Chapter XI: Polynomial 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 largely based on recent lecture notes of H. Fawzi [1].

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

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

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 $\mathcal{P}_d^{\text{pos}}$:

$$\inf_{x \in \mathbb{R}} p(x) \ = \ \sup\ \{\gamma \in \mathbb{R} : \ p - \gamma \text{ is nonnegative}\} = \sup_{\gamma \in \mathbb{R}} \quad \gamma \qquad \boldsymbol{p} - \gamma \boldsymbol{e}_0 \succeq_{\mathcal{P}_d^{\text{pos}}} \boldsymbol{0}.$$

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 $\mathcal{P}_d^{\text{sos}} \subset \mathbb{R}^{d+1}$ the set of (vectors of coefficients of) sum of squares polynomials of degree at most d.

From the definition, it is clear that $\mathcal{P}_d^{\text{sos}} \subseteq \mathcal{P}_d^{\text{pos}}$. 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 $\mathcal{P}_{2d}^{pos}$ are proper.*

Note that when d is odd, $\mathcal{P}_d^{\text{pos}}$ and $\mathcal{P}_d^{\text{sos}}$ 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 $\mathcal{P}_{2d}^{\text{sos}}$ and $\mathcal{P}_{2d}^{\text{pos}}$:

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

$$\mathcal{P}_{2d}^{pos} = \mathcal{P}_{2d}^{sos}.$$

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. □

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

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

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

Now, assume that $p(x) = \boldsymbol{x}^T M \boldsymbol{x}$ 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 $p(x) = \boldsymbol{x}^T P^T P \boldsymbol{x} = \|P\boldsymbol{x}\|^2 = \sum_{i=1}^{m} (\boldsymbol{p}_i^T \boldsymbol{x})^2$, 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 $p_i(x) := \boldsymbol{p}_i^T \boldsymbol{x}$ is a polynomial of degree $\le d$.

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

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 $\sup\ \{\gamma \in \mathbb{R} :\ \boldsymbol{p} - \gamma \boldsymbol{e}_0 \in \mathcal{P}_6^{\text{sos}}\}$, where $\boldsymbol{p} = [1, -2, 0, 2, -17, 0, 1]^T$. Then, we can use Theorem 3 to obtain the following SDP formulation:

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

Polynomial optimization over a real interval can also be handled via sum of squares. For example, we mention the following result:

Theorem 4. *Let $a < b$. A polynomial p of even degree $2d$ is nonnegative over the interval $[a, b]$ iff*

$$\exists s_1 \in \mathcal{P}^{sos}_{2d}, s_2 \in \mathcal{P}^{sos}_{2d-2} : \quad p(x) = s_1(x) + (b - x)(x - a)s_2(x).$$

A polynomial p of odd degree $2d + 1$ is nonnegative over the interval $[a, b]$ iff

$$\exists s_1 \in \mathcal{P}^{sos}_{2d}, s_2 \in \mathcal{P}^{sos}_{2d} : \quad p(x) = (x - a)s_1(x) + (b - x)s_2(x).$$

2 Multivariate Polynomials

A polynomial $p \in \mathbb{R}_d[x_1, \dots, x_n]$ 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

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

The cone $\mathcal{P}^{\text{pos}}_{n,d}$ of $n-$variate nonnegative polynomials of degree $\leq d$ can be defined as in the univariate case, as well as the cone $\mathcal{P}^{\text{sos}}_{n,d}$ of $n-$variate sum of squares of degree $\leq d$. Note that the dimension of these cones is $s(n, d) := |\Delta(n, d)| = \binom{n+d}{d}$.

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

Proposition 5. *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 $p \notin \mathcal{P}^{\text{sos}}_{n,d}$ will be given after Theorem 7. □

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

Theorem 6 (Hilbert)**.**

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

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

Theorem 7. *The polynomial $p \in \mathbb{R}_{2d}[x_1, \ldots, x_n]$, where $p(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha} \in \Delta(n,d)} p_{\boldsymbol{\alpha}}\ \boldsymbol{x}^{\boldsymbol{\alpha}}$ 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 $\boldsymbol{z} = [\boldsymbol{x}^{\boldsymbol{\alpha}}]_{\boldsymbol{\alpha} \in \Delta(n,d)} \in \mathbb{R}^{s(n,d)}$, which contains all monomials of degree $\leq d$. Direct calculation shows that $\boldsymbol{z}^T M \boldsymbol{z} = p(\boldsymbol{x})$ if and only if the equality conditions (1) are satisfied.

Then, when (1) holds, we have

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

□

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 5 is not a sum of square. In that case, we have $2d = 6$ and $n = 2$, so the matrix M of Theorem 7 is of size $s(2,3) = 10$; The Motzkin polynomial is a sum of squares iff $\exists M \in \mathbb{S}_+^{10}$ such that

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

This is, in fact, a feasibility SDP problem. To show that this problem is infeasible, we can derive its dual, and show that the dual is unbounded, by exhibiting a "dual improving direction". Formally, this is in fact equivalent to finding a vector $\boldsymbol{y}$ such that the statement on the right-hand-side is satisfied: #2

$$\boldsymbol{p} \notin \mathcal{P}_{2,6}^{\text{sos}} \iff \exists \boldsymbol{y} \in (\mathcal{P}_{2,6}^{\text{sos}})^* : \langle \boldsymbol{p}, \boldsymbol{y} \rangle < 0.$$

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 $y_{\boldsymbol{\alpha}} = 0$ for all the other multi-indices $\boldsymbol{\alpha} \in \Delta(2,6)$ 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, which is known as the sum of squares hierarchy, or the Lasserre hierarchy:

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

Proposition 8. *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

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

□

Under some mild conditions, it can be shown that the hierarchy converges. There are even cases where one can ensure *finite convergence* of the Lasserre hierarchy. This chapter in only a very short introduction to the fascinating world of polynomial optimization and sum of squares. To go further, we refer the reader to the comprehensive review [2]. In particular, further topics include:

- The study of the dual cones of $\mathcal{P}_{n,d}^{\mathrm{pos}}$ and $\mathcal{P}_{n,d}^{\mathrm{sos}}$ (which we evoked at the end of Example #2). These cones are related to the *moment cone*, which characterize sequences $(y_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\Delta(n,d)}$ having a *representing measure*, that is, a nonnegative measure μ such that $\int_{\boldsymbol{x}} \boldsymbol{x}^{\boldsymbol{\alpha}}\ d\mu(\boldsymbol{x}) = y_{\boldsymbol{\alpha}}$, for all $\boldsymbol{\alpha} \in \Delta(n,d)$.
- The possibility to use sum-of-squares relaxations to approach constrained polynomial problems, where the variable $\boldsymbol{x} \in \mathbb{R}^n$ is constrained in a *semi-algebric set*, i.e., a set defined by $\{\boldsymbol{x} \in \mathbb{R}^n : g_i(\boldsymbol{x}) \leq 0, \forall i \in [m]\}$ for some polynomials $g_1, \ldots, g_m$.
- Sufficient conditions for the convergence of the Lasserre hierarchy within a finite number of steps.

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 157–270. Springer, 2009.