in Russian).
- [FB85] A. Fettweis and S. Basu. On discrete scattering Hurwitz polynomials. *International J. Circuit Theory Appl.*, 13:47–59, 1985.
- [FKG71] C. Fortuin, P. Kasteleyn, and P. Ginibre. Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.*, 22:89–103, 1971.
- [FM92] T. Feder and M. Mihail. Balanced matroids. *Annual ACM Symposium on Theory of Computing*, pages 26–38, 1992.
- [Går51] L. Gårding. Linear hyperbolic partial differential equations with constant coefficients. *Acta Math.*, 85:1–62, 1951.
- [Går59] L. Gårding. An inequality for hyperbolic polynomials. *J. Math. Mech.*, 8:957–965, 1959.

- [GJ83] I. P. Goulden and D. M. Jackson. *Combinatorial enumeration*. A Wiley-Interscience Publication. John Wiley & Sons Inc., New York, 1983. With a foreword by Gian-Carlo Rota, Wiley-Interscience Series in Discrete Mathematics.
- [GK71] C. Gruber and H. Kunz. General properties of polymer systems. *Comm. Math. Phys.*, 22:133–161, 1971.
- [GKZ94] I. Gelfand, M. Kapranov, and A. Zelevinsky. *Discriminants, Resultants and Multidimensional Determinants*. Birkhäuser, Boston-Basel-Berlin, 1994.
- [Gül97] O. Güler. Hyperbolic polynomials and interior point methods for convex programming. *Math. Oper. Res.*, 22:350–377, 1997.
- [Gur06] L. Gurvits. The van der Waerden conjecture for mixed discriminants. *Adv. Math.*, 200:435–454, 2006.
- [Gur08] L. Gurvits. Van der Waerden/Schrijver-Valiant like conjectures and stable (aka hyperbolic) homogeneous polynomials: one theorem for all. *Elec. J. Combin.*, 15:26, 2008.
- [Hen91] Peter Henrici. *Applied and Computational Complex Analysis. Vol. 2*. John Wiley & Sons Inc., New York, 1991. Special functions—integral transforms—asymptotics—continued fractions, Reprint of the 1977 original, A Wiley-Interscience Publication.
- [HKPV09] J. B. Hough, M. Krishnapur, Y. Peres, and B. Virag. *Zeros of Gaussian analytic functions and determinantal point processes*, volume 51 of *University Lecture Series*. American Mathematical Society, Providence, RI, 2009.
- [HL72] O. Heilmann and E. Lieb. Theory of monomer-dimer systems. *Comm. Math. Phys.*, 25:190–232, 1972.
- [Hör83] Lars Hörmander. *The analysis of linear partial differential operators. I*. Springer-Verlag, Berlin, 1983. Distribution theory and Fourier analysis.
- [HOW99] J. Haglund, K. Ono, and D. G. Wagner. Theorems and conjectures involving rook polynomials with only real zeros. In *Topics in Number Theory*, volume 467 of *Math. Appl.*, pages 207–221. Kluwer Academic Publishers, Dordrecht, 1999.
- [Hur96] A. Hurwitz. On the conditions under which an equation has only roots with negative real parts. In *Stability Theory*, volume 121 of *International Series of Numerical Mathematics*, pages 239–249, Berlin, 1996. Birkhäuser. Reprinted from *Math. Ann.* **46** (1895).
- [Kar68] S. Karlin. *Total Positivity, vol. 1*. Stanford University Press, Stanford, CA, 1968.
- [Lig85] T. Liggett. *Interacting Particle Systems*, volume 276 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, New York, 1985.

- [Lig09] T. Liggett. Distributional limits for the symmetric exclusion process. *Stoch. Proc. Appl.*, 119:1–15, 2009.
- [LS81] E. Lieb and A. Sokal. A general Lee-Yang theorem for one-component and multicomponent ferromagnets. *Comm. Math. Phys.*, 80:153–179, 1981.
- [LS04] E. Lieb and R. Seiringer. Equivalent forms of the Bessis-Moussa-Villani conjecture. *J. Stat. Phys.*, 115:185–190, 2004.
- [LY52] T.D. Lee and C.N. Yang. Statistical theory of equations of state and phase transitions, II: lattice gas and Ising model. *Phys. Rev.*, 87:410–419, 1952.
- [Lyo03] R. Lyons. Determinantal probability measures. *IHES*, 98:167–212, 2003.
- [MS10] P. McNamara and B. Sagan. Infinite log-concavity: developments and conjectures. *Adv. Appl. Math.*, 44:1–15, 2010.
- [Nij76] N. Nijenhuis. On permanents and the zeros of rook polynomials. *J. Comb. Theory, ser. A*, 21:240–244, 1976.
- [NN93] Yu. Nesterov and A. Nemirovskii. *Interior point polynomial methods in convex programming*, volume 13 of *SIAM Studies in Applied Mathematics*. SIAM Publications, Philadelphia, PA, 1993.
- [Pem91] R. Pemantle. Choosing a spanning tree for the integer lattice uniformly. *Ann. Probab.*, 19:1559–1574, 1991.
- [Pem00] R. Pemantle. Toward a theory of negative dependence. *J. Math. Phys.*, 41:1371–1390, 2000.
- [Pet45] I. G. Petrovsky. On the diffusion of waves and the lacunas for hyperbolic equations. *Mat. Sb.*, 17:289–370, 1945.
- [PP11] R. Pemantle and Y. Peres. Concentration of Lipschitz functionals of determinantal and other strong Rayleigh measures. *arxiv*, 2011.
- [PS14] G. Pólya and J. Schur. Über zwei Arten von Faktorenfolgen in der Theorie der algebraischen Gleichungen. *J. Reine Angew. Math.*, 144:89–113, 1914.
- [PS97] A. Panconesi and A. Srinivasan. Randomized distributed edge coloring via an extension of the Chernoff–Hoeffding bounds. *SIAM J. Comput.*, 26:350–368, 1997.
- [PW02] R. Pemantle and M.C. Wilson. Asymptotics of multivariate sequences. I. Smooth points of the singular variety. *J. Combin. Theory Ser. A*, 97(1):129–161, 2002.
- [PW04] R. Pemantle and M.C. Wilson. Asymptotics of multivariate sequences, II. Multiple points of the singular variety. *Combin. Probab. Comput.*, 13:735–761, 2004.

- [PW08] R. Pemantle and M.C. Wilson. Twenty combinatorial examples of asymptotics derived from multivariate generating functions. *SIAM Review*, 50:199–272, 2008.
- [RS01] G. Royle and A. Sokal. The Brown-Colbourn conjecture on zeros of reliability polynomials is false. *J. Comb. Theory, ser. B*, 94:345–360, 2001.
- [RW08] Alexander Raichev and Mark C. Wilson. A new approach to asymptotics of Maclaurin coefficients of algebraic functions. Report CDMTCS-322, Centre for Discrete Mathematics and Theoretical Computer Science, University of Auckland, New Zealand, April 2008. <http://www.cs.auckland.ac.nz/CDMTCS/researchreports/322alexmcw.pdf>.
- [Saf00] K. Safonov. On power series of algebraic and rational functions. *J. Math. Anal. Appl.*, 243(2):261–277, 2000.
- [Sch47] I. Schoenberg. On totally positive functions, Laplace integrals, and entire functions of the Laguerre–Pólya–Schur type. *Proc. Nat. Acad. Sci. USA*, 33:11–17, 1947.
- [Sch98] A. Schrijver. Counting 1-factors in regular bipartite graphs. *J. Comb. Theory, ser. B*, 72:122–135, 1998.
- [Sok01] A. Sokal. Bounds on the complex zeros of (di)chromatic polynomials and Potts model partition functions. *Comb. Prob. Comput.*, 10:41–77, 2001.
- [Sok05] A. Sokal. The multivariate Tutte polynomial (alias Potts model) for graphs and matroids. In *Surveys in Combinatorics*, volume 327 of *London Math. Soc. Lecture Note Series*, pages 173–226. Cambridge University Press, Cambridge, 2005.
- [Sok11] A. Sokal. Roots of a formal power series, with applications to graph enumeration and q -series, lectures at Queen Mary University of London. <http://www.maths.qmul.ac.uk/~pjc/csgnotes/sokal/>, March–April 2011.
- [Sta89] R. Stanley. Log-concave and unimodal sequences in algebra, combinatorics and geometry. *Annals of the New York Academy of Sciences*, 576:500–535, 1989.
- [Sta11] H. Stahl. Proof of the BMV conjecture. *arXiv*, <http://front.math.ucdavis.edu/1107.4875>: 56 pages, 2011.
- [SV80] A. Schrijver and W. Valiant. On lower bounds for permanents. *Indag. Math.*, 42:425–427, 1980.
- [Sve68] L. Svensson. Necessary and sufficient conditions for the hyperbolicity of polynomials with hyperbolic principal part. *Ark. Mat.*, 8:145–162, 1968.
- [Wag00] D. G. Wagner. Zeros of reliability polynomials and f -vectors of matroids. *Comb. Prob. Comput.*, 9:167–190, 2000.

- [Wag06] D. G. Wagner. Matroid inequalities from electrical network theory. *Elec. J. Combin.*, 11:17, 2004/2006.
- [Wag08] D. G. Wagner. Negatively correlated random variables and mason's conjecture for independent sets in matroids. *Ann. Comb.*, 12:211–239, 2008.
- [Wag11] D. G. Wagner. Multivariate stable polynomials: theory and application. *Bull. Amer. Math. Soc.*, 48:53–84, 2011.
- [Wid41] D. Widder. The laplace transform. page 406, 1941.
- [Wu88] F. Wu. Potts model and graph theory. *J. Stat. Phys.*, 52:99–112, 1988.
- [WW09] D. Wagner and Y. Wei. A criterion for the half-plane property. *Disc. Math.*, 309:1385–1390, 2009.
- [YL52] C.N. Yang and T.D. Lee. Statistical theory of equations of state and phase transitions, I: Theory of condensation. *Phys. Rev.*, 87:404–409, 1952.