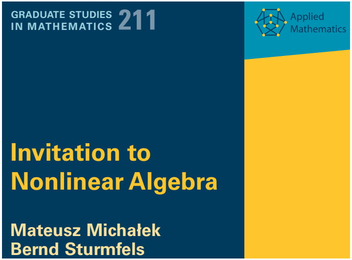
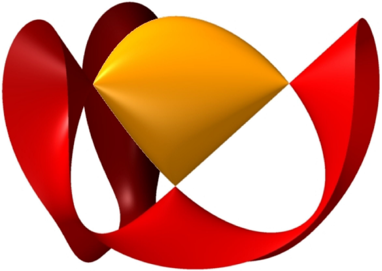
BEYOND LINEAR ALGEBRA

Bernd Sturmfels

MPI Leipzig and UC Berkeley





International Congress of Mathematicians,
July 8, 2022

Undergraduate Linear Algebra

All undergraduate students learn about *Gaussian elimination*, a general method for solving linear systems of algebraic equations:

Input:

$$\begin{aligned} x + 2y + 3z &= 5 \\ 7x + 11y + 13z &= 17 \\ 19x + 23y + 29z &= 31 \end{aligned}$$

Output:

$$\begin{aligned} x &= -35/18 \\ y &= 2/9 \\ z &= 13/6 \end{aligned}$$

Solving very large linear systems is central to applied mathematics.

Undergraduate Non-Linear Algebra

Lucky undergraduate students also learn about *Gröbner bases*, a general method for non-linear systems of algebraic equations:

Input:

$$\begin{aligned} x^2 + y^2 + z^2 &= 2 \\ x^3 + y^3 + z^3 &= 3 \\ x^4 + y^4 + z^4 &= 4 \end{aligned}$$

Output:

$$3z^{12} - 12z^{10} - 12z^9 + 12z^8 + 72z^7 - 66z^6 - 12z^4 + 12z^3 - 1 = 0$$

$$\begin{aligned} 4y^2 + (36z^{11} + 54z^{10} - 69z^9 - 252z^8 - 216z^7 + 573z^6 + 72z^5 \\ -12z^4 - 99z^3 + 10z + 3) \cdot y + 36z^{11} + 48z^{10} - 72z^9 \\ -234z^8 - 192z^7 + 564z^6 - 48z^5 + 96z^4 - 96z^3 + 10z^2 + 8 &= 0 \end{aligned}$$

$$\begin{aligned} 4x + 4y + 36z^{11} + 54z^{10} - 69z^9 - 252z^8 - 216z^7 \\ +573z^6 + 72z^5 - 12z^4 - 99z^3 + 10z + 3 &= 0 \end{aligned}$$

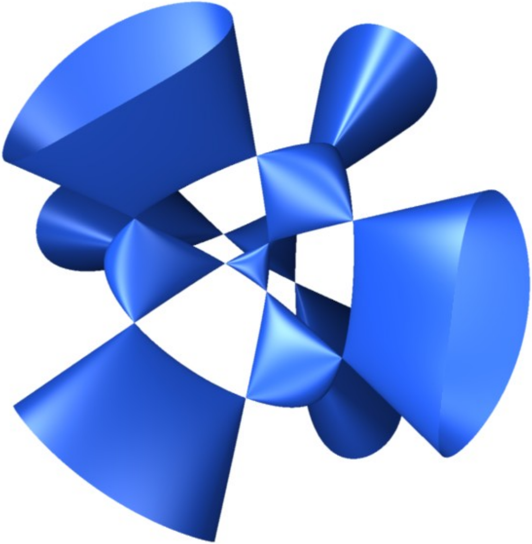
Numerical homotopy methods solve polynomial systems reliably.

Julia


3264
CONICS
IN A
SECOND
Paul Breiding
Bernd Sturmfels
Sascha Timme
In 1848 Jakob Steiner asked
«How many conics are tangent to five conics?»
In 2019 we ask
«Which conics are tangent to your five conics?»
Curious to know the answer?
Find out at:
juliahomotopycontinuation.org/do-it-yourself/
MAX-PLANCK-GESELLSCHAFT

Nonlinear Shapes

Many models in the sciences and engineering are solution sets of polynomial equations. Such a set is an algebraic variety $X \subset \mathbb{R}^n$.



Nearest Points on Varieties

Fix a variety X in $\mathbb{R}^n$, ideal $I_X = \langle f_1, \ldots, f_k \rangle$, and $c = \mathrm{codim}(X)$.
The $k \times n$ Jacobian matrix $\mathcal{J} = (\partial f_i / \partial x_j)$ has rank $\leq c$ on X.

A point $x \in X$ is *nonsingular* if rank $= c$.

Assumption: nonsingular real points are Zariski dense in X.

The following optimization problem arises in many applications:
Given a data point $u \in \mathbb{R}^n$, compute the distance to the model X.

We seek x^* in X that is closest to u.

The answer depends on the metric. We begin with

The Euclidean distance (ED) problem:

$$\text{minimize} \ \sum_{i=1}^{n} (x_i - u_i)^2 \ \text{subject to} \ x \in X.$$

Euclidean Distance Degree

The *augmented Jacobian* is the $(k+1) \times n$ matrix with the extra row $(x_1 - u_1, \ldots, x_n - u_n)$ atop the Jacobian matrix $\mathcal{J}$.

Form its ideal of $(c+1) \times (c+1)$ minors, add I_X, and saturate by the $c \times c$ minors of $\mathcal{J}$. This results in the *critical ideal* $\mathcal{C}_{X,u}$.

This defines the set of critical points of the ED problem. For random data u, this set is finite and contains the solution x^*, provided x^* is a nonsingular point of X.

Euclidean Distance Degree

The *augmented Jacobian* is the $(k+1) \times n$ matrix with the extra row $(x_1 - u_1, \ldots, x_n - u_n)$ atop the Jacobian matrix $\mathcal{J}$.

Form its ideal of $(c+1) \times (c+1)$ minors, add I_X, and saturate by the $c \times c$ minors of $\mathcal{J}$. This results in the *critical ideal* $\mathcal{C}_{X,u}$.

This defines the set of critical points of the ED problem. For random data u, this set is finite and contains the solution x^*,

provided x^* is a nonsingular point of X.

The *ED degree* of X is the number of complex zeros of $\mathcal{C}_{X,u}$. This measures the difficulty of solving the ED problem for the model X.

[Draisma, Horobeţ, Ottaviani, St, Thomas, 2016]

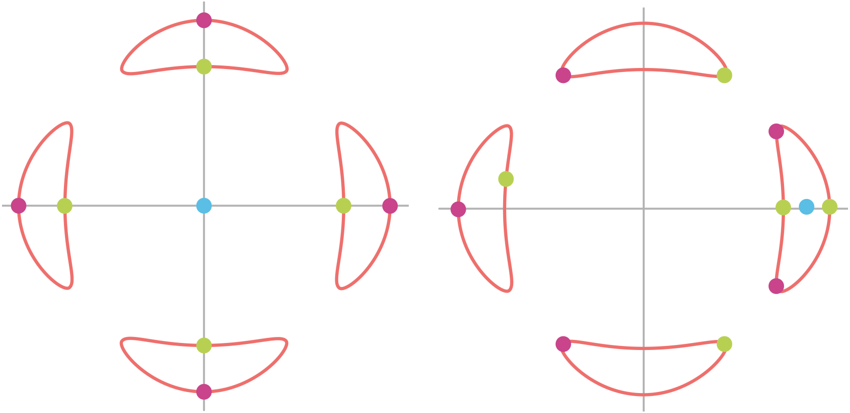
Proposition

If $f_1, \ldots, f_k$ have degrees $d_1 \geq \cdots \geq d_k$ then

$$EDdegree(X) \;\leq\; d_1 d_2 \cdots d_c \cdot \sum_{i_1+i_2+\cdots+i_c \leq n-c} (d_1-1)^{i_1}(d_2-1)^{i_2} \cdots (d_c-1)^{i_c}.$$

Trott Curve

If X is a quartic curve in $\mathbb{R}^2$ then $\mathrm{EDdegree}(X) = 16$. Not all critical points are real. For example, consider the *Trott curve*:



$$144(x^4 + y^4) - 225(x^2 + y^2) + 350x^2y^2 + 81 = 0$$

Space Curves

Fix $n = 3$ and X a curve defined by general polynomials p, q of degrees d_1, d_2 in x, y, z. The augmented Jacobian matrix is

$$\mathcal{AJ} \;=\; \begin{pmatrix} x - u_1 & y - u_2 & z - u_3 \\ \partial p/\partial x & \partial p/\partial y & \partial p/\partial z \\ \partial q/\partial x & \partial q/\partial y & \partial q/\partial z \end{pmatrix}.$$

For random data $u \in \mathbb{R}^3$, the ideal $\mathcal{C}_{X,u} = \langle p,\, q,\, \det(\mathcal{AJ}) \rangle$ has $d_1 d_2 (d_1 + d_2 - 1)$ zeros in $\mathbb{C}^3$. This is the ED degree of X.

Proposition

If X is a general smooth curve of degree d and genus g in $\mathbb{R}^n$ then

$$EDdegree(X) \;=\; 3d + 2g - 2.$$

Curve has degree $d = d_1 d_2$ and genus $g = \frac{1}{2}(d_1^2 d_2 + d_1 d_2^2) - 2 d_1 d_2 + 1$.

General Formula

Theorem

If $X \subset \mathbb{R}^n$ satisfies $\bullet\bullet\bullet$ then EDdegree(X) equals the sum of the polar degrees of the projective closure of X in $\mathbb{P}^n$.

Hypothesis $\bullet\bullet\bullet$ holds for all X after linear change of variables.

General Formula

Theorem

If $X \subset \mathbb{R}^n$ satisfies $\bullet\bullet\bullet$ then EDdegree(X) equals the sum of the polar degrees of the projective closure of X in $\mathbb{P}^n$.

Hypothesis $\bullet\bullet\bullet$ holds for all X after linear change of variables.

Points h in the dual space $(\mathbb{P}^n)^\vee$ are hyperplanes in $\mathbb{P}^n$. Consider pairs (x, h) in $\mathbb{P}^n \times (\mathbb{P}^n)^\vee$ with $x \in X$ nonsingular and h tangent to X at x. Zariski closure is the *conormal variety* $N_X \subset \mathbb{P}^n \times (\mathbb{P}^n)^\vee$.

Facts: N_X is irreducible of dim $n-1$, $N_X = N_{X^\vee}$ and $(X^\vee)^\vee = X$.

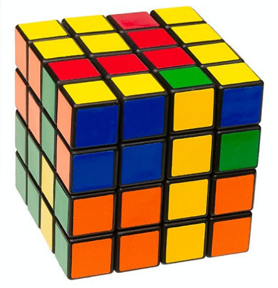
Polar degrees of X are coefficients of the cohomology class

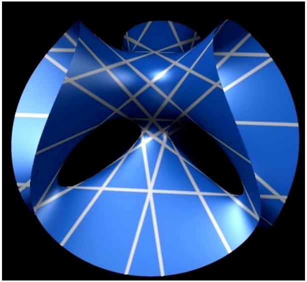
$$[N_X] = \delta_1(X)s^n t + \delta_2(X)s^{n-1}t^2 + \cdots + \delta_n(X)st^n.$$

in $H^*(\mathbb{P}^n \times (\mathbb{P}^n)^\vee, \mathbb{Z}) = \mathbb{Z}[s,t]/\langle s^{n+1}, t^{n+1}\rangle$.

Interpretation

Polar degrees satisfy $\delta_i(X) = \#(\,N_X \cap (L_{n+1-i} \times L_i)\,)$, where $L_{n+1-i} \subset \mathbb{P}^n$ and $L_i \subset (\mathbb{P}^n)^\vee$ are general subspaces.





Example (General Surfaces in 3-space)

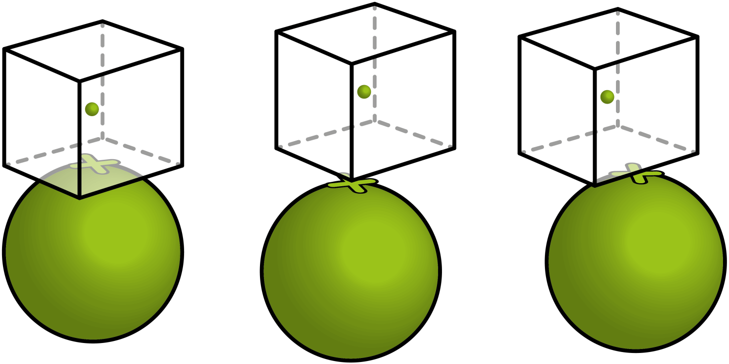
If X has degree d in $\mathbb{P}^3$ then N_X is a surface in $\mathbb{P}^3 \times (\mathbb{P}^3)^\vee$, with

$$[N_X] \;=\; d(d-1)^2\, s^3 t \;+\; d(d-1)\, s^2 t^2 \;+\; d\, s t^3 .$$

The sum of polar degrees equals $\mathrm{EDdegree}(X) = d^3 - d^2 + d$.

Polyhedral Norms (Wasserstein)

Scale the *unit ball* until it meets the *variety* X.



Suppose the optimal face of the unit ball has codimension i.

Proposition

The polar degree $\delta_i(X)$ is the number of critical points of a linear form ℓ on $L \cap X$, where L is a general affine space of codimension $i-1$ in $\mathbb{R}^n$.

Likelihood Geometry (Discrete Models)

State space $\{0, 1, \ldots, n\}$. The probability simplex is

$$\Delta_n \;=\; \{\, p \in \mathbb{R}^{n+1} \,:\, p_0 + p_1 + \cdots + p_n = 1 \text{ and } p_0, p_1, \ldots, p_n > 0 \,\}.$$

Model $X \subset \Delta$. **Data** $u \in \mathbb{N}^{n+1}$. The *log-likelihood function* is

$$\ell_u : \Delta_n \to \mathbb{R}\,,\;\; p \;\mapsto\; u_0 \cdot \log(p_0) \,+\, u_1 \cdot \log(p_1) \,+\, \cdots + \, u_n \cdot \log(p_n).$$

The *ML degree* of X is the number of critical points of:

Maximize $\ell_u(p)$ subject to $p \in X$.

Example

If $n = 3$ and X is our curve $\{p = q = 0\}$ then

$$\mathrm{ML\ degree}(X) \;=\; d_1 d_2 \,(d_1 + d_2 + 1).$$

Likelihood Geometry (ASCB)

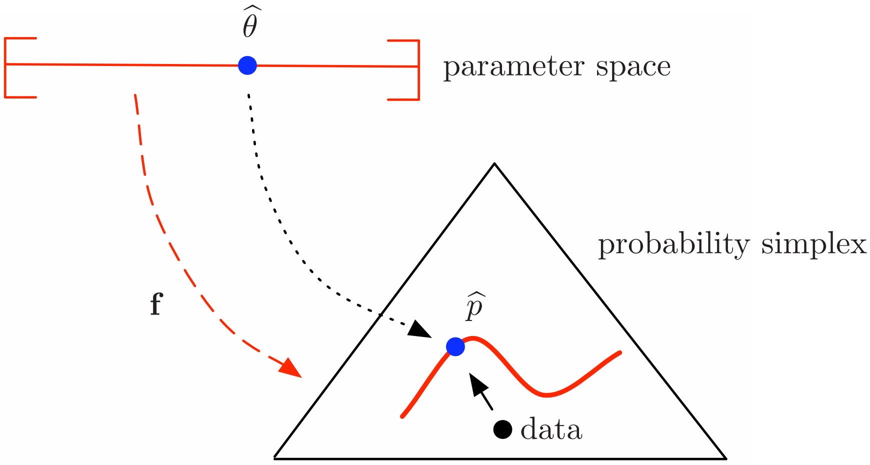


Fig. 3.2. The geometry of maximum likelihood estimation.

Euler Characteristic

View the model X in the complex projective space $\mathbb{P}^n$, and let

$$X^o \;=\; X \setminus \{p_0 p_1 \cdots p_n (\sum_{i=0}^{n} p_i) = 0\}.$$

Theorem

Suppose the very affine variety X^o is non-singular.
The ML degree of the model X equals the signed Euler characteristic $(-1)^{dim(X)} \cdot \chi(X^o)$ of the manifold X^o.

polar degrees for ED $\longleftrightarrow$ Euler characteristic for MLE

Theorem (Huh, Duarte-Marigliano-St)

If the model X has ML degree one, then each coordinate of p is an alternating product of linear forms in u with positive coefficients.

Particle Physics (Scattering Equations)

The *CEGM model*, due to Cachazo et al., is the space of m points in general position in $\mathbb{P}^{k-1}$:

$$X^o \quad = \quad \mathrm{Gr}(k,m)^o/(\mathbb{C}^*)^m.$$

The data u are the Mandelstam invariants.

Proposition

X^o is very affine; coordinates are $k \times k$ minors of

$$\begin{bmatrix} 0 & 0 & 0 & \dots & 0 & (-1)^k & 1 & 1 & 1 & \dots & 1 \\ 0 & 0 & 0 & \dots & (-1)^{k-1} & 0 & 1 & x_{1,1} & x_{1,2} & \dots & x_{1,m-k-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & -1 & \dots & 0 & 0 & 1 & x_{k-3,1} & x_{k-3,2} & \dots & x_{k-3,m-k-1} \\ 0 & 1 & 0 & \dots & 0 & 0 & 1 & x_{k-2,1} & x_{k-2,2} & \dots & x_{k-2,m-k-1} \\ -1 & 0 & 0 & \dots & 0 & 0 & 1 & x_{k-1,1} & x_{k-1,2} & \dots & x_{k-1,m-k-1} \end{bmatrix}$$

Theorem

For $k = 2$, the ML degree equals $(m-3)!$ for all $m \geq 4$.
For $k = 3$ and $m = 5, 6, 7, 8, 9$, it is $2, 26, 1272, 188112, 74570400$.
For $k = 4$, $m = 8$ it equals 5211816.

Likelihood Geometry (Gaussian Models)

Statistical model is a variety X in $\mathrm{PD}_n \subset \mathrm{Sym}_2(\mathbb{R}^n)$. MLE means:

Maximize $\Sigma \mapsto \log\det\Sigma^{-1} - \mathrm{trace}(\mathbf{S}\,\Sigma^{-1})$ subject to $\Sigma \in X$.

Statistical model is a variety X in $\mathrm{PD}_n \subset \mathrm{Sym}_2(\mathbb{R}^n)$. MLE means:

Maximize $\Sigma \mapsto \log\det\Sigma^{-1} - \mathrm{trace}(\mathbf{S}\,\Sigma^{-1})$ subject to $\Sigma \in X$.

Linear Space of Symmetric Matrices $\mathcal{L} \subset \mathrm{Sym}_2(\mathbb{R}^n)$ gives a model

$$X = \{\Sigma : \Sigma^{-1} \in \mathcal{L}\} = \{K^{-1} : K \in \mathcal{L}\}.$$

Proposition

Critical equations for maximum likelihood estimation are

$$K \in \mathcal{L} \text{ and } K\Sigma = Id_n \text{ and } \Sigma - \mathbf{S} \in \mathcal{L}^{\perp}.$$

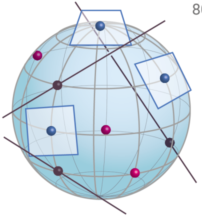
The ML degree of $\mathcal{L}$ is the number of complex critical points.

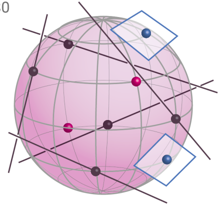
Example: For general LSSM with $n = 4$ we have

$k = \dim(\mathcal{L})$:	2	3	4	5	6	7	8	9
ML degree :	3	9	17	21	21	17	9	3

Schubert Calculus

1 3 9 17 21 21 17 9 3 1
2 6 18 34 42 34 18 6 2
4 12 36 68 68 36 12 4
8 24 72 **104** 72 24 8
16 48 112 112 48 16
32 80 **128** 80 32
56 104 104 56
80 104 80
92 92
92





Theorem (Michałek et al.)

The ML degree of $\mathcal{L}$ is the number of quadrics in $\mathbb{P}^{n-1}$ through $\binom{n+1}{2} - k$ general points and tangent to $k-1$ general hyperplanes.

For fixed k, this is a polynomial in n of degree $k-1$.

Nonlinear Algebra meets Linear PDE

The following two things are the same:

- **Polynomials**:

$$X^2 - T^2 \; = \; (X - T)(X + T)$$

- Linear homogeneous **partial differential equations** with constant coefficients:

$$\phi_{xx}(x,t) \; - \; \phi_{tt}(x,t) \;\; = \;\; 0.$$

Nonlinear Algebra meets Linear PDE

The following two things are the same:

- **Polynomials**:

$$X^2 - T^2 \; = \; (X-T)(X+T)$$

- Linear homogeneous **partial differential equations** with constant coefficients:

$$\phi_{xx}(x,t) \; - \; \phi_{tt}(x,t) \;\; = \;\; 0.$$

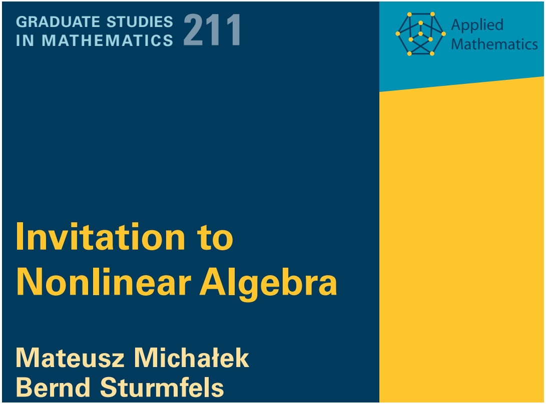
D'Alembert (1747): General solution is superposition of traveling waves:

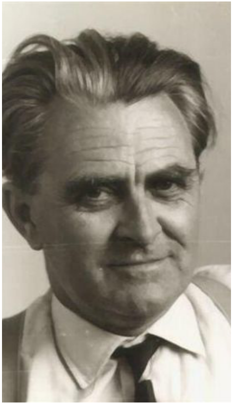
$$\phi(x,t) \; = \; f(x+t) \, + \, g(x-t),$$

where f and g are twice differentiable functions, or distributions.

Current work: **any ideal** *in a polynomial ring, and any* **module**.

Section 3.3 Wolfgang Gröbner





Theorem 3.27. *Let I be a zero-dimensional ideal in $\mathbb{C}[x_1, \ldots, x_n]$, here interpreted as a system of linear PDEs. The space of holomorphic solutions has dimension equal to the degree of I. There exist nonzero polynomial solutions if and only if the maximal ideal $M = \langle x_1, \ldots, x_n \rangle$ is an associated prime of I. In that case, the polynomial solutions are precisely the solutions to the system of PDEs given by the M-primary component $(I : (I : M^\infty))$.*

Cayley's Cubic Surface

This picture is the logo of the Nonlinear Algebra group at MPI Leipzig:

The elliptope is a PDE constraint for $\phi : \mathbb{R}^4 \to \mathbb{C}^3$:

$$\begin{bmatrix} \partial_1 & \partial_2 & \partial_3 \\ \partial_2 & \partial_1 & \partial_4 \\ \partial_3 & \partial_4 & \partial_1 \end{bmatrix} \bullet \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Quiz: What does the command `solvePDE` in `Macaulay2` tell us?

Epilog

Linear algebra is ubiquitous in the mathematical universe.

It plays a foundational role for models in the sciences and engineering; its numerical tools are a driving force for today's technologies. The power of linear algebra stems from calculus, i.e. our ability to approximate nonlinear shapes by linear spaces.

Yet, the world is nonlinear. Polynomials are a natural ingredient in mathematical models for the real world.

In our view, the nonlinear nature of a phenomenon should be respected as long as possible. We advocate against the practice of passing to a linear approximation right away.

Many Thanks for Listening