MIT 6.7220/15.084 — Nonlinear Optimization (Spring ‘25)

Lecture 10
Polynomial optimization

Instructor: Prof. Gabriele Farina

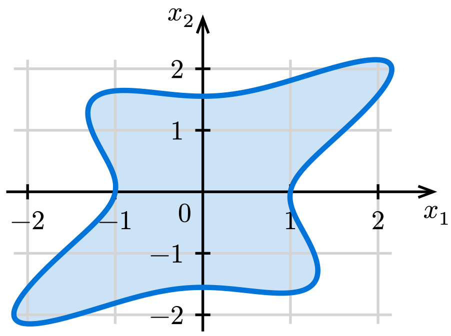
As mentioned in Lecture 9, semidefinite programming has established itself as an extremely important tool, with disparate applications in combinatorial optimization, control theory, and machine learning. Under the appropriate assumptions (namely strict feasibility), semidefinite programs can be solved efficiently, and strong commercial solvers are available. In this lecture, we will see how semidefinite programming can be used to relax a large class of nonconvex problems that includes combinatorial problems as well, and find globally optimal solutions.

L10.1 Polynomial optimization problems

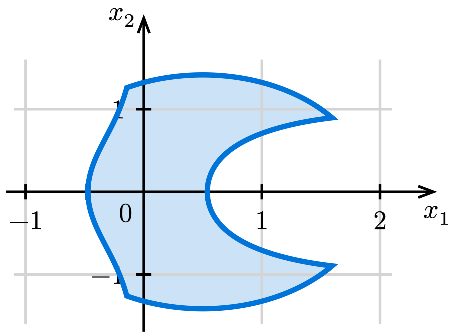
In this lecture, we focus on *polynomial optimization* problems,[1] that is, problems of the form

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & g_j(x) \geq 0 \qquad j = 1, ..., m \\ & x \in \mathbb{R}^n, \end{aligned}$$

where f and the g_j are n-variate polynomials with real coefficients, *i.e.*, $f, g_j \in \mathbb{R}[x]$. This class of problems is very expressive. To start, it contains nonconvex problems. For example, these are just two small examples of feasible sets expressible via bivariate polynomials.



$$\Omega_1 := \left\{ x_1^4 + x_2^4 - \frac{1}{2} x_1^3 x_2 - 2x_2^2 - x_1^2 x_2^2 \leq 1 \right\}$$



$$\Omega_2 := \begin{cases} 1 - 4x_1^2 + 4x_1 x_2^2 + x_1^3 \geq 0 \\ \left(x_1 - \frac{1}{2}\right)^2 + x_2^2 \leq 2. \end{cases}$$

⋆These notes are class material that has not undergone formal peer review. The TAs and I are grateful for any reports of typos.

[1] We consider constraints in the "≥" direction instead of the usual "≤" direction. This is just for notational alignment with the literature.

This class of problems also contains combinatorial optimization problems known to be hard, such as the maximum cut problem (see Lecture 1). The reason why combinatorial optimization problems can easily be expressed by polynomial optimization problems is that the condition $x \in \{0, 1\}$ can be equivalently written as the pair of polynomial constraints $x(1-x) \le 0$, $-x(1-x) \le 0$.

The key takeaway of this lecture is that polynomial optimization problems can, under mild assumptions, be converted into semidefinite programs. Two observations are immediate:

- Polynomial optimization problems are nonconvex, but semidefinite programs are convex conic problems. It is a priori not obvious at all that the latter can be used to solve the former.
- There is no free lunch: polynomial optimization problems are very expressive, and can capture computationally hard problems. In contrast, semidefinite programs can be solved in time polynomial in the dimension (under suitable hypotheses). The catch is that the semidefinite program relaxations we will consider need to be, in the worst case, of dimension exponential in the degree and number of variables of the polynomial problem considered. In practice, however, very good results are achieved already for relaxations of dimension way lower than what the theory predicts.

L10.2 The cone of nonnegative polynomials

To begin our exploration, let's focus on the simplest of settings: unconstrained optimization. In particular, suppose that we would like to find the minimum of a polynomial $p(x)$ in $\mathbb{R}^n$. (Note that the question is meaningful only if the degree of p is even, as otherwise p is not lower bounded.)

Computationally, the problem of computing the minimum of a generic polynomial $p(x)$ is equivalent to the question of being able to check whether a generic polynomial $q(x)$ satisfies $q(x) \ge 0$ for all $x \in \mathbb{R}^n$. In one direction, if we can compute $\min q(x)$, then we can answer whether $q(x) \ge 0$ by simply checking if the minimum is below or above 0. In the other direction, if we can check whether $q(x) \ge 0$ for all $x \in \mathbb{R}^n$, then we can compute $\min p(x)$ by performing a binary search over t, until we find the largest t such that $p(x) - t \ge 0$ for all $x \in \mathbb{R}^n$. For this reason, we will be interested in understanding more about the following object.

Definition L10.1 (Cone of nonnegative polynomials)**.** The *cone of nonnegative polynomials* is the set of all polynomials $p \in \mathbb{R}[x]$ such that $p(x) \ge 0$ for all $x \in \mathbb{R}^n$. We will denote this set by $\mathbb{R}[x]_{\ge 0}$.

It is immediate to check [▷ You should check!] that the set $\mathbb{R}[x]_{\ge 0}$ is a closed convex cone.

L10.2.1 Complexity considerations

Unfortunately, deciding whether a polynomial is nonnegative in $\mathbb{R}^n$ is as hard as deciding membership in the copositive cone, which is NP-hard [DG14].

Theorem L10.1. Deciding membership in the cone of nonnegative polynomials $\mathbb{R}[x]_{\geq 0}$ for any polynomial of degree ≥ 4 with $n \geq 2$ variables is computationally intractable.

Proof. By reduction from the problem of deciding is a symmetric matrix $M \in \mathbb{R}^{n\times n}$ is copositive (see Lecture 9), that is,

$$z^\top M z \geq 0 \qquad \forall z \in \mathbb{R}^n_{\geq 0}.$$

Indeed, consider the n-variate polynomial of degree 4 defined as

$$p(x_1, ..., x_n) := \begin{pmatrix} x_1^2 \\ \vdots \\ x_n^2 \end{pmatrix}^\top M \begin{pmatrix} x_1^2 \\ \vdots \\ x_n^2 \end{pmatrix}.$$

If M is copositive, this polynomial is nonnegative on $\mathbb{R}^n$, since the vector of squares $(x_1^2, ..., x_n^2)$ is always nonnegative. Conversely, if p is nonnegative on $\mathbb{R}^n$, then since $(x_1^2, ..., x_n^2)$ spans all of $\mathbb{R}^n_{\geq 0}$, we have that M is copositive. □

As a side remark, this is the second *convex* set we see for which membership—let alone optimization—is intractable (the first being the copositive cone). This should serve as a reminder that equating convexity with tractability can be a dangerous oversimplification.

L10.3 Sum-of-squares polynomials

Theorem L10.1 shows that even deciding *membership* in the cone of nonnegative polynomials is hard in general, let alone optimizing over it. This begs the question of what is the largest set of polynomials for which deciding nonnegativity is easy. The following cone will provide an extremely important step in this direction.

Definition L10.2 (Sum-of-squares polynomials, $\Sigma[x]$)**.** An n-variate polynomial $p \in \mathbb{R}[x]$ is said to be a *sum-of-squares* polynomial (or SOS for short) if it can be written as

$$p(x) = \sum_k p_k(x)^2, \qquad x \in \mathbb{R}^n$$

for appropriate polynomials $p_k \in \mathbb{R}[x]$.

Remark L10.1. Only polynomials of even degree can be sum-of-squares. Furthermore, if p has degree $2d$, then each of the p_k's can only have degree at most d.

It is immediate to check [▷ you should check!] that the set of sum-of-squares polynomials is a closed, convex, nonempty cone. Furthermore, every sum-of-squares polynomial is clearly a nonnegative polynomial, so $\Sigma[x] \subseteq \mathbb{R}[x]_{\geq 0}$. In general, the inclusion is strict, as there exist polynomials that are nonnegative but not sum-of-squares. We will see a well-known example, called the Motzkin polynomial, later in this lecture in Example L10.2.

L10.3.1 The connection to the positive semidefinite cone

Unlike the cone of nonnegative polynomials, the cone of sum-of-squares polynomials is easy to characterize. The following theorem provides a very useful criterion for deciding whether a polynomial is sum-of-squares.

Theorem L10.2. Membership in the cone of sum-of-squares polynomials can be determined by checking the feasibility of a semidefinite program. In particular, let $p(x)$ be an arbitrary polynomial of degree $2d$ in n variables, and let v_d be the vector of all monomials of degree up to d that can be constructed using the variables x.

Then, $p(x) \in \Sigma[x]$ if and only if there exists a positive semidefinite matrix Q such that

$$p(x) = (v_d(x))^\top Q v_d(x),$$

Proof. ($\Longrightarrow$) Any polynomial of degree up to d can be written as a linear combination of the monomials in v_d. Hence, any sum-of-squares polynomial can be written in the form

$$p(x) = \sum_k \left(v_d(x)^\top \alpha_k\right)^2$$

for some appropriate coefficient vectors α_k. But then, we can write

$$p(x) = \left\| \underbrace{\begin{pmatrix} -\alpha_1^\top- \\ -\alpha_2^\top- \\ \vdots \end{pmatrix}}_{=:C} v_d(x) \right\|^2 = (v_d(x))^\top C^\top C v_d(x). \tag{1}$$

Since a matrix of the form $C^\top C$ is positive semidefinite, this direction holds.

($\Longleftarrow$) Conversely, if $p(x) = (v_d(x))^\top Q v_d(x)$ for some positive semidefinite matrix Q, we can write $Q = C^\top C$ for some C. But then, we can use (1) from right to left and extract from the rows of C the coefficients α_k for the sum-of-squares representation of $p(x)$. □

Example L10.1. A bivariate polynomial $p(x_1, x_2)$ of degree 4 is a sum-of-squares polynomial if and only if there exists $Q \succeq 0$ such that

$$\begin{aligned} p(x_1,x_2) &= \begin{pmatrix} x_1^2 \\ x_1x_2 \\ x_2^2 \\ x_1 \\ x_2 \\ 1 \end{pmatrix}^\top \begin{pmatrix} q_{00} & q_{01} & q_{02} & q_{03} & q_{04} & q_{05} \\ q_{01} & q_{11} & q_{12} & q_{13} & q_{14} & q_{15} \\ q_{02} & q_{12} & q_{22} & q_{23} & q_{24} & q_{25} \\ q_{03} & q_{13} & q_{23} & q_{33} & q_{34} & q_{35} \\ q_{04} & q_{14} & q_{24} & q_{34} & q_{44} & q_{45} \\ q_{05} & q_{15} & q_{25} & q_{35} & q_{45} & q_{55} \end{pmatrix} \begin{pmatrix} x_1^2 \\ x_1x_2 \\ x_2^2 \\ x_1 \\ x_2 \\ 1 \end{pmatrix} \\ &= (q_{00})x_1^4 + (2q_{01})x_1^3x_2 + (q_{11} + 2q_{02})x_1^2x_2^2 + (2q_{12})x_1x_2^3 + (q_{22})x_2^4 \\ &\quad + (2q_{03})x_1^3 + (2q_{04} + 2q_{13})x_1^2x_2 + (2q_{23} + 2q_{14})x_1x_2^2 + (2q_{24})x_2^3 \\ &\quad + (q_{33} + 2q_{05})x_1^2 + (2q_{15} + 2q_{34})x_1x_2 + (q_{44} + 2q_{25})x_2^2 \\ &\quad + (2q_{35})x_1 + (2q_{45})x_2 + (q_{55}). \end{aligned}$$

In other words, $p(x_1, x_2)$ is SOS if and only if its coefficients match the expressions computed in the previous line, for an appropriate positive semidefinite Q.

Note that the coefficients of the polynomial $v(x)^\top Q v(x)$ are always a linear function of the entries of the matrix Q. Hence, we obtain the following corollary.

Corollary L10.1. The cone of n-variate sum-of-squares polynomials of degree $2d$ is a *linear transformation of the cone of $s \times s$ positive semidefinite matrices*, where $s = \binom{n+d}{d}$ is the dimension of the vector $v_d(x)$.

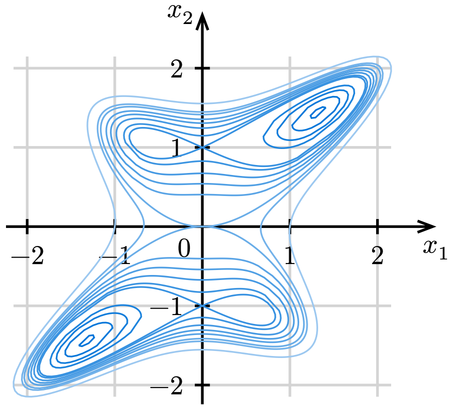
L10.3.2 An approach to unconstrained polynomial optimization

By replacing the cone of nonnegative polynomials with the cone of sum-of-squares polynomials, we can obtain an approximation for the minimum of a polynomial over $\mathbb{R}^n$. The key idea is to find a value t such that we can certify that $p(x) - t \in \Sigma[x]$. In that case, it immediately follows that $p(x) \geq t$ for all $x \in \mathbb{R}^n$.

As a concrete example, suppose we would like to approximate the minimum of the bivariate polynomial of fourth degree

$$p(x_1, x_2) := x_1^4 + x_2^4 - \frac{1}{2}x_1^3 x_2 - 2x_2^2 - x_1^2 x_2^2,$$

whose level sets are shown on the right.



If we could find a value t such that we can certify that $p(x_1, x_2) - t \in \Sigma[x]$, that would immediately prove that $p(x_1, x_2) \geq t$ for all $x \in \mathbb{R}^n$.

To certify that $p(x_1, x_2) - t \in \Sigma[x]$, we know from Theorem L10.2 and Example L10.1 that it suffices to find a positive semidefinite matrix $Q \succeq 0$ that enables writing $p(x_1, x_2) - t$ in the form $v_2(x)^\top Q v_2(x)$. More explicitly, using the expansion from Example L10.1, we are interested in solving the following semidefinite program.

max t

s.t.

- $q_{00} = 1$ (coeff. of x_1^4)
- $2q_{01} = -\frac{1}{2}$ (coeff. of $x_1^3 x_2$)
- $q_{11} + 2q_{02} = -1$ (coeff. of $x_1^2 x_2^2$)
- $2q_{12} = 0$ (coeff. of $x_1^1 x_2^3$)
- $q_{22} = 1$ (coeff. of x_2^4)
- $2q_{03} = 0$ (coeff. of x_1^3)
- $2q_{04} + 2q_{13} = 0$ (coeff. of $x_1^2 x_2$)
- $2q_{23} + 2q_{14} = 0$ (coeff. of $x_1^1 x_2^2$)
- $2q_{24} = 0$ (coeff. of x_2^3)
- $q_{33} + 2q_{05} = 0$ (coeff. of x_1^2)
- $2q_{15} + 2q_{34} = 0$ (coeff. of $x_1 x_2$)
- $q_{44} + 2q_{25} = -2$ (coeff. of x_2^2)
- $2q_{35} = 0$ (coeff. of x_1)
- $2q_{45} = 0$ (coeff. of x_2)
- $q_{55} = -t$ (constant)
- $Q = [q_{ij}] \succeq 0$.

Feeding the semidefinite program into a solver,[2] we find that the optimal objective has value $t^\star \approx -2.08053$, obtained by the positive semidefinite matrix

$$Q^\star \approx \begin{pmatrix} 1.0 & -0.25 & -0.579 & 0.0 & 0.0 & -0.0734 \\ -0.25 & 0.159 & 0.0 & 0.0 & 0.0 & 0.135 \\ -0.579 & 0.0 & 1.0 & 0.0 & 0.0 & -1.062 \\ 0.0 & 0.0 & 0.0 & 0.146 & -0.135 & 0.0 \\ 0.0 & 0.0 & 0.0 & -0.135 & 0.124 & 0.0 \\ -0.073 & 0.135 & -1.062 & 0.0 & 0.0 & 2.080 \end{pmatrix} \succeq 0.$$

This means that the polynomial $p(x_1, x_2) - t^\star$ is proved to belong in $\Sigma[x]$, and therefore $p(x_1, x_2) \geq -2.08053$ for all $x_1, x_2 \in \mathbb{R}$. The following two questions are natural:

1. Is this value optimal (*i.e.*, does it indeed match the minimum of p over $\mathbb{R}^n$)?
2. And if not, what could we do to improve the bound?

We will address these questions in the next section.

L10.4 Characterizing global nonnegativity using the SOS cone

The SOS cone provides a computationally viable relaxation of the cone of nonnegative polynomials. But how much do we lose by shifting our focus from nonnegative to SOS polynomials? As it turns out, the set of nonnegative polynomials of degree up to d can be approximated arbitrarily closely by the using SOS polynomials of degrees potentially exponential in d.

L10.4.1 When are the SOS and nonnegative polynomial cone equal?

In fact, in some cases, the cone of sum-of-squares polynomials is exactly *equal* to the cone of nonnegative polynomials. This is the content of Hilbert's theorem.

Theorem L10.3 (Hilbert)**.** $\mathbb{R}[x]_{\geq 0} = \Sigma[x]$ if and only if:

- $n = 1$, no matter the degree d; or
- $d = 2$, no matter the number of variables n; or
- $n = 2$ and $d = 4$.

The cases for $n = 1$ can be shown easily starting from a factorization of the polynomial [▹ You should try to prove that special case!].

Remark L10.2. Hilbert's theorem shows that in the example of Section L10.3.2 we were not just randomly lucky: the polynomial we considered was bivariate and of degree 4; hence, the condition $p(x) - t \in \Sigma[x]$ was exactly equivalent to $p(x) - t \in \mathbb{R}[x]_{\geq 0}$.

Other than the three settings above, in general the cone of sum-of-squares polynomials is a strict subset of the cone of nonnegative polynomials. Explicit examples of polynomials that are nonnegative and yet not sum-of-squares are known, as we show next.

[2] In my case, I used Clarabel (`https://github.com/oxfordcontrol/Clarabel.rs`).

Example L10.2 (Motzkin's polynomial)**.** The polynomial

$$M(x_1, x_2) \coloneqq x_1^4 x_2^2 + x_1^2 x_2^4 - 3x_1^2 x_2^2 + 1$$

is not a sum-of-squares polynomial, but it is nonnegative on $\mathbb{R}^2$, since from the arithmetic-geometric mean inequality we have

$$\frac{x_1^4 x_2^2 + x_1^2 x_2^4 + 1}{3} \geq \sqrt[3]{x_1^4 x_2^2 \cdot x_1^2 x_2^4 \cdot 1} = x_1^2 x_2^2.$$

[▷ As an exercise, show that the Motzkin polynomial is not a sum-of-squares polynomial. One approach is to write the semidefinite program certifying membership of the polynomial in $\Sigma[x]$, and showing that it is infeasible.]

L10.4.2 Hilbert's conjecture and Artin's theorem

As discussed above, being sum-of-squares is sufficient but not necessary for a polynomial to be nonnegative in general. One might then try to consider other sufficient conditions, hoping to cast a wider and wider net encompassing all nonnegative polynomials.

For example, even if a polynomial $p(x)$ is not sum-of-squares, if we could find *two* sum-of-squares polynomials $q(x), r(x)$ such that

$$p(x) = \frac{q(x)}{r(x)}$$

and $r(x)$ divides $q(x)$.

Example L10.3. The Motzkin polynomial can be written as (replacing $x \coloneqq x_1, y \coloneqq x_2$)

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

where both the numerator and the denominator are sum-of-squares polynomials. [▷ You should verify this!]

In his famous 17th problem, Hilbert conjectured that the condition above is necessary for a polynomial to be nonnegative. This conjecture was proved by Artin in 1927 via the Artin-Schreier theory. A celebrated result by Polyá later showed that for homogeneous polynomials (*i.e.*, those where every term has the same degree), without loss of generality one can focus on denominators in the form $r(x) = (x_1^2 + \cdots + x_n^2)^{d'}$ for some $d' \in \mathbb{N}$ to check positivity in $\mathbb{R}^n \setminus \{0\}$. In general, to check positivity one can search for $q, r \in \Sigma[x]$ such that $r(x)p(x) = 1 + q(x)$. To check nonnegativity, one can search for $q, r \in \Sigma[x]$ and $\ell \in \mathbb{N}$ such that $r(x)p(x) = p(x)^{2\ell} + q(x)$.

Combined, these results show that checking nonnegativity of polynomials of degree up to d can be approximated arbitrarily closely by using SOS polynomials of degrees potentially exponential in d. Concrete bounds are known (but are very pessimistic). For today, the important thing is the message: SOS polynomials provide a complete characterization of nonnegative polynomials, as long as one is willing to look into high-degree SOS polynomials.

L10.5 The constrained case: Putinar's Positivstellensatz

Having understood more of the unconstrained case, we can now focus on the more general case of polynomial optimization problems with constraints.

For simplicity, we focus on the case of polynomial optimization problems in which at least one of the constraints $g_j(x) \geq 0$ defines a feasible set $\{x \in \mathbb{R}^n : g_j(x) \geq 0\}$ that is compact. In this setting, the following important result serves as a bridge for how to use SOS polynomials to solve polynomial optimization problems.

Theorem L10.4 (Putinar's Positivstellensatz)**.** Let

$$\Omega := \{x \in \mathbb{R}^n : g_j(x) \geq 0 \quad \forall j = 1, ..., m\},$$

where $\{x \in \mathbb{R}^n : g_j(x) \geq 0\}$ is compact for at least one $j \in \{1, ..., m\}$. Any polynomial $p \in \mathbb{R}[x]$ positive on Ω can be written in the form

$$p = \sigma_0 + \sum_{j=1}^{m} \sigma_j g_j$$

for appropriate SOS polynomials $\sigma_j \in \Sigma[x]$, $j = 0, ..., m$.

More general versions that apply without the compactness restriction above are also known (*e.g.*, Stengle's Positivstellensatz and Schmüdgen's Positivstellensatz).

Again, bounds are known for the degree of the SOS polynomials that are needed in Putinar's theorem. In practice, however, very good results are achieved already for relaxations of dimension way lower than what the theory predicts. In particular, to check whether $p(x) > t$, we can try to see if there exist SOS polynomials σ_j of degree up to some $\gamma \in \mathbb{N}$, such that $p(x) - t = \sigma_0 + \sum_{j=1}^{m} \sigma_j g_j$. For each σ_j, we will be looking for a corresponding positive semidefinite matrix Q_j that enables writing $\sigma_j(x) = v_\gamma(x)^\top Q_j v_\gamma(x)$. Overall, for any given γ we will be writing a positive semidefinite program.

L10.5.1 An example

As a final example, suppose we want to find the minimum x-coordinate of any point (x_1, x_2) in the set

$$\Omega_2 := \begin{cases} 1 - 4x_1^2 + 4x_1 x_2^2 + x_1^3 \geq 0 \\ \left(x - \frac{1}{2}\right)^2 + y^2 \leq 2. \end{cases}$$

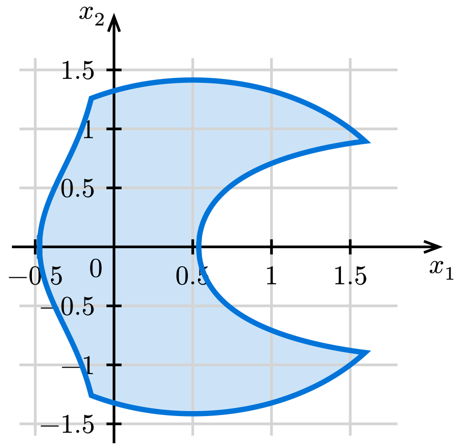
shown in Section L10.1. In this case, we have

$$g_1(x_1, x_2) := x_1^3 + 4x_1 x_2^2 - 4x_1^2 + 1, \qquad g_2(x_1, x_2) := -\left(x_1 - \frac{1}{2}\right)^2 - x_2^2 + 2.$$

If we look for sum-of-square polynomials $\sigma_0, \sigma_1, \sigma_2 \in \Sigma[x]$ of degrees $2, 4, 6, 8, 10$, we find the following objective values.

Degree of σ_j	**Objective value**
2	infeasible
4	-0.91421
6	-0.91421
8	-0.47283
10	-0.47283

The value -0.47283 is indeed the correct answer, as you can see on the magnified plot on the right.



L10.6 Further readings

The field of polynomial optimization is vast and has seen a lot of activity in the last few decades. Several connections between polynomial optimization and real algebraic geometry run deep. Several excellent books and surveys are available, including those of Lasserre, J. B. [Las15], Blekherman, G., Parrilo, P. A., & Thomas, R. R. [BPT12], and Nie, J. [Nie23].

Bibliography for this lecture

[DG14] Dickinson, P. J., & Gijben, L. (2014). On the computational complexity of membership problems for the completely positive cone and its dual. *Computational Optimization and Applications*, *57*, 403–415.

[Las15] Lasserre, J. B. (2015). *An introduction to polynomial and semi-algebraic optimization* (Vol. 52). Cambridge University Press.

[BPT12] Blekherman, G., Parrilo, P. A., & Thomas, R. R. (2012). *Semidefinite optimization and convex algebraic geometry*. SIAM.

[Nie23] Nie, J. (2023). *Moment and Polynomial Optimization*. SIAM.

Changelog

- Mar 13, 2025: Fixed typo (thanks Jonathan Huang!)
- Mar 13, 2025: Changed “sufficient” → “necessary” (thanks Anon. Helix! https://piazza.com/class/m6lg9aspoutda/post/87)
- Mar 15, 2025: Fixed typo in Section L10.3.2: $v_4 \to v_2$ (thanks Khizer Shahid!)
- Mar 19, 2025: Changed “reduction to” → “reduction from” in proof of Theorem L10.1 (thanks Alina Yang for the suggestion!)