Exponential varieties

Piotr Zwiernik
(j.w. Mateusz Michałek, Bernd Sturmfels, and Caroline Uhler)

Universitat Pompeu Fabra

Algebraic Geometry Seminar, UPC
26 May 2017

Motivation 1: Toric Geometry

A central theme in algebraic statistics is the tight connection between toric varieties and discrete exponential families.

Example (Independence of two binary random variables)

The Segre variety $\mathcal{V} = \mathbb{P}^1 \times \mathbb{P}^1 \subset \mathbb{P}^3$ is defined by

$$\det \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} = 0.$$

The moment map takes $\mathcal{V}$ onto K = the square $= \Delta_1 \times \Delta_1$.

The restriction to the model computes the sufficient statistics

$$\mathcal{V}_{\geq 0} \longrightarrow K \qquad (\text{where } \mathcal{V}_{\geq 0} = \mathcal{V} \cap \mathbb{RP}^3_{\geq 0})$$

This map is invertible. Its inverse is the maximum likelihood estimator.

Motivation 2: Gaussian geometry

Let $\mathrm{Sym}_2(\mathbb{R}^m)$ be the space of real symmetric $m \times m$ matrices; and $\mathcal{L}$ be its linear subspace. Consider the variety

$$\mathcal{L}^{-1} = \overline{\{\sigma \in \mathbb{P}(\mathrm{Sym}_2(\mathbb{R}^m)) : \ \sigma^{-1} \in \mathcal{L}\}}.$$

The Gaussian model is the subset of covariance matrices

$$\mathcal{L}^{-1}_{\succ 0} = \{\sigma \in \mathcal{L}^{-1} : \ \sigma \text{ is positive definite}\}.$$

Example (Gaussian graphical models)

$\mathcal{L}$ encodes the sparsity of an undirected graph with m nodes.
The map dual to $\mathcal{L} \hookrightarrow \mathbb{P}(\mathrm{Sym}_2(\mathbb{R}^m))$ computes sufficient statistics.
Its restriction maps the model to the dual of the spectrahedron:

$$\mathcal{L}^{-1}_{\succ 0} \ \longrightarrow \ K \ = \ (\mathcal{L}_{\succ 0})^{\vee}.$$

This map is invertible. Its inverse is the maximum likelihood estimator.

(Statistical) Exponential families

- $(\mathcal{X}, \nu, T)$ where $\mathcal{X}$ sample space, ν a measure on $\mathcal{X}$, $T : \mathcal{X} \to \mathbb{R}^d$.
- Exponential family is a parametric statistics model

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

- The identity $\int_{\mathcal{X}} p_\theta(x) \nu(\mathrm{d}x) = 1$ implies

$$A(\theta) \;\; = \;\; \log \int_{\mathcal{X}} \exp\big(-\langle \theta, T(x) \rangle\big) \, \nu(\mathrm{d}x)$$

The following two sets are convex:

- canonical parameters $\;C := \{\theta \in \mathbb{R}^d : \; A(\theta) < +\infty\}$.
- mean parameters $\qquad K := \mathrm{conv}(T(\mathcal{X})) \subset \mathbb{R}^d$

Theorem: Suppose C is open and K spans $\mathbb{R}^d$. Then the gradient map:

$$F : \; \mathbb{R}^d \longrightarrow \mathbb{R}^d, \qquad \theta \; \mapsto \; -\nabla A(\theta)$$

is an analytic bijection between C and the $\mathrm{int}(K)$.

$\nabla A(\theta)$ is a linear form on $\mathbb{R}^d$ and so naturally lives in the dual space.

Gaussian Exponential family

Let $\mathcal{X} = \mathbb{R}^m$, where ν is the Lebesgue measure, and set

$$T(x) \;=\; \frac{1}{2}x \cdot x^T \;\in\; \mathrm{Sym}_2(\mathbb{R}^m) \simeq \mathbb{R}^d, \quad d = \binom{m+1}{2}.$$

We have $\langle \theta, T(x) \rangle = \frac{1}{2}x^T\theta x$ and, if θ is positive definite,

$$\int_{\mathbb{R}^m} e^{-\frac{1}{2}x^T\theta x}\mathrm{d}x \;=\; (2\pi)^{m/2} \det \theta^{-1/2}.$$

- $A(\theta) \;=\; \log \int_{\mathbb{R}^d} e^{-\frac{1}{2}x^T\theta x}\mathrm{d}x \;=\; \frac{p}{2}\log(2\pi) - \frac{1}{2}\log\det(\theta)$

The two convex sets

- $C =$ positive definite matrices
- $K = \mathrm{conv}(T(\mathcal{X})) =$ positive semidefinite matrices

The gradient map, $\nabla A(\theta) = \frac{1}{2}\theta^{-1}$

- $F : \mathrm{Sym}_2(\mathbb{R}^m) \to \mathrm{Sym}_2(\mathbb{R}^m), \quad \theta \mapsto \frac{1}{2}\theta^{-1}.$

From Analysis toward Algebra

- Our exponential families satisfy

$$A(\theta) = -\alpha \cdot \log(f(\theta)),$$

  where $f(\theta)$ is a homogeneous polynomial and $\alpha > 0$.

- The gradient of the log-partition function is the rational function

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

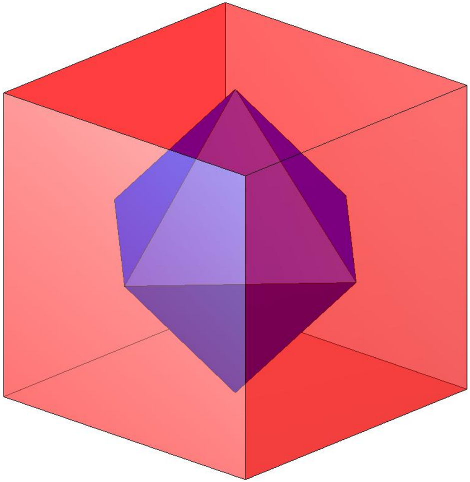
- Algebraic geometers will prefer to look at

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

The partition function $f(\theta)^\alpha$ admits a nice integral representation.
Which polynomials $f(\theta)$ and convex sets $C, K \subset \mathbb{R}^d$ are possible?

Duality of polytopes

Example (How to morph a cube into an octahedron?)



[Sturmfels,Uhler 2010, Example 3.5]

Duality of polytopes

Example (How to morph a cube into an octahedron?)

Fix the product of *linear* forms

$$f(\theta) \;=\; (\theta_1^2 - \theta_4^2)(\theta_2^2 - \theta_4^2)(\theta_3^2 - \theta_4^2).$$

The space of canonical parameters is

$$C \quad = \quad \text{cone over the 3-cube } \{|\theta_i| < 1 : \; i = 1, 2, 3\}.$$

The space of sufficient statistics is

$$K \quad = \quad \text{cone over the octahedron } \mathrm{conv}\{\pm e_1, \pm e_2, \pm e_3\}.$$

The gradient map $F = \nabla f \,:\, \mathbb{P}^3 \dashrightarrow \mathbb{P}^3$ gives a bijection between C and $\mathrm{int}(K)$. Its inverse is an algebraic function of degree 7.

Question: What is $(\mathcal{X}, \nu, T)$ in this case?

Answer: Non-trivial to construct: generalized hypergeometric functions.

Hyperbolic polynomials

- S symmetric matrix, then $g(t) = \det(S - tI)$ has only real roots

A homogeneous polynomial $f \in \mathbb{R}[\theta_1, \ldots, \theta_d]$ is hyperbolic with respect $u \in \mathbb{R}^d$ if and only if $f(u) > 0$ and for every $\theta \in \mathbb{R}^d$ the univariate polynomial

$$g(t) = f(\theta - tu) \in \mathbb{R}[t]$$

has only real roots. The connected component C of u in $\mathbb{R}^d \setminus \{f = 0\}$ is the hyperbolicity cone. It is convex.

Theorem (Gårding 1951 . . . Scott-Sokal 2015)

If $\alpha > d$, there exists a measure ν on the cone $K = C^\vee$ such that

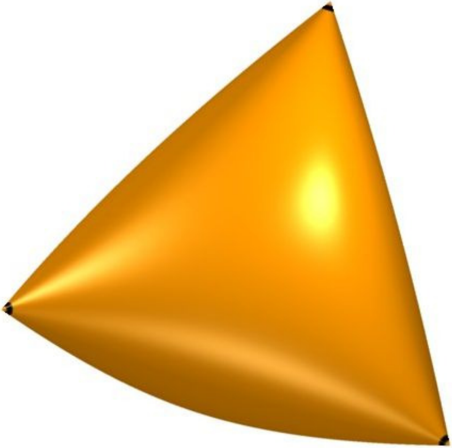
$$f(\theta)^{-\alpha} \;=\; \int_K \exp(-\langle \theta, \sigma \rangle)\nu(\mathrm{d}\sigma) \qquad \text{for all } \theta \in C.$$

Furthermore, this property characterizes hyperbolic polynomials.
This is related to hyperbolic programming in convex optimization.

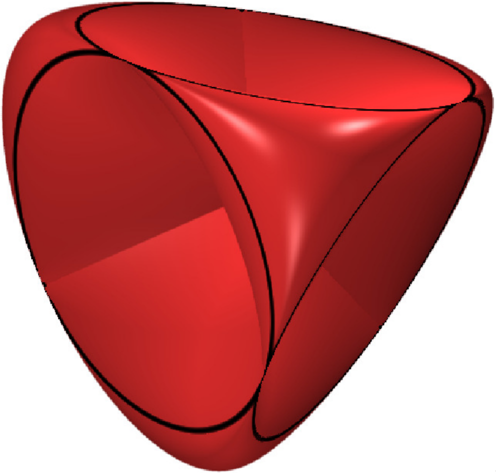
Hyperbolic Exponential Families: An Example

The space of canonical parameters C is the hyperbolicity cone of

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



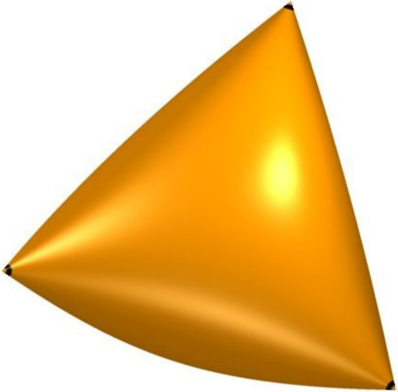
Its dual $K = C^\vee$ is the space of sufficient statistics:

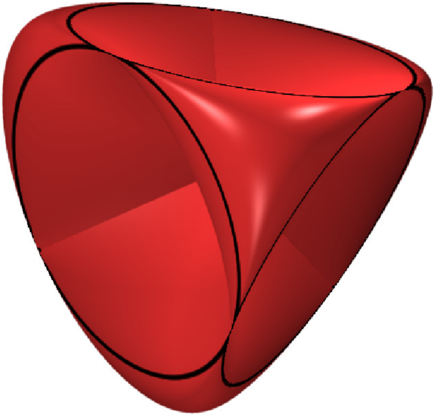


Steiner surface

$$\sum \sigma_i^4 - 4\sum \sigma_i^3\sigma_j + 6\sum \sigma_i^2\sigma_j^2 + 4\sum \sigma_i^2\sigma_j\sigma_k - 40\sigma_1\sigma_2\sigma_3\sigma_4$$

Gradient map $\nabla f : \mathbb{P}^3 \to \mathbb{P}^3$ gives a bijection between C and K:





Gaussian Exponential Family is Hyperbolic

Let $\mathcal{X} = \mathbb{R}^m$, where ν is the Lebesgue measure, and set

$$T(x) = \frac{1}{2} x \cdot x^T \in \mathrm{Sym}_2(\mathbb{R}^m) \simeq \mathbb{R}^d.$$

The symmetric determinant $f(\theta) = \det(\theta)$ is a hyperbolic polynomial in $d = \binom{m}{2}$ unknowns. Its hyperbolicity cone C consists of positive definite matrices. The cone is self dual, so

$$K = C^\vee = \mathrm{conv}(T(\mathcal{X})) \simeq C^{\mathrm{cl}}.$$

Since

$$A(\theta) \; = \; \frac{p}{2} \log(2\pi) - \frac{1}{2} \log\det(\theta),$$

the gradient map is the matrix inversion $\theta \mapsto \frac{1}{2}\theta^{-1}$.

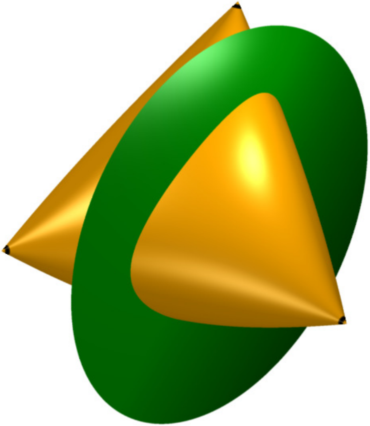
The measure that represents $f(\theta)^{-1/2}$ comes from the Wishart distribution, i.e. the distribution of the sample covariance matrix ...

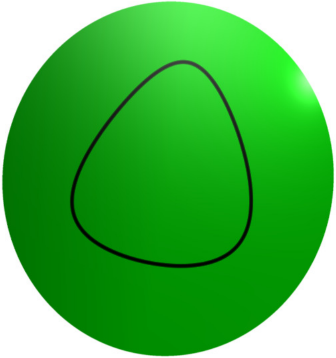
Intersection with a subspace

Fix an exponential family with rational gradient map $F : C \to K$.

- Main case: $F = \nabla f$ where f is a hyperbolic polynomial.

Consider any linear subspace $\mathcal{L} \subset \mathbb{R}^d$ with $C_{\mathcal{L}} := \mathcal{L} \cap C \neq \emptyset$:

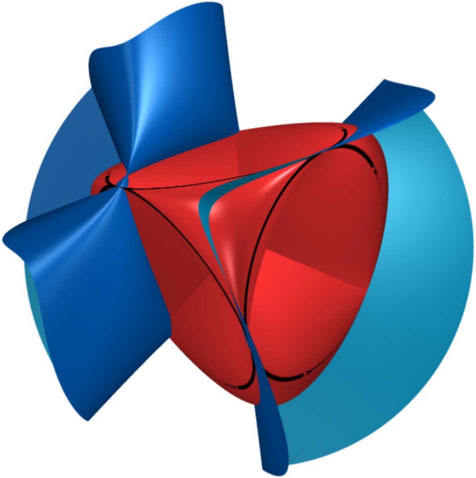


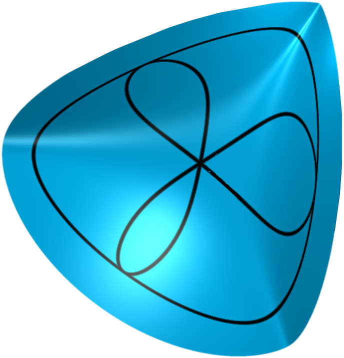


Exponential varieties

The **exponential variety** is the image under the gradient map:

$$\mathcal{L}^F \;:=\; \overline{F(\mathcal{L})} \;\subset\; \mathbb{P}^{d-1}.$$





Its positive part $\mathcal{L}^F_{\succ 0} := F(C_{\mathcal{L}})$ lives in K.

Convexity and positivity

The canonical bijection between $C_{\mathcal{L}}$ and $K_{\mathcal{L}}$ factors through $\mathcal{L}^F_{\succ 0}$.

Theorem

Let $(\mathcal{X}, \nu, T)$ be an exponential family with rational gradient map $F : \mathbb{R}^d \dashrightarrow \mathbb{R}^d$, and $\mathcal{L} \subset \mathbb{R}^d$ a linear subspace. The gradient map $F_{\mathcal{L}}$ of the restriction to $\mathcal{L}$ can be written as a sequence of maps

$$C_{\mathcal{L}} \quad \subset \quad C \quad \xrightarrow{F} \quad K \quad \xrightarrow{\pi_{\mathcal{L}}} \quad K_{\mathcal{L}}.$$

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

Maximum Likelihood Estimation for an exponential variety means inverting these two bijections, by solving polynomials.

Question: What is the algebraic degree of this inversion?

Hankel Matrices

Fix the Gaussian exponential family $f = \det(\theta)$ and let $\mathcal{L}$ be the space of $m \times m$ Hankel matrices. Hence $d = \binom{m+1}{2}$, $c = 2m-1$, and $C_{\mathcal{L}}$  is the cone of positive definite Hankel matrices.

$$\begin{bmatrix} \theta_1 & \theta_2 & \theta_3 & \theta_4 \\ \theta_2 & \theta_3 & \theta_4 & \theta_5 \\ \theta_3 & \theta_4 & \theta_5 & \theta_6 \\ \theta_4 & \theta_5 & \theta_6 & \theta_7 \end{bmatrix} \qquad m = 4, c = 7$$

Identify $\mathbb{R}^c$ with the space of polynomials of degree $2m-2$ in x.

The map $\pi_{\mathcal{L}} : \mathrm{Sym}_2(\mathbb{R}^m) \to \mathbb{R}[x]_{\leq 2m-2}$ is

$$\sigma \;\mapsto\; (1, x, x^2, \ldots, x^{m-1}) \cdot \sigma \cdot (1, x, x^2, \ldots, x^{m-1})^T.$$

The image $K_{\mathcal{L}}$  of the PSD cone $K = C^\vee$ under $\pi_{\mathcal{L}}$ is the cone of nonnegative polynomials.

Question: Who is the middleman  in these bijections:

$$\text{pd Hankel} \;=\; C_{\mathcal{L}} \xrightarrow{\nabla f} \mathcal{L}^{\nabla f}_{\succ 0} \xrightarrow{\pi_{\mathcal{L}}} K_{\mathcal{L}} \;=\; \text{nonnegative polynomials} \quad ?$$

The Other Positive Grassmanian

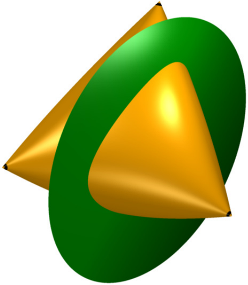
Theorem

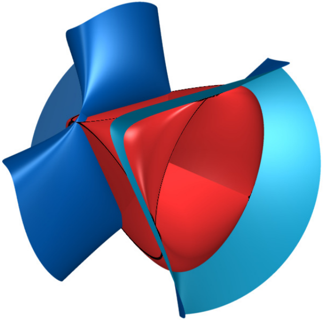
After a linear change of coordinates, the exponential variety $\mathcal{L}^{-1}$ of inverse Hankel matrices equals the Grassmanian $\mathrm{Gr}(2, m+1)$ in its Plücker embedding in $\mathbb{P}^{d-1}$.

$$\begin{bmatrix} \theta_1 & \theta_2 & \theta_3 & \theta_4 \\ \theta_2 & \theta_3 & \theta_4 & \theta_5 \\ \theta_3 & \theta_4 & \theta_5 & \theta_6 \\ \theta_4 & \theta_5 & \theta_6 & \theta_7 \end{bmatrix}^{-1} = \begin{bmatrix} p_{12} & p_{13} & p_{14} & p_{15} \\ p_{13} & p_{14}+p_{23} & p_{15}+p_{24} & p_{25} \\ p_{14} & p_{15}+p_{24} & p_{25}+p_{34} & p_{35} \\ p_{15} & p_{25} & p_{35} & p_{45} \end{bmatrix}, p_{ij}p_{kl} - p_{ik}p_{jl} + p_{il}p_{jk} = 0.$$

The positive Grassmanian $\mathrm{Gr}(2, m+1)_{\succ 0}$ consists of all positive definite Bézout matrices. These represent pairs of polynomial in x of degree $m-1$ such that the roots are all real and interlace.

Invitation to read





Abstract: Exponential varieties arise from exponential families in statistics. These real algebraic varieties have strong positivity and convexity properties, generalizing those of toric varieties and their moment maps. Another special class, including Gaussian graphical models, are varieties of inverses of symmetric matrices satisfying linear constraints. We develop a general theory of exponential varieties, with focus on those defined by hyperbolic polynomials and their integral representations on the hyperbolicity cone. We compare multidegrees and ML degrees of the gradient map for such polynomials.