

Exponential Varieties

Mateusz Michałek, Bernd Sturmfels,
Caroline Uhler, and Piotr Zwiernik

Abstract

Exponential varieties arise from exponential families in statistics. These real algebraic varieties have strong positivity and convexity properties, familiar from toric varieties and their moment maps. Among them are varieties of inverses of symmetric matrices satisfying linear constraints. This class includes Gaussian graphical models. We develop a general theory of exponential varieties. These are derived from hyperbolic polynomials and their integral representations. We compare the multidegrees and ML degrees of the gradient map for hyperbolic polynomials.

1 Introduction

An exponential variety is a real algebraic variety in an affine or projective space that has a distinguished set of positive points. This positive part maps bijectively to a convex body. The inverse to this map is an algebraic function from the convex body to the variety.

This situation is familiar to geometers from the theory of toric varieties. The convex body is a polytope or polyhedral cone which is in bijection with the positive part of the toric variety via the moment map. The inverse of the moment map is an algebraic function whose degree is the Euler characteristic of an associated hypersurface in the torus [23, Thm. 3.2].

A similar situation happens for the image \mathcal{L}^{-1} of a linear subspace $\mathcal{L} \subset \mathbb{R}^n$ under the coordinatewise inversion map $(x_1, \dots, x_n) \mapsto (x_1^{-1}, \dots, x_n^{-1})$. Suppose the convex polyhedron $\mathcal{L}_{>0} = \mathbb{R}_{>0}^n \cap \mathcal{L}$ is non-empty. The *reciprocal linear space* \mathcal{L}^{-1} is a subvariety of \mathbb{R}^n whose positive part $\mathcal{L}_{>0}^{-1}$ admits a canonical bijection to the polyhedron dual to $\mathcal{L}_{>0}$. By composing with the inversion map $\mathcal{L} \dashrightarrow \mathcal{L}^{-1}$, and by passing from \mathbb{R}^n to the projective space \mathbb{RP}^{n-1} , one obtains a canonical bijection between the interior of any polytope and its dual; see Example 3.2. In [36, Example 3.5], this process was described as “how to morph a cube into an octahedron”.

In this paper we present a general theory that explains these examples and many more. Our point of departure is the theory of *exponential families* in statistics [6]. Exponential varieties arise from exponential families. Known results about the latter explain the remarkable positivity properties of the former. In Section 2 we review basics on exponential families, such as the bijection between the canonical parameters and the sufficient statistics (Theorem 2.2). Two key examples are the multivariate Gaussian family and the full discrete exponential family. New points in Section 2 are the connection to barrier functions in convex

optimization (cf. [17]) and our definition of *rational exponential families*. These statistical models have the property that the gradient of their *log-partition* function is given by rational functions, a property much stronger than that required by Drton and Sullivant [12] for *algebraic exponential families*.

In Section 3 we define *hyperbolic exponential families*. These are associated with hyperbolic polynomials and their hyperbolicity cones [18, 26]. Work of Scott and Sokal [34] implies that every exponential family whose canonical parameters form a convex cone and whose partition function is the power of a homogeneous polynomial must be hyperbolic (Theorem 3.3). We conjecture that the converse holds as well, namely that one can build a (statistical) exponential family, i.e. with an underlying measure, from any hyperbolic polynomial (Conjecture 3.5).

The formal definition of *exponential varieties* appears in Section 4. They arise by fixing a linear subspace \mathcal{L} in the space of canonical parameters. The variety is the Zariski closure of the image of \mathcal{L} under the gradient map of the log-partition function. In the discrete case, \mathcal{L} has to be defined over \mathbb{Q} in order to get an algebraic object. We focus on the hyperbolic case, where every linear subspace that meets the parameter cone defines an exponential variety. This includes the Gaussian case, and we recover the *linear concentration models* of [1, 36].

In Section 5 we examine the gradient multidegree of a hyperbolic exponential family. This is the cohomology class of the graph of the gradient map. Its coefficients characterize the degrees of the exponential varieties defined by generic linear subspaces \mathcal{L} . Theorem 5.5 furnishes inequalities relating these degrees and the ML degree for special subspaces \mathcal{L} . Elementary symmetric form an important family of hyperbolic polynomials and Section 6 offers a detailed study of the exponential varieties that they define.

Section 7 is devoted to Hankel models. We show that the Grassmannian of lines, in its Plücker embedding, has the structure of an exponential variety. This arises from the Gaussian case by taking the linear space \mathcal{L} of all Hankel matrices. The extension to (generalized) Hankel matrices of multivariate polynomials is related to optimization via sums of squares [4].

We close the introduction with a running example that will illustrate the general theory.

Example 1.1. Consider the elementary symmetric polynomials in four parameters:

$$\begin{aligned} E_3(\theta) &= \theta_1\theta_2\theta_3 + \theta_1\theta_2\theta_4 + \theta_1\theta_3\theta_4 + \theta_2\theta_3\theta_4, \\ E_2(\theta) &= \theta_1\theta_2 + \theta_1\theta_3 + \theta_1\theta_4 + \theta_2\theta_3 + \theta_2\theta_4 + \theta_3\theta_4, \\ E_1(\theta) &= \theta_1 + \theta_2 + \theta_3 + \theta_4. \end{aligned}$$

The polynomial E_3 is hyperbolic with respect to $C = \{\theta \in \mathbb{R}^4 : E_i(\theta) > 0 \text{ for } i = 1, 2, 3\}$. This set C is the convex cone over the 3-dimensional body known as the *elliptope*, and shown on the left in Figure 1. The closed dual cone $K = C^\vee$ has as its base the convex hull of the *Steiner surface*, shown in the middle of Figure 1, which is the zero set of the quartic

$$Q = \sum \sigma_i^4 - 4 \sum \sigma_i^3 \sigma_j + 6 \sum \sigma_i^2 \sigma_j^2 + 4 \sum \sigma_i^2 \sigma_j \sigma_k - 40 \sigma_1 \sigma_2 \sigma_3 \sigma_4, \quad (1)$$

where $\sigma_1, \sigma_2, \sigma_3, \sigma_4$ are coordinates for the dual \mathbb{R}^4 .

Here, our algebraic exponential family is defined by the gradient map of the cubic E_3 :

$$\nabla E_3 : \mathbb{R}^4 \rightarrow \mathbb{R}^4, \theta \mapsto (\theta_2\theta_3 + \theta_2\theta_4 + \theta_3\theta_4, \theta_1\theta_3 + \theta_1\theta_4 + \theta_3\theta_4, \theta_1\theta_2 + \theta_1\theta_4 + \theta_2\theta_4, \theta_1\theta_2 + \theta_1\theta_3 + \theta_2\theta_3).$$

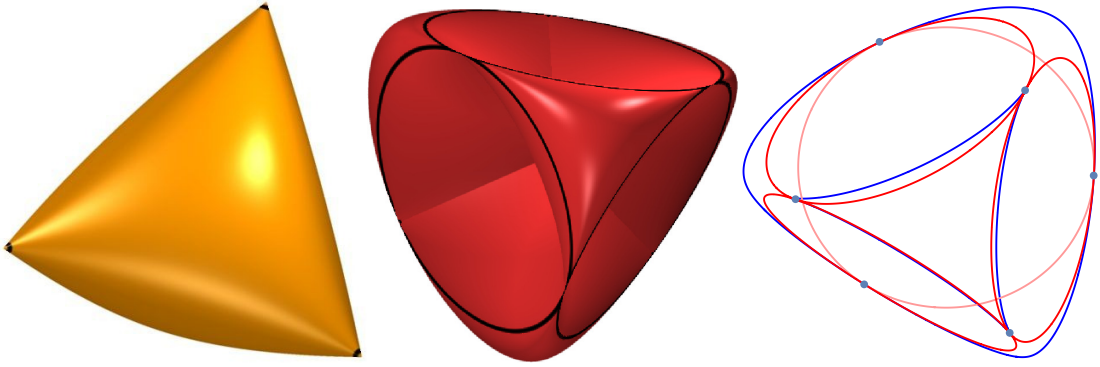


Figure 1: The gradient of the elementary symmetric polynomial E_3 gives a bijection between the convex bodies of canonical parameters C (left) and of sufficient statistics K (middle). The sextic branch curve of a projection of K in the plane touches the four ellipses (right).

This map restricts to a bijection between C and the interior of K . The degree of ∇E_3 is four, i.e. generic points $\sigma \in \mathbb{CP}^3$ have four preimages under the induced map $\nabla E_3 : \mathbb{CP}^3 \dashrightarrow \mathbb{CP}^3$. But if σ lies in the interior of K , then precisely one of these preimages lies in $C = \text{int}(K^\vee)$.

The map ∇E_3 transforms the boundary of C into the boundary of K as follows. The four circles in ∂K are created by blowing up the four vertices of ∂C , and the six points where they touch are obtained by blowing down the six edges on ∂C . The tetrahedron spanned by the vertices of C is mapped to the octahedron spanned by these six touching points in ∂K .

We now fix a hyperplane \mathcal{L} in \mathbb{R}^4 that intersects the interior of C ; see Figure 2. The inclusion of $\mathcal{L} \simeq \mathbb{R}^3$ in \mathbb{R}^4 is dual to a projection $\pi_{\mathcal{L}} : \mathbb{R}^4 \rightarrow \mathbb{R}^3$. The image of the cone K under this map is dual to that seen in C by intersecting with \mathcal{L} , i.e. $\pi_{\mathcal{L}}(K) = (\mathcal{L} \cap C)^\vee$. In the picture, we are cutting the body C by a plane, and dually we project K into a plane.

The image of \mathcal{L} under ∇E_3 is a quartic surface in \mathbb{CP}^3 ; see Figure 4. That complex surface is our *exponential variety*, denoted $\mathcal{L}^{\nabla E_3}$. The intersection curve $\{Q = 0\} \cap \mathcal{L}^{\nabla E_3}$ decomposes into two components having degrees 6 and 10. The sextic curve is smooth when \mathcal{L} is generic. It is the ramification locus of the Steiner surface $\{Q = 0\}$ under the map $\pi_{\mathcal{L}} : \mathbb{CP}^3 \dashrightarrow \mathbb{CP}^2$.

We now come to the punchline. The exponential variety $\mathcal{L}^{\nabla E_3}$ has a distinguished positive part, namely the semialgebraic set $\mathcal{L}_{>0}^{\nabla E_3} = \mathcal{L}^{\nabla E_3} \cap \text{int}(K)$. This comes with natural bijections

$$\mathcal{L} \cap C \longleftrightarrow \mathcal{L}_{>0}^{\nabla E_3} \longleftrightarrow \text{int}(\pi_{\mathcal{L}}(K)),$$

to be derived in Theorem 4.5. The intersection of the left body C with a plane in Figure 1 maps bijectively to a quartic surface inside the middle body K . That surface maps bijectively onto the two-dimensional convex set shown on the right in Figure 1. The branch curve is a plane curve of degree six, namely the image of the ramification sextic. The four ellipses in Figure 1 (right) are the projections of the circles in Figure 1 (middle). The six points are the projection of the vertices of the octahedron. We shall return to this in Example 6.8 and in Figure 2. \diamond

2 Exponential Families

A statistical model is a family of probability distribution functions or probability density functions on a sample space (\mathcal{X}, ν, T) . Here ν is the base measure on the set \mathcal{X} , and $T : \mathcal{X} \rightarrow \mathbb{R}^d$ is a measurable function that represents the *sufficient statistics*. We fix an inner product $\langle \cdot, \cdot \rangle$ on \mathbb{R}^d . In the standard basis, $\langle \theta, \sigma \rangle = \theta^T S \sigma$ for some positive definite symmetric matrix S . An *exponential family* is a parametric statistical model on (\mathcal{X}, ν, T) of the form

$$p_\theta(x) = h(x) \cdot \exp(\langle \theta, T(x) \rangle - A(\theta)).$$

Here $h : \mathcal{X} \rightarrow \mathbb{R}$ is a weight function which serves to modify the given base measure ν . Without loss of generality, we can replace ν by $h(x) \cdot \nu$. Afterwards we have $h(x) \equiv 1$. To simplify notation in later sections, the following non-standard formulation will now be used. We replace $T(x)$ by $-T(x)$ and write all our exponential families in the form

$$p_\theta(x) = \exp(-\langle \theta, T(x) \rangle - A(\theta)). \quad (2)$$

The sample space \mathcal{X} itself is less important than its image $T(\mathcal{X})$ in \mathbb{R}^d . The reason is that in (2) we see the samples $x \in \mathcal{X}$ only through their sufficient statistics $T(x)$. So, since ν induces a measure on $T(\mathcal{X})$, we might as well regard the set $T(\mathcal{X})$ as our space of data.

The function $A(\theta)$ is known as the *log-partition function*, or as the *cumulant generating function*. It is uniquely determined by (\mathcal{X}, ν, T) via the requirement that $\int_{\mathcal{X}} p_\theta(x) \nu(dx) = 1$. This equation implies the following integral representation of the log-partition function:

$$A(\theta) = \log \int_{\mathcal{X}} \exp(-\langle \theta, T(x) \rangle) \nu(dx). \quad (3)$$

The exponential of $A(\theta)$ is the *partition function* of the exponential family.

We are interested in the set of all parameters θ for which this integral is finite:

$$C := \{\theta \in \mathbb{R}^d : A(\theta) < +\infty\}. \quad (4)$$

By [6, Theorem 1.13], the set C is convex in \mathbb{R}^d . The function $A : C \rightarrow \mathbb{R}$ is convex, and it is analytic on the interior of C . The convex set C is the *space of canonical parameters*. An exponential family is said to be *regular* if C is open in \mathbb{R}^d . It is called *minimal* if $T(\mathcal{X})$ spans the whole \mathbb{R}^d . See [6, Chapter 1] for standard definitions and basic properties of exponential families. We restrict our attention to exponential families that are minimal and regular.

Let $DA(\theta)$ be the differential of $A(\theta)$. We view this as a linear form on \mathbb{R}^d . As such it lives in the dual \mathbb{R}^d with coordinates $\sigma = (\sigma_1, \dots, \sigma_d)$ and plays a crucial role in this paper.

Definition 2.1 (*Gradient map*). Let $\theta^* \in \mathbb{R}^d$ be the unique vector such that $(DA(\theta))(\eta) = \langle \eta, -\theta^* \rangle$ for all $\eta \in \mathbb{R}^d$. The map $F : \mathbb{R}^d \rightarrow \mathbb{R}^d$, $\theta \mapsto \theta^*$ is called the *gradient map* of $A(\theta)$.

If $\langle \cdot, \cdot \rangle$ is the standard scalar product then $\theta^* = -\nabla A(\theta)$, which explains the name. Formally, the gradient map maps \mathbb{R}^d to its dual. The left \mathbb{R}^d has coordinates θ , the right \mathbb{R}^d

has coordinates σ , and the pairing is the inner product $\langle \theta, \sigma \rangle$. The sufficient statistics $T(x)$ are elements of (\mathbb{R}^d, σ) . In that space we consider the convex hull of all sufficient statistics:

$$K := \text{conv}(T(\mathcal{X})) \subset \mathbb{R}^d. \quad (5)$$

The following result is standard in the theory of exponential families (see [6, Theorem 3.6]).

Theorem 2.2. *Let (\mathcal{X}, ν, T) be an exponential family that is minimal and regular. Then the gradient map F defines an analytic bijection from C to the interior of K .*

For extra clarity, let us paraphrase: the space C of canonical parameters is a convex set in the primal (\mathbb{R}^d, θ) , while the space K of sufficient statistics is a convex set in the dual (\mathbb{R}^d, σ) . The gradient of the log-partition function defines a bijection between C and $\text{int}(K)$. This paper concerns situations when this bijection has desirable algebraic properties.

Definition 2.3 (*Rational exponential family*). We say that an exponential family (\mathcal{X}, ν, T) is *rational* if the gradient map F is a rational map $\mathbb{R}^d \dashrightarrow \mathbb{R}^d$.

The case of primary interest to us arises when the partition function is a negative power of a homogeneous polynomial f in the unknowns $\theta = (\theta_1, \dots, \theta_d)$. Suppose that $\langle \cdot, \cdot \rangle$ is the standard scalar product. In that case we have $A(\theta) = -\alpha \log f(\theta)$, where $\alpha > 0$ is a constant, and the gradient map is the rational function

$$F : \mathbb{R}^d \dashrightarrow \mathbb{R}^d : \theta \mapsto \frac{\alpha}{f(\theta)} \cdot \left(\frac{\partial f}{\partial \theta_1}, \frac{\partial f}{\partial \theta_2}, \dots, \frac{\partial f}{\partial \theta_d} \right). \quad (6)$$

For an algebraic geometer it is more natural to view this as a map of projective spaces:

$$F : \mathbb{CP}^{d-1} \dashrightarrow \mathbb{CP}^{d-1} : \theta \mapsto \left(\frac{\partial f}{\partial \theta_1} : \frac{\partial f}{\partial \theta_2} : \dots : \frac{\partial f}{\partial \theta_d} \right). \quad (7)$$

Example 2.4. If $d = 4$ and $f = E_3$ is the third elementary symmetric polynomial, then this is Example 1.1. Here C and K are dual convex cones in \mathbb{R}^4 , visualized in Figure 1 by their 3-dimensional cross sections. These live in the real part \mathbb{RP}^3 of \mathbb{CP}^3 . Theorem 2.2 tells us that the map (7) restricts to a bijection between the interiors of the two convex bodies. \diamond

We next present two models that are fundamental in statistics and its applications.

Example 2.5 (*Multivariate Gaussian family*). Let $\mathcal{X} = \mathbb{R}^m$, where ν is the Lebesgue measure, and set $T(x) = \frac{1}{2}xx^T \in \mathbb{R}^{m \times m}$. Here $d = \binom{m+1}{2}$ and we identify \mathbb{R}^d with the space \mathbb{S}^m of real symmetric $m \times m$ matrices with inner product $\langle A, B \rangle = \text{tr}(AB)$. So, the sufficient statistics are symmetric matrices of rank ≤ 1 . For this exponential family, the convex set C in (4) is the cone of positive definite matrices in \mathbb{S}^m , which we denote by PD_m . This cone is open and selfdual. The set K in (5) is its closure — the cone of positive semidefinite matrices in \mathbb{S}^m . The integral in (3) is the standard multivariate Gaussian integral. We find

$$A(\theta) = -\frac{1}{2} \log \det(\theta) + \frac{m}{2} \log(2\pi).$$

Hence the multivariate Gaussian distribution on \mathbb{R}^m forms a rational exponential family. Its defining polynomial $f(\theta)$ is the determinant of a symmetric $m \times m$ -matrix, and we have $\alpha = 1/2$. The gradient map F in (6) is $\theta \mapsto \frac{1}{2}\theta^{-1}$. Theorem 2.2 says that matrix inversion is an involution on the open cone $C = \text{int}(K)$ of positive definite matrices in \mathbb{S}^m . \diamond

Example 2.6 (*Full discrete family*). We consider discrete random variables with finite state space $\mathcal{X} = \{1, 2, \dots, d\}$. Here ν is the counting measure on \mathcal{X} , and the sufficient statistics are the unit vectors $T(i) = \mathbf{e}_i$ in \mathbb{R}^d . The probability distribution function (2) has the form

$$p_\theta(i) = \exp(-\theta^T \cdot T(i) - A(\theta)) = \frac{e^{-\theta_i}}{e^{A(\theta)}}.$$

The log-partition function equals

$$A(\theta) = \log(e^{-\theta_1} + e^{-\theta_2} + \dots + e^{-\theta_d}). \quad (8)$$

The gradient map F satisfies

$$F(\theta) = -\nabla A(\theta) = \frac{1}{\sum_{i=1}^d e^{-\theta_i}} \cdot (e^{-\theta_1}, e^{-\theta_2}, \dots, e^{-\theta_d}). \quad (9)$$

The convex set in (4) is the entire space, $C = \mathbb{R}^d$, so this family is regular. The space of sufficient statistics, defined in (5), is the closed $(d-1)$ -dimensional probability simplex

$$K = \text{conv}(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_d) = \{\sigma \in \mathbb{R}_{\geq 0}^d : \sigma_1 + \sigma_2 + \dots + \sigma_d = 1\}.$$

The gradient map takes C to the interior of K . But this is not a bijection. Theorem 2.2 does not hold here because this exponential family is not minimal. Following [6, Example 1.3], we can turn this into a *minimal* exponential family by replacing d with $d-1$ and projecting $T(x)$ on the first $d-1$ coordinates. Now $C = \mathbb{R}^{d-1}$, the formulas (8) and (9) hold with $\theta_d = 0$, and the gradient map (9) gives a bijection between C and the open simplex $\text{int}(K)$.

The gradient map is analytic but not algebraic, so this is not a rational exponential family. This can be fixed by replacing the parameter space $C = \mathbb{R}^{d-1}$ with the open orthant $\mathbb{R}_{>0}^{d-1}$, by way of the bijection $r : \theta \mapsto (e^{-\theta_1}, \dots, e^{-\theta_{d-1}})$. The gradient map is the composition of r and a rational function. This will allow us in Example 4.3 to obtain *discrete exponential families* from lattices \mathcal{L} . As noted in [12, 23, 32] and elsewhere, these are precisely *toric varieties*. \diamond

We close this section by pointing out a connection between statistics and optimization which seems to have not yet been stated explicitly in the mathematical literature.

Remark 2.7. In his 1997 paper [17], Güler presented a framework for *interior point methods* in convex optimization based on canonical barrier functions. This theory can be viewed as a special case of exponential family theory. The following discussion explains how this works.

Fix a d -dimensional pointed convex cone K in \mathbb{R}^d , and let ν be the Lebesgue measure on K . To match the notation above, set $\mathcal{X} = K$ and let T be the identity on K . Then, in the setting of [17, Section 3], the log-partition function $A(\theta)$ in (3) becomes the logarithm

of the *characteristic function* of the dual cone $K^\vee = \overline{C}$. This is, up to an additive constant term $\log d!$, the *universal barrier function* of C :

$$A(\theta) = \log \int_K \exp(-\langle \theta, \sigma \rangle) d\sigma. \quad (10)$$

For example, let C be the positive definite cone PD_m . In Güler's theory [17, 18], the integral (10) corresponds to interior point methods in *semidefinite programming* and the induced rational exponential family is precisely that in Example 2.5. In other words, the multivariate Gaussian family comes from the universal barrier function of PD_m .

Our gradient map F is known as the *duality mapping* in convex optimization [17, Section 5]. That it defines a bijection between the interiors of C and K is shown in [17, Theorem 5.1]. This is a special case of what statisticians know about exponential families (Theorem 2.2).

The optimization problem of most interest in statistics is that of finding the maximum likelihood estimator (MLE). For an exponential family the solution to this problem can be described as follows. Suppose we are given n independent observations x_1, \dots, x_n in \mathcal{X} . We record these observations by the average value of their sufficient statistics. This is the point

$$\hat{\sigma} = \frac{1}{n}(T(x_1) + T(x_2) + \dots + T(x_n)) \in K.$$

The MLE is defined as the canonical parameter vector $\hat{\theta}$ that makes the data x_1, \dots, x_n most likely. By [6, Theorem 5.5], the MLE exists and is unique precisely when $\hat{\sigma}$ lies in the interior of K . If this holds, then $\hat{\theta}$ is the unique inverse image of $\hat{\sigma}$ in C under the gradient map.

In other words, computing the MLE for an exponential family means intersecting the fiber $F^{-1}(\hat{\sigma})$ of the gradient map F with the convex set C . For a rational exponential family we can study this fiber and its degree using methods from algebraic geometry. This was pioneered in [36] for the special case of linear Gaussian concentration models. It will be developed in full generality in this paper. The pertinent algebraic complexity measures, namely the gradient multidegree and the maximum likelihood degree (ML degree), will be introduced in Section 5.

3 Hyperbolic Polynomials and Riesz Kernels

In this section, we introduce a new class of rational exponential families associated with hyperbolic polynomials and their hyperbolicity cones. Recall (e.g. from [18, 26]) that a homogeneous polynomial f of degree p in $\theta = (\theta_1, \dots, \theta_d)$ is *hyperbolic* if there exists a point τ in $\mathbb{R}^d \setminus \{f = 0\}$ such that every line through τ meets the hypersurface $\{f = 0\}$ in p real points, counting multiplicities. Assume that this holds and let C be the connected component of $\mathbb{R}^d \setminus \{f = 0\}$ that contains τ . It is known that C is an open convex cone in \mathbb{R}^d . This cone is called the *hyperbolicity cone* of f . As seen in Example 3.2, it may depend on τ .

Example 3.1. The standard example of a hyperbolic polynomial is the determinant of a symmetric $m \times m$ -matrix of unknowns. Here $d = \binom{m+1}{2}$, the hyperbolicity cone C consists

of the positive definite matrices in \mathbb{S}^m , where the witness point τ can be taken to be the identity matrix. Hence, we recover the Gaussian family in Example 2.5. \diamond

Example 3.2. Let C be a convex polyhedral cone in \mathbb{R}^d , defined by s linear inequalities $\ell_1(\theta) > 0, \dots, \ell_s(\theta) > 0$. The product of linear forms $f(\theta) = \prod_{i=1}^s \ell_i(\theta)$ is a hyperbolic polynomial. The hyperbolicity cone C can here be replaced with any of the other chambers in the arrangement of hyperplanes $\{\ell_j = 0\}_{j=1, \dots, s}$. In the setting of [17, 18], this models interior point methods for *linear programming*, with barrier function $A(\theta) = -\sum_{j=1}^s \log(\ell_j(\theta))$. Its gradient map defines a bijection between C and the interior of the dual cone $K = C^\vee$. \diamond

The following recent result by Scott and Sokal [34] motivates studying hyperbolic polynomials in the context of exponential families. It shows that any homogeneous polynomial that is (up to a negative power) the partition function of an exponential family must be hyperbolic.

Theorem 3.3. *Let f be a homogeneous polynomial in $\mathbb{R}[\theta_1, \dots, \theta_d]$ that is strictly positive on an open convex cone C . If there exists a positive number α and a measure ν such that*

$$f(\theta)^{-\alpha} = \int_{C^\vee} \exp(-\langle \theta, \sigma \rangle) \nu(d\sigma) \quad \text{for all } \theta \in C,$$

then f is hyperbolic with respect to each point in the cone C .

Proof. A real analytic function g is *completely monotone* on a convex cone C if and only if

$$(-1)^k \frac{\partial^k g}{\partial_{u_1} \dots \partial_{u_k}} \geq 0 \quad \text{on } C \quad \text{for all } k \in \mathbb{N} \text{ and } u_1, \dots, u_k \in C. \quad (11)$$

There is a beautiful characterization of completely monotone functions, known as the Bernstein-Hausdorff-Widder-Choquet Theorem [8, 34]. Namely, an analytic function g defined on C is completely monotone if and only if there exists a measure μ on the dual cone C^\vee such that:

$$g(\theta) = \int_{C^\vee} \exp(-\langle \theta, \sigma \rangle) \mu(d\sigma).$$

By our assumptions it follows that the function $f(\theta)^{-\alpha}$ is completely monotone on the convex cone C . Then, by [34, Corollary 2.3], we can conclude that f has no zeros in the subset $C + i\mathbb{R}^d$ of \mathbb{C}^d . This means that the polynomial f is hyperbolic with respect to each point τ in C . \square

The following theorem, essentially due to Gårding [14], suggests a construction of exponential families from complete hyperbolic polynomials. A hyperbolic polynomial is called *complete* if the hyperbolicity cone C is pointed, that is, $C \cap (-C) = \{\mathbf{0}\}$ in \mathbb{R}^d . In what follows, both $d\eta$ and $d\sigma$ denote the Lebesgue measure on \mathbb{R}^d . The appearance of the non-real scalar $i = \sqrt{-1}$ may look puzzling at first, but it is essential for Gårding's theory.

Theorem 3.4. *Let f be a complete hyperbolic polynomial with hyperbolicity cone $C \subset \mathbb{R}^d$, and fix $\alpha > d$. The following integral converges for any $\theta \in C$, its value does not depend on the choice of θ , and it is supported on $K = C^\vee$:*

$$q_\alpha(\sigma) = (2\pi)^{-d} \int_{\mathbb{R}^d} f(\theta + i\eta)^{-\alpha} \exp(\langle \theta + i\eta, \sigma \rangle) d\eta. \quad (12)$$

The polynomial f can be recovered via the formula

$$f(\theta)^{-\alpha} = \int_K \exp(-\langle \theta, \sigma \rangle) q_\alpha(\sigma) d\sigma \quad \text{for all } \theta \in C. \quad (13)$$

Proof. It follows directly from Theorem 3.1 in Gårding's paper [14] that the formula $f(\theta)^{-\alpha} = \int_K \exp(-\langle \theta, \sigma \rangle) q_\alpha(\sigma) d\sigma$ holds for all $\theta \in C$, that $q_\alpha(\sigma)$ does not depend on the choice $\theta \in C$, and that $q_\alpha(\sigma)$ vanishes for $\sigma \notin K$. In general, the expression for $q_\alpha(\sigma)$ may not be a well-defined function. However, it is well-defined if $\alpha > d$. See also [18, Theorem 6.4]. As we shall see later, the condition $\alpha > d$ is only sufficient but not necessary. \square

The function $q_\alpha(\sigma)$ in Theorem 3.4 is known as the *Riesz kernel*. It is uniquely determined (up to a set with Lebesgue measure zero) by the integral representation (12). If $q_\alpha \geq 0$, then $\nu_\alpha(d\sigma) = q_\alpha(\sigma)d\sigma$ defines a measure on K and allows the construction of *hyperbolic exponential families*. Various authors suggested that q_α is nonnegative, see e.g. [18, p 371]. However, we are not aware of any proof of this fact and we therefore state this as a conjecture:

Conjecture 3.5. *For $\alpha \gg 0$ the Riesz kernel $q_\alpha(\sigma)$ in (12) takes nonnegative values on K .*

In the following we slightly modify the terminology relative to what is standard in statistics. Namely, we define hyperbolic exponential families with respect to a possibly signed measure:

Definition 3.6 (*(Hyperbolic exponential families)*). Let $f \in \mathbb{R}[\theta_1, \dots, \theta_d]$ be a complete hyperbolic polynomial with hyperbolicity cone $C \subset \mathbb{R}^d$. The corresponding *hyperbolic exponential family* (\mathcal{X}, ν, T) is a rational exponential family with canonical parameter space C , sample space $\mathcal{X} = C^\vee$, sufficient statistics $T(x) = x$, and ν given by the Riesz kernel (12).

Remark 3.7. *It is known from optimization theory [31] that the bijection between C and the interior of $K = C^\vee$ described in Theorem 2.2 holds for complete hyperbolic polynomials f with hyperbolicity cone $C \subset \mathbb{R}^d$. Namely, the gradient function $F(\theta) = -\log f(\theta)$ defines a bijection between C and the interior of $K = C^\vee$. This follows from the fact that $F(\theta)$ is a self-concordant barrier function of C and does not require non-negativity of the Riesz kernel; see [18, Section 6] for details. This justifies our slight misuse of vocabulary in the above definition.*

Our introduction of hyperbolic exponential families appears to be a novel contribution to statistics. In order for these to become useful for data analysis, it is essential to gain a better understanding of the Riesz kernel. In general, it is a hard problem to compute $q_\alpha(\sigma)$ in an explicit form. In the remainder of this section, we discuss some concrete examples, starting with the Gaussian family in Example 3.1. These are situations where the hyperbolic

polynomial f has a symmetric determinantal representation. In this case, the kernel $q_\alpha(\sigma)$ is closely related to the *Wishart distribution* on the positive definite cone PD_m , i.e. the sampling distribution for covariance matrices in the Gaussian family (see the proof of Proposition 3.8 below). We refer to the article of Scott and Sokal [34] for further examples, many results, and open problems.

Proposition 3.8. *Let f be the determinant of a symmetric $m \times m$ -matrix θ and $C = \text{PD}_m$ the cone of positive definite matrices. For $\alpha > \frac{m-1}{2}$, the corresponding Riesz kernel equals*

$$q_\alpha(\sigma) = \frac{1}{\Gamma_m(\alpha)} \det(\sigma)^{\alpha - \frac{m+1}{2}}. \quad (14)$$

The denominator $\Gamma_m(\alpha)$ in equation (14) denotes the *multivariate gamma function*

$$\Gamma_m(\alpha) := \pi^{m(m-1)/4} \cdot \prod_{j=1}^m \Gamma\left(\alpha - \frac{1-j}{2}\right).$$

Proof. We compute the Riesz kernel from the Wishart distribution. This is a probability distribution over the positive definite cone PD_m with two parameters: the degrees of freedom $n \in (m-1, \infty)$ and a scale matrix $\Lambda \in \text{PD}_m$. The density function of the Wishart distribution is

$$\frac{1}{2^{mn/2} \det(\Lambda)^{n/2} \Gamma_m(n/2)} \det(\sigma)^{\frac{n-m-1}{2}} \exp\left(-\frac{1}{2} \text{tr}(\Lambda^{-1} \sigma)\right). \quad (15)$$

We replace the parameters n and Λ with the new parameters $\alpha = n/2$ and $\theta = (2\Lambda)^{-1}$, and we use the fact that the integral of (15) over $\sigma \in \text{PD}_m$ is equal to 1. This implies

$$\det(\theta)^{-\alpha} = \frac{1}{\Gamma_m(\alpha)} \int_{\text{PD}_m} \det(\sigma)^{\alpha - \frac{m+1}{2}} \exp(-\langle \theta, \sigma \rangle) d\sigma \quad \text{for } \alpha > \frac{m-1}{2}, \quad (16)$$

where $\langle \cdot, \cdot \rangle$ denotes the trace inner product on symmetric matrices. By comparing this expression with (13) and by using the uniqueness of this integral representation, we see that the Riesz kernel of $f = \det(\theta)$ for $\alpha > \frac{m-1}{2}$ is given by the right hand side of (14). \square

Using an analog of Proposition 3.8 for Hermitian matrices, one can compute the Riesz kernel of the elementary symmetric polynomial $E_2(\theta)$. This was carried out in [34, Corollary 5.8]. We here go over this for the smallest instance. This is also featured in [34, Proposition 3.8].

Example 3.9. Consider $f = \theta_1\theta_2 + \theta_1\theta_3 + \theta_2\theta_3$ with hyperbolicity cone $C \subset \mathbb{R}^3$. Note that

$$f(\theta) = \det \begin{pmatrix} \theta_1 + \theta_3 & \theta_3 \\ \theta_3 & \theta_2 + \theta_3 \end{pmatrix}.$$

By Proposition 3.8, the following integral representation is valid for $\alpha > 1/2$:

$$f(\theta)^{-\alpha} = \frac{1}{\Gamma_2(\alpha)} \int_{\text{PD}_2} \exp\left(-(\theta_1 + \theta_3)\sigma_{11} - (\theta_2 + \theta_3)\sigma_{22} - 2\theta_3\sigma_{12}\right) (\sigma_{11}\sigma_{22} - \sigma_{12}^2)^{\alpha-3/2} d\sigma.$$

By setting $\sigma_1 = \sigma_{11}$, $\sigma_2 = \sigma_{22}$, and $\sigma_3 = \sigma_{11} + \sigma_{22} + 2\sigma_{12}$, this transforms into an integral over the dual $K = C^\vee$ to the hyperbolicity cone. Namely, we get the integral representation

$$f(\theta)^{-\alpha} = \frac{2^{2-2\alpha}}{\Gamma_2(\alpha)} \int_K \exp(-\theta_1\sigma_1 - \theta_2\sigma_2 - \theta_3\sigma_3) (2(\sigma_1\sigma_2 + \sigma_1\sigma_3 + \sigma_2\sigma_3) - (\sigma_1^2 + \sigma_2^2 + \sigma_3^2))^{\alpha-3/2} d\sigma.$$

By comparing this expression with (13), we see that the Riesz kernel of f is equal to

$$q_\alpha(\sigma) = \frac{2^{2-2\alpha}}{\Gamma_2(\alpha)} (2(\sigma_1\sigma_2 + \sigma_1\sigma_3 + \sigma_2\sigma_3) - (\sigma_1^2 + \sigma_2^2 + \sigma_3^2))^{\alpha-3/2}.$$

Note that the polynomial in the parenthesis vanishes on the boundary of K ; it is the quadric dual to f . See [34, equation (4.16)] for the analogous formula for $E_2(\theta)$ in four variables. \diamond

We next examine $f(\theta) = \theta_1\theta_2 \cdots \theta_m$. In the context of Proposition 3.8, this hyperbolic polynomial is the determinant of a diagonal matrix. Its hyperbolicity cone is the positive orthant $C = \mathbb{R}_{\geq 0}^m$. This cone is self-dual, so that $K = \mathbb{R}_{\geq 0}^m$. We consider the case $\alpha = 1/2$.

Proposition 3.10. *The Riesz kernel for determinants of $m \times m$ diagonal matrices equals*

$$q_{1/2}(\sigma) = \pi^{-m/2} (\sigma_1\sigma_2 \cdots \sigma_m)^{-1/2}. \quad (17)$$

Proof. Let θ denote the diagonal matrix $\text{diag}(\theta_1, \dots, \theta_m)$. We use the fact that the Gaussian density on \mathbb{R}^m with mean $\mathbf{0}$ and inverse covariance matrix θ integrates to 1. This implies

$$\det(\theta)^{-1/2} = \frac{1}{(2\pi)^{m/2}} \int_{\mathbb{R}^m} \exp\left(-\frac{1}{2} x^T \theta x\right) dx = (2/\pi)^{m/2} \int_{\mathbb{R}_{\geq 0}^m} \exp\left(-\frac{1}{2} \sum_{i=1}^m \theta_i x_i^2\right) dx.$$

By making the change of variables $\frac{1}{2}x_i^2 \rightarrow \sigma_i$, we obtain

$$\det(\theta)^{-1/2} = \pi^{-m/2} \int_{\mathbb{R}_{\geq 0}^m} \exp\left(-\sum_{i=1}^m \theta_i \sigma_i\right) (\sigma_1 \cdots \sigma_m)^{-1/2} d\sigma. \quad (18)$$

By comparing this expression with (13), we see that the Riesz kernel equals (17). \square

In principle, the computation of Riesz kernels as in Propositions 3.8 and 3.10 can be restricted to cases when the parameters satisfy some linear constraints. The resulting hyperbolic polynomials $f(\theta)$ admit a symmetric determinantal representation. This requires integrating out the complementary parameters, a task that is very difficult to do. Even the case of products of linear forms (Example 3.2) is challenging, as the following example shows.

Example 3.11. Let $f(\theta) = \theta_1\theta_2(\theta_1 + \theta_2)(\theta_1 - \theta_2)$, and take the hyperbolicity cone to be $C = \{\theta_1 > \theta_2 > 0\}$. Then its dual is $K = \{\sigma_1 \geq 0, \sigma_1 + \sigma_2 \geq 0\}$. By applying (18) we obtain

$$\det(\theta)^{-1/2} = \pi^{-2} \int_{\mathbb{R}_{\geq 0}^4} \exp(-\theta_1 y_1 - \theta_2 y_2 - (\theta_1 + \theta_2)y_3 - (\theta_1 - \theta_2)y_4) (y_1 y_2 y_3 y_4)^{-1/2} dy.$$

By replacing $\sigma_1 = y_1 + y_3 + y_4$ and $\sigma_2 = y_2 + y_3 - y_4$, and defining

$$K' = \left\{ (y_3, y_4) \in \mathbb{R}_{\geq 0}^2 \mid y_3 + y_4 \leq \sigma_1 \text{ and } y_3 - y_4 \leq \sigma_2 \right\},$$

we find

$$\det(\theta)^{-1/2} = \pi^{-2} \int_K \exp(-\theta_1 \sigma_1 - \theta_2 \sigma_2) \left(\int_{K'} ((\sigma_1 - y_3 - y_4)(\sigma_2 - y_3 + y_4)y_3 y_4)^{-1/2} dy \right) d\sigma.$$

Hence the Riesz kernel is

$$q_{1/2}(\sigma) = \pi^{-2} \int_{K'} ((\sigma_1 - y_3 - y_4)(\sigma_2 - y_3 + y_4)y_3 y_4)^{-1/2} dy.$$

The integral over K' can be expressed in terms of generalized hypergeometric functions. Performing this computation requires a case distinction, and we here consider the case when $\sigma_2 \geq \sigma_1$. Then we may substitute $y_3 \rightarrow \sigma_1 v$ and $y_4 \rightarrow \sigma_1 u$ and obtain

$$\begin{aligned} & \int_{K'} ((\sigma_1 - y_3 - y_4)(\sigma_2 - y_3 + y_4)y_3 y_4)^{-1/2} dy \\ &= \sigma_1^{\frac{1}{2}} \sigma_2^{-\frac{1}{2}} \int_0^1 \int_0^{1-v} u^{-1/2} v^{-1/2} (1 - u - v)^{-1/2} \left(1 + \frac{\sigma_1}{\sigma_2} u - \frac{\sigma_1}{\sigma_2} v\right)^{-1/2} du dv \\ &= 2\pi \sigma_1^{\frac{1}{2}} \sigma_2^{-\frac{1}{2}} F_1\left(\frac{1}{2}; \frac{1}{2}, \frac{1}{2}; \frac{3}{2}; -\frac{\sigma_1}{\sigma_2}, \frac{\sigma_1}{\sigma_2}\right), \end{aligned}$$

where F_1 is the first *Appell hypergeometric function* in two unknowns. Here we are using formula (6) in <http://mathworld.wolfram.com/AppellHypergeometricFunction.html>.

A similar analysis can be carried out for the other case $\sigma_1 > \sigma_2$, but this is omitted here. \diamond

Suppose now that $f(\theta)$ is an arbitrary product of linear forms as in Example 3.2, and we wish to find integral representations of $f(\theta)^{-\alpha}$. The Riesz kernel $q_\alpha(\sigma)$ is expressible in terms of Aomoto-Gel'fand hypergeometric functions [2, 38]. The corresponding hypergeometric D-modules should be useful for representing Riesz kernels. More generally, it would be interesting to develop the *holonomic gradient method* [19] for hyperbolic exponential families.

4 Restricting to Linear Subspaces

Many important statistical models arise by restricting a given exponential family (\mathcal{X}, ν, T) to a linear subspace $\mathcal{L} \subset \mathbb{R}^d$ of the canonical parameters. Figure 2 is meant to illustrate this. We shall assume $C \cap \mathcal{L} \neq \emptyset$ and $\dim \mathcal{L} = c$. Fix a basis of \mathcal{L} and let $L \in \mathbb{R}^{c \times d}$ denote a matrix whose rows are the elements of that basis. The map $\pi_{\mathcal{L}} : \mathbb{R}^d \rightarrow \mathbb{R}^d / \mathcal{L}^\perp \simeq \mathbb{R}^c$ is given in coordinates by the matrix L . In what follows we identify $L = \pi_{\mathcal{L}}$. Later on, we often abuse notation and we write \mathcal{L} also for the projectivization of the linear subspace $\mathcal{L} \subset \mathbb{R}^d$. In that context, \mathcal{L} will be a plane of dimension $c - 1$ in \mathbb{RP}^{d-1} or \mathbb{CP}^{d-1} . Note that all results in this section hold when ν is a signed measure and hence apply in particular to hyperbolic exponential families.

For every $\theta \in C \cap \mathcal{L}$ there exists a vector $\tau \in \mathbb{R}^c$ such that $\theta = L^T \tau$, and hence

$$\langle \theta, T(x) \rangle = \langle L^T \tau, T(x) \rangle = \langle \tau, (L \circ T)(x) \rangle. \quad (19)$$

By restricting C to $C \cap \mathcal{L}$ we obtain an exponential family $(\mathcal{X}, \nu, T_{\mathcal{L}})$ with sample space \mathcal{X} , canonical parameter space $C_{\mathcal{L}} = C \cap \mathcal{L}$, sufficient statistics $T_{\mathcal{L}}(x) = (L \circ T)(x)$, space of sufficient statistics $K_{\mathcal{L}} = \text{conv}(T_{\mathcal{L}}(\mathcal{X}))$, log-partition function $A_{\mathcal{L}}$, and gradient map $F_{\mathcal{L}}$.

Lemma 4.1. *The exponential family $(\mathcal{X}, \nu, T_{\mathcal{L}})$ is invariant under the choice of basis for \mathcal{L} .*

Proof. Let $M \in \mathbb{R}^{c \times d}$ denote another basis for \mathcal{L} . There exists an invertible matrix $U \in \mathbb{R}^{c \times c}$ such that $L = UM$. Let $\eta = U^T \tau$ be another coordinate system on \mathcal{L} . Then $L^T \tau = M^T \eta$ and, by (19), for every $\theta \in C \cap \mathcal{L}$ we find $\langle \tau, (L \circ T)(x) \rangle = \langle \theta, T(x) \rangle = \langle \eta, (M \circ T)(x) \rangle$. \square

In what follows we fix the basis consisting of the rows of L , and we identify \mathbb{R}^c with \mathcal{L} via $\tau \mapsto L^T \tau$. This means that the convex set $C_{\mathcal{L}} = C \cap \mathcal{L}$ lives in \mathbb{R}^c with coordinates τ .

Lemma 4.2. *The gradient map of the restricted exponential family $(\mathcal{X}, \nu, T_{\mathcal{L}})$ is given by*

$$F_{\mathcal{L}} = L \circ F \circ L^T. \quad (20)$$

Proof. A computation similar to (19) shows that all canonical parameters $\tau \in C_{\mathcal{L}}$ satisfy

$$A_{\mathcal{L}}(\tau) = \int_{\mathcal{X}} \exp(-\langle \tau, (L \circ T)(x) \rangle) \nu(dx) = \int_{\mathcal{X}} \exp(-\langle L^T \tau, T(x) \rangle) \nu(dx) = A(L^T \tau). \quad (21)$$

Thus the derivative of $A_{\mathcal{L}}(\tau)$ with respect to τ_i is equal to the directional derivative of $A(\theta)$ in the direction given by the i -th row of L . This implies the following identity which proves (20):

$$\nabla_{\tau} A_{\mathcal{L}}(\tau) = L \nabla_{\theta} A(\theta) \Big|_{\theta=L^T \tau}. \quad \square$$

Example 4.3 (*Toric varieties from discrete exponential families*). We build on Example 2.6 and consider linear restrictions of the parameter space in the full discrete exponential family. Let \mathcal{L} be a linear subspace of \mathbb{R}^d that is defined over the rational numbers, and take $L \in \mathbb{Z}^{c \times d}$. We assume that $(1, 1, \dots, 1) \in \mathcal{L}$. Here $C = \mathbb{R}^d$ and thus $C \cap \mathcal{L} = \mathcal{L}$. The set $K_{\mathcal{L}}$ is the image of the $(d-1)$ -dimensional simplex K under the linear map $L : \mathbb{R}^d \rightarrow \mathbb{R}^c$. Thus $K_{\mathcal{L}}$ is the $(c-1)$ -dimensional convex polytope obtained as the convex hull of the columns of L .

Consider the image $F(\mathcal{L})$ of \mathcal{L} under the gradient map F in (9). This image is a semialgebraic set of dimension $c-1$ inside the $(d-1)$ -dimensional simplex K . To be explicit,

$$F(\mathcal{L}) = \left\{ (\sigma_1, \sigma_2, \dots, \sigma_d) \in K : \prod_{i=1}^d \sigma_i^{u_i} = 1 \text{ for all } u \in \mathcal{L}^{\perp} \right\}.$$

Since L has integer entries, we can here replace the space \mathcal{L}^{\perp} with the lattice $\mathcal{L}_{\mathbb{Z}}^{\perp} = \mathcal{L}^{\perp} \cap \mathbb{Z}^d$.

We write \mathcal{L}^F for the Zariski closure of $F(\mathcal{L})$ in \mathbb{CP}^{d-1} , regarded as the complexification of K . This is the *projective toric variety* associated with the polytope $K_{\mathcal{L}}$. To be explicit,

$$\mathcal{L}^F = \left\{ (\sigma_1 : \sigma_2 : \dots : \sigma_d) \in \mathbb{CP}^{d-1} : \prod_{i:u_i>0} \sigma_i^{u_i} = \prod_{j:u_j<0} \sigma_j^{-u_j} \text{ for all } u \in \mathcal{L}_{\mathbb{Z}}^{\perp} \right\}.$$

In algebraic statistics, one refers to $F(\mathcal{L})$ as the *toric model* of the lattice $\mathcal{L}_{\mathbb{Z}}$. The linear map L takes this model bijectively onto the polytope $K_{\mathcal{L}}$ of sufficient statistics. In geometry, this bijection is known as the *moment map* of the toric variety \mathcal{L}^F . Given any point in $\text{int}(K_{\mathcal{L}})$, its unique preimage in $F(\mathcal{L})$ is known as the *Birch point* or the *MLE*. For further reading on toric models see [29, Section 1.2.2], [32], and other introductions to algebraic statistics. \diamond

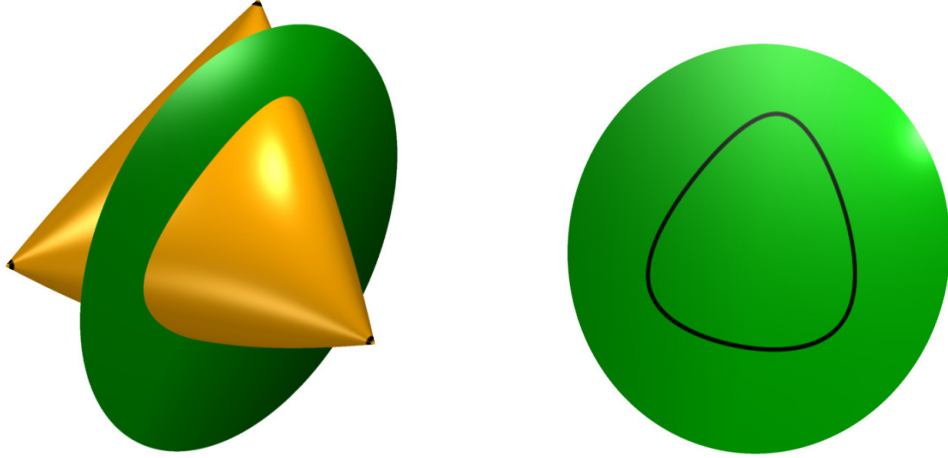


Figure 2: The convex body C of canonical parameters is intersected with a linear space \mathcal{L} . Dual to the intersection $C \cap \mathcal{L}$ is the projection of the exponential variety shown in Figure 4.

We now finally come to the definition of exponential varieties. They arise when restricting a hyperbolic exponential family to a subspace \mathcal{L} . They are analogous to the toric varieties above.

Definition 4.4 (*Exponential variety*). Let $f(\theta)$ be a hyperbolic polynomial, with hyperbolicity cone $C \subset \mathbb{R}^d$ and dual cone $K = C^\vee$. The corresponding *exponential variety*, denoted $\mathcal{L}^F = \overline{F(C \cap \mathcal{L})}$, is the Zariski closure of the image of \mathcal{L} under the gradient map $F = -\nabla \log(f)$.

Here the Zariski closure could be taken in \mathbb{R}^d , but we usually prefer to work in \mathbb{CP}^{d-1} instead. In this case, like in (7), we can equivalently study the image of ∇f . If $c = \dim(\mathcal{L})$ then \mathcal{L}^F will be a $(c-1)$ -dimensional projective variety in \mathbb{CP}^{d-1} . The *positive exponential variety* is the semialgebraic set $\mathcal{L}_{>0}^F = F(C \cap \mathcal{L})$. This is analogous to the toric model. It lives in the space $K \subset \mathbb{R}^d$ of sufficient statistics, or preferably, in its image in \mathbb{RP}^{d-1} .

A key property of the exponential variety \mathcal{L}^F is that the canonical bijection between $C_{\mathcal{L}}$ and $\text{int}(K_{\mathcal{L}})$ factors through the positive part $\mathcal{L}_{>0}^F$. This is the content of the following result.

Theorem 4.5. *Fix a hyperbolic exponential family, with f, C, K and $F = -\nabla \log(f)$ as above, and let \mathcal{L} be a linear subspace of \mathbb{R}^d that intersects C . Then the gradient map $F_{\mathcal{L}}$ of the restricted exponential family on \mathcal{L} can be written as a sequence of maps*

$$C_{\mathcal{L}} \subset C \xrightarrow{F} K \xrightarrow{L} K_{\mathcal{L}}. \quad (22)$$

The convex set $C_{\mathcal{L}}$ of canonical parameters maps bijectively to the positive exponential variety $\mathcal{L}_{>0}^F$, and $\mathcal{L}_{>0}^F$ maps bijectively to the interior of the convex set $K_{\mathcal{L}}$ of sufficient statistics.

Proof. Let τ be the coordinates on $C_{\mathcal{L}}$ and θ the coordinates on C . The projection $\pi_{\mathcal{L}} : K \rightarrow K_{\mathcal{L}}$ is defined by the $c \times d$ -matrix L , and the embedding $C_{\mathcal{L}} \subset C$ is given by $\theta = L^T \tau$. From (20) we know that $L \circ F \circ L^T = F_{\mathcal{L}}$. This is precisely what is being asserted in (22). By Theorem 2.2 and Remark 3.7, the two gradient maps F and $F_{\mathcal{L}}$ are both bijections. Since $C_{\mathcal{L}} \subset C$ is an embedding, the projection $\pi_{\mathcal{L}}$ must be a bijection when restricted to $F(C_{\mathcal{L}}) = \mathcal{L}_{>0}^F$. \square

Corollary 4.6. *The dimension of the exponential variety \mathcal{L}^F equals that of the space \mathcal{L} .*

Remark 4.7. *In defining exponential varieties we chose to restrict to hyperbolic polynomials. A more inclusive definition would start with an arbitrary rational exponential family (\mathcal{X}, ν, T) with gradient map $F : C \rightarrow K$. For any subspace $\mathcal{L} \subset \mathbb{R}^d$ with $C \cap \mathcal{L} \neq \emptyset$, one might then refer to $\mathcal{L}^F = \overline{F(C \cap \mathcal{L})}$ as an exponential variety. Theorem 4.5 remains true in that setting. But, even that definition does not yet include toric varieties. Indeed, the full discrete exponential family is not rational, and that is why L was an integer matrix in Example 4.3. It could make sense to work with a notion of exponential variety that includes toric varieties, for instance by working in the setting of algebraic exponential families as in [12]. However, we chose not to pursue that path in the present paper. We decided to use the term ‘exponential variety’ exclusively in the sense of Definition 4.4. Our focus is entirely on hyperbolic polynomials $f(\theta)$.*

Given a hyperbolic polynomial f , its restriction $f|_{\mathcal{L}}$ is a hyperbolic polynomial on the linear subspace $\mathcal{L} \subset \mathbb{R}^d$. The hyperbolicity cone of $f_{\mathcal{L}}$ equals $C_{\mathcal{L}} = C \cap \mathcal{L}$, and its dual is the cone

$$K_{\mathcal{L}} = (C_{\mathcal{L}})^{\vee} = (C \cap \mathcal{L})^{\vee} = \pi_{\mathcal{L}}(C^{\vee}) = \pi_{\mathcal{L}}(K). \quad (23)$$

From the perspective of algebraic geometry, we could also define the exponential variety $\mathcal{L}^{\nabla f}$ for any homogeneous $f \in \mathbb{R}[\theta_1, \dots, \theta_d]$ and any subspace $\mathcal{L} \subset \mathbb{CP}^{d-1}$. Namely, $\mathcal{L}^{\nabla f}$ is the closure of the image of \mathcal{L} under the gradient map in (7). This level of generality makes perfect sense for the study of algebraic degrees as in Section 5. However, in order for $\mathcal{L}^{\nabla f}$ to have a distinguished positive part $\mathcal{L}_{>0}^{\nabla f}$, with its remarkable role as the “middleman” in the bijection $F_{\mathcal{L}} : C_{\mathcal{L}} \rightarrow \text{int}(K_{\mathcal{L}})$ between two dual convex sets, we need that f is hyperbolic.

We note that Conjecture 3.5 remains valid after restricting to a linear section.

Proposition 4.8. *Let f be a hyperbolic polynomial with hyperbolicity cone C . Suppose that the Riesz kernel $q_{\alpha}(\sigma)$ is nonnegative for f and some $\alpha > 0$. Fix a linear subspace \mathcal{L} that intersects the interior of C . Then the Riesz kernel for $f|_{\mathcal{L} \cap C}$ is also nonnegative for α .*

Proof. By the Bernstein-Hausdorff-Widder-Choquet Theorem we know that $f^{-\alpha}$ is completely monotone on C . The inequalities (11) remain true after restricting to the subcone $\mathcal{L} \cap C$, hence $f^{-\alpha}|_{\mathcal{L} \cap C}$ is completely monotone. Again, by Bernstein-Hausdorff-Widder-Choquet we know that there exists a measure μ on the dual cone $K_{\mathcal{L}} = (C \cap \mathcal{L})^\vee$ such that

$$f(\theta)^{-\alpha} = \int_{(\mathcal{L} \cap C)^\vee} \exp(-\langle \theta, \sigma \rangle) \mu(d\sigma).$$

By the uniqueness of the inverse Laplace transform, the measure μ must be induced by the Riesz kernel $q_\alpha(\sigma)$ as in (13). This means that $q_\alpha(\sigma)$ cannot take negative values. \square

The most prominent hyperbolic polynomial is the determinant of a symmetric matrix of linear forms. Conjecture 3.5 holds for such Gaussian models, by Propositions 3.8 and 4.8.

Example 4.9 (*Linear Gaussian concentration models*). Consider the multivariate Gaussian family described in Example 2.5, with $C = \text{PD}_m$ and $K = \overline{\text{PD}}_m$ in the space \mathbb{S}^m of real symmetric $m \times m$ -matrices. A linear subspace \mathcal{L} with $\mathcal{L} \cap C \neq \emptyset$ defines a *linear Gaussian concentration model*. The corresponding exponential variety $\mathcal{L}^{\nabla f}$ was studied in [36], where it was denoted by \mathcal{L}^{-1} . Its positive part consists of all covariance matrices in the model, i.e.,

$$\mathcal{L}_{>0}^{-1} = \{\sigma \in \text{PD}_m \mid \sigma^{-1} \in \mathcal{L}\}.$$

The instances of most interest are the *Gaussian graphical models* in [10, 27] where \mathcal{L} is defined by the vanishing of some of the off-diagonal entries θ_{ij} . A natural extension is the class of colored Gaussian graphical models in [22]. The case when \mathcal{L} consists of diagonal matrices is studied in [36, Section 3]. For a concrete example, let $m = 4$ and consider the subspace

$$\mathcal{L} = \{\text{diag}(\tau_1, \tau_2, \tau_1 - \tau_2, \tau_1 + \tau_2) : \tau \in \mathbb{R}^2\}.$$

The exponential variety \mathcal{L}^{-1} is a cubic curve in the \mathbb{CP}^3 of diagonal matrices σ , namely

$$\mathcal{L}^{-1} = \{\sigma \in \mathbb{CP}^3 : \sigma_1\sigma_2 - \sigma_1\sigma_4 - \sigma_2\sigma_4 = \sigma_1\sigma_3 + \sigma_1\sigma_4 - 2\sigma_3\sigma_4 = \sigma_2\sigma_3 - \sigma_2\sigma_4 - 2\sigma_3\sigma_4 = 0\}.$$

The positive part $\mathcal{L}_{>0}^{-1}$ consists of $\sigma \in \mathcal{L}$ where all σ_i are real and positive. Writing L for the linear map given by $\rho_1 = \sigma_1 + \sigma_3 + \sigma_4$ and $\rho_2 = \sigma_2 - \sigma_3 + \sigma_4$, the bijections in (22) are

$$C_{\mathcal{L}} = \{\tau \in \mathbb{R}^2 : \tau_1 > \tau_2 > 0\} \xrightarrow{\nabla f} \mathcal{L}_{>0}^{-1} \xrightarrow{L} K_{\mathcal{L}} = \{\rho \in \mathbb{R}^2 : \rho_1 \geq 0, \rho_1 + \rho_2 \geq 0\}.$$

The Riesz kernel for this particular exponential variety was computed in Example 3.11. \diamond

Example 4.10 (*The central path is an exponential curve*). Let f be a hyperbolic polynomial and C its hyperbolicity body in \mathbb{RP}^{d-1} . If $c = 2$ then $\mathcal{L}_{>0}^{\nabla f}$ is a curve inside the dual body K . This curve represents a *central path* (cf. [18]) for linear optimization over K . The case of linear programming was studied in [11]. Here K is a convex polytope and $K_{\mathcal{L}}$ the line segment of possible values of the objective function. The central path and the exponential curve are shown in Figure 3 for the case when $d = 3$, $c = 2$, and $f(\theta)$ is a product of six linear forms. \diamond

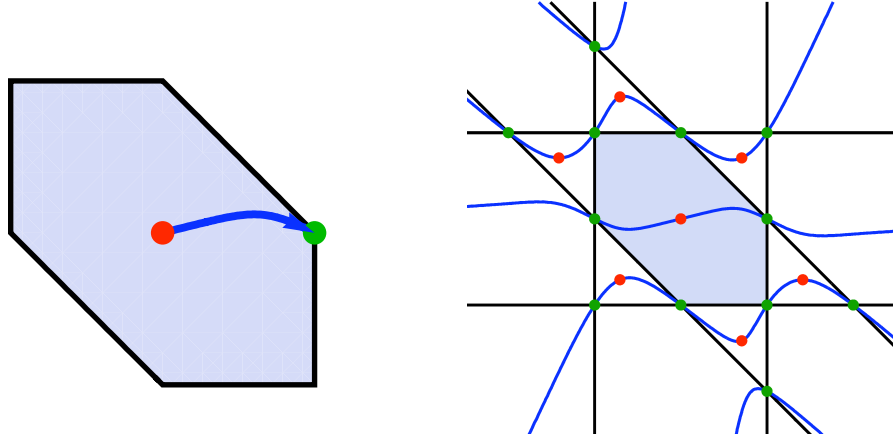


Figure 3: Central curves in linear and semidefinite programming are exponential varieties.

We discuss two further examples of combinatorial interest. These are based on [5, 25, 34].

Example 4.11 ((*Graphs*)). We fix a connected simple graph G with vertex set $\{0, 1, \dots, m\}$ and edge set $E(G)$. The *reduced Laplacian* of G is a symmetric $m \times m$ matrix Λ_G with rows and columns indexed by $\{1, \dots, m\}$. Its nonzero off-diagonal entries are $-\theta_{ij}$, and the i -th row sum of Λ_G equals θ_{0i} . By convention, $\theta_{ij} = 0$ if $\{i, j\} \notin E(G)$. The *spanning tree polynomial* $f(\theta) = \det(\Lambda_G)$ is a complete hyperbolic polynomial. We fix the exponential family for f .

Now, let H be any subgraph of G that is also connected. We consider the linear subspace

$$\mathcal{L} = \{\theta \in \mathbb{R}^{E(G)} : \theta_{ij} = 0 \text{ for } \{i, j\} \in E(G) \setminus E(H)\}.$$

The exponential variety $\mathcal{L}^{\nabla f}$ is an invariant of the pair $H \subset G$. It lives in $\mathbb{CP}^{|E(G)|-1}$ and its dimension is $|E(H)| - 1$. The players of Theorem 4.5 are quite interesting in this case.

For a concrete example, let $m = 3$ and G the complete graph K_4 , with reduced Laplacian

$$\Lambda_{K_4} = \begin{bmatrix} \theta_{01} + \theta_{12} + \theta_{13} & -\theta_{12} & -\theta_{13} \\ -\theta_{12} & \theta_{02} + \theta_{12} + \theta_{23} & -\theta_{23} \\ -\theta_{13} & -\theta_{23} & \theta_{03} + \theta_{13} + \theta_{23} \end{bmatrix}. \quad (24)$$

The cubic polynomial $f(\theta) = \det(\Lambda_{K_4})$ has 16 monomials, one for each spanning tree of K_4 . Every connected subgraph H of K_4 defines an exponential variety $\mathcal{L}^{\nabla f}$ in \mathbb{CP}^5 . For instance, let H be the 4-cycle, with $\mathcal{L} = \{\theta_{02} = \theta_{13} = 0\}$. Then $\mathcal{L}^{\nabla f}$ is a threefold of degree 4, obtained as the complete intersection of two quadrics, that is singular along three lines. On the other hand, if H is the 4-chain with $\mathcal{L} = \{\theta_{02} = \theta_{13} = \theta_{03} = 0\}$, then the exponential variety is a plane, namely $\mathcal{L}^{\nabla f} = \{\sigma \in \mathbb{CP}^5 : \sigma_{12} - \sigma_{13} + \sigma_{23} = \sigma_{02} - \sigma_{03} + \sigma_{23} = \sigma_{01} - \sigma_{03} + \sigma_{13} = 0\}$. \diamond

Example 4.12 ((*Vámos matroid*)). Let $d = 8$ and $f(\theta) = \sum \theta_i \theta_j \theta_k \theta_l$ where the sum runs over all quadruples in $\{1, 2, 3, 4, 5, 6, 7, 8\}$ other than 1256, 1278, 3456, 3478, 5678. So, $f(\theta)$ has 65 terms. This is the basis generating function of the *Vámos matroid*, a popular

hyperbolic polynomial. The specialization studied by Kummer [25] is the restriction of f to the subspace $\mathcal{L} = \{\theta \in \mathbb{R}^8 : \theta_1 = \theta_2, \theta_3 = \theta_4, \theta_5 = \theta_6, \theta_7 = \theta_8\}$. The main result in [25] states that the 3-dimensional body $C_{\mathcal{L}}$ is a spectrahedron. The bijection from $C_{\mathcal{L}}$ to the interior of $K_{\mathcal{L}} = C_{\mathcal{L}}^{\vee}$ factors through the positive variety $\mathcal{L}_{>0}^{\nabla f} \subset \mathbb{RP}^7$. On this particular plane \mathcal{L} , the gradient map ∇f is an involution, that is, $\mathcal{L}^{\nabla f} = \{\sigma_1 = \sigma_2, \sigma_3 = \sigma_4, \sigma_5 = \sigma_6, \sigma_7 = \sigma_8\}$. However, for generic choices of \mathcal{L} , the exponential variety $\mathcal{L}^{\nabla f}$ is a non-linear threefold in \mathbb{CP}^7 . For instance, if we replace θ_8 by $-\theta_8$ in \mathcal{L} then $\mathcal{L}^{\nabla f}$ is a hypersurface of degree 18 in a 4-plane in \mathbb{CP}^7 . If we further replace θ_6 by $-\theta_6$ then $\mathcal{L}^{\nabla f}$ has codimension 2 and degree 10 in a 5-plane in \mathbb{CP}^7 . \diamond

In Examples 4.11 and 4.12 we saw that the exponential variety $\mathcal{L}^{\nabla f}$ can be linear again when \mathcal{L} is chosen in a very special manner. This raises the question for which pairs (f, \mathcal{L}) this happens. The remainder of this section is devoted to what we know about this question.

The case of linear Gaussian concentration models, where $f(\theta) = \det(\theta)$, $C = \text{PD}_m$ and $\mathcal{L}^{\nabla f} = \mathcal{L}^{-1}$, was studied by Jensen in [24]. He proved that the exponential variety \mathcal{L}^{-1} is linear if and only if \mathcal{L} is a *Jordan algebra*, i.e., $\theta^2 \in \mathcal{L}$ for all $m \times m$ matrices θ in \mathcal{L} . Moreover, Jensen showed that if \mathcal{L}^{-1} is linear then $\mathcal{L}^{-1} = \mathcal{L}$ under the identification of σ with θ .

The situation is not as nice and clean for hyperbolic polynomials $f(\theta)$ other than the symmetric determinant. As we have seen in Example 4.11, a linear exponential variety $\mathcal{L}^{\nabla f}$ can be quite different from its underlying \mathcal{L} . It would be very desirable to find a characterization that generalizes Jensen's results on Gaussian families to hyperbolic exponential families.

As a first start, we examine the special case of codimension 1. The following result explicates the conditions under which an exponential variety $\mathcal{L}^{\nabla f}$ happens to be a hyperplane.

Proposition 4.13. *Fix a hyperbolic polynomial f and suppose that the hyperplane $\{\sum_{i=1}^d a_i \sigma_i = 0\}$ is an exponential variety for f . Then the polynomial $\sum_{i=1}^d a_i (\partial f / \partial \theta_i)$ is divisible by a linear form $\ell(\theta)$. Moreover, if the hyperplane $\mathcal{L} = \{\ell(\theta) = 0\}$ satisfies the conclusion of Corollary 4.6 – for instance because $\mathcal{L} \cap C \neq \emptyset$ – then $\mathcal{L}^{\nabla f} = \{\sum_{i=1}^d a_i \sigma_i = 0\}$.*

Proof. Let \mathcal{H} be the hyperplane given by $\sum_{i=1}^d a_i \sigma_i = 0$. Suppose that \mathcal{H} is an exponential variety $\mathcal{L}^{\nabla f}$, where \mathcal{L} is yet to be specified. The polynomial $\sum_{i=1}^d a_i \frac{\partial f}{\partial \theta_i}$ defines a hypersurface polar to the hyperbolic polynomial f . A point $\theta \in \mathbb{CP}^{d-1}$ lies on that polar hypersurface if and only if $F(\theta) \in \mathcal{H}$. Since the hyperplane \mathcal{L} maps to \mathcal{H} , it is a subset of the polar hypersurface. This means that $\sum_{i=1}^d a_i \frac{\partial f}{\partial \theta_i}$ is divisible by the linear equation that defines \mathcal{L} .

For the second statement, note that if the linear form $\ell(\theta)$ divides $\sum_{i=1}^d a_i \frac{\partial f}{\partial \theta_i}$, then the hyperplane \mathcal{L} defined by $\ell(\theta)$ maps to \mathcal{H} . Assuming that the image of \mathcal{L} has codimension one in $(\mathbb{CP}^{d-1}, \sigma)$, as expected, then the closure of that image must coincide with \mathcal{H} . \square

We now show how these conditions can be checked for the cubic in our running example.

Example 4.14. Consider the elementary symmetric polynomial E_3 in four variables, as in

Examples 1.1, 2.4, 5.2, 6.8. We need to analyze the quadrics that are polar to the cubic E_3 :

$$\sum_{i=1}^4 a_i \frac{\partial E_3}{\partial \theta_i} = \frac{1}{2} \begin{pmatrix} \theta_1 & \theta_2 & \theta_3 & \theta_4 \end{pmatrix} \begin{pmatrix} 0 & a_3 + a_4 & a_2 + a_4 & a_2 + a_3 \\ a_3 + a_4 & 0 & a_1 + a_4 & a_1 + a_3 \\ a_2 + a_4 & a_1 + a_4 & 0 & a_1 + a_2 \\ a_2 + a_3 & a_1 + a_3 & a_1 + a_2 & 0 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{pmatrix}$$

The quadric has a linear factor if and only if all 3×3 minors of this symmetric matrix vanish. Up to scaling and permuting coordinates, this happens only for two points, namely $a = (-1 : 1 : 1 : 1)$ and $a = (1 : -1 : 0 : 0)$. In the first case, the quadric is $\theta_1(\theta_2 + \theta_3 + \theta_4)$, providing us with two planes \mathcal{L} that yield the same exponential variety $\mathcal{L}^{\nabla E_3} = \{\sigma_1 = \sigma_2 + \sigma_3 + \sigma_4\}$. In the second case the quadric is $(\theta_2 - \theta_1)(\theta_3 + \theta_4)$. Each factor yields $\mathcal{L}^{\nabla E_3} = \{\sigma_1 = \sigma_2\}$.

It is straightforward to generalize this construction to other polynomials of degree 3. \diamond

The situation is more difficult when the given polynomial f has degree $m \geq 4$. To find all hyperplanes that are exponential for f , we must eliminate the quantifiers in the formula

$$\exists \ell \text{ linear} : \ell(\theta) \text{ divides } a_1 \frac{\partial f}{\partial \theta_1} + \cdots + a_d \frac{\partial f}{\partial \theta_d}.$$

To this end, we consider the variety of all homogeneous polynomials of degree $m - 1$ in d unknowns that have a linear factor. This variety is denoted $\text{Chow}_{1,m-2}$. It is a variant of the *Chow variety* of polynomials that are products of linear forms. This is a classical topic in algebraic geometry with recent connections to computational complexity [7, Section 6.4].

For a given f , we must intersect the linear space spanned by the partial derivatives of f with $\text{Chow}_{1,m-2}$. Each intersection point furnishes a candidate for an exponential hyperplane. For instance, let $d = 3$ and $m = 4$, so f is a ternary quartic. Then $\frac{\partial f}{\partial \theta_1}, \frac{\partial f}{\partial \theta_2}, \frac{\partial f}{\partial \theta_3}$ span a 2-plane in the \mathbb{CP}^9 of plane cubics. We must intersect that plane with $\text{Chow}_{1,2}$, which is a variety of codimension 2 and degree 21 in \mathbb{CP}^9 . Its prime ideal is generated by 35 polynomials of degree 8. The 21 points $(a_1 : a_2 : a_3)$ in that intersection represent all lines $\{a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3 = 0\}$ in \mathbb{CP}^2 that can have the structure of an exponential variety with respect to the quartic f .

5 Gradient multidegree and ML degree

In this section we introduce algebraic complexity measures for hyperbolic exponential families. The partition function is $f^{-\alpha}$ where $\alpha > 0$ and f is a hyperbolic polynomial. We are interested in the rational map $F : \mathbb{CP}^{d-1} \dashrightarrow \mathbb{CP}^{d-1}$ defined by the gradient of the log-partition function $A(\theta) = -\alpha \cdot \log f(\theta)$. We shall work in the convenient setting of projective geometry. This means that the logarithm and the constant α are ignored. From now on, F is simply the rational map (7) defined by the gradient of the polynomial $f(\theta)$.

The following definition makes sense for any homogeneous polynomial $f(\theta)$. However, in our context we always assume that $f(\theta)$ is hyperbolic, that the hyperbolicity cone C is pointed in (\mathbb{R}^d, θ) and that its dual K is a pointed cone in (\mathbb{R}^d, σ) . We regard C and K as convex bodies in (affine charts of) the real projective spaces $(\mathbb{RP}^{d-1}, \theta)$ and $(\mathbb{RP}^{d-1}, \sigma)$. The gradient map $F = \nabla f$ defines a bijection between the interiors of these two convex bodies. Its inverse is an algebraic function, and we are interested in the degree of that function.

Definition 5.1. Let f be a homogeneous polynomial in $\theta_1, \dots, \theta_d$, and let X_f be the graph of its gradient map $F = \nabla f$. Thus X_f is the closure in $\mathbb{CP}^{d-1} \times \mathbb{CP}^{d-1}$ of the set of points (θ, σ) , where θ lies in \mathbb{CP}^{d-1} but is not in the singular locus of the hypersurface $\{f = 0\}$, and $\sigma = F(\theta)$. By construction, the graph X_f is an irreducible subvariety of dimension $d - 1$ in $\mathbb{CP}^{d-1} \times \mathbb{CP}^{d-1}$. We define the *gradient multidegree* of the polynomial f to be the class $[X_f]$ of that variety in the integral cohomology ring of the product of the two projective spaces:

$$H^*(\mathbb{CP}^{d-1} \times \mathbb{CP}^{d-1}; \mathbb{Z}) = \mathbb{Z}[s, t] / \langle s^d, t^d \rangle.$$

The ring generators are the divisor classes $s = [\mathbb{CP}^{d-1} \times \{\sigma_i = 0\}]$ and $t = [\{\theta_j = 0\} \times \mathbb{CP}^{d-1}]$.

In concrete geometric terms, the gradient multidegree of f is the generating function

$$[X_f] = \alpha_d s^{d-1} + \alpha_{d-1} s^{d-2} t + \alpha_{d-2} s^{d-3} t^2 + \dots + \alpha_2 s t^{d-2} + \alpha_1 t^{d-1},$$

where α_i is the cardinality of the finite set $X_f \cap (L_{i-1} \times M_{d-i})$ where L_{i-1} and M_{d-i} are general subspaces in \mathbb{CP}^{d-1} of dimensions $i - 1$ and $d - i$ respectively. The leading coefficient α_d of $[X_f]$ is the degree of the map (7). We call this the *gradient degree* of f .

In statistical applications, the quantity α_d measures the algebraic complexity of the function $\sigma \mapsto \theta$ that take the sufficient statistics to the natural parameters. It plays the role of the ML degree for the exponential family in question. However, in this paper we reserve the term “ML degree” for exponential varieties, as in Definition 5.4 below. This is consistent with the usage in [36]. The number $\alpha_d = \text{degree}(F)$ is the *gradient degree of the hyperbolic exponential family*.

Example 5.2. Let $d = 4$ and fix the elementary symmetric polynomial as in Example 1.1,

$$f = \theta_1 \theta_2 \theta_3 + \theta_1 \theta_2 \theta_4 + \theta_1 \theta_3 \theta_4 + \theta_2 \theta_3 \theta_4.$$

The gradient multidegree of f is the following class in the integral cohomology of $\mathbb{CP}^3 \times \mathbb{CP}^3$:

$$[X_f] = 4s^3 + 4s^2 t + 2s t^2 + t^3. \quad (25)$$

We see that the gradient degree of the polynomial f is 4. Hence we can use Cardano’s formula to express the MLE $\hat{\theta}$ in terms of radicals in the sufficient statistics σ of the data.

To be explicit, the graph X_f is the subvariety of $\mathbb{CP}^3 \times \mathbb{CP}^3$ defined by the 2×2 -minors of

$$\begin{pmatrix} \sigma_1 & \sigma_2 & \sigma_3 & \sigma_4 \\ \theta_2 \theta_3 + \theta_2 \theta_4 + \theta_3 \theta_4 & \theta_1 \theta_3 + \theta_1 \theta_4 + \theta_3 \theta_4 & \theta_1 \theta_2 + \theta_1 \theta_4 + \theta_2 \theta_4 & \theta_1 \theta_2 + \theta_1 \theta_3 + \theta_2 \theta_3 \end{pmatrix}, \quad (26)$$

over the nonsingular locus of $\{f = 0\}$, and taking the Zariski closure of the resulting set.

If we replace the unknowns $\sigma_1, \sigma_2, \sigma_3, \sigma_4$ by fixed generic real numbers, then we get a system of equations in $\theta = (\theta_1 : \theta_2 : \theta_3 : \theta_4)$ that has four complex solutions. Assuming that σ is in the interior of the red convex body K , precisely one of these solutions lies in the yellow convex body C . The colors refer to Figure 1. This unique solution is the MLE $\hat{\theta}$. \diamond

Remark 5.3. If $f = \det(\theta)$ is the symmetric $m \times m$ -determinant, representing the full Gaussian family in Example 2.5, then $[X_f]$ is a binary form that is known in some cases. By [36, Theorem 2.3], it is symmetric under swapping s and t . For instance, for $m = 4$,

$$[X_f] = 1s^9 + 3s^8t + 9s^7t^2 + 17s^6t^3 + 21s^5t^4 + 21s^4t^5 + 17s^3t^6 + 9s^2t^7 + 3st^8 + 1t^9.$$

At present, no general formula is known for the gradient multidegree of the symmetric $m \times m$ -determinant. Expressions for some of its coefficients are derived in [36] after Theorem 2.3.

In the computer algebra system `Macaulay2` [16], we can compute $[X_f]$ from a given hyperbolic polynomial f using the built-in command `multidegree`. See [28, Section 8.5] for an explanation of how multidegrees in commutative algebra represent cohomology classes on toric varieties. Here is a piece of `Macaulay2` code that generates (26) and the output (25):

```
R = QQ[t1,t2,t3,t4,s1,s2,s3,s4,
      Degrees => {{1,0},{1,0},{1,0},{1,0},{0,1},{0,1},{0,1},{0,1}}];
f = t1*t2*t3 + t1*t2*t4 + t1*t3*t4 + t2*t3*t4 ;
gradf = {diff(t1,f),diff(t2,f),diff(t3,f),diff(t4,f)};
M = matrix { {s1,s2,s3,s4}, gradf };
multidegree saturate( minors(2,M), ideal(gradf) )
```

This method works only for small examples because the command `saturate` is slow. To compute the gradient multidegree for larger polynomials $f(\theta)$, it is better to use the formula

$$\alpha_i = \#(X_f \cap (L_{i-1} \times M_{d-i})). \quad (27)$$

To count the number of points in this intersection, one can either use Gröbner bases over a finite field, or the homotopy methods supplied by numerical algebraic geometry [3].

We now turn our discussion to exponential varieties. Let f be a fixed hyperbolic polynomial with hyperbolicity cone $C \subset \mathbb{R}^d$. We consider any linear subspace \mathcal{L} which intersects the interior of C , and we set $c = \dim(\mathcal{L})$. The corresponding exponential variety in \mathbb{CP}^{d-1} is $\mathcal{L}^{\nabla f} := \overline{F(\mathcal{L})}$. It follows from the results in the previous section that $\dim(\mathcal{L}^{\nabla f}) = c - 1$ and the positive exponential variety $\mathcal{L}_{>0}^{\nabla f} := \mathcal{L}^{\nabla f} \cap \text{int}(K)$ maps bijectively onto the interior of $(C \cap \mathcal{L})^\vee = \pi_{\mathcal{L}}(K)$ under the linear map $\pi_{\mathcal{L}}$. The inverse of that map is an algebraic function. We are interested in the degree. This can alternatively be described as follows.

Definition 5.4. Let $\pi_{\mathcal{L}} : \mathbb{CP}^{d-1} \dashrightarrow \mathbb{CP}^{c-1}$ denote the projection with center \mathcal{L}^\perp . This restricts to a finite-to-one map from the exponential variety $\mathcal{L}^{\nabla f}$ onto \mathbb{CP}^{c-1} . We define

$$\text{MLdegree}(\mathcal{L}^{\nabla f}) := \text{degree}(\mathcal{L}^{\nabla f} \dashrightarrow \mathbb{CP}^{c-1}).$$

Thus the *ML degree* of the exponential variety $\mathcal{L}^{\nabla f}$ is the cardinality of the generic fiber of $\pi_{\mathcal{L}}$ restricted to $\mathcal{L}^{\nabla f}$. This is the algebraic degree of the function that takes points in $\pi_{\mathcal{L}}(K)$ back to their unique preimage in the nonnegative variety $\mathcal{L}_{\geq 0}^{\nabla f}$.

Our main result in this section relates the various notions of degrees:

Theorem 5.5. *The following inequalities hold for all exponential varieties:*

$$\text{MLdegree}(\mathcal{L}^{\nabla f}) \leq \text{degree}(\mathcal{L}^{\nabla f}) \leq \text{the coefficient } \alpha_c \text{ of } s^{c-1}t^{d-c} \text{ in } [X_f]. \quad (28)$$

Moreover, the left inequality is an equality if and only if $\mathcal{L}^{\nabla f} \cap \mathcal{L}^\perp = \emptyset$.

Proof. Recall from (27) that α_c is obtained by counting the points of the intersection of the graph of ∇f with the product of a generic $(c-1)$ -dimensional plane L_{c-1} and a $(d-c)$ -dimensional plane M_{d-c} . If we replace L_{c-1} with a special subspace while keeping M_{d-c} generic, the number of points in the intersection is finite because $\dim \mathcal{L}^{\nabla f} = \dim \mathcal{L}$. It can only go down compared to the completely generic case. This proves the right inequality.

The left inequality is derived from the following well-known fact in algebraic geometry. When the center of the projection is disjoint from the variety, then the degree of the variety equals the product of the degree of the map and the degree of the image. Moreover, when the center of the projection intersects the variety, the product of the cardinality of the generic fiber times the degree of the image is strictly smaller than the degree of the variety. In our situation, the center of the projection $\pi_{\mathcal{L}}$ is \mathcal{L}^\perp , which completes the proof of the theorem. \square

We now offer an illustration of this result with two combinatorial examples.

Example 5.6. Let $d = 6$ and $f(\theta)$ be the Laplacian determinant (24) of the complete graph K_4 . Using Macaulay2 as above, we find that the gradient multidegree equals

$$[X_f] = 1s^5 + 2s^4t + 4s^3t^2 + 4s^2t^3 + 2st^4 + 1t^5.$$

For instance, we conclude that all subspaces \mathcal{L} of codimension 2 or 3 satisfy $\text{degree}(\mathcal{L}^{\nabla f}) \leq 4$. In Example 4.11, this bound is attained for the 4-cycle but not for the 4-chain. By contrast, let g be the polynomial obtained from f by adding $\theta_{01}\theta_{02}\theta_{12} + \theta_{01}\theta_{03}\theta_{13} + \theta_{02}\theta_{03}\theta_{23} + \theta_{12}\theta_{13}\theta_{23}$. Equivalently, g is the elementary symmetric polynomial in six unknowns. We have

$$[X_g] = 26s^5 + 16s^4t + 8s^3t^2 + 4s^2t^3 + 2st^4 + 1t^5. \quad (29)$$

This implies, for instance, that $\text{MLdegree}(\mathcal{L}^{\nabla g}) \leq \text{degree}(\mathcal{L}^{\nabla g}) \leq 16$ for all hyperplanes \mathcal{L} in \mathbb{CP}^5 . We especially note that the gradient map ∇g is 26-to-1 whereas ∇f is birational. \diamond

Example 5.7. All cases of equality in Theorem 5.5 occur for *Gaussian graphical models* on four nodes. Here $d = 10$ and $f(\theta)$ is the determinant of a symmetric 4×4 -matrix (θ_{ij}) .

- Let $c = 10$ and $\mathcal{L} = \mathbb{R}^{10}$ the complete model K_4 . The degrees in (28) are $1 = 1 = 1$.
- Let $c = 8$ and $\mathcal{L} = \{\theta_{13}=\theta_{24}=0\}$ the 4-cycle model. The degrees in (28) are $5 < 9 = 9$.
- Let $c = 3$ and fix the first model in [36, Table 2]. The degrees in (28) are $5 = 5 < 9$.
- Let $c = 7$ and $\mathcal{L} = \{\theta_{13}=\theta_{24}=\theta_{14}=0\}$ the 4-chain. The degrees in (28) are $1 < 5 < 17$.

In the remainder of this section, we examine the genericity conditions under which equality holds in the two inequalities in (28). It is immediate from the definition of the gradient multidegree that $\deg(\mathcal{L}^{\nabla f}) = \alpha_c$ for generic subspaces \mathcal{L} . From the proof of Theorem 5.5 we know that the left inequality is an equality if and only if $\mathcal{L}^{\nabla f} \cap \mathcal{L}^\perp = \emptyset$. We believe that this holds for all polynomials f when \mathcal{L} is generic, but we are presently unable to prove it.

Conjecture 5.8. *Let f be a hyperbolic polynomial in $\theta_1, \dots, \theta_d$. Then the exponential variety $\mathcal{L}^{\nabla f}$ defined by a generic linear subspace \mathcal{L} in \mathbb{CP}^{d-1} has its degree equal to its ML degree. Equivalently, for generic \mathcal{L} we have $\mathcal{L}^{\nabla f} \cap \mathcal{L}^\perp = \emptyset$.*

In the following, we provide a partial result concerning the set-theoretic image, here denoted $(\nabla f)(\mathcal{L})$. This image is a dense subset of $\mathcal{L}^{\nabla f}$, but in general they may differ.

Proposition 5.9. *The image $(\nabla f)(\mathcal{L})$ is disjoint from \mathcal{L}^\perp if and only if \mathcal{L} is not contained in the hyperplane tangent to the hypersurface $\{f = 0\}$ at a smooth point belonging to \mathcal{L} .*

Proof. For any $\tau \in \mathbb{CP}^{d-1}$ that does not belong to the singular locus of $\{f = 0\}$, we have

$$\sum_i \tau_i \frac{\partial f}{\partial \theta_i}(\tau) = 0 \iff f(\tau) = 0.$$

Writing $F = \nabla f$, we have

$$F(\tau) \in \tau^\perp \iff f(\tau) = 0.$$

To prove the ‘only if’ direction, consider a smooth point τ in $\{f = 0\}$. The tangent hyperplane at τ is $F(\tau)^\perp$. If τ lies in \mathcal{L} and \mathcal{L} is contained in $F(\tau)^\perp$, then $F(\tau) \in \mathcal{L}^\perp \cap F(\mathcal{L})$. For the reverse direction let $\sigma \in F(\mathcal{L}) \cap \mathcal{L}^\perp$. Then $\sigma = F(\tau)$ for some $\tau \in \mathcal{L}$. Hence, $F(\tau) \in \tau^\perp$ and therefore $f(\tau) = 0$. This means that \mathcal{L} is contained in the tangent hyperplane to $\{f = 0\}$ at τ . \square

This proposition implies that the image $(\nabla f)(\mathcal{L})$ is disjoint from \mathcal{L}^\perp if and only if the singular locus of the intersection $\mathcal{L} \cap \{f = 0\}$ is contained in the singular locus of $\{f = 0\}$. This is true for generic \mathcal{L} by Bertini’s Theorem. The following corollary identifies cases when the gradient map ∇f is regular on \mathcal{L} . In such cases, the set-theoretic image $\nabla(f)(\mathcal{L})$ will be closed and hence equal to $\mathcal{L}^{\nabla f}$, so we can conclude that it is generically disjoint from \mathcal{L}^\perp .

Corollary 5.10. *If the subspace \mathcal{L} is generic of dimension strictly smaller than the codimension of the singular locus of $\{f = 0\}$, then the ML degree equals the degree of $\mathcal{L}^{\nabla f}$.*

6 Elementary Symmetric Polynomials

Elementary symmetric polynomials form an important class of hyperbolic polynomials [5]. In this section, we study the exponential varieties obtained by slicing their hyperbolicity cones by a generic \mathcal{L} . Our main result is the formula for the gradient multidegree in Theorem 6.2.

We write E_m for the elementary symmetric polynomial of degree m in the d unknowns $\theta_1, \dots, \theta_d$, and we consider its gradient map $\nabla E_m : \mathbb{CP}^{d-1} \dashrightarrow \mathbb{CP}^{d-1}$. The i -th coordinate

of ∇E_m is denoted $E_m^i = \partial_{\theta_i} E_m$. This is the elementary symmetric polynomial of degree $m - 1$ in the $d - 1$ unknowns $\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d$. The hyperbolicity cone of E_m equals

$$C = \{ \theta \in \mathbb{R}^d : E_j(\theta) > 0 \text{ for } j = 1, 2, \dots, m \}.$$

This defines a hyperbolic exponential family as in Section 3. The geometry of this family (for $d = 4, m = 3$) was studied in Example 1.1. We shall return to this at the end of this section. See Example 3.9 and its pointer to [34] for some information on its Riesz kernel.

The exponential family discussed here is not an instance of Example 4.9. Indeed, by [26, Example 5.10], the elementary symmetric polynomial E_m is not a symmetric determinant whose entries are linear forms, provided $2 \leq m \leq d - 2$. Hence our model in this section is not a linear Gaussian concentration model. In particular, the theory in [36] is not applicable. However, a multiple of E_m is a symmetric determinant with linear entries, as shown in [5].

Definition 6.1. The *Eulerian number* $A(r, s)$ is the number of permutations of $\{1, 2, \dots, r\}$ with precisely s ascents. We have $A(r, 0) = A(r, r - 1) = 1$ and the following recursion holds:

$$A(r, s) = (r - s) \cdot A(r - 1, s - 1) + (s + 1) \cdot A(r - 1, s) \quad \text{for } 1 \leq s \leq r - 2.$$

The Eulerian numbers for small r are $A(3, \bullet) = 1, 4, 1$, $A(4, \bullet) = 1, 11, 11, 1$, $A(5, \bullet) = 1, 26, 66, 26, 1$, $A(6, \bullet) = 1, 57, 302, 302, 57, 1$, and $A(7, \bullet) = 1, 120, 1191, 2416, 1191, 120, 1$.

The gradient multidegree of E_m is given by a formula in terms of Eulerian numbers:

Theorem 6.2. *The gradient multidegree of the elementary symmetric polynomial equals*

$$[X_{E_m}] = \sum_{i=1}^d \alpha_i s^{i-1} t^{d-i}, \quad \text{where}$$

$$\alpha_i = \begin{cases} (m - 1)^{i-1} & \text{for } i < d - m + 3 \\ \sum_{j=0}^{d-m} (d - m + 1 - j) \binom{d-1-j}{d-i} (m - 1)^j A(i - 2 - j, m - d + i - 3) & \text{for } i \geq d - m + 3. \end{cases}$$

Moreover, the gradient degree of E_m is just the Eulerian number $\alpha_d = A(d - 1, m - 2)$.

For example, the formula for $d = 6, m = 3$ is in (29). For $d = 7$ we find

$$\begin{aligned} [X_{E_2}] &= 1s^6 + 1s^5t + 1s^4t^2 + 1s^3t^3 + 1s^2t^4 + 1st^5 + 1t^6, \\ [X_{E_3}] &= \mathbf{57}s^6 + 32s^5t + 16s^4t^2 + 8s^3t^3 + 4s^2t^4 + 2st^5 + 1t^6, \\ [X_{E_4}] &= 302s^6 + \mathbf{222}s^5t + 81s^4t^2 + 27s^3t^3 + 9s^2t^4 + 3st^5 + 1t^6, \\ [X_{E_5}] &= 302s^6 + 422s^5t + \mathbf{221}s^4t^2 + 64s^3t^3 + 16s^2t^4 + 4st^5 + t^6, \\ [X_{E_6}] &= 57s^6 + 157s^5t + 170s^4t^2 + \mathbf{90}s^3t^3 + 25s^2t^4 + 5st^5 + 1t^6, \\ [X_{E_7}] &= 1s^6 + 6s^5t + 15s^4t^2 + 20s^3t^3 + \mathbf{15}s^2t^4 + 6st^5 + 1t^6. \end{aligned} \tag{30}$$

The numbers α_i in Theorem 6.2 are instances of the *mixed Eulerian numbers*. We use the following geometric interpretation of these numbers. The k -th *hypersimplex* is the polytope

$$\Delta_k = \text{conv}\{\mathbf{e}_{i_1} + \mathbf{e}_{i_2} + \dots + \mathbf{e}_{i_d} : 1 \leq i_1 < \dots < i_k \leq d\} \subset \mathbb{R}^d.$$

This is a $(d-1)$ -dimensional polytope with $\binom{d}{k}$ vertices. For instance, Δ_1 is a $(d-1)$ -simplex. For $s, t > 0$, the Minkowski sum $s\Delta_k + t\Delta_1$ is a polytope of dimension $d-1$. We consider the volume of $s\Delta_k + t\Delta_1$ with respect to the normalized Lebesgue measure that is defined by $\text{vol}(\Delta_1) = 1$. This volume is a homogeneous polynomial in s and t of degree $d-1$.

Lemma 6.3. *The combinatorial numbers α_i in Theorem 6.2 satisfy*

$$\text{vol}(s\Delta_{m-1} + t\Delta_1) = \sum_{i=1}^d \alpha_i \binom{d-1}{i-1} s^{i-1} t^{d-i}.$$

Equivalently, α_i equals the mixed volume of the polytopes $(\Delta_{m-1}, \dots, \Delta_{m-1}, \Delta_1, \dots, \Delta_1)$, where the hypersimplex Δ_{m-1} appears $i-1$ times and the simplex Δ_1 appears $d-i$ times.

Proof. For $i = d$ this implies $\text{vol}(\Delta_k) = A(d-1, k-1)$. In words, the normalized volume of the hypersimplex is the Eulerian number. This is a classical result due to Laplace. Mixed volumes of hypersimplices were studied recently in [9, 30]. The formula we need is that the mixed volume of $i-1$ copies of Δ_{m-1} and $d-i$ copies of Δ_1 for $i \geq d-m+3$ equals

$$\sum_{j=0}^{d-m} (d-m+1-j) \binom{d-1-j}{d-i} (m-1)^j A(i-2-j, m-d+i-3), \quad (31)$$

where Δ_{m-1} appears $i-1$ times and Δ_1 appears $d-i$ times. Croitoru gave a recursive formula for mixed Eulerian numbers in [9, Theorem 2.4.6]. One checks that the sum in (31) satisfies both, Croitoru's recursion and initial conditions. This proves the lemma. \square

Thus, the content of Theorem 6.2 is that the coefficients of the gradient multidegree $[X_{E_m}]$ are the mixed volumes in Lemma 6.3. We shall prove this using toric geometry [13].

We decompose the gradient map as $\nabla E_m = \pi \circ f$, where $f : \mathbb{CP}^{d-1} \dashrightarrow \mathbb{CP}^{(m-1)^d-1}$ is the map given by all square-free monomials of degree $m-1$, and $\pi : \mathbb{CP}^{(m-1)^d-1} \dashrightarrow \mathbb{CP}^{d-1}$ is the linear projection given by summing the monomials in $\nabla E_m = (E_m^1, E_m^2, \dots, E_m^d)$. The closure of the image of f is the toric variety $X_{\Delta_{m-1}}$ associated to the hypersimplex Δ_{m-1} .

Lemma 6.4. *The linear projection π restricts to a regular map $X_{\Delta_{m-1}} \rightarrow \mathbb{CP}^{d-1}$.*

Proof. We first argue that π has no base points on the dense torus of $X_{\Delta_{m-1}}$. Equivalently, the homogeneous equations $E_m^1(\theta) = E_m^2(\theta) = \dots = E_m^d(\theta) = 0$ have no solutions θ with all coordinates nonzero. We use induction on m , the case $m = 1$ being trivial. Assume the statement is true for m and consider the equations $E_{m+1}^i(\theta) = 0$ for $i = 1, \dots, d$. Summing all the equations we obtain $E_m(\theta) = 0$. Notice that $E_m(\theta) - E_{m+1}^i(\theta) = \theta_i E_m^i(\theta)$. Hence, if $\theta_i \neq 0$ we obtain $E_m^i(\theta) = 0$, which allows us to conclude the proof by induction.

We next consider the other orbits on the toric variety $X_{\Delta_{m-1}}$. Each of these corresponds to a proper face H of Δ_{m-1} . Each face H is obtained by fixing some of the coordinates at 0 or 1, so it is a hypersimplex of smaller dimension. Suppose the torus orbit corresponding to H meets the center of projection of π in a point. Then, by considering its nonzero coordinates, we obtain an inconsistent system of equations, just like in the first paragraph above. \square

Corollary 6.5. *The singular locus of the hypersurface $\{E_m = 0\}$ is the union of all $\binom{d}{m-2}$ coordinate planes of dimension $m-3$ in \mathbb{CP}^{d-1} . Its ideal $\langle E_m^1, E_m^2, \dots, E_m^d \rangle$ defines a scheme that is nonreduced for $m \geq 4$, but smooth outside of the intersection of any two components.*

Proof. The fact that the scheme is smooth outside of the intersection of any two components follows from the computation of the tangent space. The scheme is nonreduced, because the ideal generators span a d -dimensional vector space, while the vector space of forms of degree $m-1$ in the radical ideal of the coordinate subspace arrangement has dimension $\binom{d}{m-1}$. \square

Example 6.6. Let $d = 5$ and $m = 4$. The gradient ideal $\langle \nabla E_4 \rangle = \langle E_4^1, E_4^2, E_4^3, E_4^4, E_4^5 \rangle$ is the intersection of its ten minimal primes $\langle \theta_i, \theta_j, \theta_k \rangle$ and the five embedded primary ideals

$$\langle \theta_i^2, \theta_j^2, \theta_k^2, \theta_l^2, \theta_i\theta_j + \theta_i\theta_k + \theta_j\theta_k, \theta_i\theta_j + \theta_i\theta_l + \theta_j\theta_l, \theta_i\theta_k + \theta_i\theta_l + \theta_k\theta_l, \theta_j\theta_k + \theta_j\theta_l + \theta_k\theta_l \rangle.$$

This defines the ten coordinate lines in \mathbb{CP}^4 , with embedded points at their intersections. \diamond

of Theorem 6.2. Fix the polytope $P = \Delta_{m-1} + \Delta_1$. This is the generalized permutohedron [30] which is the convex hull of all points $(2, 1^{m-2}, 0^{d-m+1}) \in \mathbb{R}^d$. Both polytopes P and Δ_{m-1} are normal, since they are the base polytopes of polymatroids [21, Theorem 6.1]. Consider the three projective toric varieties X_P , $X_{\Delta_{m-1}}$ and $X_{\Delta_1} \simeq \mathbb{CP}^{d-1}$. We have natural rational maps between these varieties, induced from the identification of their dense tori.

Let L_1 be the very ample line bundle on $X_{\Delta_{m-1}}$ that is given by the hypersimplex Δ_{m-1} . Let S be the linear system on \mathbb{CP}^{d-1} that is spanned by the d partial derivatives E_m^i . This induces a linear system S_1 on $X_{\Delta_{m-1}}$. The linear system S_1 is contained in $H^0(X_{\Delta_{m-1}}, L_1)$. From Lemma 6.4 we know that S_1 is base point free.

The normal fan of P is a common refinement of the normal fans of Δ_{m-1} and Δ_1 . Hence, there are regular toric maps from X_P to $X_{\Delta_{m-1}}$ and X_{Δ_1} . Pulling back the line bundle L_1 (resp. $\mathcal{O}_{\mathbb{P}^{d-1}}(1)$) from $X_{\Delta_{m-1}}$ (resp. X_{Δ_1}), we obtain a line bundle \tilde{L}_1 (resp. \tilde{L}_2) on X_P . The pull-back of the system S_1 (resp. $H^0(\mathcal{O}(1))$) is a base point free system \tilde{S}_1 (resp. \tilde{S}_2).

As both, \tilde{S}_1 and \tilde{S}_2 , are base point free, the intersection of the zero locus of $i-1$ generic sections of \tilde{S}_1 and $d-i$ generic sections of \tilde{S}_2 must be contained in the dense torus of X_P . The tori in all three toric varieties are naturally isomorphic. Hence the common zero locus of these sections corresponds to points of the torus of \mathbb{CP}^{d-1} where $i-1$ generic sections of S intersect with $d-i$ generic hyperplanes. By Bertini's Theorem, the intersection in the torus consists of smooth points only. Their number is the coefficient of $s^{i-1}t^{d-i}$ in $[X_{E_m}]$.

From Bernstein's Theorem [13, Section 5.4] we know that the intersection number of these divisors on X_P is equal to the mixed volume of $d-i$ copies of the simplex Δ_1 and $i-1$ copies of the hypersimplex Δ_{m-1} . With this, Lemma 6.3 concludes the proof of Theorem 6.2. \square

It is instructive to examine the various cases in our formula in Theorem 6.2. For $i = d$, we are just looking at $d-1$ copies of the hypersimplex Δ_{m-1} , and the mixed volume becomes $\alpha_d = \text{vol}(\Delta_{m-1}) = A(d-1, m-2)$. At the extreme, for $i = 1$, we obtain $\alpha_1 = \text{vol}(\Delta_1) = 1$.

The case distinction in our formula for α_i is understood via the singular locus of $\{E_m = 0\}$. According to Corollary 6.5, this has dimension $m-3$. If $d-i > m-3$ then the $d-i$ generic hyperplanes in \mathbb{CP}^{d-1} will miss this singular locus. In this case, Bézout's Theorem applies

to the $i - 1$ generic sections of the linear system S in \mathbb{CP}^{d-1} . They will meet in $(m - 1)^{i-1}$ distinct reduced points. This explains the formula $\alpha_i = (m - 1)^{i-1}$ for $i < d - m + 3$.

Of special interest is the borderline case $i = d - m + 3$. Here the $m - 4$ generic hyperplanes meet the singular locus in $\binom{d}{m-2}$ distinct reduced points, one for each coordinate subspace, by Corollary 6.5. This number gets subtracted from the Bézout number, and therefore

$$\alpha_{d-m+3} = (m - 1)^{d-m+2} - \binom{d}{m-2}. \quad (32)$$

This explains the bold face numbers along the borderline diagonal for $m = 7$ in (30).

For $i > d - m + 3$ the intersection is not proper in \mathbb{CP}^{d-1} . It contains the base locus of S , which is the singular locus of $\{E_m = 0\}$, whose dimension strictly dominates the dimension of the other components. This is an instance of *excess intersection*. The same issue occurs for other hyperbolic polynomials $f(\theta)$. For the Gaussian case, when $f(\theta)$ is the symmetric determinant, this is addressed in [36, Theorem 2.3]. For the elementary symmetric polynomial, we resolved this problem by passing to the toric variety $X_{\Delta_{m-1}}$. This worked well here, thanks to Lemma 6.4. However, for other $f(\theta)$, the toric approach will not suffice.

Example 6.7. Let $d = 6$, $m = 3$ and revisit Example 5.6. The gradient degree of $g(\theta) = E_3(\theta)$ is $\text{vol}(\Delta_2) = A(5, 1) = 26$, which is the leading coefficient in (29). The Laplacian determinant $f(\theta)$ is obtained from $g(\theta)$ by deleting four of the 20 terms. Both ∇f and ∇g have the same Newton polytope, namely the second hypersimplex Δ_2 . However, ∇f is birational: the gradient degree of $f(\theta)$ is 1. Lemma 6.4 holds for $g(\theta)$ but it fails dramatically for $f(\theta)$. \diamond

Theorem 6.2 provides an upper bound for the degree and ML degree of the exponential variety $\mathcal{L}^{\nabla E_m}$ defined by any linear subspace $\mathcal{L} \subset \mathbb{R}^d$, by Theorem 5.5. The more special we choose \mathcal{L} , the further away we expect to be from the upper bound α_d . In the next example we explore the range of possibilities for the exponential family seen in Example 1.1.

Example 6.8. Let $d = 4$ and $m = 3$. The cubic surface $\{E_3 = 0\}$ has four singular points. It bounds the convex set C on the left in Figure 1. The dual body K , on the right in Figure 1, is the convex hull of the surface $\{Q = 0\}$ in (1). The gradient multidegree of E_3 equals

$$[X_3] = 4s^3 + 4s^2t + 2st^2 + t^3.$$

The leading coefficient tells us that the map $C \rightarrow K$ is 4-to-1. The second coefficient reveals that the surface $\mathcal{L}^{\nabla E_3}$ has degree 4 whenever \mathcal{L} is a random plane in \mathbb{RP}^3 . However, when the plane \mathcal{L} meets the singular locus of $\{E_3 = 0\}$ then the degree of that surface will drop.

- The plane $\mathcal{L} = \{\theta_1 + \theta_2 = 2\theta_3\}$ contains one singular point. The variety $\mathcal{L}^{\nabla E_3}$ is a cubic surface that is singular along the line $\{4\sigma_1 - \sigma_3 + \sigma_4 = 4\sigma_2 - \sigma_3 + \sigma_4 = 0\}$.
- The plane $\mathcal{L} = \{\theta_1 + \theta_2 + \theta_3 = 0\}$ intersects $\{E_3 = 0\}$ in exactly the same singular point. But now the resulting exponential surface is a plane: $\mathcal{L}^{\nabla E_3} = \{\sigma_1 + \sigma_2 + \sigma_3 = \sigma_4\}$.

- The plane $\mathcal{L} = \{\theta_1 = 2\theta_2\}$ meets $\{E_3 = 0\}$ in two singular points. Now $\mathcal{L}^{\nabla E_3}$ is the quadric $\{45\sigma_1^2 - 63\sigma_1\sigma_2 + 18\sigma_2^2 + 18\sigma_1\sigma_3 - 9\sigma_2\sigma_3 + \sigma_3^2 + 18\sigma_1\sigma_4 - 9\sigma_2\sigma_4 - 2\sigma_3\sigma_4 + \sigma_4^2 = 0\}$.
- The plane $\mathcal{L} = \{\theta_1 = 0\}$ spanned by three singular points maps to the facet $\mathcal{L}^{\nabla E_3} = \{\sigma_1 = \sigma_2 + \sigma_3 + \sigma_4\}$ of the octahedron inside K that was mentioned in Example 1.1.

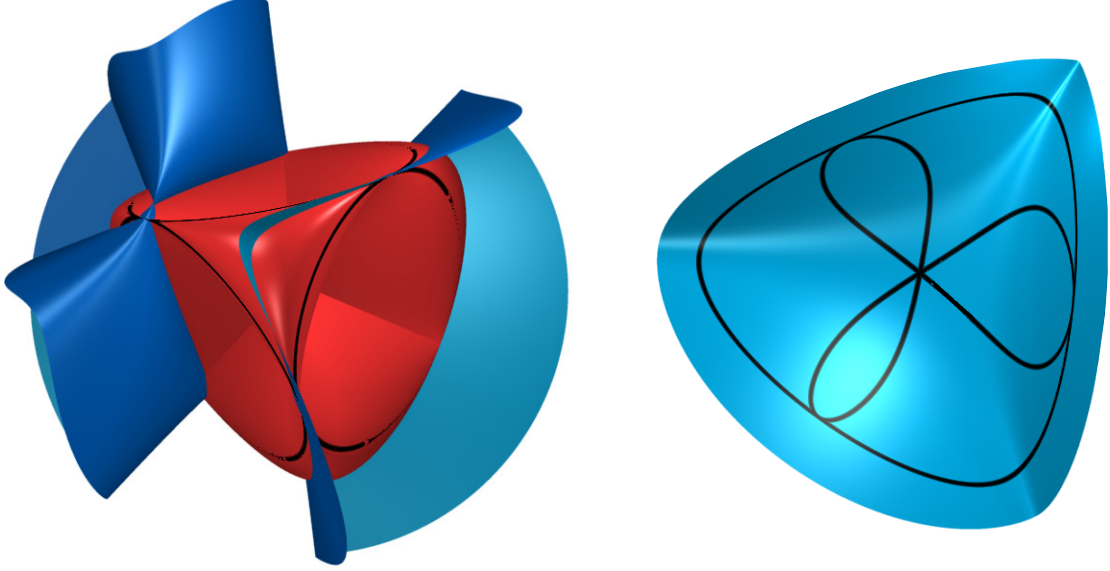


Figure 4: The positive exponential variety is one connected component in the intersection of the exponential variety (blue) with the convex body of sufficient statistics (red).

The generic case is exhibited by the plane in Figure 2, which is $\mathcal{L} = \{\theta_1 + 2\theta_2 + 3\theta_3 = 6\theta_4\}$. Figure 4 is the dual of Figure 2. The exponential variety $\mathcal{L}^{\nabla E_3}$ is the quartic surface shown in blue. We see that $\mathcal{L}^{\nabla E_3}$ intersects the convex body K (shown in red in Figure 4) in two connected components. The positive exponential variety $\mathcal{L}_{>0}^{\nabla E_3}$ is the component that cuts off the rightmost red circle in Figure 4 (left) from the other three. That component is shown in Figure 4 (right), together with the positive exponential variety $\mathcal{L}_{>0}^{\nabla E_3}$. The curve indicates where $\mathcal{L}^{\nabla E_3}$ intersects the Steiner surface. It has degree six. The outer curve bounds $\mathcal{L}_{>0}^{\nabla E_3}$. The projection $\pi_{\mathcal{L}}$ flattens out $\mathcal{L}_{>0}^{\nabla E_3}$ and maps it bijectively onto the planar convex set $\mathcal{K}_{\mathcal{L}}$.

Next consider the case when \mathcal{L} is a line in \mathbb{CP}^3 . Here $[X_3]$ tells us that the curve $\mathcal{L}^{\nabla E_3}$ is expected to have degree 2. This drops to 1 if \mathcal{L} contains one of the four singular points. However, the image $\mathcal{L}^{\nabla E_3}$ can also degenerate to a double line for special lines \mathcal{L} that are disjoint from the singular locus of $\{E_3 = 0\}$. For such a line \mathcal{L} , the restriction of the map ∇E_3 to \mathcal{L} has degree 2. One such special line is $\mathcal{L} = \{\theta_1 = \theta_2, \theta_3 = \theta_4\}$, which maps to $\mathcal{L}^{\nabla E_3} = \{\sigma_1 = \sigma_2, \sigma_3 = \sigma_4\}$. Finally, if \mathcal{L} is one of the six coordinate lines, then $\mathcal{L}^{\nabla E_3}$ is a point. In our picture this means that the edges of the tetrahedron in C contract to the vertices of the octahedron in K . \diamond

Example 6.9. Let $d = 5$. It is instructive to ponder the geometry of the convex bodies C and K and how these 4-dimensional bodies encompass their various exponential varieties.

First consider $m = 4$. Here the situation is analogous to Examples 1.1 and 6.8. The body K consists of all points $\sigma \in \mathbb{RP}^4$ where σ_5 is the sum of all entries in a positive semidefinite 4×4 -matrix with diagonal $(\sigma_1, \sigma_2, \sigma_3, \sigma_4)$. The map from C to K is 11-to-1, since

$$[X_4] = 11s^4 + 17s^3t + 9s^2t^2 + 3st^3 + t^4.$$

The main component in the algebraic boundary of K is the hypersurface dual to $\{E_4 = 0\}$. This threefold has degree 8 and it touches each of the five facets of the ambient 4-simplex in a Steiner surface (1). For a general hyperplane $\mathcal{L} \subset \mathbb{CP}^4$, the exponential variety $\mathcal{L}^{\nabla E_4}$ is a threefold of degree 17. If \mathcal{L} is a general 2-plane, then $\mathcal{L}^{\nabla E_4}$ is a surface of degree 9.

Next consider $m = 3$. Now K is strictly contained in the previous body, and ∂K meets each tetrahedral facet of the ambient 4-simplex in an inscribed sphere, such as $\{\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \sigma_4^2 = \sigma_1\sigma_2 + \sigma_1\sigma_3 + \sigma_1\sigma_4 + \sigma_2\sigma_3 + \sigma_2\sigma_4 + \sigma_3\sigma_4\}$. The map from C to K is 11-to-1, since

$$[X_3] = 11s^4 + 8s^3t + 4s^2t^2 + 2st^3 + t^4.$$

The main component in ∂K is the hypersurface dual to $\{E_3 = 0\}$, which is defined by a polynomial of degree 14 with 3060 terms. Setting $\sigma_5 = 0$, the resulting surface decomposes as twice the inscribed sphere times an irreducible surface of degree 10. This suggests that there is lots of interesting geometry yet to be discovered in the cones of sufficient statistics for the hyperbolic exponential families given by elementary symmetric polynomials for $d \geq 5$. \diamond

7 Hankel Models

The paradigm of an exponential variety is the set of symmetric matrices whose inverses satisfy linear constraints. These arise from the exponential family structure on the multivariate normal distributions. Here the hyperbolic polynomial $f(\theta)$ is the determinant of a symmetric $m \times m$ -matrix of unknowns. The resulting *linear Gaussian concentration models* were studied in [1] and their geometry in [36]. In this paper we encountered them in Example 4.9.

In this section we present a beautiful instance that is related to numerous topics in the mathematical sciences. Let $\mathbb{S}^m \simeq \mathbb{R}^{\binom{m+1}{2}}$ denote the space of symmetric $m \times m$ -matrices, and $C = \text{PD}_m$ the positive definite cone. Let \mathcal{L} be the $(2m - 1)$ -dimensional subspace of all *Hankel matrices* $(\theta_{i+j})_{1 \leq i, j \leq m}$. Then $C_{\mathcal{L}} = C \cap \mathcal{L}$ is the cone of positive definite Hankel matrices.

Proposition 7.1. *The closure of $C_{\mathcal{L}}$ consists of the vectors of $2m - 1$ first moments of all probability distributions on \mathbb{R} . Its dual $K_{\mathcal{L}} = (C_{\mathcal{L}})^{\vee}$ is the cone of all nonnegative polynomials of degree $2m - 2$ in one variable x . The polynomials in $K_{\mathcal{L}}$ are precisely the sums of squares.*

Proof. The first statement is [4, Theorem 3.146]. For the second statement see [4, Section 3.5.1]. Every nonnegative real polynomial in one variable x is a sum of squares in $\mathbb{R}[x]$. \square

We shall see in Proposition 7.2 that the exponential variety $\mathcal{L}^{\nabla f}$ equals the Grassmannian $\text{Gr}(2, m + 1)$ of 2-dimensional linear subspaces of \mathbb{R}^{m+1} , in its Plücker embedding in

$\mathbb{RP}^{\binom{m+1}{2}-1}$. The positive Grassmannian $\mathcal{L}_{>0}^{\nabla f} = \text{Gr}(2, m+1)_{>0}$ is in natural bijection with the cone $K_{\mathcal{L}}$.

Hankel matrices are closely related to *Toeplitz matrices*. In fact, under renaming the coordinates in \mathbb{S}^m , they represent the same subspace \mathcal{L} . See [15] for an introduction to Toeplitz matrices from the engineering perspective. In statistics, Toeplitz matrices are an instance of a *colored Gaussian graphical model* as in [22] and [36, Section 5]. The results in this section are stated for Hankel matrices, but the story would be isomorphic for Toeplitz matrices.

To match the notation with Section 4, we now have $d = \binom{m+1}{2}$ and $c = 2m - 1$. The map $\pi_{\mathcal{L}} : \mathbb{R}^d \rightarrow \mathbb{R}^c$ that is dual to the inclusion $\mathcal{L} \subset \mathbb{R}^d$ can be described as follows. We identify the domain \mathbb{R}^d with \mathbb{S}^m , now regarded as the ambient space for covariance matrices $\Sigma = (\sigma_{ij})$. We identify the target \mathbb{R}^c with the space $\mathbb{R}[x]_{\leq 2m-2}$ of univariate polynomials of degree at most $2m - 2$. Each symmetric matrix Σ is mapped to such a polynomial as follows:

$$\pi_{\mathcal{L}} : \Sigma \mapsto (1, x, x^2, \dots, x^{m-1}) \cdot \Sigma \cdot (1, x, x^2, \dots, x^{m-1})^T. \quad (33)$$

The image $K_{\mathcal{L}}$ of the positive semidefinite cone $K = \overline{\text{PD}}_m$ under the map $\pi_{\mathcal{L}}$ consists of all polynomials that are sums of squares in $\mathbb{R}[x]$. These are precisely the nonnegative polynomials.

The following result is a reformulation of a well-known result in linear algebra, stating that the inverses of Hankel matrices are precisely the Bézout matrices; see [20, Section 6].

Proposition 7.2. *After a linear change of coordinates in $\mathbb{R}^d \simeq \mathbb{S}^m$, the exponential variety \mathcal{L}^{-1} of all inverse Hankel matrices coincides with (the affine cone over) the Grassmannian $\text{Gr}(2, m+1)$. In particular, the degree of \mathcal{L}^{-1} in \mathbb{CP}^{d-1} is the Catalan number $\frac{1}{m} \binom{2m-2}{m-1}$.*

Proof. We here write $\text{Gr}(2, m+1)$ for the cone over the Grassmannian. This is an affine variety of dimension $2m - 1$ in the ambient space $\mathbb{R}^d = \mathbb{R}^{\binom{m+1}{2}}$. The Plücker coordinates are denoted p_{ij} for $0 \leq i < j \leq m$, and these satisfy the quadratic Plücker relations

$$p_{il}p_{jk} - p_{ik}p_{jl} + p_{ij}p_{kl} = 0 \quad \text{for } 0 \leq i < j < k < l \leq m. \quad (34)$$

The degree of $\text{Gr}(2, m+1)$ is the *Catalan number* $\frac{1}{m} \binom{2m-2}{m-1}$, e.g. by [35, Proposition 3.7.4].

For each Plücker vector $p \in \text{Gr}(2, m+1)$ we write the corresponding Bézout matrix as $\Sigma = (\sigma_{ij})$. This $m \times m$ matrix is symmetric and its entries are thought of as covariances:

$$\sigma_{ij} = \sum_{k=\max(1, i+j-m)}^i p_{k, i+j+1-k} \quad \text{for } 1 \leq i \leq j \leq m. \quad (35)$$

These linear equations have an upper triangular structure, so they can be inverted to express the Plücker coordinates p_{ij} in terms of the Bézout coordinates σ_{ij} . This is the promised linear change of coordinates in $\mathbb{R}^{\binom{m+1}{2}}$. In particular, we can write the quadrics (34) in the σ_{ij} . \square

This defines the structure of an exponential variety on the Grassmannian $\text{Gr}(2, m+1)$. The positive exponential variety $\text{Gr}(2, m+1)_{>0}$ consists of all solutions to (34) such that $\Sigma = (\sigma_{ij})$ is positive definite. This model is in bijection with the cone $K_{\mathcal{L}}$ of nonnegative polynomials of degree at most $2m - 2$ under the sufficient statistics map $\pi_{\mathcal{L}}$ in (33).

Example 7.3. Let $m = 4$. The Plücker coordinates in terms of the Bézout coordinates are

$$\begin{aligned} p_{12} &= \sigma_{12}, & p_{13} &= \sigma_{13}, & p_{14} &= \sigma_{14}, & p_{15} &= \sigma_{15}, & p_{23} &= \sigma_{23} - \sigma_{14}, \\ p_{24} &= \sigma_{24} - \sigma_{15}, & p_{25} &= \sigma_{25}, & p_{34} &= \sigma_{34} - \sigma_{25}, & p_{35} &= \sigma_{35}, & p_{45} &= \sigma_{45}. \end{aligned}$$

After this substitution, the ideal of Plücker relations for $m = 4$ equals

$$\langle \sigma_{11}\sigma_{24} - \sigma_{11}\sigma_{33} - \sigma_{12}\sigma_{14} + \sigma_{12}\sigma_{23} + \sigma_{13}^2 - \sigma_{13}\sigma_{22}, \sigma_{13}\sigma_{44} - \sigma_{14}\sigma_{34} - \sigma_{22}\sigma_{44} + \sigma_{23}\sigma_{34} + \sigma_{24}^2 - \sigma_{24}\sigma_{33}, \\ \sigma_{11}\sigma_{34} - \sigma_{12}\sigma_{24} - \sigma_{13}\sigma_{14} + \sigma_{14}\sigma_{22}, \sigma_{11}\sigma_{44} - \sigma_{13}\sigma_{24} - \sigma_{14}^2 + \sigma_{14}\sigma_{23}, \sigma_{12}\sigma_{44} - \sigma_{13}\sigma_{34} - \sigma_{14}\sigma_{24} + \sigma_{14}\sigma_{33} \rangle.$$

Thus, for the space \mathcal{L} of 4×4 -Hankel matrices, we have $\mathcal{L}^{-1} = \text{Gr}(2, 5)$. The positive exponential variety $\mathcal{L}_{>0}^{-1} = \text{Gr}(2, 5)_{>0}$ is our positive Grassmannian. It consists of all positive definite symmetric 4×4 -matrices $\Sigma = (\sigma_{ij})$ that satisfy these five quadrics. The bijection

$$\Sigma \mapsto \sum_{i=1}^4 \sum_{j=1}^4 \sigma_{ij} x^{i+j-2}$$

takes the set $\text{Gr}(2, 5)_{>0}$ to the 7-dimensional convex cone of all nonnegative sextics. \diamond

Maximum likelihood estimation means inverting the map from the positive Grassmannian $\text{Gr}(2, 5)_{>0}$ onto $K_{\mathcal{L}}$. This inverse takes a nonnegative polynomial of degree $2m - 2$ to the analytic center of its *Gram spectrahedron* [37]. We prove that this algebraic function from $K_{\mathcal{L}}$ to $\mathcal{L}_{>0}^{-1}$ has the expected degree, meaning that the left inequality in (28) holds with equality.

Proposition 7.4. *The ML degree of the Grassmannian $\text{Gr}(2, m + 1)$, in its guise as exponential variety \mathcal{L}^{-1} of $m \times m$ inverse Hankel matrices, is the Catalan number $\frac{1}{m} \binom{2m-2}{m-1}$.*

Proof. Our claim states that equality holds for the left inequality in Theorem 5.5 when f is the symmetric determinant and \mathcal{L} is the space of Hankel matrices. In order to prove this, we must show that $\mathcal{L}^{\nabla f} = \text{Gr}(2, m + 1)$ is disjoint from \mathcal{L}^{\perp} . Note that $\mathcal{L}^{\perp} = \{\ell_1 = \ell_2 = \dots = \ell_{2m-1} = 0\}$, where ℓ_i is the sum of all unknowns on the i -th antidiagonal of the matrix σ . Thus $\ell_i = 2 \sum_{j=1}^{i/2} \sigma_{j, i-j+1}$ when i is even, and $\ell_i = \sigma_{\frac{i+1}{2}, \frac{i+1}{2}} + 2 \sum_{j=1}^{\lfloor i/2 \rfloor} \sigma_{j, i-j+1}$ when i is odd.

We now pass to Plücker coordinates p_{ij} for the Grassmannian $\text{Gr}(2, m + 1)$. The Plücker relations (34) are homogeneous with respect to the \mathbb{N} -grading given by $\text{degree}(p_{jk}) = j + k$. By (35), the linear form ℓ_i is a positive integer linear combination of all Plücker coordinates of degree i . Equivalently, in the graded poset that underlies the straightening law for $\text{Gr}(2, m + 1)$ as in [28, Section 14.3], the linear form ℓ_i is a positive sum of all poset elements of height i . This implies that $\{\ell_1, \dots, \ell_{2m-1}\}$ is a linear system of parameters modulo the Stanley-Reisner ideal of that poset. That ideal is generated by the underlined leading monomials in (34). From this it follows that $\{\ell_1, \dots, \ell_{2m-1}\}$ is a linear system of parameters modulo the ideal of Plücker relations. This is equivalent to $\text{Gr}(2, m + 1) \cap \mathcal{L}^{\perp} = \emptyset$ in $\mathbb{CP}^{\binom{m+1}{2}-1}$, as desired. \square

It would be very interesting to extend this analysis to Hankel matrices of polynomials in more than one variable. While we do not yet have any results to report on this problem, let us still close our paper by describing the corresponding linear Gaussian concentration model. In our view, this may become an important object for convex algebraic geometry [4, 36].

We fix two arbitrary positive integers r and s , and we set $m = \binom{r+s-1}{s-1}$ and $c = \binom{2r+s-1}{s-1}$. These are the numbers of monomials in s variables of degree r and $2r$ respectively. Each matrix θ in \mathbb{S}^m has its rows and columns labeled with the monomials of degree r . A matrix $\theta \in \mathbb{S}^m$ is a *Hankel matrix* if its entries are indexed by the product of the row label and the column label. The set \mathcal{L} of all Hankel matrices is a subspace of dimension c in \mathbb{S}^m .

For instance, for $r = 2, s = 3, m = 6, c = 15$, with row and column labels written additively as $(200), (020), (002), (110), (101), (011)$, the space of Hankel matrices equals

$$\mathcal{L} = \left\{ \begin{pmatrix} \theta_{400} & \theta_{220} & \theta_{202} & \theta_{310} & \theta_{301} & \theta_{211} \\ \theta_{220} & \theta_{040} & \theta_{022} & \theta_{130} & \theta_{121} & \theta_{031} \\ \theta_{202} & \theta_{022} & \theta_{004} & \theta_{112} & \theta_{103} & \theta_{013} \\ \theta_{310} & \theta_{130} & \theta_{112} & \theta_{220} & \theta_{211} & \theta_{121} \\ \theta_{301} & \theta_{121} & \theta_{103} & \theta_{211} & \theta_{202} & \theta_{112} \\ \theta_{211} & \theta_{031} & \theta_{013} & \theta_{121} & \theta_{112} & \theta_{022} \end{pmatrix} : \theta \in \mathbb{R}^{15} \right\}. \quad (36)$$

Thus $C_{\mathcal{L}}$ is the spectrahedral cone consisting of all positive definite Hankel matrices.

The dual space $\mathbb{R}^d/\mathcal{L}^\perp \simeq \mathbb{R}^c$ is identified with the space of homogeneous polynomials of degree $2r$ in s variables. The dual cone $K_{\mathcal{L}} = (C_{\mathcal{L}})^\vee$ consists of those polynomials that are sums of squares of polynomials of degree r , by [4, Corollary 4.36].

Hilbert's classical theorem (see [4, Theorem 4.3]) states that the sum-of-squares cone $K_{\mathcal{L}}$ coincides with the cone of polynomials that are nonnegative on \mathbb{R}^s if and only if $r = 1$, $s = 2$, or $(r = 2 \text{ and } s = 3)$. In all other cases, the latter cone strictly contains $K_{\mathcal{L}}$. The exponential variety \mathcal{L}^{-1} consists of all inverse Hankel matrices, and its positive part $\mathcal{L}_{>0}^{-1}$ maps bijectively onto the cone $K_{\mathcal{L}}$. The inverse map takes each sum-of-squares polynomial p to a distinguished sum-of-squares representation \hat{p} . Namely, \hat{p} is the analytic center in the Gram spectrahedron of p , which is the fiber over p of the map $\pi_{\mathcal{L}}$ from $\overline{\text{PD}}_m$ onto $K_{\mathcal{L}}$. The following problem asks for a higher-dimensional generalization of Propositions 7.2 and 7.4.

Problem 7.5. *Determine the degree and the ML degree of the exponential variety \mathcal{L}^{-1} of inverse Hankel matrices for $s \geq 3$ variables, and study its defining polynomial equations.*

For example, for $r = 2, s = 3$, we start with the linear subspace $\mathcal{L} \subset \mathbb{RP}^{20}$ given in (36). Its exponential variety \mathcal{L}^{-1} consists of the inverses of all such Hankel matrices. This is a projective variety of dimension 14 in \mathbb{CP}^{20} , whose positive part maps bijectively onto the cone $K_{\mathcal{L}}$ of all nonnegative ternary quartics. That cone was described in [33, Theorem 6.2].

Acknowledgments. All four authors were hosted by the Simons Institute for the Theory of Computing in Berkeley during the Fall 2014 program *Algorithms and Complexity in Algebraic Geometry*. We thank Petter Brändén, June Huh, Mario Kummer, Gregorio Malajovich and Donald Richards for helpful discussions.

References

- [1] T.W. Anderson: *Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices*, In Bose, R. C., Chakravarti, I. M., Maha-

- Ianobis, P. C., Rao, C. R., and Smith, K. J. C., editors, *Essays in Probability and Statistics* (1970) 1–24.
- [2] K. Aomoto: *On the structure of integrals of power products of linear functions*, Scientific Papers of the College of General Education, University of Tokyo, **27** (1977) 49–61.
 - [3] D. Bates, J. Hauenstein, A. Sommese and C. Wampler: *Numerically Solving Polynomial Systems with Bertini*, Software, Environments and Tools, **25**, SIAM, Philadelphia, 2013.
 - [4] G. Blekherman, P. Parrilo and R. Thomas: *Semidefinite Optimization and Convex Algebraic Geometry*, MOS-SIAM Series on Optimization, **13**, SIAM, Philadelphia, 2013.
 - [5] P. Brändén: *Hyperbolicity cones of elementary symmetric polynomials are spectrahedral*, Optimization Letters, **8** (2014) 1773–1782.
 - [6] L.D. Brown: *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*, Institute of Mathematical Statistics, Hayward, CA, 1986.
 - [7] P. Bürgisser, J.M. Landsberg, L. Manivel, and J. Weyman: *An overview of mathematical issues arising in the geometric complexity theory approach to VP vs. VNP*, SIAM Journal on Computing **40** (2011) 1179–1209.
 - [8] G. Choquet: *Deux exemples classiques de représentation intégrale*, Enseignement Math. **15** (1969) 63–75.
 - [9] D. Croitoru: *Mixed Volumes of Hypersimplices, Root Systems and Shifted Young Tableaux*, PhD Dissertation, Massachusetts Institute of Technology, 2010.
 - [10] A.P. Dempster: *Covariance selection*, Biometrics **28** (1972) 157–175.
 - [11] J. De Loera, B. Sturmfels and C. Vinzant: *The central curve in linear programming*, Foundations of Computational Mathematics **12** (2012) 509–540.
 - [12] M. Drton and S. Sullivant: *Algebraic statistical models*, Statistica Sinica **17** (2007) 1273–1297.
 - [13] W. Fulton: *Introduction to Toric Varieties*, Princeton University Press, 1993.
 - [14] L. Gårding: *Linear hyperbolic partial differential equations with constant coefficients*, Acta Mathematica **85** (1951) 1–62.
 - [15] R.M. Gray: *Toeplitz and circulant matrices: A review*. Foundations and Trends in Communications and Information Theory **2** (2006) 155–239.
 - [16] D.R. Grayson and M.E. Stillman: *Macaulay2, a software system for research in algebraic geometry*, available at <http://www.math.uiuc.edu/Macaulay2/>.
 - [17] O. Güler: *Barrier functions in interior point methods*, Mathematics of Operations Research **21** (1996) 860–885.
 - [18] O. Güler: *Hyperbolic polynomials and interior point methods for convex programming*, Mathematics of Operations Research **22** (1997) 350–377.

- [19] H. Hashiguchi, Y. Numata, N. Takayama and A. Takemura: *The holonomic gradient method for the distribution function of the largest root of a Wishart matrix*, Journal of Multivariate Analysis **117** (2013) 296–312.
- [20] U. Helmke and P. Fuhrmann: *Bezoutians*, Linear Algebra and its Applications **122** (1989) 1039–1097.
- [21] J. Herzog and T. Hibi: *Discrete polymatroids*, Journal of Algebraic Combinatorics **16** (2002) 239–268.
- [22] S. Højsgaard and S.L. Lauritzen: *Graphical Gaussian models with edge and vertex symmetries*, Journal of the Royal Statistical Society: Series B (Statistical Methodology) **5** (2008) 1005–1027.
- [23] J. Huh and B. Sturmfels: *Likelihood geometry*, in: Combinatorial Algebraic Geometry (eds. Aldo Conca et al.), Lecture Notes in Mathematics 2108, Springer, (2014) 63–117.
- [24] S. T. Jensen: *Covariance hypotheses which are linear in both the covariance and the inverse covariance*, Annals of Statistics **16** (1988) 302–322.
- [25] M. Kummer: *A note on the hyperbolicity cone of the specialized Vámos polynomial*, [arXiv:1306.4483](#).
- [26] M. Kummer, D. Plaumann and C. Vinzant: *Hyperbolic polynomials, interlacers, and sums of squares*, Mathematical Programming, Series B, to appear, [arXiv:1212.6696](#).
- [27] S.L. Lauritzen: *Graphical Models*, Oxford University Press, 1996.
- [28] E. Miller and B. Sturmfels: *Combinatorial Commutative Algebra*, Graduate Texts in Mathematics, Springer Verlag, New York, 2004.
- [29] L. Pachter and B. Sturmfels: *Algebraic Statistics for Computational Biology*, Cambridge University Press, 2005.
- [30] A. Postnikov: *Permutohedra, associahedra, and beyond*, International Mathematics Research Notices **6** (2009) 1026–1106.
- [31] J. Renegar: *A mathematical view of interior-point methods in convex optimization*, MOS-SIAM Series on Optimization, SIAM, 2001.
- [32] J. Rauh, T. Kahle and N. Ay: *Support sets in exponential families and oriented matroid theory*, International Journal of Approximate Reasoning **52** (2011) 613–626.
- [33] R. Sanyal, F. Sottile and B. Sturmfels: *Orbitopes*, Mathematika **57** (2011) 275–314.
- [34] A. Scott and A. Sokal: *Complete monotonicity for inverse powers of some combinatorially defined polynomials*, Acta Mathematica **213** (2014) 323–392.
- [35] B. Sturmfels: *Algorithms in Invariant Theory*, Springer, Vienna and New York, 1993.
- [36] B. Sturmfels and C. Uhler: *Multivariate Gaussians, semidefinite matrix completion, and convex algebraic geometry*, Annals of the Institute of Statistical Mathematics **62** (2010) 603–638.

- [37] E. Tsukerman: *Dimension of Gram spectrahedra of univariate polynomials*, [arXiv:1504.04101](#).
- [38] V. A. Vasil'ev, I. M. Gel'fand and A. V. Zelevinskii: *General hypergeometric functions on complex Grassmannians*, *Functional Analysis and Its Applications* **21** (1987) 19–31.

Author's addresses:

Mateusz Michałek, wajcha@berkeley.edu
 Mathematical Institute, Polish Academy of Sciences, Warsaw, 00-656, Poland
 Department of Mathematics, University of California, Berkeley, CA 94720-3840, USA

Bernd Sturmfels, bernd@berkeley.edu
 Department of Mathematics, University of California, Berkeley, CA 94720-3840, USA

Caroline Uhler, caroline.uhler@ist.ac.at
 Institute of Science and Technology Austria, 3400 Klosterneuburg, Austria
 Massachusetts Institute of Technology, EECS, Cambridge, MA 02139-4307, USA

Piotr Zwiernik, piotr.zwiernik@gmail.com
 Dipartimento di Matematica, Università di Genova, 16146 Genova, Italy
 Department of Economics and Business, Universitat Pompeu Fabra, 08005 Barcelona, Spain