CS498: Algorithmic Engineering

Semidefinite Programming and Applications II

Guest lecture: Chandra Chekuri

University of Illinois Urbana-Champaign

Outline

SDP: Formulation

Input: matrices $C, D_1, \ldots, D_k \in M_n$ and scalars $d_1, \ldots, d_k$.
Variable: a matrix $Y \in M_n$.

$$
\begin{aligned}
\max \quad & C \cdot Y \\
\text{s.t.} \quad & D_i \cdot Y = d_i \quad 1 \le i \le k \\
& Y \succeq 0
\end{aligned}
$$

SDP is special case of convex programming and efficiently solvable to within arbtirary good accuracy.

SDP and Vector Programming

If $Y \succeq 0$, write $Y = W^T W$ (Cholesky) and let $v_1, \ldots, v_n$ be the columns of W. Then $Y_{ij} = v_i \cdot v_j$.

This gives an equivalent **vector program**:

$$
\begin{aligned}
\max \quad & \sum_{ij} C_{ij}\, v_i \cdot v_j \\
\text{s.t.} \quad & \sum_{ij} D_{ij}^{(k)}\, v_i \cdot v_j = d_k \quad \forall\, k \\
& v_i \in \mathbb{R}^n
\end{aligned}
$$

SDP $\longleftrightarrow$ Vector Programming. Vector programming is often more natural for graph and metric problems.

Metric Spaces: Definition

A *metric space* is a pair (V, d) where $d : V \times V \to \mathbb{R}_{\geq 0}$ satisfies:

- **Symmetry:** $d(u, v) = d(v, u)$ for all $u, v \in V$,
- **Triangle inequality:** $d(u, w) \leq d(u, v) + d(v, w)$ for all $u, v, w \in V$,
- **Identity of indiscernibles:** $d(u, v) = 0 \iff u = v$

Metric Spaces: Examples

In $\mathbb{R}^m$ (points $u = (u_1, \ldots, u_m)$, $v = (v_1, \ldots, v_m)$):

- **Euclidean (ℓ_2):** $d(u, v) = \left(\sum_{i=1}^{m} (u_i - v_i)^2\right)^{1/2}$
- **Manhattan (ℓ_1):** $d(u, v) = \sum_{i=1}^{m} |u_i - v_i|$
- **Chebyshev (ℓ_∞):** $d(u, v) = \max_{1 \leq i \leq m} |u_i - v_i|$

Graph metric: given $G = (V, E)$ with non-negative edge lengths, $d(u, v) =$ length of shortest path from u to v. Satisfies all metric axioms.

Embeddings and Distortion

Goal: given a finite metric (V, d), find an *embedding* $f : V \to W$ into a "nicer" target metric (W, ρ) that approximately preserves pairwise distances.

The *distortion* of f is the smallest $c \geq 1$ such that, for some scale factor $\alpha > 0$:

$$\alpha \cdot d(u, v) \; \leq \; \rho(f(u), f(v)) \; \leq \; \alpha \cdot c \cdot d(u, v) \qquad \forall \, u \neq v$$

- $c = 1$: *isometric* (perfect) embedding
- Larger c means more distortion

Two Target Metrics

We study embeddings into two targets arising from vectors $v_1, \ldots, v_n \in \mathbb{R}^k$:

- **Euclidean (ℓ_2):**

$$\rho(i,j) = \|v_i - v_j\|_2$$

Standard Euclidean distance between the image points.

- **Squared Euclidean (negative type):**

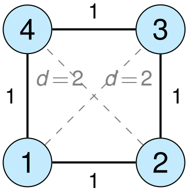
$$\rho(i,j) = \|v_i - v_j\|_2^2$$

Useful for graph partitioning applications (ARV sparsest cut).

For each target: minimizing distortion $\Rightarrow$ **SDP**.

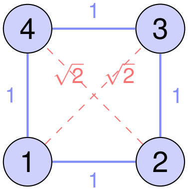
Example: Graph Metric and 2-D Embedding

Graph: 4-cycle C_4, vertices $\{1, 2, 3, 4\}$, unit edge weights. **Embedding** $f : V \to \mathbb{R}^2$: corners of the unit square.



d	1	2	3	4
1	0	1	2	1
2	1	0	1	2
3	2	1	0	1
4	1	2	1	0

$f(1) = (0, 0)$, $f(2) = (1, 0)$, $f(3) = (1, 1)$, $f(4) = (0, 1)$

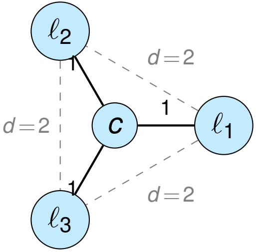


Distortion analysis:

- Adjacent pairs: distortion factor 1
- Diagonal pairs: distortion factor $\sqrt{2}$

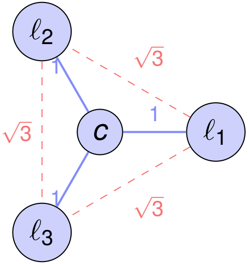
Example: Star Graph $K_{1,3}$ – Less Symmetric Metric

Graph: star $K_{1,3}$: center c, leaves ℓ_1, ℓ_2, ℓ_3. All edges weight 1; leaves are *not* adjacent.



$d(c, \ell_i) = 1; \quad d(\ell_i, \ell_j) = 2$

Embedding: leaves at 120 on unit circle.



Distortion: $2/\sqrt{3}$

Finding best ℓ_2 embedding

Given a finite metric (V, d) (say a graph metric). Find the smallest distortion embedding into ℓ_2

Finding best ℓ_2 embedding

Given a finite metric (V, d) (say a graph metric). Find the smallest distortion embedding into ℓ_2

- NP-Hard if you fix the dimension of embedding (say into one dimension or 2 dimensions etc)
- Can solve it via SDP if dimension is not fixed!

Finding best ℓ_2 embedding: Vector Program

Find vectors $v_1, \ldots, v_n \in \mathbb{R}^n$. Normalize so no pair contracts; minimize expansion. Check if expansion c is feasible and then do binary search for smallest c.

$$d(i,j)^2 \leq \|v_i - v_j\|_2^2 \leq c^2 \cdot d(i,j)^2 \quad \forall\, i \neq j$$
$$v_i \in \mathbb{R}^n$$

All constraints are linear in inner products $v_a \cdot v_b \Rightarrow$ *vector program* $\Rightarrow$ **SDP**.

The ℓ_2 triangle inequality holds *automatically* for any vectors, so no extra constraints are needed.

Finding best ℓ_2 embedding: Vector Program

Find vectors $v_1, \ldots, v_n \in \mathbb{R}^n$ to see if expansion with c is feasible.

$$d(i,j)^2 \leq \|v_i - v_j\|_2^2 \leq c^2 \cdot d(i,j)^2 \quad \forall i \neq j$$
$$v_i \in \mathbb{R}^n$$

In standard form:

$$Y_{ij} \geq d(i,j)^2 \quad \forall i \neq j$$
$$Y_{ij} \leq c^2 \cdot d(i,j)^2 \quad \forall i \neq j$$
$$Y \succeq 0$$

Note that c is a fixed constant that we are trying.

Two Results on ℓ_2-embeddings

Theorem (Bourgain)

Every n-point metric embeds into ℓ_2 with distortion $O(\log n)$.

Hence optimum value of SDP solution will never be worse than $O(\log n)$. There exist metrics for which this amount of distortion is needed.

Dimensionality reduction of Euclidean metrics:

Lemma (Johnson-Lindenstrauss)

Every n-point Euclidean metric (in any number of dimensions) can be embedded into $\mathbb{R}^d$ for $d = O(\log n/\epsilon^2)$ with distortion at most $(1 + \epsilon)$.

Best Squared-ℓ_2 Embedding: Vector Program

Find vectors $v_1, \ldots, v_n$ such that $\|v_i - v_j\|_2^2$ approximates $d(i,j)$ up to factor c:

$$\begin{aligned} \min \quad & c \\ \text{s.t.} \quad & d(i,j) \leq \|v_i - v_j\|_2^2 \leq c \cdot d(i,j) \quad \forall\, i \neq j \\ & \|v_i - v_j\|^2 + \|v_j - v_k\|^2 \geq \|v_i - v_k\|^2 \quad \forall\, i,j,k \\ & v_i \in \mathbb{R}^n \end{aligned}$$

Unlike ℓ_2, squared Euclidean distances do *not* satisfy the triangle inequality automatically — so it must be added explicitly.

Why add the triangle inequality?

For ℓ_2 distances, triangle inequality holds for any vectors. For *squared* ℓ_2, it fails.

Counterexample: collinear points at $0, 1, 2$:

$$\|v_1 - v_3\|^2 = 4 \quad \not\leq \quad 1 + 1 = \|v_1 - v_2\|^2 + \|v_2 - v_3\|^2$$

Lemma

$v_1, v_2, \ldots, v_n$ satisfy the triangle inequality for squared Euclidean lengths iff any three vectors induce an acute angled triangle.

The constraint $\|v_i - v_j\|^2 + \|v_j - v_k\|^2 \geq \|v_i - v_k\|^2$ expands to

$$v_j \cdot v_j - v_i \cdot v_j - v_j \cdot v_k + v_i \cdot v_k \geq 0$$

which is *linear* in inner products $v_a \cdot v_b \Rightarrow$ still a **vector program**.

Minimum Bisection: The Problem

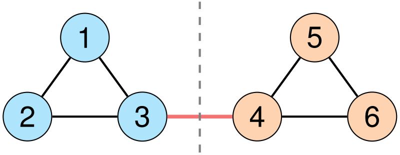
Input: $G = (V, E)$, $|V| = n$ (assume n even).

Goal: find $S \subseteq V$ with $|S| = n/2$, minimize crossing edges $|E(S, \bar{S})|$.

Example graph: two triangles joined by one bridge edge; $n = 6$, seek $|S| = 3$.

Optimal bisection: cost $= 1$

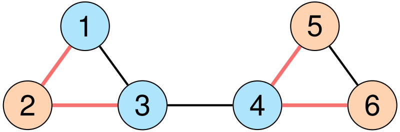
$S = \{1, 2, 3\}$ $\bar{S} = \{4, 5, 6\}$



Only $\{3, 4\}$ crosses the cut.

Non-optimal bisection: cost $= 4$

$S = \{1, 3, 4\}$ $\bar{S} = \{2, 5, 6\}$

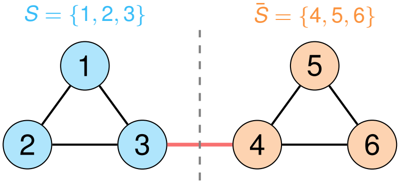


Crossing: $\{1, 2\}, \{2, 3\}, \{4, 5\}, \{4, 6\}$.

Bisection Example: Optimal vs. Non-Optimal

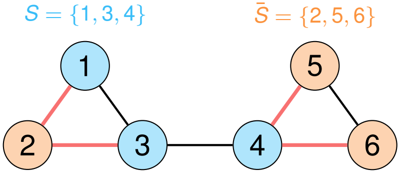
Graph: two triangles joined by one bridge edge; $n = 6$, seek $|S| = 3$.

Optimal bisection: cost $= 1$



Only $\{3, 4\}$ crosses the cut.

Non-optimal bisection: cost $= 4$



Crossing: $\{1, 2\}, \{2, 3\}, \{4, 5\}, \{4, 6\}$.

Minimum Bisection: The Problem

Problem is NP-Hard. Important in practice, many heuristics.

Quadratic program (label each vertex $x_i \in \{-1, +1\}$):

$$\min \frac{1}{2} \sum_{(i,j) \in E} (1 - x_i x_j) \quad \text{s.t.} \quad \sum_{i<j} (x_i - x_j)^2 = n^2$$

- Each term $\frac{1}{2}(1 - x_i x_j)$ is 0 if same side, 1 if opposite
- Balance constraint enforces $|S| = n/2$

Vector Programming Formulation

Relax: replace $x_i \in \{-1, +1\}$ with unit vectors $v_i \in \mathbb{R}^n$, and $x_i x_j \mapsto v_i \cdot v_j$:

$$
\begin{aligned}
\min \quad & \frac{1}{4} \sum_{(i,j) \in E} \|v_i - v_j\|^2 \\
\text{s.t.} \quad & \|v_i\|^2 = 1 \quad \forall i \in V \\
& \sum_{i<j} \|v_i - v_j\|^2 = n^2 \qquad \text{(balance)} \\
& v_i \in \mathbb{R}^n
\end{aligned}
$$

All constraints are **linear in inner products** $v_i \cdot v_j \Rightarrow$ **SDP**.

Vector Programming Formulation: Add triangle inequality

Relax: replace $x_i \in \{-1, +1\}$ with unit vectors $v_i \in \mathbb{R}^n$, and $x_i x_j \mapsto v_i \cdot v_j$:

$$\min \quad \frac{1}{4} \sum_{(i,j) \in E} \|v_i - v_j\|^2$$

$$\text{s.t.} \quad \|v_i\|^2 = 1 \quad \forall i \in V$$

$$\sum_{i<j} \|v_i - v_j\|^2 = n^2 \qquad \text{(balance)}$$

$$\|v_i - v_j\|^2 + \|v_j - v_k\|^2 \geq \|v_i - v_k\|^2 \qquad \text{triangle inequalities}$$

$$v_i \in \mathbb{R}^n$$

We are asking for an embedding into squared Euclidean distance!

Vector Programming Formulation: Add triangle inequality

Relax: replace $x_i \in \{-1, +1\}$ with unit vectors $v_i \in \mathbb{R}^n$, and $x_i x_j \mapsto v_i \cdot v_j$:

$$\begin{aligned}
\min \quad & \frac{1}{4} \sum_{(i,j)\in E} \|v_i - v_j\|^2 \\
\text{s.t.} \quad & \|v_i\|^2 = 1 \quad \forall i \in V \\
& \sum_{i<j} \|v_i - v_j\|^2 = n^2 \qquad \text{(balance)} \\
& \|v_i - v_j\|^2 + \|v_j - v_k\|^2 \geq \|v_i - v_k\|^2 \qquad \text{triangle inequalities} \\
& v_i \in \mathbb{R}^n
\end{aligned}$$

We are asking for an embedding into squared Euclidean distance! Valid because it holds when vectors are in $\{-1, 1\}$.

Vector Programming Formulation: Add triangle inequality

Arora-Rao-Vazirani' 2004 breakthrough work that used squared triangle inequality based SDP to obtain an $O(\sqrt{\log n})$-approximation for sparsest cut, (near)-balanced partition etc.

Improved the previous $O(\log n)$-approximation

Led to many other results and ideas. After Max-Cut this is one of the big successes of SDP in (approximation) algorithms. Both won the Fulkerson award.

Optimizing Non-Convex Functions

Unconstrained optimization: given smooth function $f : \mathbb{R}^d \rightarrow \mathbb{R}$

$$\min_{x \in \mathbb{R}^d} f(x)$$

Saw gradient descent methods to achieve a *local optimum*. For convex functions, local optumum is also a *global optimum*.

What if f is not convex? Can we find global optimum in some structured way even if it takes exponential time?

What are interesting classes of non-convex functions?

Optimizing Polynomials

Univariate polynomials

- $f(x) = 2x^2 + 3x - 10$
- $f(x) = x^3 + 10x^2 + x - 100$
- $f(x) = 10x^4 + x^3 - 53x^2$

Multivariate polynomials

- $f(x) = x^3y^2 + x^2y^1 - 3x^2y^2 + 1$ (degree is 5)
- $f(x) = x^4y^2 + x^2y^4 - 3x^2y^2 + 1$ (degree is 6)

Optimizing Polynomials

Univariate polynomials

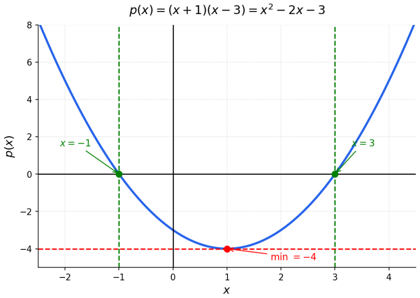
- $f(x) = 2x^2 + 3x - 10$
- $f(x) = x^3 + 10x^2 + x - 100$
- $f(x) = 10x^4 + x^3 - 53x^2$

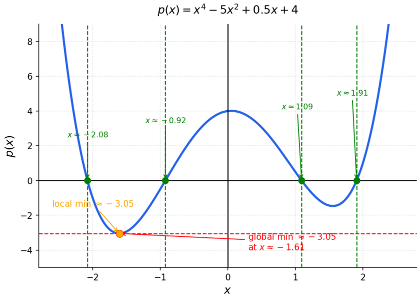
Multivariate polynomials

- $f(x) = x^3y^2 + x^2y^1 - 3x^2y^2 + 1$ (degree is 5)
- $f(x) = x^4y^2 + x^2y^4 - 3x^2y^2 + 1$ (degree is 6)

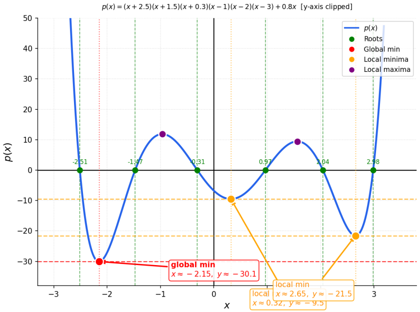
If degree is odd then global minimum is $-\infty$ (why?) and hence main interest for *unconstrained* minimization is even degree polynomials

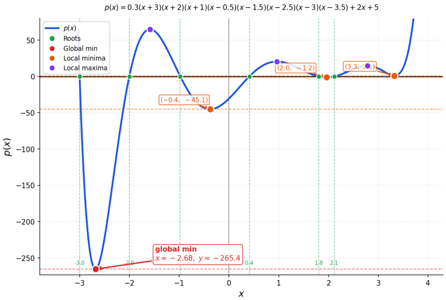
Examples: Univariate Polynomials



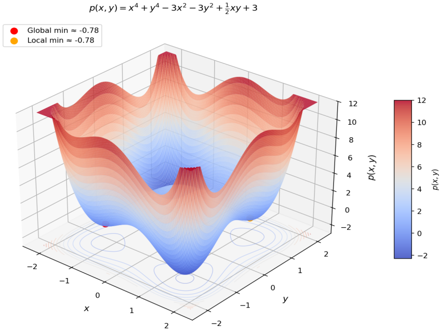


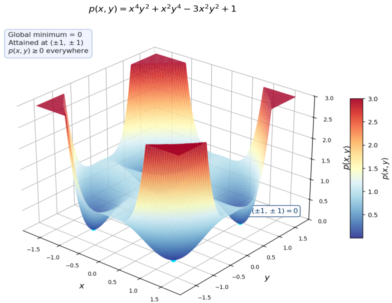
Examples: Multivariate Polynomials





Examples: Multivariate Polynomials





When is a polynomial non-negative?

Question: given $p(x_1, \ldots, x_n)$ of even degree is $p(\mathbf{x}) \geq 0$ for *all* $\mathbf{x} \in \mathbb{R}^n$?

When is a polynomial non-negative?

Question: given $p(x_1, \ldots, x_n)$ of even degree is $p(\mathbf{x}) \geq 0$ for *all* $\mathbf{x} \in \mathbb{R}^n$?

Suppose we can answer the above. Then finding the minimum of p is finding

$$\max \lambda \text{ such that } p(x_1, \ldots, x_n) - \lambda \geq 0$$

When is a polynomial non-negative?

Question: given $p(x_1, \ldots, x_n)$ of even degree is $p(\mathbf{x}) \geq 0$ for *all* $\mathbf{x} \in \mathbb{R}^n$?

Suppose we can answer the above. Then finding the minimum of p is finding

$$\max \lambda \text{ such that } p(x_1, \ldots, x_n) - \lambda \geq 0$$

- Computational. Can we solve it efficiently?
- Structural/mathematical: Is there a nice and easy to understand reason that a polynomial is always non-negative?

When is a polynomial non-negative?

Question: given $p(x_1, \ldots, x_n)$ of even degree is $p(\mathbf{x}) \geq 0$ for *all* $\mathbf{x} \in \mathbb{R}^n$?

Suppose we can answer the above. Then finding the minimum of p is finding

$$\max \lambda \text{ such that } p(x_1, \ldots, x_n) - \lambda \geq 0$$

- Computational. Can we solve it efficiently?
- Structural/mathematical: Is there a nice and easy to understand reason that a polynomial is always non-negative?

Definition

A polynomial p is a sum of squares (SOS) polynomial if there exist polynomials $q_1, q_2, \ldots, q_k$ such that $p(\mathbf{x}) = \sum_{i=1}^{k} q_i(\mathbf{x})^2$.

When is a polynomial non-negative?

Question: given $p(x_1, \ldots, x_n)$ of even degree is $p(\mathbf{x}) \geq 0$ for *all* $\mathbf{x} \in \mathbb{R}^n$?

Suppose we can answer the above. Then finding the minimum of p is finding

$$\max \lambda \text{ such that } p(x_1, \ldots, x_n) - \lambda \geq 0$$

- Computational. Can we solve it efficiently?
- Structural/mathematical: Is there a nice and easy to understand reason that a polynomial is always non-negative?

Definition

A polynomial p is a sum of squares (SOS) polynomial if there exist polynomials $q_1, q_2, \ldots, q_k$ such that $p(\mathbf{x}) = \sum_{i=1}^{k} q_i(\mathbf{x})^2$.

Clearly SOS $\Rightarrow$ nonneg. Is the converse true?

Quadratic Case in One Variable

Consider $p(x) = ax^2 + bx + c$ with $a > 0$.

When is $p(x) \geq 0$ for all $x \in \mathbb{R}$?

Complete the square:

$$p(x) = a\left(x + \frac{b}{2a}\right)^2 + \left(c - \frac{b^2}{4a}\right)$$

So $p \geq 0$ iff $b^2 - 4ac \leq 0$ (discriminant condition).

When $p \geq 0$, it is explicitly a sum of (at most two) squares:

$$p(x) = \left(\sqrt{a}\,x + \frac{b}{2\sqrt{a}}\right)^2 + \left(\sqrt{c - \frac{b^2}{4a}}\right)^2$$

Conclusion: for univariate quadratics, $p \geq 0 \iff p$ is SOS.

Univariate Polynomials of Higher Degree

Consider $p(x)$ of even degree $2d$ in one variable.

Theorem

A univariate polynomial $p(x)$ of even degree is nonnegative for all $x \in \mathbb{R}$ if and only if it is a sum of squares of polynomials.

Proof sketch: Over $\mathbb{C}$, factor $p = c\prod_k(x - r_k)$. Since p has real coefficients, complex roots come in conjugate pairs $(x - \alpha)(x - \bar{\alpha}) = (x - a)^2 + b^2$. Real roots of even multiplicity contribute $(x - r)^{2m}$. $p \geq 0$ forces all real roots to have even multiplicity, so p is a product of nonneg quadratics — hence a sum of squares.

So far: nonneg $=$ SOS for all univariate polynomials. Does this extend to multiple variables?

Multivariate Case

Do nonneg multivariate polynomials have to be SOS?

No! Hilbert proved in 1888 that nonneg $\neq$ SOS in general. Specifically, in all cases *except*:

The three cases where nonneg = SOS (Hilbert, 1888)

(i) univariate ($n = 1$), any even degree (ii) quadratics ($\deg = 2$), any n (iii) quartics ($\deg = 4$), $n = 2$

In all other cases ($n \geq 2$, $\deg \geq 4$ except the case above), there exist nonneg polynomials that are *not* SOS.

Hilbert's proof was non-constructive. An explicit example had to wait until **1967**.

The Motzkin Polynomial

The first *explicit* nonneg polynomial that is not SOS, due to **Motzkin (1967)**:

$$M(x, y) = x^4y^2 + x^2y^4 - 3x^2y^2 + 1$$

$M \geq 0$ **for all** (x, y)**:** Apply AM-GM inequality to three terms:

$$\frac{x^4y^2}{3} + \frac{x^2y^4}{3} + \frac{1}{3} \geq \left(x^4y^2 \cdot x^2y^4 \cdot 1\right)^{1/3} = x^2y^2$$

So $x^4y^2 + x^2y^4 + 1 \geq 3x^2y^2$, hence $M \geq 0$.

M **is not SOS:** we will see why shortly.

Is a polynomial SOS?

Definition

A polynomial p is a sum of squares (SOS) polynomial if there exist polynomials $q_1, q_2, \ldots, q_k$ such that $p(\mathbf{x}) = \sum_{i=1}^{k} q_k(\mathbf{x})^2$.

Question: given $p(x_1, \ldots, x_n)$ of even degree is $p(\mathbf{x})$ an SOS polynomial?

Is a polynomial SOS?

Definition

A polynomial p is a sum of squares (SOS) polynomial if there exist polynomials $q_1, q_2, \ldots, q_k$ such that $p(\mathbf{x}) = \sum_{i=1}^{k} q_k(\mathbf{x})^2$.

Question: given $p(x_1, \ldots, x_n)$ of even degree is $p(\mathbf{x})$ an SOS polynomial?

Efficient algorithm via SDP!

Checking whether a polynomial is SOS

Let $p(x)$ be a univariate polynomial of degree $2d$.

$$p(x) = p_0 + p_1 x + p_2 x^2 + \ldots + p_{2d} x^{2d}$$

View p as a vector of size $2d + 1$.

Let $z(x) = [1 \quad x \quad x^2 \quad \ldots \quad x^d]^T$ be a vector of monomials ($(d+1)$ dimensional)

Theorem

$p(x)$ is SOS iff there exists a psd matrix $Q \in M_{d+1}$ such that $z(x)^T Q z(x) = p(x)$.

Degree-4 Univariate SOS

Given: $p(x) = p_0 + p_1 x + p_2 x^2 + p_3 x^3 + p_4 x^4$. Use monomial vector $z = [1,\ x,\ x^2]^T$ (degree ≤ 2).

Step 1. Expand $p(x) = z^T Q\, z$ for an unknown 3×3 symmetric Q:

$$z^T Q z = Q_{11} \cdot 1 + 2Q_{12} \cdot x + (Q_{22} + 2Q_{13}) \cdot x^2 + 2Q_{23} \cdot x^3 + Q_{33} \cdot x^4$$

Step 2. Match each monomial coefficient with p:

$$Q_{11} = p_0, 2Q_{12} = p_1, (Q_{22} + 2Q_{13}) = p_2, 2Q_{23} = p_3, Q_{33} = p_4$$

Step 3. SDP: find any $Q \succeq 0$ satisfying the 5 equations. If feasible: Cholesky $Q = LL^T$ gives $p = \sum_k (l_k^T z)^2$.

Example: $p(x) = x^4 - 2x^3 + 3x^2 - 2x + 1$

Coefficients: $p_0 = 1,\ p_1 = -2,\ p_2 = 3,\ p_3 = -2,\ p_4 = 1$.

Constraints (Step 2):

$$Q_{11} = 1, \quad Q_{12} = -1, \quad Q_{23} = -1, \quad Q_{33} = 1, \quad Q_{22} = 3 - 2Q_{13}$$

Set $Q_{13} = q$ (free parameter). The full matrix is:

$$Q(q) = \begin{pmatrix} 1 & -1 & q \\ -1 & 3 - 2q & -1 \\ q & -1 & 1 \end{pmatrix}$$

PSD conditions Want $Q \succeq 0$ and it works for any $q \in [-\frac{1}{2}, 1]$. Choose $q = 1$.

SOS decomposition: $(1 - x + x^2)^2 = 1 - 2x + 3x^2 - 2x^3 + x^4$

Checking whether a polynomial is SOS

Let $p(x_1, x_2, \ldots, x_n)$ be a multivariate polynomial with n variables and degree $2d$.

Each monomial is $x_1^{a_1} x^{a_2} \ldots x_n^{a_n}$ s.t all $a_i \geq 0$ and $0 \leq \sum_i a_i \leq 2d$. Hence, $\binom{n+2d}{2d}$

Let $z(\mathbf{x})$ be the vector of all monomials of degree $\leq d$:

$$z(\mathbf{x}) = \left[1,\ x_1,\ \ldots,\ x_n,\ x_1^2,\ x_1 x_2,\ \ldots,\ x_n^d\right]^T$$

The dimension of z is $s = \binom{n+d}{d}$.

Theorem

p is SOS if and only if there exists $Q \in \mathbb{R}^{s \times s}$ with $Q \succeq 0$ such that

$$p(\mathbf{x}) = z(\mathbf{x})^T Q\, z(\mathbf{x})$$

Example: Motzkin Polynomial via SDP

$M(x, y) = x^4y^2 + x^2y^4 - 3x^2y^2 + 1$ has degree 6.

Monomial vector for degree ≤ 3 in 2 variables:

$$z = [1,\ x,\ y,\ x^2,\ xy,\ y^2,\ x^3,\ x^2y,\ xy^2,\ y^3]^T \quad (s = 10)$$

So Q is 10×10.

Matching coefficients: e.g. coefficient of x^4y^2 in z^TQz must equal 1. Coefficient of x^3 must equal 0. And similarly for other coefficients.

Result: the linear system on the entries of Q, together with $Q \succeq 0$, is *infeasible* $\Rightarrow$ M is **not SOS**.

Proof of Theorem

Theorem

p is SOS iff there exists $Q \in \mathbb{R}^{s \times s}$ with $Q \succeq 0$ such that $p(\mathbf{x}) = z(\mathbf{x})^T Q\, z(\mathbf{x})$.

- Suppose there is $Q \succeq 0$ such that $z(x)^T Qz(x) = p(x)$.
- Write $Q = W^T W$ and $w_1, w_2, \ldots, w_s$ be *rows* of W.
- $z^T Qz = \sum_{i=1}^{s} (w_i \cdot z(x))^2 = p(x)$.
- $q_i(x) = w_i \cdot z(x)$ is a polynomial of degree d. Thus $p(x) = \sum_i q_i(x)^2$.

Proof of Theorem

Theorem

p is SOS iff there exists $Q \in \mathbb{R}^{s \times s}$ with $Q \succeq 0$ such that $p(\mathbf{x}) = z(\mathbf{x})^T Q\, z(\mathbf{x})$.

- Suppose there is $Q \succeq 0$ such that $z(x)^T Q z(x) = p(x)$.
- Write $Q = W^T W$ and $w_1, w_2, \ldots, w_s$ be *rows* of W.
- $z^T Q z = \sum_{i=1}^{s} (w_i \cdot z(x))^2 = p(x)$.
- $q_i(x) = w_i \cdot z(x)$ is a polynomial of degree d. Thus $p(x) = \sum_i q_i(x)^2$.

Other direction.

- Suppose $p(x) = \sum_{i=1}^{k} q_i(x)^2$. Each q_i is of degree $\leq d$.
- Write $q_i(x) = c_i^T \cdot z(x)$ for some coefficient vector c_i
- Let Q_i be psd matrix $c_i c_i^T$ (outer product)
- Then $p(x) = \sum_{i=1}^{k} z(x)^T Q_i z(x) = z(x)^T (\sum_i Q_i) z(x)$. And $Q = \sum_i Q_i$ is psd.

Checking SOS via SDP

Checking whether p is SOS reduces to an SDP!

Expand $z(\mathbf{x})^T Q\, z(\mathbf{x})$ and match each monomial coefficient with p:

- Each monomial x^γ in p gives one *linear equation* in the entries of Q:

$$\sum_{\alpha+\beta=\gamma} Q_{\alpha\beta} = p_\gamma \quad \forall\, \gamma$$

- Seek Q satisfying all these linear constraints *and* $Q \succeq 0$

$$\begin{array}{rl} \text{Find} & Q \in \mathbb{R}^{s\times s} \\ \text{s.t.} & \sum_{\alpha+\beta=\gamma} Q_{\alpha\beta} = p_\gamma \quad \forall\, \gamma \\ & Q \succeq 0 \end{array}$$

Feasible $\Rightarrow$ p is SOS (Cholesky of Q gives decomposition).
Infeasible $\Rightarrow$ p is not SOS (Motzkin polynomial!).

Summary so far

- A multivariate polynomial p can be non-negative but may not be SOS
- Given an even degree multivariate polynomial p one can check whether it is SOS via SDP.

Question: Is there a way to certify that $p \geq 0$?

Hilbert's 17th Problem

Even though nonneg $\not\Rightarrow$ SOS, Hilbert asked a related question:

Hilbert's 17th Problem (1900)

If $p(x_1, \ldots, x_n) \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$, can p be written as a *sum of squares of rational functions*?

$$p(\mathbf{x}) = \sum_k \left(\frac{q_k(\mathbf{x})}{r_k(\mathbf{x})} \right)^2$$

Hilbert's 17th Problem

Even though nonneg $\not\Rightarrow$ SOS, Hilbert asked a related question:

Hilbert's 17th Problem (1900)

If $p(x_1, \ldots, x_n) \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$, can p be written as a *sum of squares of rational functions*?

$$p(\mathbf{x}) = \sum_k \left(\frac{q_k(\mathbf{x})}{r_k(\mathbf{x})} \right)^2$$

Theorem (Artin, 1927)

Yes. Every nonnegative polynomial $p \in \mathbb{R}[x_1, \ldots, x_n]$ *is a sum of squares of rational functions.*

Artin's proof was non-constructive (used model theory / real algebra). Finding an explicit representation and doing so algorithmically is where **SDP** enters.

The Motzkin Polynomial Again

$$M(x, y) = x^4y^2 + x^2y^4 - 3x^2y^2 + 1$$

M **is not SOS but non-negative**

Artin representation: $M(x, y) \cdot (x^2 + y^2 + 1)$ is SOS!

This prove that $M(x, y) \geq 0$ since $(x^2 + y^2 + 1)$ is SOS.

Gives an explicit ratio-of-SOS representation consistent with Artin's theorem.

The Motzkin Polynomial Again

$$M(x, y) = x^4y^2 + x^2y^4 - 3x^2y^2 + 1$$

M **is not SOS but non-negative**

Artin representation: $M(x, y) \cdot (x^2 + y^2 + 1)$ is SOS!

This prove that $M(x, y) \geq 0$ since $(x^2 + y^2 + 1)$ is SOS.

Gives an explicit ratio-of-SOS representation consistent with Artin's theorem.

How did we know to try $(x^2 + y^2 + 1)$? Not obvious. Need additional ideas including Laserre hierarchy etc.

The Motzkin Polynomial Again

$$M(x, y) = x^4y^2 + x^2y^4 - 3x^2y^2 + 1$$

M **is not SOS but non-negative**

Artin representation: $M(x, y) \cdot (x^2 + y^2 + 1)$ is SOS!

This prove that $M(x, y) \geq 0$ since $(x^2 + y^2 + 1)$ is SOS.

Gives an explicit ratio-of-SOS representation consistent with Artin's theorem.

How did we know to try $(x^2 + y^2 + 1)$? Not obvious. Need additional ideas including Laserre hierarchy etc.

Interesting fact: $M(x, y) - \gamma$ is not SOS for any γ!

Summary: SOS, Hilbert, and SDP

Question	Answer / Tool
$p \geq 0$ everywhere?	NP-hard for $\deg \geq 4$, $n \geq 2$
p SOS?	Feasibility SDP (Gram matrix)
nonneg $\Rightarrow$ SOS?	Yes for $n = 1$; $n = 2, \deg = 4$; $\deg = 2$
	No in general (Motzkin 1967)
nonneg $\Rightarrow$ ratio of SOS?	Yes (Artin 1927; Hilbert's 17th)
Lower bound on $\min_K p$?	SOS hierarchy SDP (Lasserre)

Key message: SDP makes Artin's theorem constructive — the SOS hierarchy provides a systematic, algorithmically tractable sequence of relaxations that converge to the true polynomial optimum.

Polynomial Optimization: The Problem

Setting: minimize a polynomial subject to polynomial constraints.

$$p^* = \min_{\mathbf{x}} p(\mathbf{x}) \qquad \text{s.t.} \quad g_i(\mathbf{x}) \geq 0, \quad i = 1, \ldots, m$$

where $p, g_1, \ldots, g_m \in \mathbb{R}[x_1, \ldots, x_n]$ are polynomials.

The feasible set

$$K = \{ \mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) \geq 0 \; \forall i \}$$

is called a *semialgebraic set*.

Examples of Feasible Sets

Many important feasible sets are semialgebraic:

- **Unconstrained:** $K = \mathbb{R}^n$ (no constraints, $m = 0$)
- **Box:** $[-1, 1]^n$ via $g_i(\mathbf{x}) = 1 - x_i^2 \geq 0$
- **Euclidean ball:** $\|\mathbf{x}\| \leq 1$ via $g_1(\mathbf{x}) = 1 - \|\mathbf{x}\|^2 \geq 0$
- **Polytope:** $\{A\mathbf{x} \leq \mathbf{b}\}$ via $g_i(\mathbf{x}) = b_i - a_i^T\mathbf{x} \geq 0$ (linear g_i)
- **Binary variables:** $x_i \in \{-1, +1\}$ via $g_i = 1 - x_i^2 \geq 0$ and $x_i^2 - 1 \geq 0$, which together force $x_i^2 = 1$

Hardness and Strategy

Hardness: Polynomial optimization is NP-hard in general.

It subsumes:

- Integer programming (binary variables are semialgebraic)
- Quadratic programming with combinatorial constraints
- Many other NP-hard problems

Strategy: Rather than solving exactly, compute the *best provable lower bound* via an algebraic certificate.

Certifying lower bounds is tractable via SDP.

Reformulation: Certifying Lower Bounds

Every lower bound $\lambda \leq p^*$ corresponds to $p(\mathbf{x}) - \lambda \geq 0$ on K.

So the optimum equals the *best certifiable lower bound*:

$$p^* = \sup\{\, \lambda : p(\mathbf{x}) - \lambda \geq 0 \ \text{ for all } \ \mathbf{x} \in K \,\}$$

Challenge: verifying $p(\mathbf{x}) - \lambda \geq 0$ on K is NP-hard.

Key insight: replace the semantic condition "≥ 0 on K" with a *syntactic algebraic certificate* that can be checked efficiently.

The SOS Certificate

Certificate structure: prove $p - \lambda \geq 0$ on K by writing it as

$$p(\mathbf{x}) - \lambda = \underbrace{\sigma_0(\mathbf{x})}_{\geq 0 \text{ everywhere}} + \sum_{i=1}^{m} \underbrace{\sigma_i(\mathbf{x})}_{\geq 0 \text{ everywhere}} \cdot \underbrace{g_i(\mathbf{x})}_{\geq 0 \text{ on } K}$$

where each $\sigma_i(\mathbf{x})$ is a **sum of squares (SOS)**.

Why valid: each term $\sigma_i \cdot g_i \geq 0$ on K, so the RHS ≥ 0 on K, confirming $\lambda \leq p^*$.

Optimization: maximize λ subject to the existence of such a certificate $\Rightarrow$ an **SDP**.

Gram Matrices for Each Multiplier

Each SOS multiplier $\sigma_i(\mathbf{x})$ is represented via a Gram matrix.

Let $z_r(\mathbf{x})$ be the vector of all monomials of degree $\leq r$. Then:

$$\sigma_i(\mathbf{x}) \;=\; z_r(\mathbf{x})^T\, Q_i\, z_r(\mathbf{x}), \qquad Q_i \succeq 0$$

Degree bound at level r: require $\deg(\sigma_i\, g_i) \leq 2r$ for all i, giving

$$\deg(\sigma_i) \;\leq\; 2r - \deg(g_i)$$

Expanding each $\sigma_i\, g_i$ and matching monomial coefficients with $p - \lambda$ yields *linear equations* in λ and the entries of $Q_0, \ldots, Q_m$.

The Level-r SDP

Combining the coefficient equations and PSD constraints gives an SDP:

Level-r SDP Relaxation (Lasserre Hierarchy)

$$\begin{aligned} \max \quad & \lambda \\ \text{s.t.} \quad & \text{coefficient equations: } p - \lambda = \sigma_0 + \textstyle\sum_i \sigma_i g_i \\ & Q_0, Q_1, \ldots, Q_m \succeq 0 \end{aligned}$$

Matrix sizes:

- Q_0 has size $\binom{n+r}{r} \times \binom{n+r}{r}$
- Each Q_i is smaller (degree of g_i reduces the size)
- Solvable in polynomial time in the SDP size

Putinar's Positivstellensatz (1993)

Question: is a certificate always achievable?

Not in general. Under a mild condition, every *strictly positive* polynomial has one.

Theorem (Putinar, 1993)

Suppose K satisfies the Archimedean condition: $R - \|\mathbf{x}\|^2$ has an SOS-multiplier certificate for some $R > 0$ (i.e. the constraint system certifies K is bounded). Then every polynomial $p > 0$ on K admits a certificate

$$p \;=\; \sigma_0 + \sum_{i=1}^{m} \sigma_i \, g_i, \quad \sigma_i \text{ SOS.}$$

Practical note: if K is bounded, add the redundant constraint $g_{m+1} = R - \|\mathbf{x}\|^2 \geq 0$ to ensure Archimedean.

The Lasserre Hierarchy

Putinar's theorem guarantees that the level-r SDP bounds converge to p^*:

- Level-r SDP produces lower bound $\lambda_r \leq p^*$
- Bounds are **non-decreasing**: $\lambda_1 \leq \lambda_2 \leq \cdots \leq p^*$
- **Convergence:** $\lambda_r \to p^*$ as $r \to \infty$ (finite convergence when $p > 0$ on K)
- In practice, often exact at level $r = 1$ or $r = 2$