The Algebraic Degree of Semidefinite Programming

BERND STURMFELS
UNIVERSITY OF CALIFORNIA, BERKELEY

joint work with
Kristian Ranestad (University of Oslo, Norway)
and Jiawang Nie (IMA Minneapolis)

Commutative Algebra and Algebraic Geometry Seminar,
UC Berkeley, November 6, 2006

Our Question

Semidefinite programming is a numerical method in convex optimization.

SDP is very efficient, both in theory and in practice.

SDP is widely used in engineering and the sciences.

Input: Several symmetric $n \times n$-matrices
Output: One symmetric $n \times n$-matrix

What is the function from the input to the output?
How does the solution depend on the data?

Let's begin with three slides on the philosophy.....

Linear Programming

is semidefinite programming for diagonal matrices

The optimal solution of the linear program

$$\text{Maximize } c \cdot x \text{ subject to } A \cdot x = b \text{ and } x \geq 0$$

is a piecewise linear function of c and b.

It is a piecewise rational function in the entries of A.

To study the function *data* $\mapsto$ *solution*, one needs geometric combinatorics, namely matroids for the dependence on A and secondary polytopes for b, c.

Universality of Nash Equilibria

Consider the following problem in game theory:

Given payoff matrices, compute the Nash equilibria.

For two players, finding fully mixed Nash equilibria is a linear problem, just like linear programming.

For three and more players, the dependence is non-linear. Datta's Universality Theorem (2003) says: Every real algebraic variety is isomorphic to the set of Nash equilibria of some three-person game.

Corollary: The coordinates of the Nash equilibria can be arbitrary algebraic functions of the payoff matrices.

Maximum Likelihood Degree

An optimization problem in algebraic statistics:

$$\text{Maximize } \; p_1(\theta)^{u_1} p_2(\theta)^{u_2} \cdots p_m(\theta)^{u_m}$$
$$\text{subject to } \; \theta \in \Theta \quad \subset \mathbb{R}^d$$

The $p_i(\theta)$ are polynomial functions that sum to one, and the u_i are positive integers (these are the data).

The optimal solution $\hat{\theta}$ is the maximum likelihood estimator. It is an algebraic function of the data:

$$(u_1, \ldots, u_m) \; \mapsto \; \hat{\theta}(u_1, \ldots, u_m)$$

This defines the ML degree of the statistical model.

Semidefinite Programming

Given: A matrix C and an m-dimensional affine subspace $\mathcal{U}$ of real symmetric $n \times n$-matrices

SDP Problem:

$$\text{Maximize} \quad \text{trace}(C \cdot X)$$
$$\text{subject to } X \in \mathcal{U} \text{ and } X \succeq 0.$$

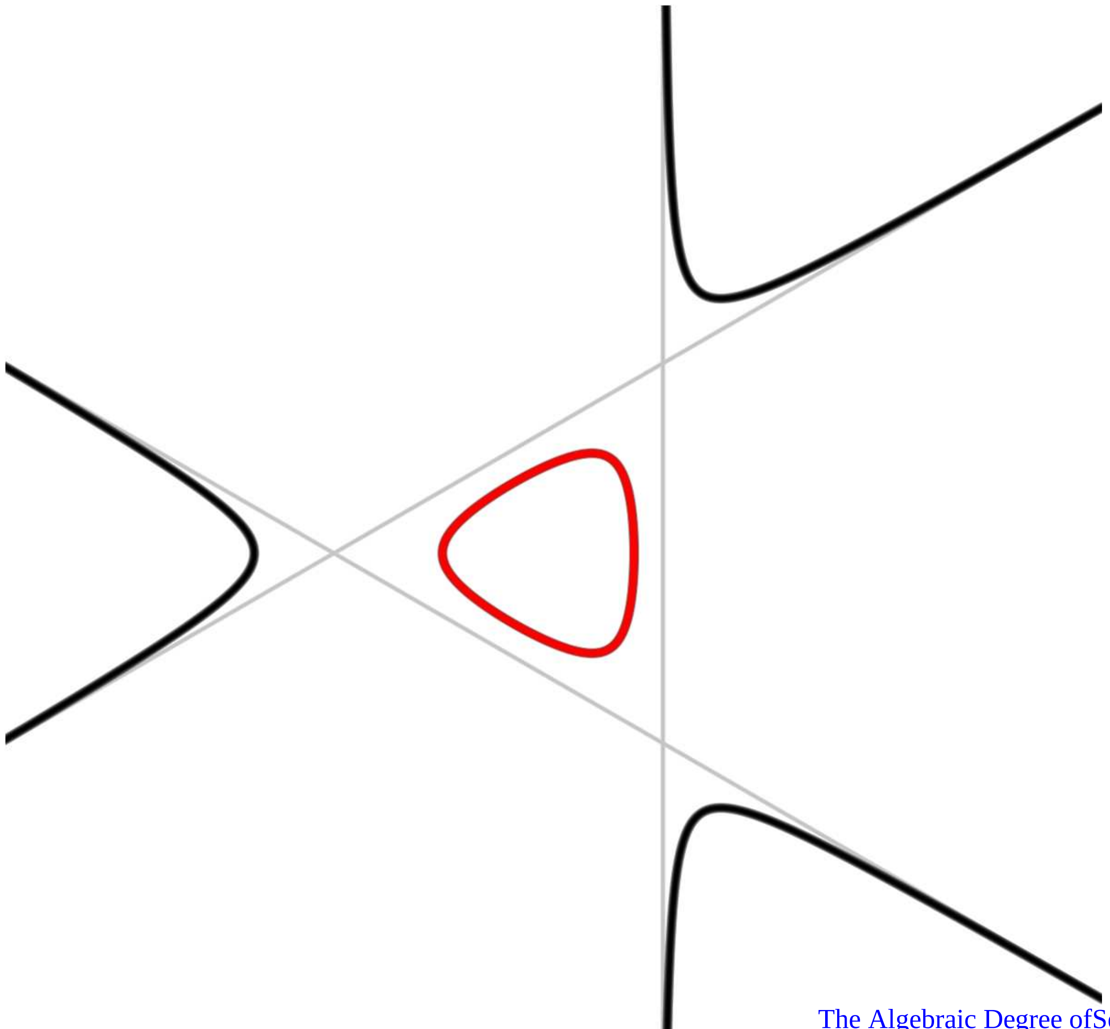
Here $X \succeq 0$ means that X is positive semidefinite.

The problem is feasible if and only if the subspace $\mathcal{U}$ intersects the cone of positive semidefinite matrices.

The optimal solution $\hat{X}$ is a (piecewise) algebraic function of the matrix C and of the subspace $\mathcal{U}$.

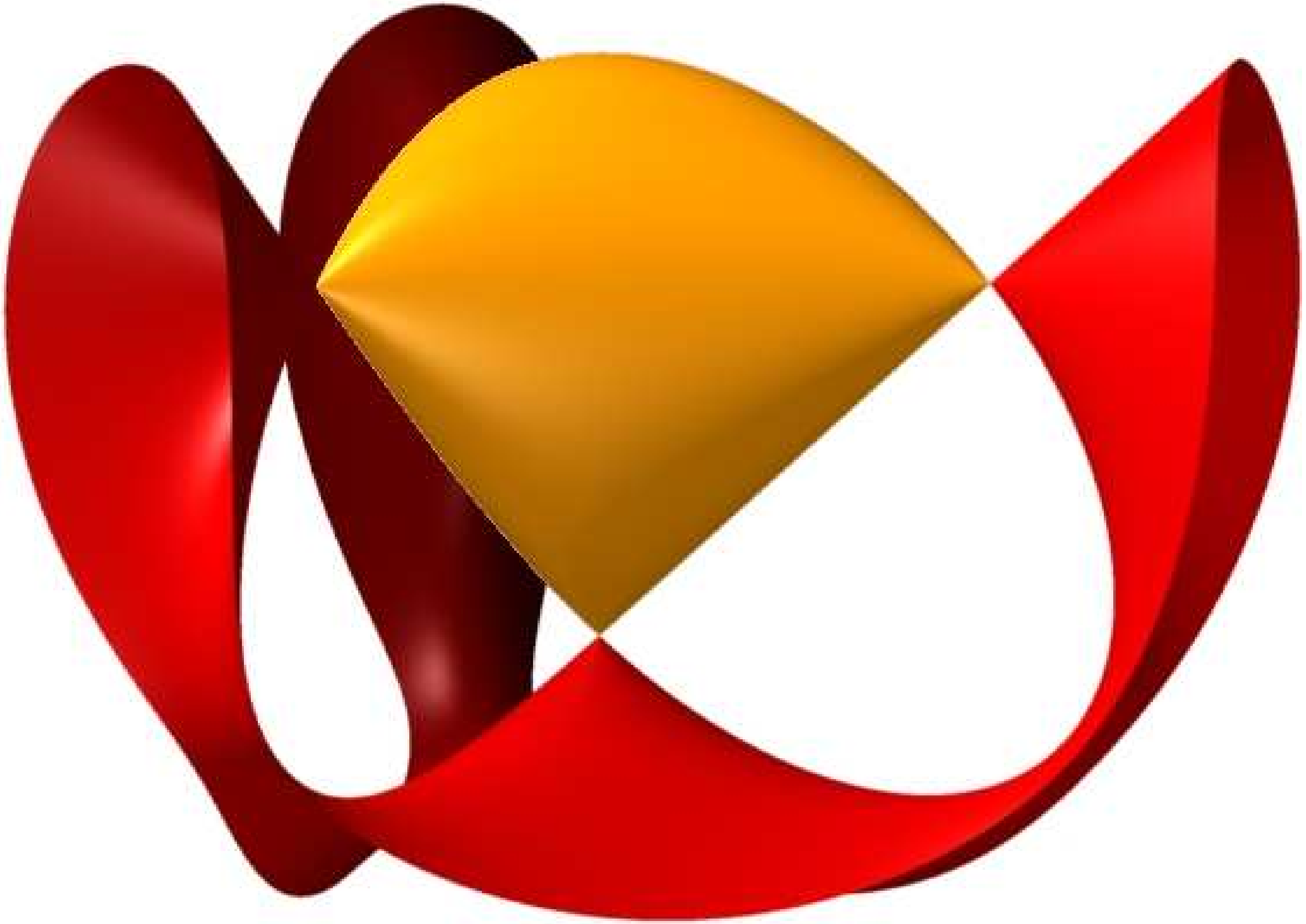
Example: Vinnikov Curves

Let $m = 2$ and $n = 3$. Then $X \succeq 0$ defines a semialgebraic convex region in $\mathcal{U} \simeq \mathbb{R}^2$. It is bounded by the elliptic Vinnikov curve $\det(X) = 0$



Example: Cayley Cubic

Let $m = n = 3$. The cubic surface $\det(X) = 0$ is a Cayley cubic, with four singular points...



Analytic Solution

Let $m = n = 3$. The cubic surface $\det(X) = 0$ is a Cayley cubic. Its dual is a quartic Steiner surface.

SDP: Maximize a linear function over the convex region $X \succeq 0$ bounded by the Cayley cubic.

We can express the optimal solution $\hat{X}$ in terms of radicals $\sqrt{\quad\quad}$ using Cardano's formula:

Either $\hat{X}$ has rank one and is one of the four singular points of the Cayley cubic, or $\hat{X}$ has rank two and is found by intersecting the Steiner surface with a line.

Pataki's Inequalities

We always assume that the data C and $\mathcal{U}$ are generic.

Theorem 1. *(Pataki, 2000) The rank r of the solution $\hat{X}$ to the SDP problem satisfies the two inequalities*

$$\binom{r+1}{2} \leq \binom{n+1}{2} - m$$

$$\binom{n-r+1}{2} \leq m.$$

For fixed m and n, all ranks r in the specified range are attained for an open set of instances $(\mathcal{U}, C)$.

Distribution of the optimal rank

n	3		4		5		6	
m	rank	percent	rank	percent	rank	percent	rank	percent
3	2	24.00%	3	35.34%	4	79.18%	5	82.78%
	1	76.00%	2	64.66%	3	20.82%	4	17.22%
4			3	23.22%	4	16.96%	5	37.42%
	1	100 %	2	76.78%	3	83.04%	4	62.58%
5					4	5.90%	5	38.42%
	1	100 %	2	100 %	3	94.10%	4	61.58%
6							5	1.32%
			2	67.24%	3	93.50%	4	93.36%
			1	32.76%	2	6.50%	3	5.32%
7			2	52.94%	3	82.64%	4	78.82%
			1	47.06%	2	17.36%	3	21.18%
8					3	34.64%	4	45.62%
			1	100 %	2	65.36%	3	54.38%
9					3	7.60%	4	23.50%
			1	100 %	2	92.40%	3	76.50%

Algebraic Degree of SDP

Fix m, n and r in the Pataki range.

The degree $\delta(m,n,r)$ of the algebraic function $(C,\mathcal{U}) \mapsto \hat{X}$ is the algebraic degree of SDP.

Examples:

$$\delta(1,n,n-1) \;=\; n$$
$$\delta(2,n,n-1) \;=\; n(n-1)$$
$$\delta(3,3,1) \;=\; \delta(3,3,2) \;=\; 4$$

Duality:

$$\delta(m,n,r) \;=\; \delta(\tbinom{n+1}{2}-m,\, n\,, n-r).$$

m	r	degree	r	degree	r	degree	r	degree	r	degree
1	1	2	2	3	3	4	4	5	5	6
2	1	2	2	6	3	12	4	20	5	30
3			2	4	3	16	4	40	5	80
			1	4	2	10	3	20	4	35
4					3	8	4	40	5	120
			1	6	2	30	3	90	4	210
5							4	16	5	96
			1	3	2	42	3	207	4	672
6									5	32
					2	30	3	290	4	1400
					1	8	2	35	3	112
7					2	10	3	260	4	2040
					1	16	2	140	3	672
8							3	140	4	2100
					1	12	2	260	3	1992
9							3	35	4	1470
					1	4	2	290	3	3812

Determinantal Varieties

Consider the complex projective space $\mathbb{P}\mathcal{U} \simeq \mathbb{P}^m$.

Let $D^r_{\mathcal{U}}$ denote the variety of all matrices of rank $\leq r$.

Theorem 2. *The codimension of* $D^r_{\mathcal{U}}$ *is* $\binom{n-r+1}{2}$. *If* $m > \binom{n-r+1}{2}$ *then* $D^r_{\mathcal{U}}$ *is irreducible.*

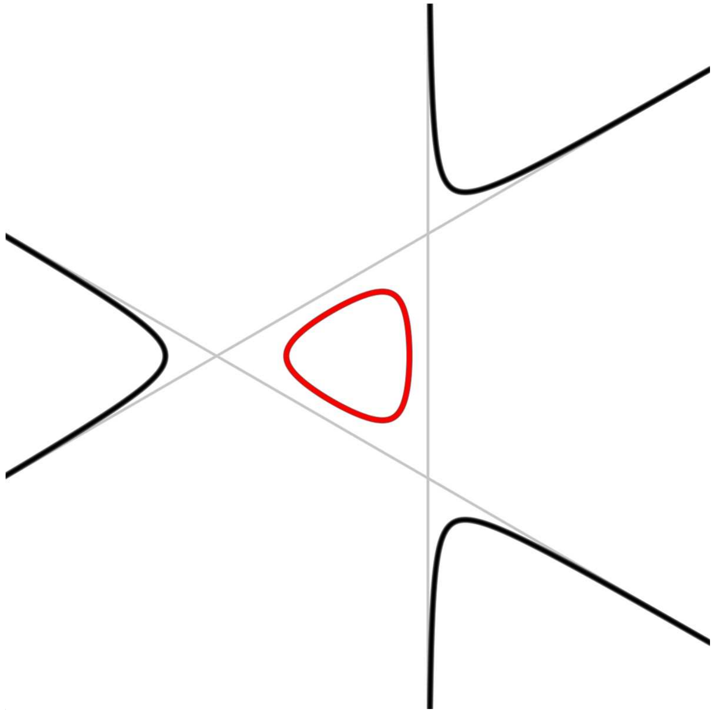
The singular locus of $D^r_{\mathcal{U}}$ *equals* $D^{r-1}_{\mathcal{U}}$, *and*

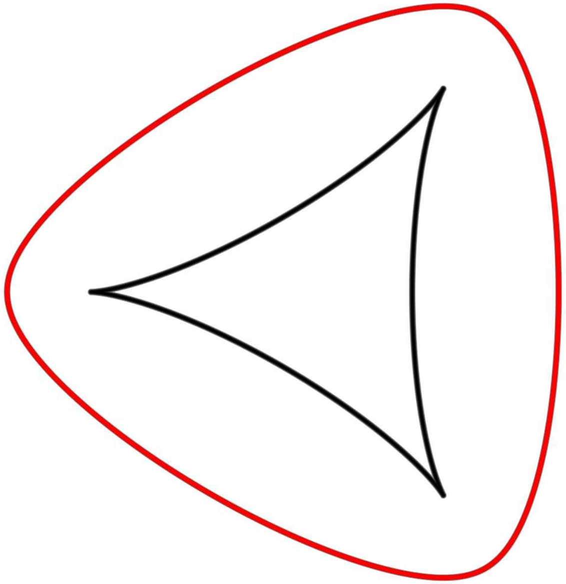
$$\mathrm{degree}(D^r_{\mathcal{U}}) \quad = \quad \prod_{j=0}^{n-r-1} \frac{\binom{n+j}{n-r-j}}{\binom{2j+1}{j}}$$

Projective Duality

Let $\mathbb{P}\mathcal{U}^*$ denote the dual projective space to $\mathbb{P}\mathcal{U}$. The points in $\mathbb{P}\mathcal{U}^*$ correspond to hyperplanes in $\mathbb{P}\mathcal{U}$.

Any variety $\mathcal{V} \subset \mathbb{P}\mathcal{U}$ has a dual variety $\mathcal{V}^* \subset \mathbb{P}\mathcal{U}^*$. $\mathcal{V}^*$ is the Zariski closure of the set of all hyperplanes in $\mathbb{P}\mathcal{U}$ that are tangent to $\mathcal{V}$ at a smooth point.

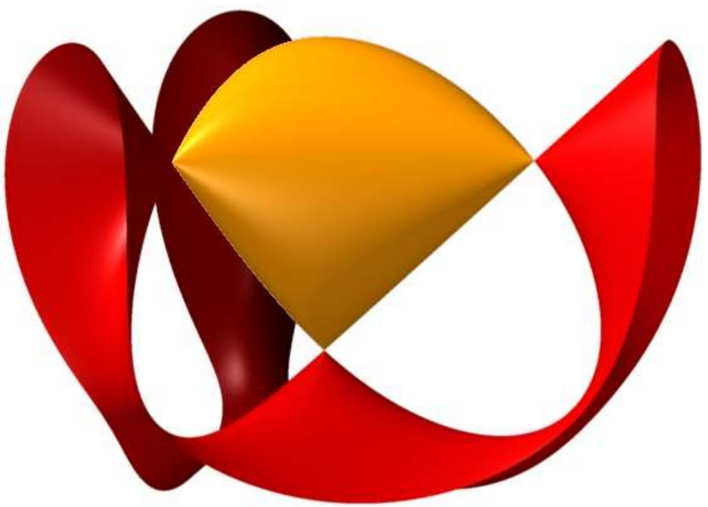




Projection is Dual to Section

Usually $\mathcal{V}^*$ is a hypersurface, but some varieties $\mathcal{V}$ are *degenerate*, in the sense that $\mathrm{codim}(\mathcal{V}^*) \geq 2$.

Lemma 3. *Suppose $\mathcal{V} \subset \mathbb{P}\mathcal{U}$ is degenerate and $\mathcal{W} = \mathcal{V} \cap H$ is a general hyperplane section. Then $\mathcal{W}^*$ is the projection of $\mathcal{V}^*$ from the point H in $\mathbb{P}\mathcal{U}^*$, and $\mathcal{W}^*$ has the same dimension and degree as $\mathcal{V}^*$.*



Connecting the Dots

Lemma 4. *If* $m = \binom{n+1}{2}$ *then the projective dual of* $D^r_{\mathcal{U}}$ *equals the complementary determinantal variety:*

$$(D^r_{\mathcal{U}})^* \quad = \quad D^{n-r}_{\mathcal{U}^*}$$

Theorem 5. *The variety* $D^r_{\mathcal{U}}$ *is non-degenerate if and only if Pataki's inequalities hold. The algebraic degree of SDP is the degree of the dual hypersurface:*

$$\delta(m, n, r) \quad = \quad \text{degree}\,(D^r_{\mathcal{U}})^*$$

m	r	degree	r	degree	r	degree	r	degree	r	degree
1	1	2	2	3	3	4	4	5	5	6
2	1	2	2	6	3	12	4	20	5	30
3			2	4	3	16	4	40	5	80
			1	4	2	10	3	20	4	35
4					3	8	4	40	5	120
			1	6	2	30	3	90	4	210
5							4	16	5	96
			1	3	2	42	3	207	4	672
6									5	32
					2	30	3	290	4	1400
					1	8	2	35	3	112
7					2	10	3	260	4	2040
					1	16	2	140	3	672
8							3	140	4	2100
					1	12	2	260	3	1992
9							3	35	4	1470
					1	4	2	290	3	3812

Rows of the table

We have a conjectured formula which expresses each row as a polynomial in n. It is based on work of Piotr Pragacz on degeneracy loci and Schur Q-functions.

Many cases have been proved, for instance:

$$\delta(6,n,3) = \frac{n+2}{30}\binom{n}{3}\binom{n+1}{2},$$

$$\delta(7,n,4) = 28\binom{n+3}{7} - 12\binom{n+2}{6}$$

$$\delta(8,n,5) = 248\binom{n+4}{8} - 320\binom{n+3}{7} + 84\binom{n+2}{6}$$

Columns of the table

Theorem 6. *Let* $\mathcal{Q}^{\{r\}}$ *denote the variety of pairs* (X, Y) *of symmetric* $n \times n$*-matrices such that* $X \cdot Y = 0$, $\operatorname{rank}(X) = r$ *and* $\operatorname{rank}(Y) = n - r$. *The bidegree of* $\mathcal{Q}^{\{r\}}$ *equals the generating function for the algebraic degree of semidefinite programming:*

$$\mathcal{C}(\mathcal{Q}^{\{r\}}; s, t) \quad = \quad \sum_{m=0}^{\binom{n+1}{2}} \delta(m, n, r) \cdot s^{\binom{n+1}{2}-m} \cdot t^m.$$

Setting $s = t = 1$ we get the scalar degree of $\mathcal{Q}^{\{r\}}$:

$$\begin{aligned} \mathcal{C}(\mathcal{Q}^{\{3\}}; 1, 1) &= 4 + 12 + 16 + 8 = 40 \\ \mathcal{C}(\mathcal{Q}^{\{2\}}; 1, 1) &= 10 + 30 + 42 + 30 + 10 = 122 \end{aligned}$$

Let’s check with Macaulay 2

```
R = QQ[x11,x12,x13,x14,x22,x23,x24,x33,x34,x44,
       y11,y12,y13,y14,y22,y23,y24,y33,y34,y44];

X = matrix {{x11, x12, x13, x14},
            {x12, x22, x23, x24},
            {x13, x23, x33, x34},
            {x14, x24, x34, x44}};

Y = matrix {{y11, y12, y13, y14},
            {y12, y22, y23, y24},
            {y13, y23, y33, y34},
            {y14, y24, y34, y44}};

Q3 = minors(1,X*Y) + minors(4,X) + minors(2,Y);
codim Q3, degree Q3
```

(10, 40)

```
Q2 = minors(1,X*Y) + minors(3,X) + minors(3,Y);
codim Q2, degree Q2
```

(10, 122)

Conclusions

Conclusion for Engineers:

Algebraic Geometry might be useful.

Conclusion for Mathematicians:

Optimization might be interesting.