Algebraic Methods for Multivariate Polynomial Optimization

Jiawang Nie

July 30, 2010

1 An overview

Consider the optimization problem

$$\begin{array}{rl} \min\limits_{x\in\mathbb{R}^n} & f(x) \\ \text{s.t.} & h_1(x) = \cdots = h_{m_1}(x) = 0 \\ & g_1(x) \geq 0, \ldots, g_{m_2}(x) \geq 0 \end{array}$$

where $f(x), g_i(x), h_j(x)$ are polynomial functions in $x \in \mathbb{R}^n$. Let S be its feasible set and f_{min} be its global minimum. We are interested in finding f_{min}. This problem is NP-hard, even when one of the polynomials is quadratic. A standard approach for solving (5.1) is semidefinite programming (SDP) and Sum of Squares (SOS) relaxations proposed in the original work by Lasserre, Parrilo and Sturmfels.

In this lecture series, we aim to understand the properties, both mathematical and computational, of semidefinite programming, sum of squares, polynomial optimization, and their applications. In particular, we want to work towards methods that will enable the solution of optimization problems with feasible sets that are defined through polynomial systems. There is a very interesting interaction between algebraic geometry and convex optimization.

To understand better what is going on, we will embark in a journey to learn a wide variety of methods used to approach these problems. Some of our stops along the way will include:

- Some backgrounds
- Semidefinite programming
- Lasserre type SDP relaxation
- Jacobian type SDP relaxation

Several software for polynomial optimization are available. Typical ones are

- GloptiPoly (using SeDuMi or SDPT3)
- SOSTOOLS (using SeDuMi or SDPT3)
- YALMIP (using SeDuMi or SDPT3)

2 Some backgrounds

2.1. Linear Algebra

A real matrix $A \in \mathbb{R}^{n\times n}$ is symmetric if $A = A^T$.

Theorem 2.1. *Suppose $A \in \mathbb{R}^{n\times n}$ is symmetric. Then we have*

- *All the eigenvalues are real.*
- *Eigenvectors for distinct eigenvalues are orthogonal to each other.*
- *There exists an orthogonal matrix Q such that*

$$Q^T AQ = diag(\lambda_1, \dots, \lambda_n)$$

where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A.

For a symmetric $A \in \mathbb{R}^{n\times n}$, denote by $\lambda_i(A)$ the i-th biggest eigenvalue of A.

Theorem 2.2. *Suppose $A \in \mathbb{R}^{n\times n}$ is symmetric. Then we have*

- $\lambda_1(A) = \max\limits_{x\neq 0} \frac{x^T Ax}{x^T x}$, *and* $\lambda_n(A) = \min\limits_{x\neq 0} \frac{x^T Ax}{x^T x}$.
- *Max-Min and Min-Max characterizations for $\lambda_i(A)$:*

$$\lambda_i(A) = \min_{\substack{S\subseteq\mathbb{R}^n \\ \dim(S)=n-i+1}} \max_{0\neq x\in S} \frac{x^T Ax}{x^T x}.$$

$$\lambda_i(A) = \max_{\substack{S\subseteq\mathbb{R}^n \\ \dim(S)=i}} \min_{0\neq x\in S} \frac{x^T Ax}{x^T x}.$$

- *Each eigenvalue of A is a critical value of optimization problem*

$$\min \quad x^T Ax \quad \textit{subject to} \quad x^T x = 1.$$

If a symmetric $A \in \mathbb{R}^{n\times n}$ has eigenvalues such that

$$\lambda_1(A) \geq \cdots \lambda_r(A) > 0 = \lambda_{r+1}(A) = \cdots = \lambda_{r+s}(A) > \lambda_{r+s+1}(A) \geq \cdots \geq \lambda_{r+s+t}(A),$$

we say the triple (r, s, t) is the inertia of A, and $r - t$ is the signature of A.

For two symmetric A, B, if there exists invertible P such that $A = P^T BP$, we say A is congruent to B.

Theorem 2.3. *If A, B are symmetric and congruent to each other, then they have the same inertia and signature.*

Proof. Use the min-max or max-min characterization of eigenvalues. $\square$

Definition 2.4. *A symmetric matrix $A \in S\mathbb{R}^{n\times n}$ is called positive semidefinite if $x^T Ax \geq 0$ for all $x \in \mathbb{R}^n$, and is called positive definite if $x^T Ax > 0$ for all nonzero $x \in R^n$. The set of positive semidefinite matrices is denoted $\mathcal{S}^n_+$, and the set of positive definite matrices is denoted by $\mathcal{S}^n_{++}$.*

If A is positive semidefinite (resp. positive definite), we denote $A \succeq 0$ (resp. $A \succ 0$).

Theorem 2.5. *The following statements are equivalent:*

- *The symmetric matrix A is positive semidefinite.*
- *All eigenvalues of A are nonnegative.*
- *All the principal minors of A are nonnegative.*
- *There exists B such that $A = B^T B$.*

Theorem 2.6. *The following statements are equivalent:*

- *The symmetric matrix A is positive definite.*
- *All eigenvalues of A are positive.*
- *All the leading principal minors of A are positive.*
- *There exists nonsingular square matrix B such that $A = B^T B$.*

If $X - Y \succeq 0$, then we write $X \succeq Y$.

Theorem 2.7. *For two symmetric X, Y, if $X \succeq Y$, then*

$$\lambda_i(X) \geq \lambda_i(Y), \text{ for every } i.$$

Here $\lambda_i(\cdot)$ denotes the i-th largest eigenvalue.

Proof. Use the characterization of $\lambda_i(\cdot)$. $\square$

Congruent transformations preserve positive semidefiniteness.

Theorem 2.8. *Let P be a nonsingular matrix.*

- *The $A \succeq 0$ if and only if $P^T AP \succeq 0$.*
- *The $A \succ 0$ if and only if $P^T AP \succ 0$.*

If P is singular, then $A \succeq 0$ implies $P^T AP \succeq 0$ (while the reverse might not).

Theorem 2.9 (Schur's complement)**.** *Let $A \succ 0$. Then*

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0 \quad \Longleftrightarrow \quad C - B^T A^{-1} B \succeq 0.$$

If matrices A, B satisfy

$$\begin{bmatrix} A & B \\ B^T & 0 \end{bmatrix} \succeq 0,$$

then $A \succeq 0$ and $B = 0$.

2.2. Convex sets and optimization

Definition 2.10. *A set $S \subset \mathbb{R}^n$ is convex if $x, y \in S$ implies $\lambda x + (1-\lambda)y \in S$ for all $\lambda \in [0,1]$.*

Definition 2.11. *Let $S \subset \mathbb{R}^n$ be a convex set. A convex subset F of S is called a face of S if*

$$\lambda x + (1-\lambda)y \in F, x \in S, y \in S, \lambda \in [0,1] \quad \Longrightarrow \quad x, y \in F.$$

If a face F has codimension one in S, it is called a facet. If a face $F = \{x\}$ is a singleton, x is called an extremal point of S.

For any set $S \subset \mathbb{R}^n$, its convex hull is the smallest convex set containing it, and is denoted by $conv(S)$.

Definition 2.12. *A set $S \subset \mathbb{R}^n$ is a cone if $\lambda x \in S$ for all $x \in S$ and $\lambda \geq 0$.*

A cone K is **pointed** if $K \cap -K = \{0\}$, and **solid** if the interior of K is not empty. A cone that is convex, closed, pointed and solid is called a **proper** cone. Type proper cones are

- Nonnegative orthant $\mathbb{R}^n_+$.
- Laurenz cone $\mathcal{L}_n = \{(x,t) \in \mathbb{R}^{n+1} : \|x\|_2 \leq t\}$.
- Positive semidefinite cone $\mathcal{S}^n_+$.

Definition 2.13. *The dual of a cone $K \subset \mathbb{R}^n$ is defined as*

$$K^* = \{y \in \mathbb{R}^n : \langle x, r \rangle \geq 0 \quad \forall\, x \in K\}.$$

If K is a closed convex cone, then $K = K^{**}$. The dual set of a proper cone is also a proper cone, called the **dual cone**. An element x is in the interior of the cone K if and only if

$$\langle x, y \rangle > 0, \quad \forall\, y \in K^*, y \neq 0.$$

A standard linear conic optimization is

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & c^T x \\ \text{s.t.} \quad & Ax = b, \quad x \in K \end{aligned}$$

where $A \in \mathbb{R}^{m\times n}$ and K is a proper cone. Its standard dual is

$$\begin{aligned} \max_{y\in\mathbb{R}^m} \quad & b^T y \\ \text{s.t.} \quad & c - A^T y \in K^*. \end{aligned}$$

Here K^* is the dual cone of K.

- When $K = \mathbb{R}^n_+$, it reduces to LP.
- When $K = \mathcal{L}_n$, it reduces to second order cone programming (SOCP).
- When $K = \mathcal{S}^n_+$, it reduces to semidefinite programming.

3 Semidefinite programming

A very important notion in modern optimization is convexity. Significant advances have been made in this area. Efficient numerical methods have been developed.

A standard semidefinite programming (SDP) problem is

$$(P): \qquad \begin{cases} \min\limits_{X} & C \bullet X \\ \text{s.t.} & A_i \bullet X = b_i, \quad i = 1, \dots, m \\ & X = X^T \succeq 0 \end{cases}$$

where all $A_i \in S\mathbb{R}^{n\times n}, b \in \mathbb{R}^m, C \in S\mathbb{R}^{n\times n}$, $\bullet$ denotes the standard Frobenius inner product, and $X \succeq 0$ means X is positive semidefinite.

- When all A_i and C are diagonal, (P) reduces to LP.
- When X is not symmetric, it is equivalent to symmetric case.
- Problem (P) is convex, but its boundary is typically highly nonlinear.

The dual problem of (P) is

$$(D): \qquad \begin{cases} \max\limits_{y} & b^T y \\ \text{s.t.} & C - \sum_{i=1}^m A_i \succeq 0. \end{cases}$$

Example 3.1. Consider the primal problem

$$\begin{aligned} \min_X \quad & I \bullet X \\ \text{s.t.} \quad & \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \bullet X = 2, \quad X = X^T \succeq 0 \end{aligned}$$

Its dual is

$$\begin{aligned} \max_y \quad & 2y \\ \text{s.t.} \quad & \begin{bmatrix} 1 & -y \\ -y & 1 \end{bmatrix} \succeq 0. \end{aligned}$$

Example 3.2. Consider the primal problem

$$\begin{aligned} \min_X \quad & \textstyle\sum_{i \neq j} X_{ij} \\ \text{s.t.} \quad & X_{11} = \cdots X_{nn} = 1, \quad X = X^T \succeq 0 \end{aligned}$$

Its dual is

$$\begin{aligned} \max_y \quad & y_1 + \ldots + y_n \\ \text{s.t.} \quad & \begin{bmatrix} -y_1 & 1 & \cdots & 1 \\ 1 & -y_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 1 & \cdots & 1 & -y_n \end{bmatrix} \succeq 0. \end{aligned}$$

The condition that $e^T X e \geq 0$ implies that any feasible X satisfy

$$\sum_{i \neq j} X_{ij} \geq -n.$$

So the optimal value is $-n$.

Theorem 3.3 (Weak Duality)**.** *For any X feasible for (P) and y feasible for (D), it holds that*

$$C \bullet X \geq b^T y.$$

- (P) can be formulated in (D), and (D) can be formulated in (P).
- If (P) is unbounded from below, (D) is not feasible.
- If (D) is unbounded from above, (P) is not feasible.
- Both (P) and (D) would be infeasible simultaneously.

Theorem 3.4 (Weak Duality)**.** *For any X feasible for (P) and y feasible for (D), it holds that*

$$C \bullet X \geq b^T y.$$

Let p^* be the optimal value of (P), and d^* be the optimal value of (D), then

$$p^* \geq d^*.$$

Some remarks about (P) and (D)

- It is possible that p^* and/or d^* are not attainable. For example, consider

$$\min \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \bullet X, \quad s.t. \quad \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix} \bullet X = -1, \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix} \bullet X = 0, \, X \succeq 0.$$

  Its dual is

$$\max -y_1 \quad s.t. \quad \begin{bmatrix} y_1 & 1 \\ 1 & y_2 \end{bmatrix} \succeq 0.$$

  It holds that $p^* = d^* = 0$ but d^* is not attainable.

- It is possible that $p^* > d^*$. Consider

$$\min \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \bullet X, \quad s.t. \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & \\ 0 & 0 & 0 \end{bmatrix} \bullet X = 0, \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 2 \end{bmatrix} \bullet X = 2, \, X \succeq 0.$$

  Any feasible X for the above has the form

$$\begin{bmatrix} 0 & \xi_1 & \xi_2 \\ \xi_1 & \xi_3 & \xi_4 \\ \xi_2 & \xi_4 & 1 - \xi_1 \end{bmatrix} \succeq 0.$$

  So the optimal value $p^* = 1$. The dual problem is

$$\max 2y_2 \quad s.t. \quad \begin{bmatrix} -y_2 & -y_2 & 0 \\ -y_2 & 0 & 0 \\ 0 & 0 & 1 - 2y_2 \end{bmatrix} \succeq 0.$$

  So the dual optimal value is $d^* = 0$. Both p^* and d^* are attainable, but $p^* > d^*$.

- It is possible that (P) is infeasible (by convention $p^* = +\infty$), but (D) is feasible and d^* is finite. Consider

$$\min \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \bullet X, \quad s.t. \quad \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \bullet X = 0, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \bullet X = 2, \, X \succeq 0.$$

  Clearly, the above primal SDP is infeasible, so $p^* = +\infty$. Its dual is

$$\max 2y_2 \quad s.t. \quad \begin{bmatrix} -y_1 & -y_2 \\ -y_2 & 0 \end{bmatrix} \succeq 0.$$

  The above dual SDP is feasible and $d^* = 0$.

Denote the feasible sets of (P) and (D) as

$$\mathcal{F}(P) = \left\{X \in S\mathbb{R}^{n\times n} : \mathcal{A}(X) = b,\, X \succeq 0\right\},$$

$$\mathcal{F}^\circ(P) = \left\{X \in S\mathbb{R}^{n\times n} : \mathcal{A}(X) = b,\, X \succ 0\right\},$$

$$\mathcal{F}(D) = \left\{y \in \mathbb{R}^m : C - \sum_{i=1}^m y_i A_i \succeq 0\right\},$$

$$\mathcal{F}^\circ(D) = \left\{y \in \mathbb{R}^m : C - \sum_{i=1}^m y_i A_i \succ 0\right\}.$$

In the above $\mathcal{A}(X) = (A_1 \bullet X, \ldots, A_m \bullet X)$.

Theorem 3.5 (Strong Duality)**.** *Suppose $\mathcal{F} \neq \emptyset$ and $\mathcal{F}^\circ(D) \neq \emptyset$. Then (P) has a nonempty compact set of minimizers, and the optimal values of (P) and (D) are equal, that is, $p^* = d^*$.*

Theorem 3.6. *If $\mathcal{F}^\circ(P) \neq \emptyset$ and $\mathcal{F}^\circ(D) \neq \emptyset$, then both (P) and (D) have a nonempty compact set of minimizers, and $p^* = d^*$.*

Theorem 3.7 (Optimality Condition)**.** *Suppose $\mathcal{F}^\circ(P) \neq \emptyset$ and $\mathcal{F}^\circ(D) \neq \emptyset$. Then X is optimal for (P) and y is optimal for (D) if and only if*

$$\begin{aligned} \mathcal{A}(X) &= b \\ \mathcal{A}^*(y) + S &= C \\ XS &= 0 \\ X, S &\succeq 0. \end{aligned}$$

4 Lasserre type SDP relaxation

Lasserre's relaxation is a typical approach to solve polynomial optimization by using semidefinite programming and sum of squares techniques. We begin with a short review of SOS polynomials.

Let $p(x) \in \mathbb{R}[x] := \mathbb{R}[x_1, \ldots, x_n]$. We say $p(x)$ is positive semidefinite (psd) or nonnegative if $p(u) \geq 0$ for every $u \in \mathbb{R}^n$. We say $p(x)$ is sum of squares (SOS) if there exist real polynomials $q_1(x), \ldots, q_r(x)$ such that

$$p(x) = q_1(x)^2 + \cdots + q_r(x)^2.$$

Clearly, if $p(x)$ is SOS, then $p(x)$ is psd.

- The polynomial $2x_1^4 + 2x_1^3x_2 - x_1^2x_2^2 + 5x_2^4$ is SOS, equaling

$$\frac{1}{2}\Big((2x_1^2 - 3x_2^2 + x_1x_2)^2 + (x_2^2 + 3x_1x_2)^2\Big).$$

- The polynomial $x_1^4 - (2x_2x_3 + 1)x_1^2 + (x_2^2x_3^2 + 2x_2x_3 + 2)$ is SOS, equaling
$$1 + x_1^2 + (1 - x_1^2 + x_2x_3)^2$$
- The polynomial $3(x_1^4 + x_2^4 + x_3^4 + x_4^4 - 4x_1x_2x_3x_4)$ is SOS, being
$$(x_1^2 + x_3^2 - x_2^2 - x_4^2)^2 + (x_1^2 + x_2^2 - x_3^2 - x_4^2)^2 + (x_1^2 + x_4^2 - x_2^2 - x_3^2)^2 + \\ 2(x_1x_2 - x_3x_4)^2 + 2(x_1x_3 - x_2x_4)^2 + 2(x_1x_4 - x_2x_3)^2.$$

Note that not every nonnegative polynomial is SOS. For example, the Motzkin polynomial
$$M(x, y, z) := x^4y^2 + x^2y^4 + z^6 - 3x^2y^2z^2$$
is psd but not SOS.

4.1. Cone of nonnegative polynomials

Denote two sets of polynomials
$$P_{n,2d} = \{f \in \mathbb{R}[x_1, \ldots, x_n] : f(x) \text{ is psd of degree } 2d\},$$
$$\Sigma_{n,2d} = \{f \in \mathbb{R}[x_1, \ldots, x_n] : f(x) \text{ is SOS of degree } 2d\}.$$

Theorem 4.1 (Hilbert, 1888)**.** *The equality $P_{n,2d} = \Sigma_{n,2d}$ holds if and only if*
$$n = 1, \quad \textit{or} \quad 2d = 2, \quad \textit{or} \quad (n, 2d) = (2, 4).$$

Let $f(x) \in \mathbb{R}[x_1, \ldots, x_n]_{2d}$. Let $[x]_d$ denote the column vector of all monomials of degree at most d. Define symmetric matrices A_α such that
$$[x]_d[x]_d^T = \sum_{|\alpha| \le 2d} A_\alpha x^\alpha.$$
Then $f(x)$ is SOS if and only if there exists $X \succeq 0$ such that
$$f(x) \equiv [x]_d^T X [x]_d,$$
or equivalently by comparing coefficients
$$f_\alpha = A_\alpha \bullet X, \quad \forall\, |\alpha| \le 2d.$$
Testing whether $f(x)$ is SOS would be done by
$$\text{finding } X \succeq 0, \quad A_\alpha \bullet X = f_\alpha, \quad \forall\, |\alpha| \le 2d.$$
This is an SDP feasibility problem.

Example 4.2. The polynomial $p(x) = 2x_1^4 + 2x_1^3x_2 - x_1^2x_2^2 + 5x_2^4$ has representation

$$\begin{bmatrix} x_1^2 \\ x_2^2 \\ x_1x_2 \end{bmatrix}^T \underbrace{\begin{bmatrix} 2 & -\alpha & 1 \\ -\alpha & 5 & 0 \\ 1 & 0 & -1+2\alpha \end{bmatrix}}_{G} \begin{bmatrix} x_1^2 \\ x_2^2 \\ x_1x_2 \end{bmatrix}$$

$$p(x) \text{ is SOS} \Longleftrightarrow \exists\ G \succeq 0$$

When $\alpha = 3$, the Gram matrix G is positive semidefinite.

For any vector y indexed by $\alpha \in \mathbb{N}^n$, define its moment matrix as

$$M_d(y) := \sum_{|\alpha| \le 2d} y_\alpha A_\alpha$$

For instance, $n = 2, d = 2$, $\Sigma^*_{2,2d}$ consists of all y such that $M_2(y) \succeq 0$, where

$$M_2(y) = \begin{bmatrix} y_{00} & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} \\ y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} \\ y_{02} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} \\ y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22} \\ y_{11} & y_{21} & y_{12} & y_{31} & y_{22} & y_{13} \\ y_{02} & y_{12} & y_{03} & y_{22} & y_{13} & y_{04} \end{bmatrix}.$$

Theorem 4.3. *The dual cone $\Sigma^*_{n,2d}$ is $\{y : M_d(y) \succeq 0\}$.*

Proof. Suppose $y \in \Sigma^*_{n,2d}$, then

$$f^T y \ge 0, \quad \forall f \in \Sigma_{n,2d}.$$

Write $f(x) = [x]_d^T X [x]_d = X \bullet ([x]_d[x]_d^T)$, then comparing coefficients gives

$$f^T y \;=\; X \bullet ([x]_d[x]_d^T) \;=\; X \bullet M_d(y) \ge 0, \quad \forall X \succeq 0.$$

We clearly have $M_d(y) \succeq 0$.

The other direction would be obtained by applying Hahn-Banach Theorem. □

Let $f(x) \in \mathbb{R}[x]$ of degree $2d$. Consider problem

$$f_{min} \quad = \quad \min_{x \in \mathbb{R}^n} \quad f(x).$$

It is NP-hard to find f_{min}. It is equivalent to

$$\begin{aligned} f_{min} \quad = \quad & \max \ \gamma \\ & s.t. \ \ f(x) - \gamma \ge 0 \, \forall x \in \mathbb{R}^n. \end{aligned}$$

Replacing psd by SOS, we get the relaxation

$$\begin{array}{rcl} f_{sos} & = & \max \ \gamma \\ & & s.t. \ \ f(x) - \gamma \text{ is SOS.} \end{array}$$

The constraint above is $f(x) - \gamma \in \Sigma_{n,2d}$. If we write $f(x) = \sum_{|\alpha| \le 2d} f_\alpha x^\alpha$, the above is equivalent to the SDP

$$(P): \left\{ \begin{array}{rcll} f_{sos} & = & \max & \gamma \\ & & s.t. & A_0 \bullet X + \gamma \ = f_0, \\ & & & A_\alpha \bullet X \qquad = f_\alpha, \quad \forall \alpha \neq 0. \end{array} \right.$$

The dual of the above SDP is

$$(D): \quad \left\{ \begin{array}{rcll} f_{mom} & = & \min & \sum_\alpha f_\alpha y_\alpha \\ & & s.t. & M_d(y) \succeq 0, y_0 = 1. \end{array} \right.$$

Theorem 4.4. *The primal (P) and dual (D) satisfy the following properties:*

(1) The dual problem (D) has nonempty interior.

(2) (P) and (D) have the same optimal value, that is, $f_{sos} = f_{mom} \le f_{min}$.

(3) (P) has a minimizer, that is, f_{sos} *is attainable in (P), while* (D) *might not achieve its max value.*

(4) If y^* *is a maximizer of (D) and rank* $M_d(y^*) = 1$*, then* $f_{sos} = f_{mom} = f_{min}$*, and a global minimizer of* $f(x)$ *can be obtained.*

Example 4.5. Consider the problem of minimizing the bivariate polynomial

$$x_1^2 + (x_1 x_2 - 1)^2.$$

The dual version of the SOS relaxation is the SDP:

$$\begin{array}{ll} \min & y_{20} + y_{22} - 2y_{11} + 1 \\ \text{s.t.} & \begin{bmatrix} 1 & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} \\ y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} \\ y_{02} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} \\ y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22} \\ y_{11} & y_{21} & y_{12} & y_{31} & y_{22} & y_{13} \\ y_{02} & y_{12} & y_{03} & y_{22} & y_{13} & y_{04} \end{bmatrix} \succeq 0. \end{array}$$

The optimal $y_{20}^* = 0, y_{11}^* = 1, y_{22}^* = 1$.

4.2. Polynomials nonnegative on a set S

Let $S \subset \mathbb{R}^n$ be a basic closed semialgebraic set, i.e., there exists polynomials $g_1(x), \ldots, g_m(x)$ such that

$$S = \{x \in \mathbb{R}^n : g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}.$$

A polynomial $f(x)$ is called nonnegative on S if

$$f(u) \geq 0 \quad \forall\, u \in S.$$

A polynomial $f(x)$ is called positive on S if

$$f(u) > 0 \quad \forall\, u \in S.$$

We are interested in sufficient and/or necessary conditions for a polynomial $f(x)$ positive or nonnegative on a set S.

Definition 4.6. *The preordering of polynomials $g_1(x), \ldots, g_m(x)$ is the set*

$$P(S) = \left\{ \sum_{\nu \in \{0,1\}^m} \sigma_\nu(x) g_1(x)^{\nu_1} \cdots g_m(x)^{\nu_m} : \quad \text{each } \sigma_\nu \text{ is SOS} \right\}$$

The quadratic module of polynomials $g_1(x), \ldots, g_m(x)$ is the set

$$M(S) = \left\{ \sum_{i=0}^{m} \sigma_i(x) g_i(x) : \quad \text{each } \sigma_i \text{ is SOS} \right\}$$

Here $g_0(x) = 1$.

Theorem 4.7. *The sets $P(g_1, \ldots, g_m)$ and $M(g_1, \ldots, g_m)$ are convex.*

- *If $f(x) \in P(g_1, \ldots, g_m)$, then $f(x)$ is nonnegative on S.*
- *If $f(x) \in M(g_1, \ldots, g_m)$, then $f(x)$ is nonnegative on S.*

Theorem 4.8 (Schmüdgen, 1991)**.** *Let $S = \{x \in \mathbb{R}^n : g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}$ be a compact set. If $f(x)$ is positive on S, then $f(x) \in P(g_1, \ldots, g_m)$.*

Example 4.9. The quadratic $f = x_1 x_2 + 1$ is positive on $x^T x \leq 1$. We have

$$x_1 x_2 + 1 = \frac{1}{2}(x_1 + x_2)^2 + \frac{1}{2} + \frac{1}{2}(1 - x_1^2 - x_2^2).$$

Example 4.10. If $f(x)$ is only nonnegative on S, then Schmüdgen's theorem may not be true. For example, consider

$$f(x) = 1 - x^2, \quad S = \{(1-x^2)^3 \geq 0\}.$$

Suppose there are SOS polynomials σ, ϕ such that

$$1 - x^2 = \sigma(x) + \phi(x)(1-x^2)^3.$$

Then -1 must be a root of $\sigma(x)$ with multiplicity 2, but it has multiplicity 1 on the left.

Theorem 4.11 (Putinar, 1993)**.** *Let $S = \{x \in \mathbb{R}^n : g_1(x) \geq 0, \dots, g_m(x) \geq 0\}$ be a compact set. Suppose there exists $N > 0$ such that*

$$N - \|x\|_2^2 \in M(g_1, \dots, g_m).$$

If $f(x)$ is positive on S, then $f(x) \in M(g_1, \dots, g_m)$.

Remark 4.12. The condition that $N - \|x\|_2^2 \in P(g_1, \dots, g_m)$ is called the archimedean condition (AC). When AC holds, the set S must be compact, but the reverse might not be true. When AC fails, the conclusion of Putinar's theorem might not be true.

Example 4.13. Not every compact set is AC. For instance, consider the set

$$S = \left\{ x \in \mathbb{R}^2 : x_1 - \frac{1}{2} \geq 0, x_2 - \frac{1}{2} \geq 0, 1 - x_1 x_2 \geq 0 \right\}.$$

Clearly, it is compact. There are no SOS polynomials σ_i such that

$$N - (x_1^2 + x_2^2) \ = \ \sigma_0 + (x_1 - \frac{1}{2})\sigma_1 + (x_2 - \frac{1}{2})\sigma_2 + (1 - x_1 x_2)\sigma_3.$$

If they exist, then the maximum degree

$$D = \max(\deg(\sigma_0), \deg(\sigma_3) + 2) \ \geq \ 1 + \max(\deg(\sigma_1), \deg(\sigma_2)).$$

When $D = 2$, it does not work. When $D > 2$, the highest even term of $\sigma_0 + (1 - x_1 x_2)\sigma_3$ must vanish, which is not possible.

- Schmüdgen's Positivstellensatz has a weaker assumption than Putinar's Positivstellensatz, but the conclusion is also weaker.
- When $f(x)$ is only nonnegative but not strictly positive on S, both Schmüdgen's and Putinar's Positivstellensatz may fail.
- When $f(x)$ is only nonnegative on S, then for any $\epsilon > 0$ we have $f(x) + \epsilon \in M(g_1, \dots, g_m)$ or $P(g_1, \dots, g_m)$. As $\epsilon \to$, typically the degree bounds of the representation of $f(x) + \epsilon$ in $M(g_1, \dots, g_m)$ or $P(g_1, \dots, g_m)$ goes to infinity.

Example 4.14. For every $\epsilon > 0$, the polynomial $1 - x^2 + \epsilon$ is positive on

$$S = \{x \in \mathbb{R} : (1 - x^2)^3 \geq 0\}.$$

By Schmüdgen's or Putinar's Positivstellensatz, there exists SOS polynomials P_ϵ, Q_ϵ such that

$$1 - x^2 + \epsilon \;=\; P_\epsilon (1 - x^2)^3 + Q_\epsilon.$$

It was shown by Stengle that

$$\deg(P_\epsilon),\, \deg(Q_\epsilon) \;\geq\; \mathcal{O}(\epsilon^{-1/2}).$$

In the definition of preordering $P(S)$ and quadratic module $M(S)$, there is no restriction on the degrees. In practice, we can only handle polynomials of finite degrees. Thus, we need to define truncated prepordering and quadratic module as

$$P^{(N)}(S) = \left\{ \sum_{\nu \in \{0,1\}^m} \sigma_\nu g_1^{\nu_1} \cdots g_m^{\nu_m} : \quad \sigma_\nu \text{ is SOS and } \deg(\sigma_\nu g_1^{\nu_1} \cdots g_m^{\nu_m}) \leq 2N \right\}$$

$$M^{(N)}(S) = \left\{ \sum_{i=0}^{m} \sigma_i(x) g_i(x) : \quad \sigma_i \text{ is SOS and } \deg(\sigma_i g_i) \leq 2N \right\}.$$

If $f \in P^{(N)}(g_1, \ldots, g_m)$ or $M^{(N)}(g_1, \ldots, g_m)$, then $f(x)$ is nonnegative on S.

If p is a polynomial such that

$$p(x)[x]_{N-k}[x]_{N-k}^T \;=\; \sum_{\alpha \in \mathbb{N}^n : |\alpha| \leq 2N} A_\alpha^{(N)} x^\alpha, \quad k = \lceil \deg(p)/2 \rceil$$

define the N*-th localizing matrix* of p as

$$L_p^{(N)}(y) \;=\; \sum_{\alpha \in \mathbb{N}^n : |\alpha| \leq 2N} A_\alpha^{(N)} y_\alpha.$$

Note that $L_1^{(N)}(y)$ is the standard moment matrix $M_N(y)$.

Theorem 4.15. *The dual cone of $M^{(N)}(S)$ is ($g_0 \equiv 1$) is the set*

$$M^{(N)}(S)^* \;=\; \left\{ y : L_{g_0}^{(N)}(y) \succeq 0, \quad L_{g_1}^{(N)}(y) \succeq 0, \quad \ldots, \quad L_{g_m}^{(N)}(y) \succeq 0 \right\}.$$

4.3. Lasserre's relaxation hierarchy

Let $f(x), g_1(x), \ldots, g_m(x)$ be polynomials in $x \in \mathbb{R}^n$. Consider problem

$$(POP): \quad \begin{cases} \min & f(x) \\ s.t. & g_1(x) \geq 0, \ldots, g_m(x) \geq 0. \end{cases}$$

Let $d_i = \lceil \deg(g_i)/2 \rceil$ and $d = \max_i d_i$. Clearly, for any $N \geq d$, (POP) is equivalent to

$$\begin{array}{rl} \max & \gamma \\ s.t. & f(x) - \gamma \text{ is nonnegative on } S \end{array}$$

Using Putinar's Positivstellensatz, we can get a lower bound for (POP) by using Sum of Squares (SOS) approach. The Lasserre's relaxation is the SOS program

$$(SOS)_N : \quad \begin{cases} f_{sos}^{(N)} := \max & \gamma \\ \qquad\qquad s.t. & f(x) - \gamma \in M^{(N)}(S) \end{cases}$$

Its dual is the SDP

$$(MOM)_N : \quad \begin{cases} f_{mom}^{(N)} := \min & \sum_\alpha f_\alpha y_\alpha \\ \qquad\qquad s.t. & L_{g_0}^{(N)}(y) \succeq 0 \\ & L_{g_1}^{(N)}(y \succeq 0 \\ & \dots \\ & L_{g_m}^{(N)}(y \succeq 0 \\ & y_0 = 1 \end{cases}$$

The above dual can be interpreted as follows. The original (POP) is the same as

$$(POP)_N : \quad \begin{array}{rl} \min & f(x) \\ s.t. & g_0(x)[x]_N[x]_N^T \succeq 0, \\ & g_1(x)[x]_{N-d_1}[x]_{N-d_1}^T \geq 0 \\ & \dots \\ & g_m(x)[x]_{N-d_1}[x]_{N-d_m}^T \geq 0. \end{array}$$

For each $k = 0, 1, \dots, m$, defined symmetric matrices $A_\alpha^{(N,i)}$ as

$$g_k(x)[x]_{N-d_k}[x]_{N-d_k}^T = \sum_{|\alpha| \leq 2N} x^\alpha A_\alpha^{(N,i)}.$$

If we replace every monomial x^α by a single parameter y_α, then $(POP)_N$ is relaxed to

$$(MOM)_N : \quad \begin{cases} f_{mom}^{(N)} := \min & \sum_\alpha f_\alpha y_\alpha \\ \qquad\qquad s.t. & \sum_\alpha y_\alpha A_\alpha^{(N,0)} \succeq 0 \\ & \sum_\alpha y_\alpha A_\alpha^{(N,1)} \succeq 0 \\ & \dots \\ & \sum_\alpha y_\alpha A_\alpha^{(N,m)} \succeq 0 \\ & y_0 = 1 \end{cases}$$

They are precisely the localizing matrices of polynomials $g_i(x)$.

Theorem 4.16. *Let $(MOM)_N, (POP)_N$ be as above. Then we have*

- *The optimal values $f_{sos}^{(N)}, f_{mom}^{(N)}$ are always lower bounds for the global minimum f_{min} of (POP).*
- *The problems $(MOM)_N$ and $(SOS)_N$ are dual to each other.*
- *When the feasible set has nonempty interior, we have $f_{sos}^{(N)} = f_{mom}^{(N)}$.*
- *If a minimizer y^* of $(MOM)_N$ has rank one, then a global minimizer x^* of (POP) can be obtained by setting*

  $$x_1^* = y_1^*, \ldots, x_n^* = y_n^*.$$

  It is feasible for (POP).
- *If the archimedean condition holds, then*

  $$\lim_{N\to\infty} f_{sos}^{(N)} = \lim_{N\to\infty} f_{mom}^{(N)} = f_{min}.$$

Example 4.17. Consider the cubic optimization

$$\begin{array}{rl} \min & x_1 x_2^2 \\ s.t. & 1 - x_1^2 - x_2^2 \geq 0. \end{array}$$

The second order Lasserre's relaxation is

$$\begin{array}{rl} \min & y_{12} \\ s.t. & \begin{bmatrix} 1 - y_{20} - y_{02} & y_{10} - y_{30} - y_{12} & y_{01} - y_{21} - y_{03} \\ y_{10} - y_{30} - y_{12} & y_{20} - y_{40} - y_{22} & y_{11} - y_{31} - y_{13} \\ y_{01} - y_{21} - y_{03} & y_{11} - y_{31} - y_{13} & y_{02} - y_{22} - y_{04} \end{bmatrix} \succeq 0, \\ & \begin{bmatrix} 1 & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} \\ y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} \\ y_{02} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} \\ y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22} \\ y_{11} & y_{21} & y_{12} & y_{31} & y_{22} & y_{13} \\ y_{02} & y_{12} & y_{03} & y_{22} & y_{13} & y_{04} \end{bmatrix} \succeq 0. \end{array}$$

5 Jacobian type SDP relaxation

Consider the optimization problem

$$\begin{array}{rl} \min\limits_{x\in\mathbb{R}^n} & f(x) \\ \text{s.t.} & h_1(x) = \cdots = h_{m_1}(x) = 0 \\ & g_1(x) \geq 0, \ldots, g_{m_2}(x) \geq 0 \end{array} \tag{5.1}$$

where $f(x), g_i(x), h_j(x)$ are polynomial functions in $x \in \mathbb{R}^n$. Let

$$m = \min(m_1 + m_2, n - 1). \tag{5.2}$$

For convenience, we denote $h(x) = (h_1(x), \dots, h_{m_1}(x))$ and $g(x) = (g_1(x), \dots, g_{m_2}(x))$. For a subset $J = \{j_1, \dots, j_k\} \subset [m_2]$, dennote

$$g_J(x) = (g_{j_1}(x), \dots, g_{j_k}(x)).$$

Let x^* be a minimizer of (5.1). If $J = \{j_1, \dots, j_k\}$ is the index set of $g_j(x^*) = 0$ and the KKT conditions hold at x^*, then there exist λ_i and $\mu_j (j \in J)$ such that

$$h(x^*) = 0, \quad g_J(x^*) = 0, \quad \nabla f(x^*) = \sum_{i \in [m_1]} \lambda_i \nabla h_i(x^*) + \sum_{j \in J} \mu_j \nabla g_j(x^*).$$

The above implies the Jacobian matrix of (f, h, g_J) is singular at x^*. For a subset $J \subset [m_2]$, denote the determinantal variety of (f, h, g_J)'s Jacobian being singular by

$$G_J = \left\{ x \in \mathbb{C}^n : \operatorname{rank} B^J(x) \le m_1 + |J| \right\}, \quad B^J(x) = \begin{bmatrix} \nabla f(x) & \nabla h(x) & \nabla g_J(x) \end{bmatrix}. \tag{5.3}$$

Then, $x^* \in V(h, g_J) \cap G_J$ where $V(h, g_J) := \left\{ x \in \mathbb{C}^n : h(x) = 0, g_J(x) = 0 \right\}$.

This motivates us to use $g_J(x) = 0$ and G_J to get tighter SDP relaxations for (5.1). To do so, a practical issue is how to get a "nice" description for G_J? An obvious description for G_J is that all its maximal minors vanish. But there are totally $\binom{n}{m_1+k+1}$ such minors (if $m_1 + k + 1 \le n$), which is huge for big n, m_1, k. Can we define G_J by a set of the smallest number of equations? Furthermore, the active index set J is usually unknown in advance. Can we get an SDP relaxation that is independent of J?

5.1. Minimum defining equations for determinantal varieties

Let $k \le n$ and $X = (X_{ij})$ be a $n \times k$ matrix of indeterminants X_{ij}. Define the determinantal variety

$$D_{t-1}^{n,k} = \left\{ X \in \mathbb{C}^{n \times k} : \operatorname{rank} X < t \right\}.$$

For any index set $I = \{i_1, \dots, i_k\} \subset [n]$, denote by $\det_I(X)$ the $(i_1, \dots, i_k) \times (1, \dots, k)$-minor of matrix X, i.e., the determinant of the submatrix of X whose row indices are $i_1, \dots, i_k$ and column indices are $1, \dots, k$. Clearly, it holds that

$$D_{k-1}^{n,k} = \left\{ X \in \mathbb{C}^{n \times k} : \det_I(X) = 0 \quad \forall\, I \in [n]_k \right\}.$$

The above has $\binom{n}{k}$ defining equations of degree k. An interesting fact is that we do not need $\binom{n}{k}$ equations to define $D_{k-1}^{n,k}$. Actually $nk - k^2 + 1$ are enough. There is very nice work on this issue. Bruns and Vetter [3] showed $nk - t^2 + 1$ equations are enough for defining $D_{t-1}^{n,k}$. Later, Bruns and Schwänzl [2] showed $nk - t^2 + 1$ is the smallest number of equations for defining $D_{t-1}^{n,k}$. Typically, $nk - t^2 + 1 \ll \binom{n}{k}$ for big n and k. A general method for constructing $nk - t^2 + 1$ defining polynomial equations for $D_{t-1}^{n,k}$ was described in Chapt. 5 of [3]. Here we briefly show how it works for $D_{k-1}^{n,k}$.

Let $\Gamma(X)$ denote the set of all k-minors of X (assume their row indices are strictly increasing). For convenience, for any $1 \le i_1 < \cdots < i_k \le n$, we just denote by

$[i_1, \dots, i_k]$ the $(i_1, \dots, i_k) \times (1, \dots, k)$-minor of X. Define a partial ordering on $\Gamma(X)$ as follows:

$$[i_1, \dots, i_k] < [j_1, \dots, j_k] \quad \Longleftrightarrow \quad i_1 \le j_1, \dots, i_k \le j_k,\ \sum_{\ell=1}^{k} i_\ell < \sum_{\ell=1}^{k} j_\ell.$$

If $I = \{i_1, \dots, i_k\}$, we also write $I = [i_1, \dots, i_k]$ as a minor in $\Gamma(X)$ for convenience. For any $I \in \Gamma(X)$, define its rank as

$$rk(I) = \max\left\{\ell : \ I = I^{(\ell)} > \cdots > I^{(1)}, \quad \text{every} \quad I^{(i)} \in \Gamma(X)\right\}.$$

The maximum minor in $\Gamma(X)$ is $[n-k+1, \dots, n]$ and has rank $nk - k^2 + 1$. For every $1 \le \ell \le nk - k^2 + 1$, define

$$\eta_\ell(X) \quad = \quad \sum_{I \in [n]_k, rk(I) = \ell} \det{}_I(X). \tag{5.4}$$

Lemma 5.1 (Lemma (5.9), Bruns and Vetter [3])**.** *It holds that*

$$D_{k-1}^{n,k} = \left\{X \in \mathbb{C}^{n \times k} : \ \eta_\ell(X) = 0,\ \ell = 1, \dots, nk - k^2 + 1\right\}.$$

When $k = 2$, $D_1^{n,2}$ would be defined by $2n - 3$ polynomials. The biggest minor is $[n-1, n]$ and has rank $2n - 3$. For each $\ell = 1, 2, \dots, 2n - 3$, we clearly have

$$\eta_\ell(X) = \sum_{1 \le i_1 < i_2 \le n:\, i_1 + i_2 = \ell + 2} [i_1, i_2].$$

Every 2-minor of X is a summand of some $\eta_\ell(X)$.

When $k = 3$, $D_2^{n,3}$ can be defined by $3n - 8$ polynomials of the form $\eta_\ell(X)$. For instance of $n = 6$, the partial ordering on $\Gamma(X)$ is shown in the following diagram.

$$\begin{tikzcd}
 & & 125 \arrow[r] \arrow[dr] & 126 \arrow[r] \arrow[dr] & 136 \arrow[r] \arrow[dr] & 146 \arrow[r] \arrow[dr] & 156 \arrow[r] & 256 \arrow[dr] & & \\
123 \arrow[r] & 124 \arrow[ur] \arrow[dr] & & 135 \arrow[r] \arrow[ur] \arrow[dr] & 145 \arrow[r] \arrow[ur] \arrow[dr] & 236 \arrow[r] \arrow[ur] \arrow[dr] & 246 \arrow[ur] \arrow[dr] & & 356 \arrow[r] & 456 \\
 & & 134 \arrow[r] \arrow[ur] & 234 \arrow[r] \arrow[ur] & 235 \arrow[r] \arrow[ur] & 245 \arrow[r] \arrow[ur] & 345 \arrow[r] & 346 \arrow[ur] & &
\end{tikzcd}$$

In the above an arrow points to a bigger minor. Clearly, we have the expressions

$$\eta_1(X) = [1, 2, 3], \quad \eta_2(X) = [1, 2, 4], \quad \eta_3(X) = [1, 2, 5] + [1, 3, 4],$$

$$\eta_4(X) = [1, 2, 6] + [1, 3, 5] + [2, 3, 4], \quad \eta_5(X) = [1, 3, 6] + [1, 4, 5] + [2, 3, 5],$$

$$\eta_6(X) = [1, 4, 6] + [2, 3, 6] + [2, 4, 5], \quad \eta_7(X) = [1, 5, 6] + [2, 4, 6] + [3, 4, 5],$$

$$\eta_8(X) = [2, 5, 6] + [3, 4, 6], \quad \eta_9(X) = [3, 5, 6], \quad \eta_{10}(X) = [4, 5, 6].$$

Every above $\eta_i(X)$ has degree 3. Note that the summands $[i_1, i_2, i_3]$ from the same $\eta_i(X)$ have a constant summation $i_1+i_2+i_3$. Thus, for each $\ell = 1, \ldots, 3n-8$, we have

$$\eta_\ell(X) = \sum_{1\le i_1<i_2<i_3\le n:\, i_1+i_2+i_3=\ell+5} [i_1, i_2, i_3].$$

When $k > 3$ is general, $D^{n,k}_{k-1}$ can be defined by $nk - k^2 + 1$ polynomials of the form $\eta_\ell(X)$. For each $\ell = 1, 2, \ldots, nk - k^2 + 1$, we similarly have the expression

$$\eta_\ell(X) = \sum_{1\le i_1<\cdots<i_k\le n:\, i_1+\cdots+i_k=\ell+\binom{k+1}{2}-1} [i_1, \ldots, i_k].$$

5.2. The exact Jacobian SDP relaxation

For every $J = \{j_1, \ldots, j_k\} \subset [m_2]$ with $k \le m - m_1$, by applying formula (5.4), let

$$\eta^J_1, \ldots, \eta^J_{len(J)} \quad \text{where} \quad len(J) = n(m_1+k+1) - (m_1+k+1)^2 + 1$$

be the set of defining polynomials for the determinantal variety G_J defined in (5.3) of the Jacobian of (f, h, g_J) being singular. For each $i = 1, \ldots, len(J)$, define

$$\varphi^J_i(x) = \eta^J_i(x) \cdot \prod_{j\in J^c} g_j(x), \quad \text{where} \quad J^c = [m_2]\backslash J. \tag{5.5}$$

For simplicity, list all possible polynomials $\varphi^J_i(x)$ in (5.5) sequentially as

$$\varphi_1(x),\ \varphi_2(x),\ \ldots,\ \varphi_r(x), \quad \text{where} \quad r = \sum_{J\subset[m_2], |J|\le m-m_1} len(J). \tag{5.6}$$

Now wefine the variety

$$W = \{x \in \mathbb{C}^n : h_1(x) = \cdots = h_{m_1}(x) = \varphi_1(x) = \cdots = \varphi_r(x) = 0\}. \tag{5.7}$$

If the minimum f_{min} of (5.1) is achieved at a KKT point, then (5.1) is equivalent to

$$\begin{array}{rl} \min\limits_{x\in\mathbb{R}^n} & f(x) \\ \text{s.t.} & h_1(x) = \cdots = h_{m_1}(x) = 0, \\ & \varphi_1(x) = \cdots = \varphi_r(x) = 0, \\ & g_\nu(x) \ge 0,\ \forall \nu \in \{0,1\}^{m_2}. \end{array} \tag{5.8}$$

Here, we denote $g_\nu(x) = g_1(x)^{\nu_1} \cdots g_{m_2}(x)^{\nu_{m_2}}$.

To construct an SDP relaxation for (5.8), we need to define localizing moment matrices. Let $q(x)$ be a polynomial with $\deg(q) \le 2N$. Define symmetric matrices $A^{(N)}_\alpha$ such that

$$q(x)[x]_d[x]_d^T \quad = \sum_{\alpha\in\mathbb{N}^n:|\alpha|\le 2N} A^{(N)}_\alpha x^\alpha, \quad \text{where} \quad d = N - \lceil \deg(q)/2 \rceil.$$

Then the N-th order localizing moment matrix of q is defined as

$$L_q^{(N)}(y) = \sum_{\alpha\in\mathbb{N}^n:|\alpha|\le 2N} A_\alpha^{(N)} y_\alpha.$$

Here y is a moment vector indexed by $\alpha \in \mathbb{N}^n$ with $|\alpha| \le 2N$. Moreover, denote

$$L_f(y) = \sum_{\alpha\in\mathbb{N}^n:|\alpha|\le \deg(f)} f_\alpha y_\alpha \quad \text{for} \quad f(x) = \sum_{\alpha\in\mathbb{N}^n:|\alpha|\le \deg(f)} f_\alpha x^\alpha.$$

The N-th order Lasserre's relaxation for (5.8) is the SDP

$$\begin{aligned} f_N^{(1)} := \min \quad & L_f(y) \\ \text{s.t.} \quad & L_{h_i}^{(N)}(y) = 0,\ i = 1, \dots, m_1, \\ & L_{\varphi_j}^{(N)}(y) = 0,\ j = 1, \dots, r, \\ & L_{g_\nu}^{(N)}(y) \succeq 0,\ \forall \nu \in \{0,1\}^{m_2},\ y_0 = 1. \end{aligned} \tag{5.9}$$

Compared to Schmüdgen type Lasserre's relaxation, by (5.6), the number of new equations in (5.9) is $r = O\Big(2^{m_2} \cdot n \cdot (m_1 + m_2)\Big)$. That is, r is of linear order in nm_1 for fixed m_2, but is exponential in m_2. So, when m_2 is small or moderately large, (5.9) is practical; but for big m_2, (5.9) becomes more difficult to solve numerically. Now we present the dual of (5.9). Define the truncated preordering $P^{(N)}$ generated by g_j as

$$P^{(N)} = \left\{ \sum_{\nu\in\{0,1\}^{m_2}} \sigma_\nu(x) g_\nu(x) : \deg(\sigma_\nu g_\nu) \le 2N \right\}, \tag{5.10}$$

and the truncated ideal $I^{(N)}$ generated by h_i and φ_j as

$$I^{(N)} = \left\{ \sum_{i=1}^{m_1} p_i(x) h_i(x) + \sum_{j=1}^{r} q_j(x)\varphi_j(x) : \begin{array}{ll} \deg(p_i h_i) \le 2N & \forall\, i \\ deg(q_j \varphi_j) \le 2N & \forall\, j \end{array} \right\}. \tag{5.11}$$

Then, as shown in Lasserre [8], the dual of (5.9) is the following SOS relaxation for (5.8):

$$\begin{aligned} f_N^{(2)} := \max \quad & \gamma \\ \text{s.t.} \quad & f(x) - \gamma \in I^{(N)} + P^{(N)}. \end{aligned} \tag{5.12}$$

Let f^* be the optimal value of (5.8). Then, by weak duality, we have the relation

$$f_N^{(2)} \le f_N^{(1)} \le f^*. \tag{5.13}$$

We are going to show that when N is big enough, (5.9) is an exact SDP relaxation for (5.8), i.e., $f_N^{(2)} = f_N^{(1)} = f^*$. For this purpose, we need the following assumption.

Assumption 5.2. *(i)* $m_1 \le n$. *(ii) For any* $u \in S$, *at most* $n - m_1$ *of* $g_1(u), \dots, g_{m_2}(u)$ *vanish. (iii) For every* $J = \{j_1, \dots, j_k\} \subset [m_2]$ *with* $k \le n - m_1$, *the variety* $V(h, g_J) = \{x \in \mathbb{C}^n : h(x) = 0,\, g_J(x) = 0\}$ *is nonsingular (its Jacobian has full rank on* $V(h, g_J)$*).*

Theorem 5.3. *Suppose Assumption 5.2 holds. Let* f^* *be the minimum of (5.8). Then there exists an integer* $N^* > 0$ *such that* $f_N^{(1)} = f_N^{(2)} = f^*$ *for all* $N \ge N^*$. *Furthermore, if the minimum* f_{min} *of (5.1) is achievable, then* $f_N^{(1)} = f_N^{(2)} = f_{min}$ *for* $N \ge N^*$.

When the feasible set S of (5.1) is compact, the minimum f_{min} is always achievable. Thus, Theorem 5.3 implies the following.

Corollary 5.4. *Suppose Assumption 5.2 holds. If* S *is compact, then* $f_N^{(1)} = f_N^{(2)} = f_{min}$ *for* N *big enough.*

A practical issue in applications is how to identify whether (5.9) is exact for a given N. This would be possible by applying the flat-extension condition (FEC) [5]. Let y^* be a minimizer of (5.9). We say y^* satisfies FEC if $\operatorname{rank} L_{g_\nu}^{(N)}(y^*) = \operatorname{rank} L_{g_\nu}^{(N-1)}(y^*)$ for every ν. When FEC holds, (5.9) is exact for (5.1), and a finite set of global minimizers would be extracted from y^*. We refer to [6] for a numerical method on how to do this. A very nice software for solving SDP relaxations from polynomial optimization is *GloptiPoly 3* [7] which also provides routines for finding minimizers if FEC holds.

References

[1] D. Bertsekas. *Nonlinear Programming*, second edition. Athena Scientific, 1995.

[2] W. Bruns and R. Schwänzl. The number of equations defining a determinantal variety. *Bull. London Math. Soc.* 22 (1990), no. 5, 439–445.

[3] W. Bruns and U. Vetter. *Determinantal rings.* Lecture Notes in Math. 1327, Springer, Berlin, 1988.

[4] D. Cox, J. Little and D. O'Shea. *Ideals, varieties, and algorithms. An introduction to computational algebraic geometry and commutative algebra.* Third edition. Undergraduate Texts in Mathematics. Springer, New York, 1997.

[5] R. Curto and L. Fialkow. The truncated complex K-moment problem. *Trans. Am. Math. Soc.*, 352, pp. 2825-2855 (2000).

[6] D. Henrion and J. Lasserre. Detecting global optimality and extracting solutions in GloptiPoly. *Positive polynomials in control* (D. Henrion, A. Garulli Eds.), Lecture Notes on Control and Information Sciences, Vol. 312, Springer, Berlin, 2005, pp. 293–310.

[7] D. Henrion, J. Lasserre and J. Loefberg. *GloptiPoly 3: moments, optimization and semidefinite programming.* `http://homepages.laas.fr/henrion/software/gloptipoly3/`

[8] J. B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optim.*, 11(3): 796-817, 2001.

[9] J. Nie, J. Demmel and B. Sturmfels. Minimizing polynomials via sum of squares over the gradient ideal. *Math. Prog.*, Series A, Vol. 106, No. 3, pp. 587–606, 2006.

[10] P. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Math. Prog.*, Ser. B, Vol. 96, No.2, pp. 293-320, 2003.

[11] P. A. Parrilo and B. Sturmfels. Minimizing polynomial functions. In S. Basu and L. Gonzalez-Vega, editors, *Algorithmic and Quantitative Aspects of Real Algebraic Geometry in Mathematics and Computer Science, volume 60 of DIMACS Series in Discrete Mathematics and Computer Science*, pages 83-99. AMS, 2003.

[12] M. Putinar. Positive polynomials on compact semi-algebraic sets, *Ind. Univ. Math. J.* 42 (1993), 203206.

[13] B. Reznick. Some concrete aspects of Hilberts 17th problem. *Contemp. Math.*, Vol. 253, pp. 251-272. American Mathematical Society, 2000.

[14] K. Schmüdgen. The K-moment problem for compact semialgebraic sets. *Math. Ann.* 289 (1991), 203206.

[15] G. Stengle. A Nullstellensatz and Positivstellensatz in semialgebraic geometry. *Mathematische Annalen* 207, 87-97, 1974.

[16] B. Sturmfels. *Solving systems of polynomial equations.* CBMS Regional Conference Series in Mathematics, 97. American Mathematical Society, Providence, RI, 2002.