Chapter IX: The Lasserre Hierarchy for Polynomial and Combinatorial Optimization

The purpose of this chapter is to give an introduction on the topic of polynomial optimization via semidefinite programming and sums of squares relaxations. This material is partly based on lecture notes and review papers from H. Fawzi [1], M. Laurent [2], M. Mevissen [4], T. Rothvoß [6] and Y. de Castro [5], as well as the book of J.-B. Lasserre [3].

In this chapter, we will study a *polynomial optimization problem* of the form

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^n}{\textbf{minimize}} \quad & p(\boldsymbol{x}) \\ s.t. \quad & g_i(\boldsymbol{x}) \geq 0, \quad (\forall i \in [m]), \end{aligned} \tag{P}$$

where $p, g_1, \ldots, g_m \in \mathbb{R}[x_1, \ldots, x_n]$ are polynomials. Clearly, Problem (P) is NP-hard in general, as binary variables $x_i \in \{0, 1\}$ can be encoded by introducting the constraints $x_i^2 = x_i$ (which are equivalent to the pair of polynomial inequalities $x_i^2 - x_i \geq 0$ and $x_i - x_i^2 \geq 0$).

Another reduction shows that it is NP-hard to minimize a quartic polynomial (i.e., of degree 4) over $\mathbb{R}^n$. In the previous lecture, we have seen that it is NP-hard to *separate* the copositive positive cone $\mathcal{C}_n$ (as otherwise we could optimize efficiently over $\mathcal{C}_n$, and solve e.g. the maximum stable set problem in polytime). So, given a symmetric matrix $Q \in \mathbb{S}^n$, it is NP-hard to decide whether $Q \in \mathcal{C}_n$, or to output a separating hyperplane H such that $\langle H, Q\rangle < 0$ and $\langle H, M\rangle \geq 0, \forall M \in \mathcal{C}_n$, i.e., $H \in \mathcal{C}_n^*$ is completely positive. Now, define the quartic polynomial $p(\boldsymbol{x}) = (\boldsymbol{x} \circ \boldsymbol{x})^T Q(\boldsymbol{x} \circ \boldsymbol{x}) = \sum_{i,j} Q_{i,j} x_i^2 x_j^2$. Clearly, $\inf_{\boldsymbol{x}\in\mathbb{R}^n} p(\boldsymbol{x}) = 0$ iff $\boldsymbol{y}^T Q \boldsymbol{y} \geq 0$ for all $\boldsymbol{y} \geq \mathbf{0}$, that is, $Q \succeq_{\mathcal{C}_n} \mathbf{0}$. On the other hand, any $\boldsymbol{x}$ such that $p(\boldsymbol{x}) < 0$ yields a separating hyperplane $H = (\boldsymbol{x} \circ \boldsymbol{x})(\boldsymbol{x} \circ \boldsymbol{x})^T \in \mathcal{C}_n^*$.

1 Nonnegative Polynomials of one variable

Definition 1 (Nonnegative Polynomial)**.** We say that a polynomial $p \in \mathbb{R}[x]$ is nonnegative if

$$p(x) \geq 0, \quad \forall x \in \mathbb{R}.$$

The set of nonnegative polynomials of degree $\leq d$ can be identified with the cone

$$\mathcal{P}_d^+ := \{\boldsymbol{p} \in \mathbb{R}^{d+1} : \sum_{i=0}^{d} p_i\, x^i \geq 0, \quad \forall x \in \mathbb{R}\} \subset \mathbb{R}^{d+1}.$$

Note that the problem of minimizing a polynomial $p \in \mathbb{R}_d[x]$ over $\mathbb{R}$ can be written as a conic program over $\mathcal{P}_d^+$:

$$\inf_{x\in\mathbb{R}} p(x) \;=\; \sup\,\{\gamma \in \mathbb{R} :\ p - \gamma \text{ is nonnegative}\} = \begin{aligned}[t] \sup_{\gamma\in\mathbb{R}} \quad & \gamma \\ & \boldsymbol{p} - \gamma \boldsymbol{e}_0 \succeq_{\mathcal{P}_d^+} \mathbf{0}. \end{aligned}$$

Definition 2 (Sum of squares)**.** We say that a polynomial $p \in \mathbb{R}[x]$ is a sum of squares if there exist polynomials $p_1, \ldots, p_m \in \mathbb{R}[x]$ such that $p = \sum_{i=1}^m p_i^2$. We denote by $\mathcal{P}_d^{\mathrm{SOS}} \subset \mathbb{R}^{d+1}$ the set of (vectors of coefficients of) sum of squares polynomials of degree at most d.

From the definition, it is clear that $\mathcal{P}_d^{\mathrm{SOS}} \subseteq \mathcal{P}_d^+$. The proof of the next proposition is left to the reader.

Proposition 1. *For all $d \in \mathbb{N}$, the cones $\mathcal{P}_{2d}^{SOS}$ and $\mathcal{P}_{2d}^+$ are proper.*

Note that when d is odd, $\mathcal{P}_d^+$ and $\mathcal{P}_d^{\mathrm{SOS}}$ are reduced to the set of nonnegative constant polynomials. So we can restrict our attention on polynomials of even degree. In fact, in the case of univariate polynomials, equality holds between $\mathcal{P}_{2d}^{\mathrm{SOS}}$ and $\mathcal{P}_{2d}^+$:

Theorem 2. *All nonnegative polynomials of one variable can be written as the sum of two squares. Hence, it holds:*

$$\mathcal{P}_{2d}^+ = \mathcal{P}_{2d}^{SOS}.$$

Proof. Let $a_1, \ldots, a_{2d} \in \mathbb{C}$ be the (complex-valued) roots of $p \in \mathbb{R}_{2d}[x]$ (counted with multiplicity). So, we have $p(x) = p_{2d} \prod_{i=1}^{2d}(x - a_i)$. Since p has real-valued coefficients, it holds $p(\bar{z}) = \overline{p(z)}$ for all $z \in \mathbb{C}$, hence z is a root of p iff $\bar{z}$ is a root. Also, if $x \in \mathbb{R}$ is a real root, then it must have even multiplicity because p is nonnegative on the whole real line. Hence, after reindexing the roots, we can write

$$p(x) = p_{2d} \prod_{i=1}^{d}(x - a_i)(x - \overline{a_i}).$$

Now, we recognize that this expression can be written as $p(x) = q(x)\, \overline{q(x)} = |q(x)|^2$, where $q(x) = \sqrt{p_{2d}} \prod_{i=1}^{d}(x - a_i)$. Finally, we have $p(x) = p_1(x)^2 + p_2(x)^2$, where the polynomials p_1 and p_2 correspond to the real and imaginary parts of q, respectively. $\square$

While checking whether a polynomial is nonnegative basically accounts to solving a polynomial optimization problem (or, as in the above proof, compute all its complex roots), we can easily check if a given polynomial is a sum of squares, by solving a linear matrix inequality:

Theorem 3. *The polynomial $p(x) = \sum_{i=0}^{2d} p_k x^k$ is a sum of squares if and only if there exists a matrix $M \succeq 0$ such that the sum of the kth antidiagonal is p_k, for each $k = 0, \ldots, 2d$:*

$$s_k(M) = \sum_{\{0 \le i,j \le d:\ i+j=k\}} M_{ij} = p_k, \quad \forall k \in \{0, \ldots, 2d\}.$$

Proof. Let $x \in \mathbb{R}$ and denote by $\boldsymbol{v}(x) = [1, x, x^2, \ldots, x^d] \in \mathbb{R}^{d+1}$ the vector of the first $(d+1)$ powers of x. Direct calculation shows that

$$\boldsymbol{v}(x)^T M\ \boldsymbol{v}(x) = \sum_{0 \le i,j \le d} x^i M_{ij} x^j = \sum_{k=0}^{2d} \sum_{\{i,j:\ i+j=k\}} M_{ij}\ x^{i+j} = \sum_{k=0}^{2d} s_k(M)\ x^k.$$

Hence, it holds $p(x) = \boldsymbol{v}(x)^T M\, \boldsymbol{v}(x)$ iff the matrix M satisfies $s_k(M) = p_k$, for all $k \in \{0, \ldots, 2d\}$.

Now, assume that $p(x) = \boldsymbol{v}(x)^T M \boldsymbol{v}(x)$ for some positive semidefinite matrix M. This means that $M = P^T P$ for some matrix $P \in \mathbb{R}^{m \times (d+1)}$, and $p(x) = \boldsymbol{v}(x)^T P^T P \boldsymbol{v}(x) = \|P\boldsymbol{v}(x)\|^2 = \sum_{i=1}^m (\boldsymbol{p}_i^T \boldsymbol{v}(x))^2$, where $\boldsymbol{p}_i^T$ is the ith row of P. We have thus shown that p is a sum of squares: $p(x) = \sum_{i=1}^m p_i(x)^2$, where $p_i(x) := \boldsymbol{p}_i^T \boldsymbol{v}(x)$ is a polynomial of degree $\le d$.

Conversely, if p is a sum of squares, then we have $p(x) = \|P\boldsymbol{v}(x)\|^2$ for some matrix P, that is, $p(x) = \boldsymbol{v}(x)^T P^T P \boldsymbol{v}(x)$. So, we obtain the result of the theorem by setting $M = PP^T \succeq 0$. $\square$

By combining the results of Theorems 2 and 3, we see that polynomial minimization problems over $\mathbb{R}$ can be formulated as an SDP.

Example:
We formulate the problem of minimizing the polynomial $p(x) = x^6 - 17x^4 + 2x^3 - 2x + 1$ over $\mathbb{R}$ as an SDP. By Theorem 2, this problem is equivalent to solving $\sup\ \{\gamma \in \mathbb{R} : \ \boldsymbol{p} - \gamma \boldsymbol{e}_0 \in \mathcal{P}_6^{\mathrm{SOS}}\}$, where $\boldsymbol{p} = [1, -2, 0, 2, -17, 0, 1]^T$. Then, we can use Theorem 3 to obtain the following SDP formulation:

$$\begin{aligned}
\underset{M \in \mathbb{S}^4}{\textbf{maximize}} \quad & \gamma \\
s.t. \quad & M_{00} = 1 - \gamma \\
& M_{10} + M_{01} = -2 \\
& M_{20} + M_{11} + M_{02} = 0 \\
& M_{30} + M_{21} + M_{12} + M_{03} = 2 \\
& M_{31} + M_{22} + M_{13} = -17 \\
& M_{32} + M_{23} = 0 \\
& M_{33} = 1 \\
& M \succeq 0.
\end{aligned}$$

#1

2 Multivariate Polynomials

A polynomial $p \in \mathbb{R}_d[x_1, \ldots, x_n]$ can be written compactly as

$$p(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha} \in \Delta(n,d)} p_{\boldsymbol{\alpha}}\ \boldsymbol{x}^{\boldsymbol{\alpha}},$$

where $\Delta(n, d)$ is the set of nonnegative integer vectors with sum $\leq d$: $\Delta(n, d) := \{\boldsymbol{\alpha} \in \mathbb{Z}^n_{\geq 0} : \sum_{i=1}^n \alpha_i \leq d\}$, and $\boldsymbol{x}^{\boldsymbol{\alpha}}$ is a compact notation for

$$\boldsymbol{x}^{\boldsymbol{\alpha}} := x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}.$$

We say that $\boldsymbol{x}^{\boldsymbol{\alpha}}$ is a monomial of total degree $|\boldsymbol{\alpha}| := \sum_{i=1}^n \alpha_i$.

The cone $\mathcal{P}^+_{n,d}$ of $n-$variate nonnegative polynomials of total degree $\leq d$ can be defined as in the univariate case, as well as the cone $\mathcal{P}^{\mathrm{SOS}}_{n,d}$ of $n-$variate sum of squares of degree $\leq d$. Note that the dimension of these cones is $s(n, d) := |\Delta(n, d)| = \binom{n+d}{d}$. As in the univariate case, it is also easy to see that a polynomial can only be nonnegative if its degree is even, so we will often write $\mathcal{P}^+_{n,2d}$.

Unlike the univariate case, $\mathcal{P}^+_{n,d} \neq \mathcal{P}^{\mathrm{SOS}}_{n,d}$ in general. A famous counter-example is the Motzkin polynomial:

Proposition 4. *The polynomial $p(x, y) = x^4y^2 + x^2y^4 + 1 - 3x^2y^2 \in \mathbb{R}_6[x, y]$ is nonnegative, but is not a sum of squares.*

Proof. Nonnegativity follows from the arithmetic-geometric mean applied to x^2y^4, x^4y^2 and 1:

$$\frac{1}{3}(x^4y^2 + x^2y^4 + 1) \geq (x^6y^6)^{1/3}.$$

A certificate for $p \notin \mathcal{P}^{\mathrm{SOS}}_{n,d}$ will be given after Theorem 6. □

In fact, the study of nonnegative polynomials and sum of squares dates back to Hilbert, who characterized the equality cases:

Theorem 5 (Hilbert)**.**

$$(\mathcal{P}_{n,d}^{+} = \mathcal{P}_{n,d}^{SOS}) \iff \Big(\ (n=1) \quad or \quad (d=2) \quad or \quad (n,d)=(2,4) \ \Big).$$

However, it is clear that the inclusion $\mathcal{P}_{n,d}^{\mathrm{SOS}} \subseteq \mathcal{P}_{n,d}^{+}$ still holds, so we obtain relaxations of polynomial optimization problems by optimizing over $\mathcal{P}_{n,d}^{\mathrm{SOS}}$ instead of $\mathcal{P}_{n,d}^{+}$. The next theorem shows that it can be done by semidefinite programming:

Theorem 6. *The polynomial $p \in \mathbb{R}_{2d}[x_1,\ldots,x_n]$, where $p(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha}\in\Delta(n,2d)} p_{\boldsymbol{\alpha}}\, \boldsymbol{x}^{\boldsymbol{\alpha}}$ is a sum of squares if and only if there exists a matrix $M \in \mathbb{S}^{s(n,d)}$ (indexed by multi-indices $\alpha,\beta \in \Delta(n,d)$) such that $M \succeq 0$ and*

$$\sum_{\substack{\boldsymbol{\alpha},\boldsymbol{\beta}\in\Delta(n,d)\\ \boldsymbol{\alpha}+\boldsymbol{\beta}=\boldsymbol{\gamma}}} M_{\boldsymbol{\alpha},\boldsymbol{\beta}} = p_{\boldsymbol{\gamma}}, \qquad \forall \boldsymbol{\gamma} \in \Delta(n,2d). \tag{1}$$

Proof. The proof is completely similar to that of Theorem 3: Take any $\boldsymbol{x} \in \mathbb{R}^n$ and introduce the vector $\boldsymbol{v}(\boldsymbol{x}) = [\boldsymbol{x}^{\boldsymbol{\alpha}}]_{\boldsymbol{\alpha}\in\Delta(n,d)} \in \mathbb{R}^{s(n,d)}$, which contains all monomials of degree $\leq d$. Direct calculation shows that $\boldsymbol{v}(\boldsymbol{x})^T M \boldsymbol{v}(\boldsymbol{x}) = p(\boldsymbol{x})$ if and only if the equality conditions (1) are satisfied.

Then, when (1) holds, we have

$$M \succeq 0 \iff M = P^T P \iff p(\boldsymbol{x}) = \boldsymbol{v}(\boldsymbol{x})^T P^T P \boldsymbol{v}(\boldsymbol{x}) = \|P\boldsymbol{v}(\boldsymbol{x})\|^2 \iff p \text{ is a sum of squares.}$$

□

Example:

We sketch how to use the above theorem to establish a certificate that the Motzkin polynomial

$$p(x_1,x_2) = x_1^4 x_2^2 + x_1^2 x_2^4 + 1 - 3x_1^2 x_2^2$$

of Proposition 4 is not a sum of square. In that case, we have $2d = 6$ and $n = 2$, so the matrix M of Theorem 6 is of size $s(2,3) = 10$; The Motzkin polynomial is a sum of squares iff $\exists M \in \mathbb{S}_+^{10}$ such that

$$\begin{array}{ll}
M_{00,00} = 1 & [p_{00} = 1]\\
2M_{11,00} + 2M_{10,01} = 1 & [p_{11} = 0]\\
2M_{21,01} + 2M_{20,02} + 2M_{12,10} + M_{11,11} = -3 & [p_{22} = -3]\\
2M_{21,03} + M_{12,12} = 1 & [p_{24} = 1]\\
(\ldots) \quad \textit{there are } s(2,6) = 28 \textit{ such constraints, one for each } p_{\boldsymbol{\alpha}} & (\ldots)
\end{array}$$

#2

This is, in fact, a feasibility SDP problem. Let $P_{\boldsymbol{\gamma}}$ denote the matrix such that the above constraints read $\langle M, P_{\boldsymbol{\gamma}}\rangle = p_{\boldsymbol{\gamma}}$, for all $\boldsymbol{\gamma} \in \Delta(2,6)$. To show that this problem is infeasible, we apply the Farkas lemma for cones (cf. Chapter 2):

$$\boldsymbol{p} \notin \mathcal{P}_{2,6}^{\mathrm{SOS}} \iff \exists \boldsymbol{y} : \sum_{\boldsymbol{\gamma}} y_{\boldsymbol{\gamma}} P_{\boldsymbol{\gamma}} \succeq 0 \ \text{ and } \ \langle \boldsymbol{p}, \boldsymbol{y}\rangle < 0.$$

Such a certificate can be obtained by using an SDP solver. For example, one can check numerically that the vector $\boldsymbol{y}$ defined by

$$\begin{array}{llll}
y_{00} = 1, & y_{20} = y_{02} = 1.1660, & y_{40} = y_{04} = 1.8484, & y_{60} = y_{06} = 9.5289,\\
 & y_{24} = y_{42} = 0.8523, & y_{22} = 0.9348, &
\end{array}$$

and $y_{\boldsymbol{\gamma}} = 0$ for all the other multi-indices $\boldsymbol{\gamma} \in \Delta(2,6)$ is a valid certificate of infeasibility. Hence, the Motzkin polynomial is not a sum of squares.

In fact, one can show that, while the Motzkin polynomial $p(x,y)$ is not a sum of squares, multiplying this polynomial by $(1+x^2+y^2)$ yields a sum of squares. Indeed, one can verify that

$$(1+x^2+y^2)\big(x^4y^2+x^2y^4+1-3x^2y^2\big) = y^2(1-x^2)^2+x^2(1-y^2)^2+(x^2y^2-1)^2+\frac{3}{4}x^2y^2(x^2+y^2-2)^2+\frac{1}{4}x^2y^2(x^2-y^2)^2.$$

For the purpose of polynomial optimization over $\mathbb{R}^n$, this motivates the study of the following hierarchy of semidefinite programming problems:

$$v_r^* := \sup\ \{\gamma \in \mathbb{R} :\ (1+x_1^2+\ldots+x_n^2)^r\ (p(\boldsymbol{x})-\gamma) \text{ is a sum of square}\}.$$

Proposition 7. *We have:*

$$v_0^* \leq v_1^* \leq v_2^* \leq \ldots \leq \inf_{\boldsymbol{x}\in\mathbb{R}^n} p(\boldsymbol{x}).$$

Proof. The inequality $v_r^* \leq v_{r+1}^*$ follows from the fact that the product of two sums of squares is a sum of squares. So, $p \in \mathcal{P}_{n,d}^{\mathrm{SOS}} \implies q \in \mathcal{P}_{n,d+2}^{\mathrm{SOS}}$, where $q(\boldsymbol{x}) = (1+x_1^2+\ldots+x_n^2)\cdot p(\boldsymbol{x})$. The inequality $v_r^* \leq \inf_{\boldsymbol{x}} p(\boldsymbol{x})$ follows from the implication

$$(1+x_1^2+\ldots+x_n^2)^r\ (p(\boldsymbol{x})-\gamma)\ \text{ is a sum of squares } \implies p(\boldsymbol{x}) \geq \gamma,\ \forall \boldsymbol{x}\in\mathbb{R}^n.$$

$\square$

Under some mild conditions, it can be shown that the hierarchy converges. In the next section, we are going to study the dual cones of $\mathcal{P}_{n,2d}^+$ and $\mathcal{P}_{n,2d}^{\mathrm{SOS}}$, which will lead to another hierarchy –The Lasserre hierarchy– for the general polynomial problem (P) presented in the introduction.

3 The moment problem

Denote by $\mathcal{M}^+(\mathbb{R}^n)$ the set of all nonnegative measures over $\mathbb{R}^n$. Given a nonnegative measure $\mu \in \mathcal{M}^+(\mathbb{R}^n)$, define its (infinite) sequence of moments $(y_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{Z}_{\geq 0}^n}$ by

$$y_{\boldsymbol{\alpha}} := \int_{\mathbb{R}^n} \boldsymbol{x}^\alpha\, \mu(d\boldsymbol{x}), \quad \forall \boldsymbol{\alpha}\in\mathbb{Z}_{\geq 0}^n.$$

For example, if μ is the probability measure corresponding to a random vector $Y \in \mathbb{R}^n$, then the y_α's correspond to the raw moments of Y: we have $y_{\mathbf{0}} = \int \mu(d\boldsymbol{x}) = 1$, and for all i it holds

$$y_{\boldsymbol{e}_i} = \int x_i\, \mu(d\boldsymbol{x}) = \mathbb{E}[Y_i] \qquad y_{2\boldsymbol{e}_i} = \int x_i^2\, \mu(d\boldsymbol{x}) = \mathbb{E}[Y_i^2] = \mathbb{V}[Y_i] + (\mathbb{E}[Y_i])^2.$$

More generally, $y_{\boldsymbol{\alpha}} = \mathbb{E}[Y_1^{\alpha_1}\cdots Y_n^{\alpha_n}]$.

Conversely, given a vector $\boldsymbol{y}\in\mathbb{R}^{s(n,d)}$, we may ask ourselves whether $\boldsymbol{y}$ has a *representing measure*, i.e., if there exists a measure $\mu\in\mathcal{M}^+(\mathbb{R}^n)$ whose moments of total degree $\leq d$ correspond to the coordinates of $\boldsymbol{y}$. This is the *moment problem.* Those vectors $\boldsymbol{y}$ with a representing measure are in the *moment cone*

$$\mathcal{M}_d^+(\mathbb{R}^n) := \left\{\ \Big(\int_{\mathbb{R}^n} \boldsymbol{x}^\alpha\, \mu(d\boldsymbol{x})\Big)_{\boldsymbol{\alpha}\in\Delta(n,d)} :\quad \mu\in\mathcal{M}^+(\mathbb{R}^n)\right\}.$$

As it turns out, we now show that $\mathcal{P}_{n,2d}^+$ and $\mathcal{M}_{2d}^+(\mathbb{R}^n)$ are dual from each other.

Proposition 8.

$$(\mathcal{P}^+_{n,2d})^* = \mathcal{M}^+_{2d}(\mathbb{R}^n)$$

.

Proof. Let $\boldsymbol{p} \in \mathcal{P}^+_{n,2d}$, and $\boldsymbol{y} \in \mathcal{M}^+_{2d}(\mathbb{R}^n)$. Then, $\boldsymbol{y}$ has a representing measure $\mu \in \mathcal{M}^+(\mathbb{R}^n)$, so

$$\langle \boldsymbol{p}, \boldsymbol{y} \rangle = \sum_{\boldsymbol{\alpha}\in\Delta(n,2d)} p_\alpha y_\alpha = \sum_{\boldsymbol{\alpha}\in\Delta(n,2d)} p_\alpha \int_{\mathbb{R}^n} \boldsymbol{x}^{\boldsymbol{\alpha}} \mu(d\boldsymbol{x}) = \int_{\mathbb{R}^n} \underbrace{\sum_{\boldsymbol{\alpha}\in\Delta(n,2d)} p_\alpha \boldsymbol{x}^{\boldsymbol{\alpha}}}_{=p(\boldsymbol{x})\geq 0} \mu(d\boldsymbol{x}) \geq 0.$$

This already proves $\mathcal{P}^+_{n,2d} \subseteq (\mathcal{M}^+_{2d}(\mathbb{R}^n))^*$. Conversely, assume $\boldsymbol{p} \notin \mathcal{P}^+_{n,2d}$, that is, $\exists \boldsymbol{z} \in \mathbb{R}^n : p(\boldsymbol{z}) < 0$. Let $\boldsymbol{y}$ be the vector of moments of degree $\leq 2d$ corresponding to the dirac measure $\mu = \delta_{\boldsymbol{z}}$ at $\boldsymbol{z}$. Then, $y_{\boldsymbol{\alpha}} = \int_{\mathbb{R}^n} \boldsymbol{x}^{\boldsymbol{\alpha}} \delta_{\boldsymbol{z}}(d\boldsymbol{x}) = \boldsymbol{z}^{\boldsymbol{\alpha}}$, so it holds $\langle \boldsymbol{p}, \boldsymbol{y} \rangle = \sum p_{\boldsymbol{\alpha}} \boldsymbol{z}^{\boldsymbol{\alpha}} = p(\boldsymbol{z}) < 0$, hence $\boldsymbol{p} \notin (\mathcal{M}^+_{2d}(\mathbb{R}^n))^*$. This shows $(\mathcal{M}^+_{2d}(\mathbb{R}^n))^* \subseteq \mathcal{P}^+_{n,2d}$. Finally, the statement of the proposition follows because the considered cones are proper, so they are equal to their bi-dual. □

As for the case of nonnegative polynomials, there is no simple condition which ensures that $\boldsymbol{y} \in \mathcal{M}^+_{2d}(\mathbb{R}^n)$. However, there is a simple linear matrix inequality that is satisfied by all $\boldsymbol{y} \in \mathcal{M}^+_{2d}(\mathbb{R}^n)$. Note that the situation is reversed compared to the case of polynomials. While an LMI allowed to give a *sufficient* condition of positivity for a polynomial (if it is an SOS), for the case of moments we obtain a *necessary* condition relying on an LMI for $\boldsymbol{y}$ to have a representing measure. Given a (possibly infinite) sequence $(\boldsymbol{y}_{\boldsymbol{\alpha}})$ containing all elements indexed by some $\boldsymbol{\alpha}$ of total degree $|\boldsymbol{\alpha}| := \sum_i \alpha_i \leq 2r$, denote by $M_r(\boldsymbol{y}) \in \mathbb{S}^{s(n,r)}$ the matrix with elements

$$\big(M_r(\boldsymbol{y})\big)_{\boldsymbol{\alpha},\boldsymbol{\beta}} = y_{\boldsymbol{\alpha}+\boldsymbol{\beta}}, \qquad \forall \boldsymbol{\alpha}, \boldsymbol{\beta} \in \Delta(n,r).$$

It can be seen that $M_r(\boldsymbol{y}) = \sum_{\boldsymbol{\gamma}\in\Delta(n,2r)} y_{\boldsymbol{\gamma}} P_{\boldsymbol{\gamma}}$, where $P_{\boldsymbol{\gamma}} \in \mathbb{S}^{s(n,r)}$ is the matrix with a 1 at each coordinate $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ such that $\boldsymbol{\alpha} + \boldsymbol{\beta} = \boldsymbol{\gamma}$, and zeros elsewhere.

Proposition 9. *Let $\boldsymbol{y} \in \mathcal{M}^+_{2d}(\mathbb{R}^n)$, and let $r \in \mathbb{N}$ such that $r \leq d$. Then, $M_r(\boldsymbol{y}) \succeq 0$.*

Proof. As $\mathcal{M}^+_{2d}(\mathbb{R}^n)$ and $\{\boldsymbol{y} : M_r(\boldsymbol{y}) \succeq 0\}$ are cones, we can rescale $\boldsymbol{y}$ and assume w.l.o.g. that $y_{\boldsymbol{0}} = 1$. Let μ be a realizing probability measure for $\boldsymbol{y}$, and let X be a random variable corresponding to μ. For an arbitrary sequence $\boldsymbol{c} \in \Delta(n,r)$, define the polynomial $c(\boldsymbol{x}) = \sum_{|\boldsymbol{\alpha}|\leq r} c_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}}$. Then, we have

$$0 \leq \mathbb{E}[c(X)^2] = \mathbb{E}[\sum_{\boldsymbol{\alpha}} c_{\boldsymbol{\alpha}} X^{\boldsymbol{\alpha}} \sum_{\boldsymbol{\beta}} c_{\boldsymbol{\beta}} X^{\boldsymbol{\beta}}] = \sum_{\boldsymbol{\alpha},\boldsymbol{\beta}} c_{\boldsymbol{\alpha}} c_{\boldsymbol{\beta}} \underbrace{\mathbb{E}[X^{\boldsymbol{\alpha}+\boldsymbol{\beta}}]}_{=y_{\boldsymbol{\alpha}+\boldsymbol{\beta}}} = \sum_{\boldsymbol{\alpha},\boldsymbol{\beta}} c_{\boldsymbol{\alpha}} c_{\boldsymbol{\beta}} \big(M_r(\boldsymbol{y})\big)_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \boldsymbol{c}^T M_r(\boldsymbol{y}) \boldsymbol{c}.$$

The above inequality holds for all vectors $\boldsymbol{c}$, which proves $M_r(\boldsymbol{y}) \succeq 0$. □

Now, let us define

$$\mathcal{M}^{\mathrm{SDP}}_{2d}(\mathbb{R}^n) := \{\boldsymbol{y} \in \mathbb{R}^{s(n,2d)} :\ M_r(\boldsymbol{y}) \succeq 0, \forall r \leq d\} = \{\boldsymbol{y} \in \mathbb{R}^{s(n,2d)} :\ M_d(\boldsymbol{y}) \succeq 0\},$$

where the equality follows from the observation that $M_r(\boldsymbol{y})$ is a principal submatrix of $M_{r+1}(\boldsymbol{y})$. Then, the above proposition simply rewrites $\mathcal{M}^+_{2d}(\mathbb{R}^n) \subseteq \mathcal{M}^{\mathrm{SDP}}_{2d}(\mathbb{R}^n)$.

We also leave the following result as an exercise:

Proposition 10. $(\mathcal{P}^{SOS}_{n,2d})^* = \mathcal{M}^{SDP}_{2d}(\mathbb{R}^n)$.

4 Polynomial optimization: The point of view of moments

We return to the polynomial optimization problem given in the introduction of this chapter:

$$\begin{aligned} p^* = \inf_{\boldsymbol{x}\in\mathbb{R}^n} \quad & p(\boldsymbol{x}) \\ s.t. \quad & g_i(\boldsymbol{x}) \geq 0, \quad (\forall i \in [m]), \end{aligned} \tag{P}$$

Problem (P) asks to minimize the polynomial $p(\boldsymbol{x})$ over $K := \{\boldsymbol{x} \in \mathbb{R}^n : g_i(\boldsymbol{x}) \geq 0,\ \forall i \in [m]\}$. We say that K, which is defined by polynomial inequalities, is a *semi-algebraic* set. It is possible to reformulate (P) as a moment problem over K:

$$\begin{aligned} p^* = \inf_{\boldsymbol{\mu}\in\mathcal{M}^+(K)} \quad & \int_K p(\boldsymbol{x})\, \mu(d\boldsymbol{x}) \\ s.t. \quad & \mu(K) = 1. \end{aligned} \tag{2}$$

The decision variable is a probabililty measure μ supported by K: we have $\mu \in \mathcal{M}^+(K)$, where

$$\mathcal{M}^+(K) := \{\mu \in \mathcal{M}^+(\mathbb{R}^n) : \mu(\mathbb{R}^n \setminus K) = 0\}.$$

We also define $\mathcal{M}_d^+(K) := \left\{ \ (\int_{\mathbb{R}^n} \boldsymbol{x}^\alpha\, \mu(d\boldsymbol{x}))_{\boldsymbol{\alpha}\in\Delta(n,d)} \ : \ \mu \in \mathcal{M}^+(K)\right\}$, i.e., the set of truncated moment sequences (up to degree d) for all nonnegative measures over K. Then, denote by $\boldsymbol{y}$ the truncated moment sequence of $\mu \in \mathcal{M}^+(K)$, and observe that $\mu(K) = \mu(\mathbb{R}^n) = y_0$ and $\int_K p(\boldsymbol{x})\, \mu(d\boldsymbol{x}) = \int_K \sum_{\boldsymbol{\alpha}\in\Delta(n,d)} p_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}}\, \mu(d\boldsymbol{x}) = \langle \boldsymbol{p}, \boldsymbol{y}\rangle$, so we obtain:

$$\begin{aligned} p^* = \inf_{\boldsymbol{y}\in\mathcal{M}_d^+(K)} \quad & \langle \boldsymbol{p}, \boldsymbol{y}\rangle \\ s.t. \quad & y_0 = 1. \end{aligned} \tag{3}$$

Thus, the problem is now to derive a tractable approximation of $\mathcal{M}_d^+(K)$. As in the previous section, we first obtain necessary conditions. Consider a polynomial $g = \sum_{|\boldsymbol{\alpha}|\leq q} g_{\boldsymbol{\alpha}} \boldsymbol{x}^\alpha$ of degree $\leq 2u$ and an integer $r \in \mathbb{N}$. Then, for a (possibly infinite) sequence $(\boldsymbol{y}_{\boldsymbol{\alpha}})$ containing all elements indexed by some $\boldsymbol{\alpha}$ of total degree $|\boldsymbol{\alpha}| \leq 2(u + r)$, define the *localizing matrix* $M_r(g\boldsymbol{y}) \in \mathbb{S}^{s(n,r)}$ with elements

$$\big(M_r(g\boldsymbol{y})\big)_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \sum_{|\boldsymbol{\gamma}|\leq 2u} g_{\boldsymbol{\gamma}}\, y_{\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\gamma}}, \qquad \forall \boldsymbol{\alpha}, \boldsymbol{\beta} \in \Delta(n, r).$$

Proposition 11. *Let $\boldsymbol{y}$ be the moment sequence of a measure $\mu \in \mathcal{M}^+(K)$. Then, for all $r \in \mathbb{N}$ we have*

$$M_r(\boldsymbol{y}) \succeq 0 \quad \text{and} \quad M_r(g_i\boldsymbol{y}) \succeq 0, \quad \forall i \in [m].$$

Proof. The fact that $M_r(\boldsymbol{y}) \succeq 0$ follows from Proposition 9, because the truncated vector $(y_{\boldsymbol{\alpha}})_{|\alpha|\leq 2r}$ is a member of $\mathcal{M}_{2r}^+(K) \subseteq \mathcal{M}_{2r}^+(\mathbb{R}^n)$. Then, we proceed similarly as in the proof of Proposition 9 to show $M_r(g_i\boldsymbol{y}) \succeq 0$: For an arbitrary sequence $\boldsymbol{c} \in \Delta(n, r)$, define the polynomial $c(\boldsymbol{x}) = \sum_{|\boldsymbol{\alpha}|\leq r} c_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}}$. The polynomial $\boldsymbol{x} \mapsto g_i(\boldsymbol{x})c(\boldsymbol{x})^2$ is nonnegative over K, so it holds $\int_K g_i(\boldsymbol{x})c(\boldsymbol{x})^2\mu(d\boldsymbol{x}) \geq 0$, where $\mu \in \mathcal{M}^+(K)$ is a representing measure for $\boldsymbol{y}$. Finally, it can be seen that $\int_K g_i(\boldsymbol{x})c(\boldsymbol{x})^2\mu(d\boldsymbol{x}) = \boldsymbol{c}^T M_r(g_i\boldsymbol{y})\boldsymbol{c} \geq 0$, and since $\boldsymbol{c}$ is arbitrary we obtain the desired result. ☐

This proposition can also be rewritten as follows: Define

$$\mathcal{M}_{2r}^{\text{SDP}}(K) := \{\boldsymbol{y} \in \mathbb{R}^{s(n,2r)} : \ M_r(\boldsymbol{y}) \succeq 0 \quad \text{and} \quad M_{r-u_i}(g_i\boldsymbol{y}) \succeq 0, \forall i \in [m]\},$$

where $u_i = \lceil \deg(g_i)/2 \rceil$. Then, $\mathcal{M}_{2r}^+(K) \subseteq \mathcal{M}_{2r}^{\text{SDP}}(K)$.

In fact, we can also construct more precise outer approximations of $\mathcal{M}_{2r}^{+}(K)$ by considering the necessary condition from Proposition 11 for moment sequences truncated at a higher degree, say $2(r+\delta)$:

$$\mathcal{M}_{2r,\delta}^{\mathrm{SDP}}(K) := \Big\{ \boldsymbol{y} \in \mathbb{R}^{s(n,2r)} : \quad \exists \tilde{\boldsymbol{y}} \in \mathbb{R}^{s(n,2(r+\delta))} \text{ such that } \begin{array}{l} \tilde{y}_{\boldsymbol{\alpha}} = y_{\boldsymbol{\alpha}}, \forall |\boldsymbol{\alpha}| \leq 2r; \\ M_{r+\delta}(\tilde{\boldsymbol{y}}) \succeq 0; \\ M_{r+\delta-u_i}(g_i \tilde{\boldsymbol{y}}) \succeq 0, \forall i \in [m] \end{array} \Big\}.$$

This definition basically says that $\boldsymbol{y} \in \mathcal{M}_{2r,\delta}^{\mathrm{SDP}}(K)$ if we can *extend* the sequence $(\boldsymbol{y}_{\boldsymbol{\alpha}})_{|\boldsymbol{\alpha}|\leq 2r}$ to obtain a sequence $(\tilde{\boldsymbol{y}}_{\boldsymbol{\alpha}})_{|\boldsymbol{\alpha}|\leq 2(r+\delta)}$ for which the moment matrix and the localizing matrices are positive semidefinite. Proposition 11 implies that $\mathcal{M}_{2r}^{+}(K) \subseteq \mathcal{M}_{2r,\delta}^{\mathrm{SDP}}(K)$ holds for all $\delta \in \mathbb{N}$. Moreover, these outer approximations are nested by construction, hence

Corollary 12.

$$\mathcal{M}_{2r}^{+}(K) \subseteq \cdots \subseteq \mathcal{M}_{2r,2}^{SDP}(K) \subseteq \mathcal{M}_{2r,1}^{SDP}(K) \subseteq \mathcal{M}_{2r,0}^{SDP}(K) = \mathcal{M}_{2r}^{SDP}(K).$$

Under some technical conditions, it can be shown that the converse of Proposition 11 is valid: This result is a consequence of Putinar's *Positivstellensatz*, which we will mention at the end of Section 5. For the sake of this lecture, we will simply say that K *satisfies the Archimedean condition* if

$$R^2 - \|\boldsymbol{x}\|^2 = \sum_{i=1}^{m} \sigma_i(\boldsymbol{x}) g_i(\boldsymbol{x})$$

for some R and SOS polynomials $\sigma_1, \ldots, \sigma_m$. This condition can be interpreted as an algebraic certificate of compactness for K. Indeed, for $\boldsymbol{x} \in K$ we have $g_i(\boldsymbol{x}) \geq 0, \forall i$, so $\sum_i \sigma_i(\boldsymbol{x}) g_i(\boldsymbol{x}) \geq 0$ for all SOS polynomials $\sigma_1, \ldots, \sigma_m$, and the Archimedean condition implies $\|\boldsymbol{x}\|^2 \leq R^2$. Note that if we know that K is included in the ball of center $\boldsymbol{0}$ and radius R, we can simply add the constraint $g_{m+1}(\boldsymbol{x}) = R^2 - \|\boldsymbol{x}\|^2 \geq 0$ into the problem, so that the Archimedean condition is automatically satisfied (consider the SOS polynomials $\sigma_i(\boldsymbol{x}) = 0$, $\forall i \in [m]$ and $\sigma_{m+1}(\boldsymbol{x}) = 1$).

Theorem 13. *Assume that K satisfies the Archimedean condition. Then, an infinite moment sequence $\boldsymbol{y}$ has a representing measure $\mu \in \mathcal{M}^{+}(K)$ if and only if for all $r \in \mathbb{N}$, $M_r(\boldsymbol{y}) \succeq 0$ and $M_r(g_i \boldsymbol{y}) \succeq 0, \forall i \in [m]$. In this case, the hierarchy of Corollary 12 is convergent:*

$$\mathcal{M}_{2r}^{+}(K) = \bigcap_{\delta \in \mathbb{N}} \mathcal{M}_{2r,\delta}^{SDP}(K).$$

As a consequence, we obtain a hierarchy of SDPs which converges to the polynomial optimization problem (P), Let $u = \max_i \ u_i$ and $r = \max\big(\lceil \deg(p)/2 \rceil, u\big)$, and replace the constraint $\boldsymbol{y} \in \mathcal{M}_{d}^{+}(K)$ in (3) by $\boldsymbol{y} \in \mathcal{M}_{2r,\delta}^{\mathrm{SDP}}(K)$ for some $\delta \in \mathbb{N}$:

$$\begin{aligned} \underset{\boldsymbol{y} \in \mathbb{R}^{s(n,2(r+\delta))}}{\textbf{minimize}} \quad & \langle \boldsymbol{p}, \boldsymbol{y} \rangle, \\ s.t. \quad & y_{\boldsymbol{0}} = 1. \\ & M_{r+\delta}(\boldsymbol{y}) \succeq 0; \\ & M_{r+\delta-u_i}(g_i \boldsymbol{y}) \succeq 0, \forall i \in [m]. \end{aligned} \tag{Las$_\delta$}$$

Note that in the above problem, $\boldsymbol{p}$ was implicitely extended to a vector in $\mathbb{R}^{s(n,2(r+\delta))}$ (so its scalar product with $\boldsymbol{y}$ is well defined), i.e., we set $p_{\boldsymbol{\alpha}} = 0$ for all $|\boldsymbol{\alpha}| > \deg(p)$. By construction, the sequence of optimal values p^*_δ of (Las$_\delta$) is nondecreasing, and Theorem (13) guarantees that p^*_δ converges to the optimal value p^* of (P):

$$p^*_0 \le p^*_1 \le p^*_2 \le \ldots \le p^* \quad \text{and} \quad \lim_{\delta\to\infty} p^*_\delta = p^*.$$

Moreover, under certain conditions (one of them is that the problem contains only *equality constraints*, we will prove a special case of this result in Section 6), it can be shown that the convergence occurs after a finite number of steps, i.e., $\exists \delta \in \mathbb{N} : p^*_\delta = p^*$. The following condition can be used to check whether convergence took place after δ rounds of the hierarchy:

Theorem 14. *Let $\boldsymbol{y}^*$ be an optimal solution of* (Las$_\delta$)*, so we have $p^*_\delta = \langle \boldsymbol{p}, \boldsymbol{y}^*\rangle$. If $\boldsymbol{y}^*$ satisfies*

$$\textbf{rank}\, M_{r+\delta}(\boldsymbol{y}^*) = \textbf{rank}\, M_{r+\delta-u}(\boldsymbol{y}^*) \quad (=: r_\delta)$$

where $u = \max_i u_i$, then $p^*_\delta = p^*$, and $\boldsymbol{y}^*$ has a representing measure $\mu^* \in \mathcal{M}^+(K)$ that solves Problem (2)*. Moreover, μ^* is r_δ-atomic, which means that μ^* can be decomposed as a convex combination of r_δ dirac measures: $\mu^* = \sum_{i=1}^{r_\delta} w_i \delta_{\boldsymbol{x}^*_i}$, with $\boldsymbol{w} \ge \boldsymbol{0}$, $\sum_i w_i = 1$, and the support points $\boldsymbol{x}^*_i$ of μ^* are global minimizers of* (P)*.*

We also point out that there is an algorithm that can be used to *extract* the support points $\boldsymbol{x}^*_i$ of the optimal measure μ^*, cf. [3]. This is particularly easy when $r_\delta = \textbf{rank}\, M_{r+\delta}(\boldsymbol{y}^*) = 1$, because in this case we have $M_{r+\delta}(\boldsymbol{y}^*) = v(\boldsymbol{x}^*)v(\boldsymbol{x}^*)^T$, where $v(\boldsymbol{x}^*) = (\boldsymbol{x}^{*\boldsymbol{\alpha}})_{|\boldsymbol{\alpha}|\le r+\delta}$ is the vector of monomials of $\boldsymbol{x}^*$. So an optimal solution $\boldsymbol{x}^*$ to Problem (P) is simply recovered by reading the coordinates of $\boldsymbol{y}^*$ corresponding to monomials of degree 1: $x^*_i = y^*_{\boldsymbol{e}_i}, \forall i \in [n]$.

Example:
Consider the polynomial optimization problem

$$\min_{x,y} \qquad p(x,y) = 11x^2 + 16xy + x - y^2 + 2y + 1 \quad \text{s.t.} \quad x^2 + y^2 \le 1.$$

If there is a solution to this problem in the interior of the unit ball ($x^2 + y^2 < 1$), then it must be a stationary point, i.e., the gradient $\nabla p = [22x + 16y + 1, 16x - 2y + 2]^T$ vanishes. This yields the point $[x_0, y_0] \simeq [-0.1133, 0.0933]$ as a good candidate, but the analysis of the eigenvalues of the hessian matrix shows that this point is in fact a saddle point of p. So the optimum must be attained on the boundary of the unit ball. Plotting $p(\cos\theta, \sin\theta)$ over $\theta \in (-\pi, \pi]$ reveals that the minimum is attained at $\theta^* \simeq -1.15043$, hence $[x^*, y^*] = [0.4080, -0.9129]$ is the solution to the above polynomial optimization problem.
The polynomial p has $n = 2$ variables and is of degree $2r = 2$, and the constraint is of degree $2u = 2$, so $r = u = 1$. For a vector $(z_\alpha)_{|\alpha|\le 2\rho}$, recall that the moment matrix is $M_\rho(\boldsymbol{z})_{\alpha,\beta} = z_{\alpha+\beta}$, $(|\alpha| \le \rho, |\beta| \le \rho)$ #3
so if we order the monomials of degree ≤ 1 as $(1, x, y)$, corresponding to the multi-indices $(00, 10, 01)$, at the level $\delta = 0$ of the hierarchy we have

$$M_{r+\delta}(\boldsymbol{z}) = M_1(\boldsymbol{z}) = \begin{bmatrix} z_{00} & z_{10} & z_{01} \\ z_{10} & z_{20} & z_{11} \\ z_{01} & z_{11} & z_{02} \end{bmatrix}.$$

And the localizing matrix $M_\rho(g\boldsymbol{z})$ has coordinates $M_\rho(g\boldsymbol{z})_{\alpha,\beta} = \sum_\gamma g_\gamma z_{\alpha+\beta+\gamma}$ for $|\alpha| \le \rho, |\beta| \le \rho$. Hence, for $g(x,y) = 1 - x^2 - y^2$,

$$M_{r+\delta-u}(g\boldsymbol{z}) = M_0(g_1\boldsymbol{z}) = (1 - z_{20} - z_{02}).$$

In this example, the localizing matrix is scalar. But if we need to go to the next level ($\delta = 1$) of the hierarchy, we will need to consider the moment matrix $M_2(\boldsymbol{z})$ indexed over $|\alpha| \le 2, |\beta| \le 2$ of dimension $s(n,2) = \binom{n+2}{2} = \binom{4}{2} = 6$, and the localizing matrix $M_1(g\boldsymbol{z})$ of dimension $s(n,1) = \binom{n+1}{1} = \binom{3}{1} = 3$.

Example (continued):
The SDP for the hierarchy at level $\delta = 0$ is therefore:

$$\begin{aligned} \underset{\boldsymbol{z}}{\textbf{minimize}} \quad & 11\, z_{20} + 16\, z_{11} + z_{10} - z_{02} + 2\, z_{01} + 1 \\ s.t. \quad & z_{00} = 1 \\ & \begin{bmatrix} z_{00} & z_{10} & z_{01} \\ z_{10} & z_{20} & z_{11} \\ z_{01} & z_{11} & z_{02} \end{bmatrix} \succeq 0 \\ & 1 - z_{20} - z_{02} \geq 0. \end{aligned}$$

#3

If we solve this SDP, we get the optimal moment matrix

$$M_1(\boldsymbol{z}^*) = \begin{pmatrix} 1. & 0.40809016 & -0.91294164 \\ 0.40809016 & 0.16653757 & -0.37256249 \\ -0.91294164 & -0.37256249 & 0.83346242 \end{pmatrix}.$$

This matrix has rank 1, so the certificate of global optimality of Theorem 14 is satisfied, as obviously, $M_{r+\delta-u}(\boldsymbol{z}^*) = M_0(\boldsymbol{z}^*) = 1$ has rank one, too. This is the easy case ($r_\delta = 1$) where the optimal solution of the polynomial optimization problem can be read directly from the vector of optimal moments $\boldsymbol{z}^*$:

$$x^* = z^*_{10} = 0.40809016 \quad \text{and} \quad y^* = z^*_{01} = -0.91294164.$$

5 Polynomial Optimization and Sum of Squares

Now, we derive the dual optimization problem of (Las$_\delta$), which will give an alternative interpretation of the hierarchy. For $\rho \geq 0$, recall that the moment matrix $M_\rho(\boldsymbol{y})$ can be written as $M_\rho(\boldsymbol{y}) = \sum_{|\boldsymbol{\gamma}|\leq 2\rho} y_{\boldsymbol{\gamma}} P_\gamma$, where P_γ is a $\{0,1\}$-symmetric matrix of size $s(n,\rho) \times s(n,\rho)$ with a 1 at all coordinates $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ such that $\boldsymbol{\alpha} + \boldsymbol{\beta} = \boldsymbol{\gamma}$. Similarly, in can be seen that the localizing matrix of a polynomial g of degree $\leq 2u$, which has elements $M_\rho(g\boldsymbol{y})_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \sum_{|\boldsymbol{\tau}|\leq 2u} g_{\boldsymbol{\tau}} y_{\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\tau}}$, can be written as $M_\rho(g\boldsymbol{y}) = \sum y_\gamma Q^g_\gamma$, where Q^g_γ is the matrix in $\mathbb{S}^{s(n,\rho)}$ such that $(Q^g_\gamma)_{\boldsymbol{\alpha},\boldsymbol{\beta}} = g_{\boldsymbol{\tau}}$ whenever $\boldsymbol{\alpha} + \boldsymbol{\beta} + \boldsymbol{\tau} = \boldsymbol{\gamma}$.

Let us now derive the dual of (Las$_\delta$):

$$\begin{aligned} p^*_\delta &= \inf_{\boldsymbol{y}} \langle \boldsymbol{p}, \boldsymbol{y}\rangle + \sup_{\substack{\lambda\in\mathbb{R} \\ \Lambda\succeq 0 \\ \Omega_i\succeq 0, \forall i\in[m]}} \lambda\cdot(1-y_{\boldsymbol{0}}) - \langle M_{r+\delta}(\boldsymbol{y}), \Lambda\rangle - \sum_{i\in[m]} \langle M_{r+\delta-u_i}(g_i\boldsymbol{y}), \Omega_i\rangle \\ d^*_\delta &= \sup_{\substack{\lambda\in\mathbb{R} \\ \Lambda\succeq 0 \\ \Omega_i\succeq 0, \forall i\in[m]}} \inf_{\boldsymbol{y}} \langle \boldsymbol{p}, \boldsymbol{y}\rangle + \lambda\cdot(1-y_{\boldsymbol{0}}) - \left\langle \sum_{|\boldsymbol{\gamma}|\leq 2(r+\delta)} y_{\boldsymbol{\gamma}} P_\gamma,\, \Lambda \right\rangle - \sum_i \left\langle \sum_{|\boldsymbol{\gamma}|\leq 2(r+\delta-u_i)} y_{\boldsymbol{\gamma}} Q^{g_i}_\gamma,\, \Omega_i \right\rangle \\ &= \sup_{\substack{\lambda\in\mathbb{R} \\ \Lambda\succeq 0 \\ \Omega_i\succeq 0, \forall i\in[m]}} \lambda + \inf_{\boldsymbol{y}} \sum_{\boldsymbol{\gamma}} y_{\boldsymbol{\gamma}} \left(p_{\boldsymbol{\gamma}} - \lambda(\boldsymbol{e_0})_{\boldsymbol{\gamma}} - \langle P_\gamma, \Lambda\rangle - \sum_i \langle Q^{g_i}_\gamma, \Omega_i\rangle \right) \\ &= \sup_{\lambda,\Lambda,(\Omega_i)_{i\in[m]}} \quad \lambda \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (\mathrm{Sos}_\delta) \\ &\qquad\quad s.t. \quad (\boldsymbol{p} - \lambda\boldsymbol{e}_0)_\gamma = \langle P_\gamma, \Lambda\rangle + \sum_{i\in[m]} \langle Q^{g_i}_\gamma, \Omega_i\rangle, \quad \forall |\boldsymbol{\gamma}| \leq 2(r+\delta) \\ &\qquad\qquad\quad\;\; \Lambda \succeq 0 \\ &\qquad\qquad\quad\;\; \Omega_i \succeq 0,\ \forall i \in [m] \end{aligned}$$

Now, let us try to understand the meaning of this dual formulation. The equality constraint tells us that the polynomial $\boldsymbol{x} \mapsto p(\boldsymbol{x}) - \lambda$ is the sum of a polynomial σ_0 with coefficients $\langle P_\gamma, \Lambda \rangle$ and some polynomials $q_1, \ldots, q_m$ with coefficients $\langle Q_\gamma^{g_i}, \Omega_i \rangle$, $\forall i \in [m]$. We know from Theorem 6 that $\Lambda \succeq 0$ is equivalent to σ_0 being a sum of squares. Similarly, we can observe that

$$q_i(\boldsymbol{x}) = \sum_{\boldsymbol{\gamma}} \boldsymbol{x}^{\boldsymbol{\gamma}} \langle Q_\gamma^{g_i}, \Omega_i \rangle = \sum_{\boldsymbol{\gamma}} \boldsymbol{x}^{\boldsymbol{\gamma}} \sum_{\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\tau}=\boldsymbol{\gamma}} (g_i)_{\boldsymbol{\gamma}} \, (\Omega_i)_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \underbrace{\sum_{\boldsymbol{\tau}} (g_i)_{\boldsymbol{\gamma}} \, \boldsymbol{x}^{\boldsymbol{\tau}}}_{=g_i(\boldsymbol{x})} \cdot \underbrace{\sum_{\boldsymbol{\alpha},\boldsymbol{\beta}} \boldsymbol{x}^{\boldsymbol{\alpha}} (\Omega_i)_{\boldsymbol{\alpha},\boldsymbol{\beta}} \; \boldsymbol{x}^{\boldsymbol{\beta}}}_{=\sigma_i(\boldsymbol{x})},$$

and $\Omega_i \succeq 0 \iff \sigma_i$ is a sum of squares, cf. proof of Theorem 6. The dual problem of the hierarchy can also be interpreted as follows:

$$\begin{aligned} d_\delta^* = \sup_{\lambda, \sigma_0, (\sigma_i)_{i \in [m]}} \quad & \lambda \\ s.t. \quad & p(\boldsymbol{x}) - \lambda = \sigma_0(\boldsymbol{x}) + \sum_{i \in [m]} g_i(\boldsymbol{x}) \cdot \sigma_i(\boldsymbol{x}) \\ & \sigma_0 \text{ is a SOS polynomial of degree } \leq 2(r+\delta) \\ & \sigma_i \text{ is a SOS polynomial of degree } \leq 2(r+\delta-u_i), \ \forall i \in [m] \end{aligned} \tag{Sos_δ}$$

It is easy to see that $p(\boldsymbol{x}) - \lambda = \sigma_0(\boldsymbol{x}) + \sum_{i \in [m]} g_i(\boldsymbol{x}) \cdot \sigma_i(\boldsymbol{x})$ for some SOS polynomials is an algebraic certificate for $p - \lambda$ to be nonnegative over K, i.e., $p(\boldsymbol{x}) \geq \lambda, \forall \boldsymbol{x} \in K$. Therefore, the optimal value $\lambda^* = d_\delta^*$ is an underestimator for the optimal value of the polynomial optimization problem p^*, which we already knew from weak duality: $d_\delta^* \leq p_\delta^* \leq p^*$.

Putinar's Positivstellensatz, the result which can be used to prove convergence of the Lasserre / sum-of-squares hierarchy, has a very similar flavour indeed:

Theorem 15. *(Putinar's Positivstellensatz) Let p be a* positive *polynomial over a set $K = \{\boldsymbol{x} \in \mathbb{R}^n : g_i(\boldsymbol{x}) \geq 0, \ \forall i \in [m]\}$ that has the Archimedean property. Then, p can be written as $p = \sigma_0 + \sum_{i \in [m]} g_i \cdot \sigma_i$ for some SOS polynomials $\sigma_0, \ldots, \sigma_m$.*

Note that we require p to be *positive* on K (nonnegative is not enough), and the result does not tell anything on the degrees of the σ_i's. The convergence of the sum-of-squares hierarchy (Sos_δ) –and hence of the Lasserre hierarchy (Las_δ)– simply follows from this theorem: $p - p^* + \epsilon$ is positive over K for all $\epsilon > 0$, so $p - p^* + \epsilon = \sigma_0 + \sum_{i \in [m]} g_i \cdot \sigma_i$ for some SOS polynomials σ_i's of degree high enough. This means that there exists $\delta \in \mathbb{N}$ such that $p^* - \epsilon$ is feasible for (Sos_δ).

6 The Lasserre Hierarchy in Combinatorial Optimization

In this section we will review a few properties of the Lasserre hierarchy applied to the integer program

$$\textbf{minimize} \quad \boldsymbol{c}^T \boldsymbol{x} \quad \text{s.t.} \quad A\boldsymbol{x} \geq \boldsymbol{b}, \quad \boldsymbol{x} \in \{0,1\}^n. \tag{IP}$$

The integer constraints can be handled by the equalities $x_i^2 = x_i$, forall $i \in [m]$. Instead of considering the pair of polynomial inequalities $x_i^2 - x_i \geq 0$ and $x_i^2 \leq x_i$, we observe that we can simplify the Lasserre hierarchy by carrying out some moment substitutions: for all nonnegative measure μ supported on the feasible set $K = \{\boldsymbol{x} \in \{0,1\}^n : A\boldsymbol{x} \geq \boldsymbol{b}\}$, it holds $\int x_i^2 \, \mu(d\boldsymbol{x}) = \int x_i \, \mu(d\boldsymbol{x})$. More generally, for all $\boldsymbol{\alpha} \in \mathbb{Z}_{\geq 0}^n$ we have

$$\int_K \boldsymbol{x}^{\boldsymbol{\alpha}} \, \mu(d\boldsymbol{x}) = \int_K \left(\prod_{i \in I} x_i \right) \mu(d\boldsymbol{x}), \qquad \text{where } I = \{i \in [n] : \alpha_i \geq 1\}.$$

This indicates that the moments $y_{\boldsymbol{\alpha}}$ do not depend on the actual values of the α_i, but only on the sparsity pattern I of $\boldsymbol{\alpha}$. Therefore, we can simplify the hierarchy and consider a vector of moments $\boldsymbol{y}$ indexed by some subsets $I \subseteq [n]$: we identify y_I with $y_{\boldsymbol{e}_I}$, where $\boldsymbol{e}_I = \sum_{k\in I} \boldsymbol{e}_k$ is the incidence vector of I. For example, the moment matrix has coordinates

$$\big(M_\rho(\boldsymbol{y})\big)_{I,J} = \int_K \left(\prod_{i\in I} x_i\right) \left(\prod_{j\in J} x_j\right) \mu(d\boldsymbol{x}) = \int_K \left(\prod_{i\in I\cup J} x_i\right) \mu(d\boldsymbol{x}) = y_{I\cup J}.$$

In the moment formulation of the problem, we are searching a probability measure μ over K, corresponding to a random variable X. Since K is finite, the distribution of X is discrete, and it can be interpreted as a *randomized algorithm* that outputs $\boldsymbol{x} \in K$ with probability $\mathbb{P}[X = \boldsymbol{x}]$. Then, we have

$$y_I = \int_K \left(\prod_{i\in I} x_i\right) \mu(d\boldsymbol{x}) = \mathbb{E}\left[\prod_{i\in I} X_i\right] = \mathbb{P}\left[\bigwedge_{i\in I} (X_i = 1)\right].$$

In particular, $y_\emptyset = 1$ and $y_{\{i\}} = \mathbb{P}[X_i = 1]$. When $\boldsymbol{x}^*$ is an optimal solution for the LP relaxation $\min\{\boldsymbol{c}^T\boldsymbol{x} : A\boldsymbol{x} \geq \boldsymbol{b}, \boldsymbol{0} \leq \boldsymbol{x} \leq \boldsymbol{1}\}$, it is customary to interprete $\boldsymbol{x}_i^*$ as the probability with which $\boldsymbol{x}_i$ should be set to 1. However, rounding all variables indepently from each other will most likely result in a non-feasible solution $\boldsymbol{x} \notin K$. So information on *joint events* is required, for example we need to know $\mathbb{P}[X_i = X_j = 1]$. One interpretation of the Lasserre hierarchy is that it gives information on the *correlation structure* of the solution, by introducing a set of variables y_I giving information on joint events of bounded cardinality $|I|$.

Let us now have a look at the problem (Las$_\delta$) applied to (IP). We only consider the linear constraints $g_i(\boldsymbol{x}) = \sum_j a_{ij}x_j - b_i \geq 0$, $\forall i \in [m]$, since we are already handling the equality constraints $x_i^2 = x_i$ thanks to the aforementioned moment substitutions. The objective function and all constraints of (IP) are linear, so $u_i = u = r = 1$. The level $\delta \geq 0$ of the hierarchy depends on the variables y_I for all subsets I of cardinality $\leq 2(\delta + 1)$ (thanks to our moments substitution, as opposed to all variables $y_{\boldsymbol{\alpha}}$ for $\boldsymbol{\alpha} \in \Delta(n, 2(\delta + r))$).

Definition 3. A vector $\boldsymbol{y} = (y_I)_{|I|\leq 2(\delta+1)}$ is said to be in the δ-th level of the Lasserre hierarchy, and we write $\boldsymbol{y} \in \mathcal{L}_\delta$, if the following LMIs hold:

$$\begin{aligned} y_\emptyset &= 1 \\ M_{\delta+1}(\boldsymbol{y}) := (y_{I\cup J})_{|I|,|J|\leq\delta+1} &\succeq 0 \\ M_\delta(g_i\boldsymbol{y}) := \left(\sum_{j\in[n]} a_{ij} y_{I\cup J\cup\{j\}} - b_i y_{I\cup J}\right)_{|I|,|J|\leq\delta} &\succeq 0, \quad \forall i \in [m]. \end{aligned}$$

Define further the set $\mathcal{L}_\delta^{proj} = \{\ [y_{\{1\}}, \ldots, y_{\{n\}}]^T \mid \ \boldsymbol{y} \in \mathcal{L}_\delta\}$, i.e., the projection of $\mathcal{L}_\delta$ onto the set of original coordinates.

Example:
The level $\delta = 0$ of the hierarchy corresponds to the *general recipe* for constructing the SDP relaxation of problems with binary variables seen in the previous chapter: Denote by $\boldsymbol{z} \in \mathbb{R}^n$ the vector with elements $z_i = y_{\{i\}}$ and denote by $Z \in \mathbb{S}^n$ the matrix with coordinates $Z_{ij} = y_{\{i,j\}}$. Then, if we order the sets I of cardinality $|I| \leq 1$ as $\big\{\emptyset, \{1\}, \ldots, \{n\}\big\}$, we have

$$M_1(\boldsymbol{y}) = \begin{bmatrix} y_\emptyset & \boldsymbol{z}^T \\ \boldsymbol{z} & Z \end{bmatrix} = \begin{bmatrix} 1 & \boldsymbol{z}^T \\ \boldsymbol{z} & Z \end{bmatrix},$$ #4

and for all i it holds $Z_{ii} = y_{\{i,i\}} = y_{\{i\}} = z_i$, that is, $\mathbf{diag}\,(Z) = \boldsymbol{z}$. The matrix $M_0(g_i\boldsymbol{y})$ is scalar, its only element is indexed by $(\emptyset, \emptyset)$ and it is equal to $\sum_{j\in[n]} a_{ij} y_{\{j\}} - b_i y_\emptyset = \boldsymbol{a}_i^T\boldsymbol{z} - b_i$. This shows:

$$\boldsymbol{y} \in \mathcal{L}_0^{proj} \iff A\boldsymbol{y} \geq \boldsymbol{b} \quad \text{and} \quad \exists Z : \begin{bmatrix} 1 & \boldsymbol{y}^T \\ \boldsymbol{y} & Z \end{bmatrix} \succeq 0,\ \mathrm{Diag}(Z) = \boldsymbol{y}.$$

By construction, the $\mathcal{L}_\delta^{proj}$ are nested and they contain K (as in Corollary 12):

$$\{\boldsymbol{x} \in \mathbb{R}^n : A\boldsymbol{x} \geq \boldsymbol{b}\} \supseteq \mathcal{L}_0^{proj} \supseteq \mathcal{L}_1^{proj} \supseteq \ldots \supseteq K = \{\boldsymbol{x} \in \{0,1\}^n : A\boldsymbol{x} \geq \boldsymbol{b}\}.$$

Since the $\mathcal{L}_\delta^{proj}$ are convex (they are SDP-representable), we can be a bit more precise and observe that the $\mathcal{L}_\delta^{proj}$ contain the *integer hull* H of Problem (IP), that is, the polytope formed by the integer vertices of K:

$$\mathcal{L}_\delta^{proj} \supseteq H := \mathbf{conv}\,\{\boldsymbol{x} \in \{0,1\}^n : A\boldsymbol{x} \geq \boldsymbol{b}\}.$$

In fact, we will now show that the hierarchy converges to H after at most $\delta = n$ rounds, that is, $\mathcal{L}_n^{proj} = H$. In other words, there exists a $\delta \leq n$ such that solving the SDP-relaxation over $\mathcal{L}_\delta^{proj}$ is equivalent to solving the original problem (IP).

Lemma 16. *Let $\boldsymbol{y} \in \mathcal{L}_\delta$. Then, for all $|I| \leq 2(\delta+1)$ it holds $y_I \in [0,1]$.*

Proof. The principal submatrix of $M_{\delta+1}(\boldsymbol{y})$ corresponding to the coordinates $\emptyset$ and I is:

$$\begin{bmatrix} 1 & y_I \\ y_I & y_I \end{bmatrix}.$$

It must positive semidefinite, so by the Schur complement lemma we have $y_I \geq 0$ and $y_i \geq y_I^2 \iff 1 \geq y_I$. □

Lemma 17. *Let $\boldsymbol{y} \in \mathcal{L}_\delta$ and assume that $0 < y_k < 1$ for some $k \in [n]$. Define the vectors $\boldsymbol{z}^{(1)}$ and $\boldsymbol{z}^{(2)}$ as follows:*

$$(\boldsymbol{z}^{(1)})_I = \frac{y_{I\cup\{k\}}}{y_k} \quad \textit{and} \quad (\boldsymbol{z}^{(2)})_I = \frac{y_I - y_{I\cup\{k\}}}{1-y_k}, \qquad \forall |I| \leq 2\delta.$$

Then, we have $\boldsymbol{z}^{(1)}, \boldsymbol{z}^{(2)} \in \mathcal{L}_{\delta-1}$, $(\boldsymbol{z}^{(1)})_{\{k\}} = 1$, $(\boldsymbol{z}^{(2)})_{\{k\}} = 0$, and the vector $\bar{\boldsymbol{y}} = (y_I)_{|I|\leq 2\delta}$ satisfies

$$\bar{\boldsymbol{y}} = y_k \boldsymbol{z}^{(1)} + (1-y_k)\boldsymbol{z}^{(2)}.$$

Proof. We have $(\boldsymbol{z}^{(1)})_{\{k\}} = \frac{y_{\{k\}}}{y_{\{k\}}} = 1$, $(\boldsymbol{z}^{(2)})_{\{k\}} = \frac{y_{\{k\}} - y_{\{k\}}}{1-y_{\{k\}}} = 0$ and for all $|I| \leq 2\delta$ it holds

$$y_k \boldsymbol{z}_I^{(1)} + (1-y_k)\boldsymbol{z}_I^{(2)} = y_{I\cup\{k\}} + y_I - y_{I\cup\{k\}} = y_I = (\bar{y})_I,$$

so the only thing left to show is $\boldsymbol{z}^{(1)}, \boldsymbol{z}^{(2)} \in \mathcal{L}_{\delta-1}$. The first easy thing to check is that $(\boldsymbol{z}^{(1)})_\emptyset = (\boldsymbol{z}^{(2)})_\emptyset = 1$. Then, the matrix $M_{\delta+1}(\boldsymbol{y})$ is positive semidefinite, so there exist vectors $(\boldsymbol{v}_I)_{|I|\leq\delta+1}$ such that $\big(M_{\delta+1}(\boldsymbol{y})\big)_{I,J} = y_{I\cup J} = \langle \boldsymbol{v}_I, \boldsymbol{v}_J\rangle$, $\forall |I|, |J| \leq \delta+1$. Now, for all $|I| \leq \delta$, define the vectors

$$\bar{\boldsymbol{v}}_I^{(1)} = \frac{\boldsymbol{v}_{I\cup\{k\}}}{\sqrt{y_k}} \quad \text{and} \quad \bar{\boldsymbol{v}}_I^{(2)} = \frac{\boldsymbol{v}_I - \boldsymbol{v}_{I\cup\{k\}}}{\sqrt{1-y_k}}.$$

We have $\langle \bar{\boldsymbol{v}}_I^{(1)}, \bar{\boldsymbol{v}}_J^{(1)}\rangle = \frac{1}{y_k}\langle \boldsymbol{v}_{I\cup\{k\}}, \boldsymbol{v}_{J\cup\{k\}}\rangle = \frac{1}{y_k} y_{I\cup J\cup\{k\}}$, which is also the element of coordinates (I,J) of the matrix $M_\delta(\boldsymbol{z}^{(1)})$. Hence, $M_\delta(\boldsymbol{z}^{(1)}) \succeq 0$.

Similarly, $\langle \bar{\boldsymbol{v}}_I^{(2)}, \bar{\boldsymbol{v}}_J^{(2)}\rangle = \frac{1}{1-y_k}\langle \boldsymbol{v}_I - \boldsymbol{v}_{I\cup\{k\}}, \boldsymbol{v}_J - \boldsymbol{v}_{J\cup\{k\}}\rangle = \frac{1}{1-y_k}(y_{I\cup J} - 2y_{I\cup J\cup\{k\}} + y_{I\cup J\cup\{k\}}) = \frac{y_{I\cup J} - y_{I\cup J\cup\{k\}}}{1-y_k}$, which is the element (I,J) of $M_\delta(\boldsymbol{z}^{(2)})$. This shows $M_\delta(\boldsymbol{z}^{(2)}) \succeq 0$.

Finally, for all $i \in [n]$, $M_{\delta-1}(g_i\boldsymbol{z}^{(1)}) \succeq 0$ and $M_{\delta-1}(g_i\boldsymbol{z}^{(2)}) \succeq 0$ can be proved in a similar manner: Let $(\boldsymbol{w}_I)_{|I|\leq\delta}$ be such that $\big(M_\delta(g_i\boldsymbol{y})\big)_{I,J} = \langle \boldsymbol{w}_I, \boldsymbol{w}_J\rangle$, and for all $|I| \leq \delta-1$ define the vectors

$$\bar{\boldsymbol{w}}_I^{(1)} = \frac{\boldsymbol{w}_{I\cup\{k\}}}{\sqrt{y_k}} \quad \text{and} \quad \bar{\boldsymbol{w}}_I^{(2)} = \frac{\boldsymbol{w}_I - \boldsymbol{w}_{I\cup\{k\}}}{\sqrt{1-y_k}}.$$

It can be verified that $\big(M_{\delta-1}(g_i\boldsymbol{z}^{(1)})\big)_{I,J} = \langle \bar{\boldsymbol{w}}_I^{(1)}, \bar{\boldsymbol{w}}_J^{(1)}\rangle$ and $\big(M_{\delta-1}(g_i\boldsymbol{z}^{(2)})\big)_{I,J} = \langle \bar{\boldsymbol{w}}_I^{(2)}, \bar{\boldsymbol{w}}_J^{(2)}\rangle$, which proves the claim and concludes this proof. □

Corollary 18. *The projection of $\mathcal{L}_\delta$ over the subset of coordinates $|I| \leq 2\delta$, $\mathcal{L}_{\delta|\delta-1} := \{(y_I)_{|I|\leq 2\delta} \,|\boldsymbol{y} \in \mathcal{L}_\delta\}$, satisfies*

$$\mathcal{L}_{\delta|\delta-1} \subseteq \mathbf{conv}\Big(\{\boldsymbol{z} \in \mathcal{L}_{\delta-1} : z_k = 0\} \cup \{\boldsymbol{z} \in \mathcal{L}_{\delta-1} : z_k = 1\}\Big).$$

Iterating the above result, we also see that for all subsets $S \subseteq [n]$, $\mathcal{L}_{\delta|\delta-|S|} := \{(y_I)_{|I|\leq 2(\delta-|S|+1)} \,|\boldsymbol{y} \in \mathcal{L}_\delta\}$ is the convex hull of all elements of $\mathcal{L}_{\delta-|S|}$ with $\{0,1\}$ elements in S:

$$\mathcal{L}_{\delta|\delta-|S|} := \{(y_I)_{|I|\leq 2(\delta-|S|+1)} \,|\boldsymbol{y} \in \mathcal{L}_\delta\} \subseteq \mathbf{conv}\,\{\boldsymbol{z} \in \mathcal{L}_{\delta-|S|} : z_i \in \{0,1\},\ \forall i \in S\}.$$

In particular, if we take $\delta = n$ and $S = [n]$, we obtain $\mathcal{L}_{n|0} \subseteq \mathbf{conv}\,\{\boldsymbol{z} \in \mathcal{L}_0 : \boldsymbol{z} \in \{0,1\}^n\}$. Then, by projecting onto the subset of original coordinates $(\{1\},\ldots,\{n\})$, we obtain $\mathcal{L}_n^{proj} \subseteq H := \mathbf{conv}\,\{\boldsymbol{x} \in \{0,1\}^n : A\boldsymbol{x} \geq \boldsymbol{b}\}$. Therefore, we have shown:

Proposition 19.

$$\{\boldsymbol{x} \in \mathbb{R}^n : A\boldsymbol{x} \geq \boldsymbol{b}\} \supseteq \mathcal{L}_0^{proj} \supseteq \mathcal{L}_1^{proj} \supseteq \ldots \supseteq \mathcal{L}_n^{proj} = \mathbf{conv}\,\{\boldsymbol{x} \in \{0,1\}^n : A\boldsymbol{x} \geq \boldsymbol{b}\}.$$

We conclude this chapter by mentioning that the study of these hierarchies is an active field of research. In particular, the SDP relaxation of (IP) over $\mathcal{L}_\delta^{proj}$ can be solved in polynomial time *for a fixed* $\delta \in \mathbb{N}$. Many of the tightest known polytime approximation algorithms for certain NP hard optimization problems involve solving $\delta = O(1/\epsilon)$ rounds of the Lasserre hierarchy.

References

[1] Hamza Fawzi. Topics in Convex Optimisation. Lecture notes availble at `http://www.damtp.cam.ac.uk/user/hf323/lent2017_topics_convex_optimisation/topics_convex_optimisation.html`, 2017.

[2] Monique Laurent. Sums of squares, moment matrices and optimization over polynomials. In Emerging applications of algebraic geometry, pages 157270. Springer, 2009.

[3] Jean-Bernard Lasserre. Moments, positive polynomials and their applications. World Scientific. 2010.

[4] Martin Mevissen. Introduction to concepts and advances in polynomial optimization. Review available at `https://inf.ethz.ch/personal/fukudak/semi/optpast/FS07/opt_abs/PolynomialOptimization.pdf`, *Institute for Operations Research, ETH, Zurich*, 2007.

[5] Yohann de Castro. A short introduction to Moment-SOS hierarchies. Review paper available at `https://ydecastro.github.io/research/paper24.pdf`, Deuxième Congès National de la SMF, 2018.

[6] Thomas Rothvoß. The Lasserre hierarchy in Approximation algorithms. Lecture Notes for the MAPSP 2013 Tutorial available at `https://sites.math.washington.edu/~rothvoss/lecturenotes/lasserresurvey.pdf`, 2013.