Semidefinite Programming

Notes by Bernd Sturmfels
for the lecture on June 26, 2018, in the
IMPRS Ringvorlesung *Introduction to Nonlinear Algebra*

The transition from linear algebra to nonlinear algebra has a natural counterpart in convex optimization, namely the transition from linear programming to semidefinite programming. This transition is the topic of the current lecture. We work over the field $\mathbb{R}$ of real numbers.

Linear programming is concerned with the solution of linear systems of inequalities, and with optimizing linear functions subject to such linear constraints. The feasible region is a *convex polyhedron*, and the optimal solutions are given by some face of that polyhedron.

In *semidefinite programming* we work in the space of symmetric $n \times n$-matrices, denoted $\mathrm{Sym}_2(\mathbb{R}^n)$. The inequality constraints now stipulate that some linear combination of symmetric matrices be positive semidefinite. The feasible region given by such constraints is a closed convex set, known as a *spectrahedron*. We again wish to optimize a linear function.

We know from the Spectral Theorem in Linear Algebra that all eigenvalues of a symmetric matrix $A \in \mathrm{Sym}_2(\mathbb{R}^n)$ are real. Moreover, there is an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A. We say that the matrix A is *positive definite* if it satisfies the following conditions. It is a basic fact about quadratic forms that these three conditions are equivalent:

(1) All n eigenvalues of A are positive real numbers.

(2) All 2^n principal minors of A are positive real numbers.

(3) Every non-zero column vector $\mathbf{u} \in \mathbb{R}^n$ satisfies $\mathbf{u}^T A \mathbf{u} > 0$.

Here, by a *principal minor* we mean the determinant of any square submatrix of A whose set of column indices agree with its set of row indices. For the empty set, we get the 0×0 minor of A, which equals 1. Next there are the n diagonal entries of A, which are the 1×1 principal minors, and finally the determinant of A, which is the unique $n \times n$ principal minor.

Each of the three conditions (1), (2) and (3) behaves as expected when we pass to the closure. This is not obvious because the closure of an open semialgebraic set $\{f > 0\}$, where $f \in \mathbb{R}[\mathbf{x}]$, is generally smaller than the corresponding closed semialgebraic set $\{f \geq 0\}$.

Example 1. Let $f = x^3 + x^2y + xy^2 + y^3 - x^2 - y^2$. The set $\{f > 0\}$ is the open halfplane above the line $x + y = 1$ in $\mathbb{R}^2$. The closure of the set $\{f > 0\}$ is the corresponding closed halfplane. It is properly contained in $\{f \geq 0\}$ which also contains the origin $(0, 0)$.

Luckily, no such thing happens with condition (2) for positive definite matrices.

Theorem 2. *For a symmetric $n \times n$ matrix A, the following three conditions are equivalent:*

(1') All n eigenvalues of A are nonnegative real numbers.

(2') All 2^n principal minors of A are nonnegative real numbers.

(3') Every non-zero column vector $\mathbf{u} \in \mathbb{R}^n$ satisfies $\mathbf{u}^T A \mathbf{u} \geq 0$.

If this holds then the matrix A is called positive semidefinite. *The semialgebraic set* PSD_n *of all positive semidefinite $n \times n$ matrices is a full-dimensional closed convex cone in* $\mathrm{Sym}_2(\mathbb{R}^n)$.

We use the notation $X \succeq 0$ to express that a symmetric matrix X is positive semidefinite. A *spectrahedron* $\mathcal{S}$ is the intersection of the cone PSD_n with an affine-linear subspace $\mathcal{L}$ of the ambient space $\mathrm{Sym}_2(\mathbb{R}^n)$. Hence, spectrahedra are closed convex semialgebraic sets.

A subspace $\mathcal{L}$ of symmetric matrices is either given parametrically, or as the solution set to an inhomogeneous system of linear equations. In the equational representation, we write

$$\mathcal{L} \quad = \quad \big\{ X \in \mathrm{Sym}_2(\mathbb{R}^n) \,:\, \langle A_1, X\rangle = b_1,\, \langle A_2, X\rangle = b_2,\, \ldots, \langle A_s, X\rangle = b_s \big\}. \tag{1}$$

Here $A_1, A_2, \ldots, A_s \in \mathrm{Sym}_2(\mathbb{R}^n)$ and $b_1, b_2, \ldots, b_s \in \mathbb{R}$ are fixed, and we employ the usual inner product in the space of square matrices, which is the trace of the matrix product:

$$\langle A, X\rangle \;:=\; \mathrm{trace}(AX) \;=\; \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_{ij}. \tag{2}$$

The associated spectrahedron $\mathcal{S} = \mathcal{L} \cap \mathrm{PSD}_n$ consists of all positive semidefinite matrices that lie in the subspace $\mathcal{L}$. If the subspace is given by a parametric representation, say

$$\mathcal{L} \quad = \quad \big\{ A_0 + x_1 A_1 + \cdots + x_s A_s \,:\, (x_1, \ldots, x_s) \in \mathbb{R}^s \big\}, \tag{3}$$

then it is customary to identify the spectrahedron with its preimage in $\mathbb{R}^s$. Hence we write

$$\mathcal{S} \quad = \quad \big\{ (x_1, \ldots, x_s) \in \mathbb{R}^s \,:\, A_0 + x_1 A_1 + \cdots + x_s A_s \succeq 0 \big\}. \tag{4}$$

Proposition 3. *Every convex polyhedron is a spectrahedron. Convex polyhedra are precisely the spectrahedra that arise when the subspace $\mathcal{L}$ consists only of diagonal $n \times n$ matrices.*

Proof. Suppose that the matrices $A_0, A_1, \ldots, A_s$ are diagonal matrices. Then (4) is the solution set in $\mathbb{R}^s$ of a system of n inhomogeneous linear inequalities. Such a set is a convex polyhedron. Every convex polyhedron in $\mathbb{R}^s$ has such a representation. We simply write its defining linear inequalities as the diagonal entries of the matrix $A_0 + x_1 A_1 + \cdots + x_s A_s$.

The formula $\mathcal{S} = \mathcal{L} \cap \mathrm{PSD}_n$ with $\mathcal{L}$ as in (1) corresponds to the standard representation of a convex polyhedron, as the set of non-negative points in an affine-linear space. Here the equations in (1) include those that require the off-diagonal entries of all matrices to be zero:

$$\langle X, E_{ij}\rangle \;=\; x_{ij} \;=\; 0 \quad \text{for } i \neq j.$$

In the other inequalities, the matrices A_i are diagonal and the b_i are typically nonzero. $\square$

Example 4. Let $\mathcal{L}$ be the space of symmetric 3×3 matrices whose three diagonal entries are all equal to 1. This is an affine-linear subspace of dimension $s = 3$ in $\mathrm{Sym}_2(\mathbb{R}^3) \simeq \mathbb{R}^6$. The spectrahedron $\mathcal{S} = \mathcal{L} \cap \mathrm{SDP}_3$ is the yellow convex body seen in the first lecture of this course. To draw this spectrahedron in $\mathbb{R}^3$, one uses the representation (4), namely

$$\mathcal{S} \quad = \quad \left\{ (x, y, z) \in \mathbb{R}^3 : \begin{pmatrix} 1 & x & y \\ x & 1 & z \\ y & z & 1 \end{pmatrix} \succeq 0 \right\}.$$

The boundary of $\mathcal{S}$ consists of all points (x, y, z) where the matrix has determinant zero and its nonzero eigenvalues are positive. The determinant is a polynomial of degree three in x, y, z, so the boundary lies in cubic surface in $\mathbb{R}^3$. This cubic surface also contains points where the three eigenvalues are positive, zero and negative. Such points are drawn in red in our picture in Lecture 1. They lie in the Zariski closure of the yellow boundary points.

We next slice our 3-dimensional spectrahedron to get a picture in the plane.

Example 5. Suppose that $\mathcal{L} \subset \mathrm{Sym}_2(\mathbb{R}^3)$ is a general plane that intersects the cone PSD_3. The spectrahedron $\mathcal{S}$ is a planar convex body whose boundary is a smooth cubic curve, drawn in red in Figure 1. On that boundary, the 3×3 determinant vanishes and the other two eigenvalues are positive. For points $(x, y) \in \mathbb{R}^2 \backslash \mathcal{S}$, the matrix has at least one negative eigenvalue. The black curve lie in the Zariski closure of the red curve. It separates points in $\mathbb{R}^2 \backslash \mathcal{S}$ whose remaining two eigenvalues are positive from those with two negative eigenvalues.

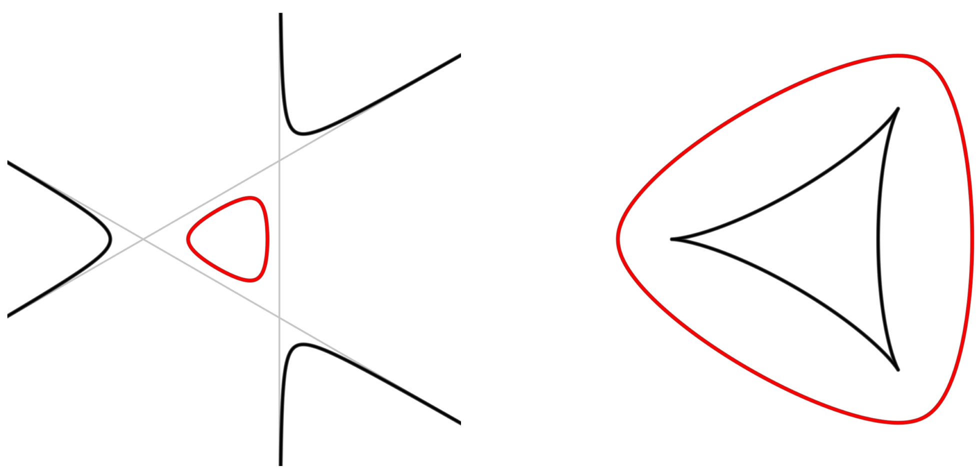

Figure 1: A plane curve of degree three (left) and its dual curve of degree six (right). The red part on the left bounds a spectrahedron while that on the right bounds its convex dual.

To be explicit, suppose that our planar cubic spectrahedron is defined as follows:

$$\mathcal{S} \;=\; \left\{ (x, y) \in \mathbb{R}^3 : \begin{pmatrix} 1 & x & x+y \\ x & 1 & y \\ x+y & y & 1 \end{pmatrix} \succeq 0 \right\}. \tag{5}$$

The cubic curve is the locus where the 3×3 matrix is singular. Its determinant equals

$$f \quad = \quad 2x^2y + 2xy^2 - 2x^2 - 2xy - 2y^2 + 1. \qquad (6)$$

The cubic curve $\{f = 0\}$ has four connected components in $\mathbb{R}^2$, one in red and three in black, as shown in Figure 1 (left). The convex curve in red is the boundary of the spectrahedron.

The picture on the right in Figure 1 shows the *dual curve*. This lives in the dual plane whose points (u, v) represent the lines $\ell = \{(x, y) : ux + vy = 1\}$ in $\mathbb{R}^2$. The points in the dual curve correspond to lines ℓ that are tangent to the original curve. The dual curve has degree six, and its equation is computed by the following ideal computation in $\mathbb{R}[x, y, u, v]$:

$$\begin{array}{c} \langle\, f(x,y)\,,\; u\cdot x + v \cdot y - 1\,,\; \partial f/\partial x \cdot v - \partial f / \partial y \cdot u \,\rangle \;\cap\; \mathbb{R}[u,v] \qquad = \\ \langle\, 8u^6 - 24u^5v + 21u^4v^2 - 2u^3v^3 + 21u^2v^4 - 24uv^5 + 8v^6 - 24u^5 + 60u^4v \\ -24u^3v^2 - 24u^2v^3 + 60uv^4 - 24v^5 + 12u^4 - 24u^3v + 36u^2v^2 - 24uv^3 \\ +12v^4 + 24u^3 - 36u^2v - 36uv^2 + 24v^3 - 24u^2 + 24uv - 24v^2 + 4\,\rangle. \end{array} \qquad (7)$$

The black points on the sextic correspond to lines that are tangent at black points of the cubic, and similarly for the red points. Moreover, the convex set enclosed by the red sextic on the right in Figure 1 is dual, in the sense of convexity, to the spectahedron on the left.

The polynomials in (**??**) and (6) have degree three and six respectively, confirming what was asserted in the caption to Figure 1. A random line L will meet the curve in three (left) or six (right) complex points. Consider the point on the other side that is dual to L. There are three (right) or six (left) complex lines through that point that are tangent to the curve.

We now finally come to *semidefinite programming* (SDP). This refers to the problem of maximizing or minimizing a linear function over a spectrahedron. Linear programming is the special case when the spectrahedron consists of diagonal matrices. If the spectrahedron is given in its standard form representation (1), then we get the SDP in its primal form:

$$\text{Minimize } \langle C, X\rangle \text{ subject to } \langle A_1, X\rangle = b_1,\ \langle A_2, X\rangle = b_2,\ \ldots, \langle A_s, X\rangle = b_s \text{ and } X \succeq 0. \quad (8)$$

Here $C = (c_{ij})$ is a matrix that represents the cost function. Every convex optimization problem has a dual problem. On first glance, it is not so easy to relate that duality to those for plane curves in Figure 1. The semidefinite problem dual to (7) takes the following form

$$\text{Maximize } b^T x \;=\; \sum_{i=1}^{s} b_i x_i \;\text{ subject to }\; C - x_1A_1 - x_2A_2 - \cdots - x_sA_s \;\succeq\; 0. \qquad (9)$$

In this formulation, the spectrahedron of all feasible points lives in $\mathbb{R}^s$, similarly to (4).

We refer to either formulation (7) or (8) as a *semidefinite program*, also abbreviated SDP. Here the term "program" is simply an old-fashioned way of saying "optimization problem". The relationship between the primal and the dual SDP is given by the following theorem:

Theorem 6 (Weak Duality)**.** *If x is any feasible solution to (8) and X is any feasible solution to (7) then $b^T x \leq \langle C, X\rangle$. If the equality $b^T x = \langle C, X\rangle$ holds then both x and X are optimal.*

The term *feasible* means only that the point x resp. X satisfies the equations and inequalities that are required in (8) resp. (7). The point is *optimal* if it is feasible and it solves the program, i.e. it attains the minimum resp. maximum value for that optimization problem.

Proof. The inner product of two positive semidefinite matrices is a non-negative real number:

$$0 \;\leq\; \langle C - \sum_{i=1}^{s} x_i A_i, X \rangle \;=\; \langle C, X \rangle - \sum_{i=1}^{s} x_i \cdot \langle A_i, X \rangle \;=\; \langle C, X \rangle - b^T x. \tag{10}$$

This shows that the optimal value of the minimization problem (7) is an upper bound for the optimal value of the maximization problem (8). If the equality is attained by a pair (X, x) of feasible solutions then X must be optimal for (7) and x must be optimal for (8). □

There is also Strong Duality Theorem which states that, under suitable hypotheses, the *duality gap* $\langle C, X \rangle - b^T x$ must attain the value zero for some feasible pair (X, x). These hypotheses are always satisfied for diagonal matrices, and we recover the Duality Theorem for Linear Programming as a special case. Interior point methods for Linear Programming are numerical algorithms that start at an interior point of the feasible polyhedron and create a path from that point towards an optimal vertex. The same class of algorithms works for Semidefinite Programming. These run in polynomial time and are well-behaved in practice.

Semidefinite Programming has a much larger expressive power than Linear Programming. Many more problems can be phrased as an SDP. We illustrate this with a simple example.

Example 7 (The largest eigenvalue)**.** Let A be a real symmetric $n \times n$ matrix, and consider the problem of computing its largest eigenvalue $\lambda_{\max}(A)$. We would like to solve this without having to write down the characteristic polynomial and extract its roots. Let $C = \mathrm{Id}$ be the identity matrix and consider the SDP problems (7) and (8) with $s = 1$ and $b = 1$. They are

(7') Minimize $\mathrm{trace}(X)$ subject to $\langle A, X \rangle = 1$.

(8') Maximize x subject to $\mathrm{Id} - xA \succeq 0$.

If x^* is the common optimal value of these two optimization problems then $\lambda_{\max}(A) = 1/x^*$.

The inner product $\langle A, X \rangle = \mathrm{trace}(A \cdot X)$ of two positive semidefinite matrices A and X can only be zero when their matrix product $A \cdot X$ is zero. We record this for our situation:

Lemma 8. *If the expression in (9) is zero then* $(C - \sum_{i=1}^{s} x_i A_i) \cdot X$ *is the zero matrix.*

This lemma allows us to state the following algebraic reformulation of SDP:

Corollary 9. *Consider the following system of s linear equations and $\binom{n+1}{2}$ bilinear equations in the $\binom{n+1}{2} + s$ unknown coordinates of the pair (X, x):*

$$\langle A_1, X \rangle = b_1\,,\; \langle A_2, X \rangle = b_2\,,\; \ldots, \langle A_s, X \rangle = b_s \quad \text{and} \quad (C - \sum_{i=1}^{s} x_i A_i) \cdot X = 0. \tag{11}$$

If $X \succeq 0$ *and* $C - \sum_{i=1}^{s} x_i A_i \succeq 0$ *holds then X is optimal for (7) and x is optimal for (8).*

The equations (10) are known as the *Karush-Kuhn-Tucker (KKT) equations.* These play a major role when one explores semidefinite programming from an algebraic perspective. In particular, they allow us to study the nature of the optimal solution as function of the data.

A key feature of the KKT system is that the two optimal matrices have complementary ranks. This follows from the *complementary slackness* condition on the right of (10):

$$\operatorname{rank}\bigl(C - \sum_{i=1}^{s} x_i A_i\bigr) \; + \; \operatorname{rank}(X) \; \leq \; n.$$

In particular, if X is known to be nonzero then the determinant of $C - \sum_{i=1}^{s} x_i A_i$ vanishes. For instance, for the eigenvalue problem in Example 7, we have $(\mathrm{Id} - xA) \cdot X = 0$ and $\langle A, X \rangle = 1$. This implies $\det(\mathrm{Id} - xA) = 0$, so $1/x$ is a root of the characteristic polynomial.

Example 10. Consider the problem of maximizing a linear function $\ell(x,y) = ux + vy$ over the spectrahedron $\mathcal{S}$ in (5). This is the primal SDP (7) with $s = 2$ and $b = (u,v)$ and

$$A_1 \; = \; - \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \qquad \text{and} \qquad A_2 \; = \; - \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

The KKT system (10) consists of eight equations in eight unknowns, with two parameters:

$$2x_{12}+2x_{13}+u = 2x_{13}+2x_{23}+v = 0 \quad \text{and} \quad \begin{pmatrix} 1 & x & x+y \\ x & 1 & y \\ x+y & y & 1 \end{pmatrix} \cdot \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{12} & x_{22} & x_{23} \\ x_{13} & x_{23} & x_{33} \end{pmatrix} \; = \; \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

By eliminating the variables x_{ij} we obtain an ideal I in $\mathbb{Q}[u,v,x,y]$ that characterizes the optimal solution (x^*, y^*) to our SDP as an algebraic function of (u,v). Let ℓ^* now be a new unknown, and consider the elimination ideal $\bigl(I + \langle ux + vy - \ell^* \rangle\bigr) \cap \mathbb{Q}[u,v,\ell^*]$. Its generator is a ternary sextic in u, v, ℓ^*. This is precisely the homogenization of the dual sextic in (6). It expresses the optimal value ℓ^* as an algebraic function of degree six in the cost (u,v).

This relationship between the dual hypersurface and the optimal value function generalizes to arbitrary polynomial optimization problems, including semidefinite programs. This is the content of [1, Theorem 5.23]. We refer to the book [1], and especially Chapter 5, for further reading on spectrahedra, semidefinite programming, and the relevant duality theory.

A fundamental task in Convex Algebraic Geometry [1] is the computation of the convex hull of a given algebraic variety or semialgebraic set. Recall that the *convex hull* of a set is the small convex set containing the given set. Spectrahedra or their linear projections, known as *spectrahedral shadows*, can be used for this task. This matters for optimization since minimizing a linear function over a set is equivalent to minimizing over its convex hull.

Example 11 (Toeplitz Spectrahedron)**.** The *Toeplitz spectrahedron* is the convex body

$$\left\{ (x,y,z) \in \mathbb{R}^3 \; : \; \begin{bmatrix} 1 & x & y & z \\ x & 1 & x & y \\ y & x & 1 & x \\ z & y & x & 1 \end{bmatrix} \succeq 0 \right\}. \tag{12}$$

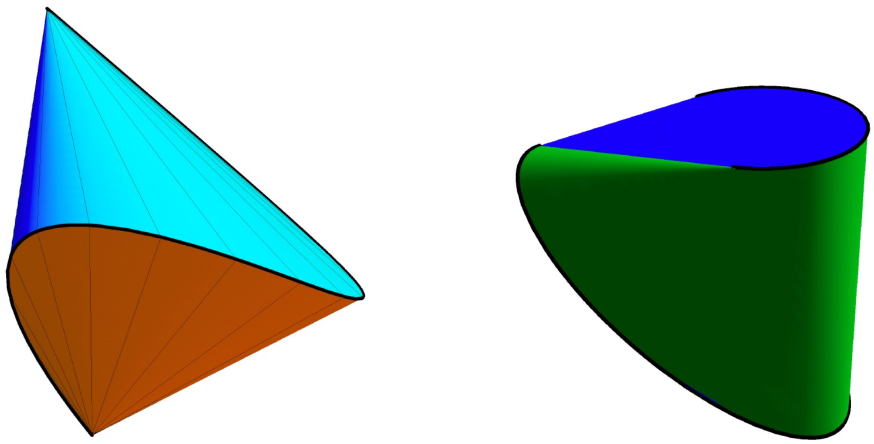

Figure 2: Toeplitz spectrahedron and its dual convex body.

The determinant of the given *Toeplitz matrix* of size 4×4 factors as

$$(x^2 + 2xy + y^2 - xz - x - z - 1)(x^2 - 2xy + y^2 - xz + x + z - 1).$$

The Toeplitz spectrahedron (11) is the convex hull of the *cosine moment curve*

$$\{ (\cos(\theta), \cos(2\theta), \cos(3\theta)) \,:\, \theta \in [0, \pi] \}.$$

The curve and its convex hull are shown on the left in Figure 2. The two endpoints, $(x, y, z) = (1, 1, 1)$ and $(x, y, z) = (-1, 1, -1)$, correspond to rank 1 matrices. All other points on the curve have rank 2. To construct the Toeplitz spectrahedron geometrically, we form the cone from each endpoint over the cosine curve, and we intersect these two quadratic cones. The two cones intersect along this curve and the line through the endpoints of the cosine curve.

Shown on the right in Figure 2 is the dual convex body. It is the set of trigonometric polynomials $1 + a_1 \cos(\theta) + a_2 \cos(2\theta) + a_3 \cos(3\theta)$ that are nonnegative on $[0, \pi]$. This convex body is not a spectrahedron because it has a non-exposed edge (cf. [1, Exercise 6.13]).

Semidefinite programming can used to model and solve arbitrary polynomial optimization problems. The key to this is the representation of nonnegative polynomials in terms of sums of squares, or, more generally, the Real Nullstellensatz (as seen on May 15). We explain this for the simplest scenario, namely the problem of unconstrained polynomial optimization.

Let $f(x_1, \ldots, x_n)$ be a polynomial of even degree $2p$, and suppose that f attains a minimal real value f^* on $\mathbb{R}^n$. Our goal is to compute f^* and a point $\mathbf{u}^* \in \mathbb{R}^n$ such that $f(\mathbf{u}^*) = f^*$. Minimizing a function is equivalent to finding the best possible lower bound λ for that function. Our goal is therefore equivalent to solving the following optimization problem:

$$\text{Maximize } \lambda \text{ such that } \quad f(\mathbf{x}) - \lambda \geq 0 \quad \text{for all } \mathbf{x} \in \mathbb{R}^n. \tag{13}$$

This is a difficult problem. Instead, we consider the following relaxation:

$$\text{Maximize } \lambda \text{ such that } \quad f(\mathbf{x}) - \lambda \quad \text{is a sum of squares in } \mathbb{R}[\mathbf{x}]. \tag{14}$$

Here *relaxation* means that we enlarged the set of feasible solutions. Indeed, every sum of squares is nonnegative, but not every nonnegative polynomial is a sum of squares of polynomials. For instance, the Motzkin polynomial $x^4y^2 + x^2y^4 + 1 - 3x^2y^2$ is nonnegative but it is not a sum of squares of polynomials (as seen on May 15).

For that reason, the optimal value of (13) is always a lower bound for the optimal value of (12), but the two values can be different in some cases. However, here is the good news:

Proposition 12. *The optimization problem (13) is a semidefinite program.*

Proof. Let $\mathbf{x}^{[p]}$ be the column vector whose entries are all monomials in $x_1, \ldots, x_n$ of degree $\leq p$. Thus $\mathbf{x}^{[p]}$ has length $\binom{n+p}{n}$. Let $G = (g_{ij})$ be a symmetric $\binom{n+p}{n} \times \binom{n+p}{n}$ matrix with unknown entries. Then $(\mathbf{x}^{[p]})^T \cdot G \cdot \mathbf{x}^{[p]}$ is a polynomial of degree $d = 2p$ in $x_1, \ldots, x_n$. We set

$$f(\mathbf{x}) - \lambda \quad = \quad (\mathbf{x}^{[p]})^T \cdot G \cdot \mathbf{x}^{[p]}. \tag{15}$$

By collecting coefficients of the $\mathbf{x}$-monomials, this translates into a system of $\binom{2p+n}{n}$ linear equations in the unknowns g_{ij} and λ. The number of unknowns is $\binom{\binom{n+p}{n}+1}{2} + 1$.

Suppose the linear system (14) has a solution (G, λ) such that G is positive semidefinite. Then we can write $G = H^T H$ where H is a real matrix with r rows and $\binom{p+n}{n}$ columns. (This is known as a *Cholesky factorization* of H.) The polynomial in (14) then equals

$$f(\mathbf{x}) - \lambda \quad = \quad (H\mathbf{x}^{[p]})^T \cdot (H\mathbf{x}^{[p]}). \tag{16}$$

This is the scalar product of a vector of length r with itself. Hence $f(\mathbf{x}) - \lambda$ is a sum of squares. Conversely, every representation of $f(\mathbf{x}) - \lambda$ as a sum of squares of polynomials uses polynomials of degree $\leq p$, and it can hence be written in the form as in (15).

Our argument shows that the optimization problem (13) is equivalent to

$$\text{Maximize } \lambda \text{ subject to } (G, \lambda) \text{ satisfying the linear equations (15) and } G \succeq 0. \tag{17}$$

This is a semidefinite programming problem, and so the proof is complete. □

If $n = 1$ or $d = 2$ or ($n = 2$ and $d = 4$) then, according to Hilbert, every nonnegative polynomial is a sum of squares. In those special cases, problems (12) and (16) are equivalent.

Example 13 ($n = 1, p = 2, d = 4$)**.** Suppose we seek to find the minimum of the degree 4 polynomial $f(x) = 3x^4 + 4x^3 - 12x^2$. Of course, we know how to do this using Calculus. However, we here present the SDP approach. The linear equations (14) have a one-dimensional space of solutions. Introducing a parameter μ for that line, the solutions can be written as

$$f(x) - \lambda \quad = \quad \begin{pmatrix} x^2 & x & 1 \end{pmatrix} \begin{pmatrix} 3 & 2 & \mu - 6 \\ 2 & -2\mu & 0 \\ \mu - 6 & 0 & -\lambda \end{pmatrix} \begin{pmatrix} x^2 \\ x \\ 1 \end{pmatrix}. \tag{18}$$

Consider the set of all pairs (λ, μ) such that the 3×3 matrix in (17) is positive semidefinite. This set is a cubic spectrahedron in the plane $\mathbb{R}^2$, just like that shown on the left in (1).

We seek to maximize λ over all points in that cubic spectrahedron. The optimal point equals $(\lambda^*, \mu^*) = (-32, -2)$. Substituting this into the matrix in (17) we obtain a positive definite matrix of rank 2. This can be factored as $G = H^T H$, where H has format 2×3. The resulting representation (15) as a sum of two squares equals

$$f(x) - \lambda^* \;=\; f(x) + 32 \;=\; \big((\sqrt{3}x - \frac{4}{\sqrt{3}}) \cdot (x+2)\big)^2 \;+\; \frac{8}{3}(x+2)^2.$$

The right hand side is nonnegative for all $x \in \mathbb{R}$, and it takes the value 0 only when $x = -2$.

Any polynomial optimization problem can be translated into a relaxation that is a semidefinite programming problem. If we are minimizing $f(\mathbf{x})$ subject to some polynomial constraints, then we seek a certificate for $f(\mathbf{x}) - \lambda < 0$ to have no solution. This certificate is promised by the Real Nullstellensatz or Postitivstellensatz. If we fix a degree bound then the existence of a certificate translates into a semidefinite program, and so does the additional requirement for λ to be minimal. This relaxation may or may not give the correct solution for some fixed degree bound. However, if one increases the degree bound then the SDP formulation is more likely to succeed, albeit at the expense of having to solve a much larger problem. This is a powerful and widely used approach to polynomial optimization, known as *SOS programming*. The term *Lasserre hierarchy* refers to varying the degree bounds.

Every spectrahedron $\mathcal{S} = \mathcal{L} \cap \mathrm{PSD}_n$ has a special point in its relative interior. This point, defined as the unique matrix in $\mathcal{S}$ whose determinant is maximal, is known as *analytic center*. Finding the analytic center of $\mathcal{S}$ is a convex optimization problem, since the function $X \mapsto \log\det(X)$ is strictly convex on the cone of positive definite matrices X. The analytic center is important for semidefinite programming because it serves as the starting point for interior point methods. Indeed, the *central path* of an SDP starts at the analytic center and runs to the optimal face. It is computed by a sequence of numerical approximations.

Example 14. The determinant function takes on all values between 0 and 1 on the spectrahedron $\mathcal{S}$ in (5). The value 1 is attained only by the identity matrix, for $(x, y) = (0, 0)$. This point is therefore the analytic center of $\mathcal{S}$.

We close by relating spectrahedra and their analytic centers to statistics. Every positive definite $n \times n$ matrix $\Sigma = (\sigma_{ij})$ is the covariance matrix of a multivariate normal distribution. Its inverse Σ^{-1} is known as the concentration matrix of that distribution.

A *Gaussian graphical model* is specified by requiring that some off-diagonal entries of Σ^{-1} are zero. These entries correspond to the non-edges of the graph. Maximum likelihood estimation for this graphical model translates into a matrix completion problem. Suppose that S is the sample covariance matrix of a given data set. We regard S as a partial matrix, with visible entries only on the diagonal and on the edges of the graph. One considers the set of all completions of the non-edge entries that make the matrix S positive definite. The set of all these completions is a spectrahedron. Maximum likelihood estimate for the data S in the graphical model amounts to maximizing the logarithm of the determinant. We hence seek to compute the analytic center of the spectrahedron of all completions.

Example 15 (Positive definite matrix completion)**.** Suppose that the eight entries σ_{ij} in the following symmetric 4×4-matrix are visible, but the entries x and y are unknown:

$$\Sigma \quad = \quad \begin{pmatrix} \sigma_{11} & \sigma_{12} & x & \sigma_{14} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} & y \\ x & \sigma_{23} & \sigma_{33} & \sigma_{34} \\ \sigma_{14} & y & \sigma_{34} & \sigma_{44} \end{pmatrix}. \tag{19}$$

This corresponds to the graphical model of the four-cycle $12, 23, 34, 41$. Given visible entries σ_{ij}, we consider the set of pairs (x, y) that make Σ positive definite. This is the interior of a planar spectrahedron $\mathcal{S}_\sigma$ bounded by a quartic curve. The MLE is the analytic center of $\mathcal{S}_\sigma$.

One is also interested in conditions on the σ_{ij} such that $\mathrm{int}(\mathcal{S}_\sigma)$ is non-empty? In other words, when can we find (x, y) that make Σ positive definite? A necessary condition is that the four diagonal entries σ_{ii} and the four visible principal 2×2-minors are positive:

$$\sigma_{11}\sigma_{22} > \sigma_{12}^2\,, \;\; \sigma_{22}\sigma_{33} > \sigma_{23}^2\,, \;\; \sigma_{33}\sigma_{44} > \sigma_{34}^2\,, \;\; \sigma_{11}\sigma_{44} > \sigma_{14}^2. \tag{20}$$

But this is not sufficient. The true answer is a cone that is bounded by the hypersurface

$$\begin{gathered} \sigma_{33}^2\sigma_{44}^2\sigma_{12}^4 - 2\sigma_{22}\sigma_{33}^2\sigma_{44}\sigma_{12}^2\sigma_{14}^2 - 2\sigma_{11}\sigma_{33}\sigma_{44}^2\sigma_{12}^2\sigma_{23}^2 - 2\sigma_{11}\sigma_{22}\sigma_{33}\sigma_{44}\sigma_{14}^2\sigma_{23}^2 + 4\sigma_{33}\sigma_{44}\sigma_{12}^2\sigma_{14}^2\sigma_{23}^2 \\ +\sigma_{11}^2\sigma_{44}^2\sigma_{23}^4 + 8\sigma_{11}\sigma_{22}\sigma_{33}\sigma_{44}\sigma_{12}\sigma_{14}\sigma_{23}\sigma_{34} - 4\sigma_{33}\sigma_{44}\sigma_{12}^3\sigma_{14}\sigma_{23}\sigma_{34} - 4\sigma_{22}\sigma_{33}\sigma_{12}\sigma_{14}^3\sigma_{23}\sigma_{34} \\ +\sigma_{22}^2\sigma_{33}^2\sigma_{14}^4 - 4\sigma_{11}\sigma_{44}\sigma_{12}\sigma_{14}\sigma_{23}^3\sigma_{34} - 2\sigma_{11}\sigma_{22}\sigma_{33}\sigma_{44}\sigma_{12}^2\sigma_{34}^2 - 2\sigma_{11}\sigma_{22}^2\sigma_{33}\sigma_{14}^2\sigma_{34}^2 + 4\sigma_{22}\sigma_{33}\sigma_{12}^2\sigma_{14}^2\sigma_{34}^2 \\ -2\sigma_{11}^2\sigma_{22}\sigma_{44}\sigma_{23}^2\sigma_{34}^2 + 4\sigma_{11}\sigma_{44}\sigma_{12}^2\sigma_{23}^2\sigma_{34}^2 + 4\sigma_{11}\sigma_{22}\sigma_{14}^2\sigma_{23}^2\sigma_{34}^2 - 4\sigma_{11}\sigma_{22}\sigma_{12}\sigma_{14}\sigma_{23}\sigma_{34}^3 + \sigma_{11}^2\sigma_{22}^2\sigma_{34}^4. \end{gathered}$$

This polynomial of degree eight is found by eliminating x and y from the determinant and its partial derivatives with respect to x and y, after saturating by the ideal of 3×3-minors. For more details on this example and its generalizations we refer to [2, Theorem 4.8].

Exercises

1. Prove Theorem 2.
2. Show that a real symmetric matrix G is positive semidefinite if and only if it admits a Cholesky factorization $G = H^T H$ over the real numbers, with H upper triangular.
3. What is the largest eigenvalue of any of the 3×3 matrices in the set $\mathcal{S}$ in (5)?
4. Maximize and minimize the linear function $13x + 17y + 23z$ over the spectrahedron $\mathcal{S}$ in Example 4. Use SDP software if you can.
5. Maximize and minimize the linear function $13x + 17y + 23z$ over the Toeplitz spectrahedron in Example 11. Use SDP software if you can.
6. Write the dual SDP and solve the KKT system for the previous two problems.
7. Determine the convex body dual to the spectrahedron $\mathcal{S}$ in Example 4.

8. Consider the problem of minimizing the univariate polynomial $x^6 + 5x^3 + 7x^2 + x$. Express this problem as a semidefinite program.

9. In the partial matrix (18) set $\sigma_{11} = \sigma_{22} = \sigma_{33} = \sigma_{44} = 5$, $\sigma_{12} = \sigma_{23} = \sigma_{34} = 1$ and $\sigma_{14} = 2$. Compute the spectrahedron $\mathcal{S}_\sigma$, draw a picture, and find the analytic center.

10. Find numerical values for the eight entries σ_{ij} in (18) that satisfy (19) but $\text{int}(\mathcal{S}_\sigma) = \emptyset$.

References

[1] G. Blekherman, P. Parrilo and R. Thomas: *Semidefinite Optimization and Convex Algebraic Geometry*, MOS-SIAM Series on Optimization **13**, 2012.

[2] B. Sturmfels and C. Uhler: *Multivariate Gaussian, semidefinite matrix completion, and convex algebraic geometry*, Ann. Inst. Statist. Math. **62** (2010) 603–638.