Convex Optimization and Applications

9 - The Lasserre Hierarchy for Polynomial optimization

Guillaume Sagnol

Technische Universität Berlin

Outline

Polynomial Optimization

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

where $p, g_1, \ldots, g_m$ are polynomials.

Polynomial Optimization

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

where $p, g_1, \ldots, g_m$ are polynomials.

- Extremely versatile

Polynomial Optimization

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

where $p, g_1, \ldots, g_m$ are polynomials.

- Extremely versatile
- NP-hard (e.g., INTEGER PROGRAMMING is a special case, consider the constraints $x_i^2 = x_i$)

Polynomial Optimization

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\textbf{minimize}} \quad p(x) \qquad \text{(P)}$$
$$s.t. \quad g_i(x) \geq 0, \quad \forall i \in [m],$$

where $p, g_1, \ldots, g_m$ are polynomials.

- Extremely versatile
- NP-hard (e.g., INTEGER PROGRAMMING is a special case, consider the constraints $x_i^2 = x_i$)
- Equivalent formulation

$$\max\{\lambda : p(\mathbf{x}) - \lambda \geq 0, \ \forall \mathbf{x} \in K\}$$

where $K = \{x : g_i(x) \geq 0, \forall i \in [m]\}$ is the feasible set.

Polynomial Optimization

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\textbf{minimize}} \quad p(\mathbf{x}) \qquad \text{(P)}$$
$$s.t. \quad g_i(\mathbf{x}) \geq 0, \quad \forall i \in [m],$$

where $p, g_1, \ldots, g_m$ are polynomials.

- Extremely versatile
- NP-hard (e.g., INTEGER PROGRAMMING is a special case, consider the constraints $x_i^2 = x_i$)
- Equivalent formulation

$$\max\{\lambda : p(\mathbf{x}) - \lambda \geq 0,\ \forall \mathbf{x} \in K\}$$

  where $K = \{\mathbf{x} : g_i(\mathbf{x}) \geq 0, \forall i \in [m]\}$ is the feasible set.
- So checking the nonnegativity of a polynomial over a *semi-algebraic* set K is NP-hard.

From Copositive Programming to Nonnegative Polynomials

$$X \succeq_{\mathcal{C}_n} 0 \iff \boldsymbol{u}^T X \boldsymbol{u} \geq 0, \quad \forall \boldsymbol{u} \geq \boldsymbol{0}$$

$\iff$ The polynomial $\boldsymbol{u} \mapsto \sum_{ij} X_{ij} u_i u_j$ is nonnegative over $\mathbb{R}_+^n$.

$\iff$ The polynomial $\boldsymbol{u} \mapsto \sum_{ij} X_{ij} u_i^2 u_j^2$ is nonnegative over $\mathbb{R}^n$.

Therefore, testing the nonnegativity of a polynomial of degree 4 is NP hard.

Outline

Univariate polynomials

We identify polynomials of degree d with their vector of coefficients:

$$p(x) = \sum_{i=0}^{d} p_i x^i = \boldsymbol{p}^T \boldsymbol{v}(x), \qquad \text{with } \boldsymbol{v}(x) := [1, x, x^2, \dots, x^d]^T.$$

Definition

We define the cone of nonnegative polynomials of degree $\leq 2d$

$$\mathcal{P}_{2d}^{+} := \{\boldsymbol{p} \in \mathbb{R}^{2d+1} : \sum_{i=0}^{2d} p_i\, x^i \geq 0, \quad \forall x \in \mathbb{R}\} \subset \mathbb{R}^{2d+1}$$

and the cone of *sum of squares*

$$\mathcal{P}_{2d}^{\text{sos}} := \{\boldsymbol{p} \in \mathbb{R}^{2d+1} : \exists \boldsymbol{q}_1, \dots, \boldsymbol{q}_r \in \mathbb{R}^{d+1}, \sum_{i=0}^{2d} p_i\, x^i = \sum_{j=1}^{r} (\boldsymbol{q}_j^T \boldsymbol{v}(x))^2\}$$

These two cones are proper, and clearly $\mathcal{P}_{2d}^{\text{sos}} \subseteq \mathcal{P}_{2d}^{+}$.

Univariate Polynomials

In fact, for univariate polynomials, nonnegative polynomials and sums of squares are the same thing!

Theorem

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

Univariate Polynomials

In fact, for univariate polynomials, nonnegative polynomials and sums of squares are the same thing!

Theorem

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

Proof:

- Let $p \in \mathcal{P}_{2d}^{+}$. Then $p(x) = p_{2d} \prod_{i=1}^{2d} (x - a_i)$ for some $a_1, \ldots, a_{2d} \in \mathbb{C}$.

Univariate Polynomials

In fact, for univariate polynomials, nonnegative polynomials and sums of squares are the same thing!

Theorem

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

Proof:

- Let $p \in \mathcal{P}_{2d}^{+}$. Then $p(x) = p_{2d} \prod_{i=1}^{2d} (x - a_i)$ for some $a_1, \ldots, a_{2d} \in \mathbb{C}$.
- The complex roots come by conjugate pairs, and the real roots are of even multiplicity (because p does not change sign over $\mathbb{R}$).

Univariate Polynomials

In fact, for univariate polynomials, nonnegative polynomials and sums of squares are the same thing!

Theorem

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

Proof:

- Let $p \in \mathcal{P}_{2d}^{+}$. Then $p(x) = p_{2d} \prod_{i=1}^{2d} (x - a_i)$ for some $a_1, \ldots, a_{2d} \in \mathbb{C}$.
- The complex roots come by conjugate pairs, and the real roots are of even multiplicity (because p does not change sign over $\mathbb{R}$).
- So after reordering the roots, $p(x) = p_{2d} \prod_{i=1}^{d} (x - a_i)(x - \overline{a_i}) = |q(z)|^2$,

  where $q(z) = \sqrt{p_{2d}} \prod_{i=1}^{d} (z - a_i) \in \mathbb{C}_d[z]$

Univariate Polynomials

In fact, for univariate polynomials, nonnegative polynomials and sums of squares are the same thing!

Theorem

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

Proof:

- Let $p \in \mathcal{P}_{2d}^{+}$. Then $p(x) = p_{2d} \prod_{i=1}^{2d} (x - a_i)$ for some $a_1, \ldots, a_{2d} \in \mathbb{C}$.
- The complex roots come by conjugate pairs, and the real roots are of even multiplicity (because p does not change sign over $\mathbb{R}$).
- So after reordering the roots, $p(x) = p_{2d} \prod_{i=1}^{d} (x - a_i)(x - \overline{a_i}) = |q(z)|^2$,

  where $q(z) = \sqrt{p_{2d}} \prod_{i=1}^{d} (z - a_i) \in \mathbb{C}_d[z]$
- Finally, $p = (\Re e\ q)^2 + (\Im m\ q)^2$.

Univariate Polynomials

- We have: $\mathcal{P}_{2d}^{\text{sos}} = \mathcal{P}_{2d}^{+}$.
- And there is a nice characterization $\mathcal{P}_{2d}^{\text{sos}}$ of based on LMIs:

Theorem

Denote by $s_k(M)$ the sum of the kth antidiagonal of $M \in \mathbb{S}^d$:

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

Then, $\boldsymbol{p} \in \mathbb{R}_d[x]$ is a sum of square iff

$$\exists M \succeq 0 : \quad s_k(M) = p_k, \forall k \in \{0,\ldots,2d\}.$$

Univariate Polynomials

- We have: $\mathcal{P}_{2d}^{\text{sos}} = \mathcal{P}_{2d}^{+}$.
- And there is a nice characterization $\mathcal{P}_{2d}^{\text{sos}}$ of based on LMIs:

Theorem

Denote by $s_k(M)$ the sum of the kth antidiagonal of $M \in \mathbb{S}^d$:

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

Then, $\boldsymbol{p} \in \mathbb{R}_d[x]$ is a sum of square iff

$$\exists M \succeq 0 : \quad s_k(M) = p_k, \forall k \in \{0,\ldots,2d\}.$$

Example:

$$p(x) = 5 - 14x + 2x^2 + 12x^3 + 4x^4 = [1,x,x^2] \overbrace{\begin{pmatrix} 5 & -7 & -4 \\ -7 & 10 & 6 \\ -4 & 6 & 4 \end{pmatrix}}^{M} [1,x,x^2]^T$$

Univariate Polynomials

- We have: $\mathcal{P}_{2d}^{\text{sos}} = \mathcal{P}_{2d}^{+}$.
- And there is a nice characterization $\mathcal{P}_{2d}^{\text{sos}}$ of based on LMIs:

Theorem

Denote by $s_k(M)$ the sum of the kth antidiagonal of $M \in \mathbb{S}^d$:

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

Then, $\boldsymbol{p} \in \mathbb{R}_d[x]$ is a sum of square iff

$$\exists M \succeq 0: \quad s_k(M) = p_k, \forall k \in \{0, \ldots, 2d\}.$$

Example:

$$p(x) = 5 - 14x + 2x^2 + 12x^3 + 4x^4 = [1, x, x^2] \overbrace{\begin{pmatrix} 5 & -7 & -4 \\ -7 & 10 & 6 \\ -4 & 6 & 4 \end{pmatrix}}^{M} [1, x, x^2]^T$$

$M \succeq 0$, so we can decompose $M = HH^T$, where $H^T = \begin{pmatrix} 1 & -1 & 0 \\ -2 & 3 & 2 \end{pmatrix}$.

Univariate Polynomials

- We have: $\mathcal{P}_{2d}^{\text{sos}} = \mathcal{P}_{2d}^{+}$.
- And there is a nice characterization $\mathcal{P}_{2d}^{\text{sos}}$ of based on LMIs:

Theorem

Denote by $s_k(M)$ the sum of the kth antidiagonal of $M \in \mathbb{S}^d$:

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

Then, $\boldsymbol{p} \in \mathbb{R}_d[x]$ is a sum of square iff

$$\exists M \succeq 0: \quad s_k(M) = p_k, \forall k \in \{0, \ldots, 2d\}.$$

Example:

$$p(x) = 5 - 14x + 2x^2 + 12x^3 + 4x^4 = [1, x, x^2] \overbrace{\begin{pmatrix} 5 & -7 & -4 \\ -7 & 10 & 6 \\ -4 & 6 & 4 \end{pmatrix}}^{M} [1, x, x^2]^T$$

$M \succeq 0$, so we can decompose $M = HH^T$, where $H^T = \begin{pmatrix} 1 & -1 & 0 \\ -2 & 3 & 2 \end{pmatrix}$.

So $p(x) = \|H^T [1, x, x^2]^T\|^2 = (1-x)^2 + (-2 + 3x + 2x^2)^2$.

Univariate Polynomials

- We have: $\mathcal{P}_{2d}^{\text{sos}} = \mathcal{P}_{2d}^{+}$.
- And there is a nice characterization $\mathcal{P}_{2d}^{\text{sos}}$ of based on LMIs:

Theorem

Denote by $s_k(M)$ the sum of the kth antidiagonal of $M \in \mathbb{S}^d$:

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

Then, $\boldsymbol{p} \in \mathbb{R}_d[x]$ is a sum of square iff

$$\exists M \succeq 0 : \quad s_k(M) = p_k, \forall k \in \{0,\ldots,2d\}.$$

Corollary 1

We can check if $p \in \mathbb{R}_{2d}[x]$ is a nonnegative polynomial by using an SDP solver.

Univariate Polynomials

- We have: $\mathcal{P}_{2d}^{\text{sos}} = \mathcal{P}_{2d}^{+}$.
- And there is a nice characterization $\mathcal{P}_{2d}^{\text{sos}}$ of based on LMIs:

Theorem

Denote by $s_k(M)$ the sum of the kth antidiagonal of $M \in \mathbb{S}^d$:

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

Then, $\boldsymbol{p} \in \mathbb{R}_d[x]$ is a sum of square iff

$$\exists M \succeq 0 : \quad s_k(M) = p_k, \forall k \in \{0, \ldots, 2d\}.$$

Corollary 2

Let $p \in \mathbb{R}_{2d}[x]$. We can reformulate the problem $\inf_{x \in \mathbb{R}} p(x)$ as an SDP:

$$\inf_{x \in \mathbb{R}} p(x) = \sup \{t : \ (p - t) \in P_{2d}^{sos}\}$$

Outline

Multivariate Polynomials and Sum-of-Squares

- Use multi-indices:

$$p(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha} \in \Delta(n,d)} p_{\boldsymbol{\alpha}}\, \boldsymbol{x}^{\boldsymbol{\alpha}}, \quad \text{where} \quad \boldsymbol{x}^{\boldsymbol{\alpha}} := x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n},$$

  where $\Delta(n,d) \subset \mathbb{Z}^n_{\geq 0}$ is the set of multi-indices with sum $\leq d$.
- Cardinality of $\Delta(n,d)$ is $s(n,d) := \binom{n+d}{d}$.

Multivariate Polynomials and Sum-of-Squares

- Use multi-indices:

$$p(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha} \in \Delta(n,d)} p_{\boldsymbol{\alpha}}\, \boldsymbol{x}^{\boldsymbol{\alpha}}, \quad \text{where} \quad \boldsymbol{x}^{\boldsymbol{\alpha}} := x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n},$$

  where $\Delta(n,d) \subset \mathbb{Z}^n_{\geq 0}$ is the set of multi-indices with sum $\leq d$.
- Cardinality of $\Delta(n,d)$ is $s(n,d) := \binom{n+d}{d}$.

Example

Let $n = 2, d = 3, \boldsymbol{x} = [y, z]^T$. The set of multi-indices is $\Delta(2,3) = [00, 10, 01, 20, 11, 02, 30, 21, 12, 03]$, corresponding to the monomials $[1, y, z, y^2, yz, z^2, y^3, y^2 zyz^2, z^3]$.
The polynomial $p(\boldsymbol{x}) = y^3 - 3yz + 2$ is represented by the vector $\boldsymbol{p}$ such that $p_{30} = 1, p_{11} = -3, p_{00} = 2$, and $p_{\boldsymbol{\alpha}} = 0$ for all others α's.

Multivariate Polynomials and Sum-of-Squares

- Use multi-indices:

$$p(x) = \sum_{\boldsymbol{\alpha} \in \Delta(n,d)} p_{\boldsymbol{\alpha}} \, \boldsymbol{x}^{\boldsymbol{\alpha}}, \quad \text{where} \quad \boldsymbol{x}^{\boldsymbol{\alpha}} := x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n},$$

  where $\Delta(n, d) \subset \mathbb{Z}^n_{\geq 0}$ is the set of multi-indices with sum $\leq d$.

- Cardinality of $\Delta(n, d)$ is $s(n, d) := \binom{n+d}{d}$.

- Define cone of nonnegative polynomials and sum of squares of degree d on n variables, $\mathcal{P}^+_{n,2d}$ and $\mathcal{P}^{\text{sos}}_{n,2d}$.

Multivariate Polynomials and Sum-of-Squares

- Use multi-indices:

$$p(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha} \in \Delta(n,d)} p_{\boldsymbol{\alpha}} \, \boldsymbol{x}^{\boldsymbol{\alpha}}, \quad \text{where} \quad \boldsymbol{x}^{\boldsymbol{\alpha}} := x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n},$$

  where $\Delta(n, d) \subset \mathbb{Z}^n_{\geq 0}$ is the set of multi-indices with sum $\leq d$.
- Cardinality of $\Delta(n, d)$ is $s(n, d) := \binom{n+d}{d}$.
- Define cone of nonnegative polynomials and sum of squares of degree d on n variables, $\mathcal{P}^+_{n,2d}$ and $\mathcal{P}^{\text{sos}}_{n,2d}$.
- **Bad news:** $n = 1$ was a stroke of luck:

Theorem (Hilbert's theorem).

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

A nonnegative polynomial which is not SOS

Example: Motzkin polynomial

$$p(x,y) = x^4y^2 + x^2y^4 + 1 - 3x^2y^2 \in \mathbb{R}_6[x,y].$$

- p is nonnegative! (Arithmetic-Geometric mean inequality for x^2y^4, x^4y^2 and 1).

A nonnegative polynomial which is not SOS

Example: Motzkin polynomial

$$p(x,y) = x^4y^2 + x^2y^4 + 1 - 3x^2y^2 \in \mathbb{R}_6[x,y].$$

- p is nonnegative! (Arithmetic-Geometric mean inequality for x^2y^4, x^4y^2 and 1).
- But we can show that p is not SOS. (Analytical proof, or we can test if $p \in \mathcal{P}_{2,6}^{\mathrm{sos}}$ by solving an SDP):

Theorem

The polynomial $p(x) = \sum_{\boldsymbol{\alpha} \in \Delta(n,2d)} p_{\boldsymbol{\alpha}}\, \boldsymbol{x}^{\boldsymbol{\alpha}}$ is a SOS iff there exists $M \in \mathbb{S}^{s(n,d)}$ (indexed by $\alpha, \beta \in \Delta(n,d)$) such that $M \succeq 0$ and

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

A first hierarchy

Good news:

$(1+x^2+y^2)\cdot p(x,y)$ **is a sum of square.**

$\rightarrow$ SDP-based certificate that Motzkin's polynomial is ≥ 0.

A first hierarchy

Good news:

$(1+x^2+y^2)\cdot p(x,y)$ **is a sum of square.**

$\rightarrow$ SDP-based certificate that Motzkin's polynomial is ≥ 0.

For $r \geq 0$, define

$$P_{n,2d}^{(\delta)} = \left\{ p \in \mathbb{R}_{2d}[x_1,\ldots,x_n] : \quad x \mapsto (1+\sum_{i=1}^{n} x_i^2)^\delta \cdot p(x) \text{ is SOS} \right\}.$$

The product of two SOS is SOS, hence:

$$\mathcal{P}_{n,2d}^{\text{sos}} = P_{n,2d}^{(0)} \subseteq P_{n,2d}^{(1)} \subseteq P_{n,2d}^{(2)} \subseteq \ldots \subseteq \mathcal{P}_{n,2d}^{+}.$$

A first hierarchy

Good news:

$(1+x^2+y^2)\cdot p(x,y)$ **is a sum of square.**

$\rightarrow$ SDP-based certificate that Motzkin's polynomial is ≥ 0.
For $r \geq 0$, define

$$P_{n,2d}^{(\delta)} = \left\{ p \in \mathbb{R}_{2d}[x_1,\ldots,x_n] : \quad x \mapsto (1+\sum_{i=1}^n x_i^2)^\delta \cdot p(x) \text{ is SOS} \right\}.$$

The product of two SOS is SOS, hence:

$$\mathcal{P}_{n,2d}^{\text{sos}} = P_{n,2d}^{(0)} \subseteq P_{n,2d}^{(1)} \subseteq P_{n,2d}^{(2)} \subseteq \ldots \subseteq \mathcal{P}_{n,2d}^{+}.$$

- For a fixed δ, checking whether a polynomial is in $P_{n,2d}^{(\delta)}$, or optimizing over $P_{n,2d}^{(\delta)}$, can be done in polytime; it's an SDP with a matrix variable $X \in \mathbb{S}_+^N$, where $N = s(n, 2(d+\delta))$.

Outline

Moment Sequences

- $\mathcal{M}^+(\mathbb{R}^n) := \{$nonnegative measures μ over $\mathbb{R}^n\}$

Moment Sequences

- $\mathcal{M}^+(\mathbb{R}^n) := \{$nonnegative measures μ over $\mathbb{R}^n\}$
- The (infinite) sequence $(y_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{Z}^n_{\geq 0}}$ of moments of μ is given by:

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

Moment Sequences

- $\mathcal{M}^+(\mathbb{R}^n) := \{$nonnegative measures μ over $\mathbb{R}^n\}$
- The (infinite) sequence $(y_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{Z}^n_{\geq 0}}$ of moments of μ is given by:

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

For example, if μ is the probability measure corresponding to a random vector $Y \in \mathbb{R}^n$, then

- $y_{\mathbf{0}} = \int \mu(d\boldsymbol{x}) = 1$

Moment Sequences

- $\mathcal{M}^+(\mathbb{R}^n) := \{$nonnegative measures μ over $\mathbb{R}^n\}$
- The (infinite) sequence $(y_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{Z}^n_{\geq 0}}$ of moments of μ is given by:

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

For example, if μ is the probability measure corresponding to a random vector $Y \in \mathbb{R}^n$, then

- $y_{\mathbf{0}} = \int \mu(d\boldsymbol{x}) = 1$
- $y_{\boldsymbol{e}_i} = \int x_i\, \mu(d\boldsymbol{x}) = \mathbb{E}[Y_i]$

Moment Sequences

- $\mathcal{M}^+(\mathbb{R}^n) := \{$nonnegative measures μ over $\mathbb{R}^n\}$
- The (infinite) sequence $(y_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{Z}^n_{\geq 0}}$ of moments of μ is given by:

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

For example, if μ is the probability measure corresponding to a random vector $Y \in \mathbb{R}^n$, then

- $y_{\mathbf{0}} = \int \mu(d\boldsymbol{x}) = 1$
- $y_{\boldsymbol{e}_i} = \int x_i\, \mu(d\boldsymbol{x}) = \mathbb{E}[Y_i]$
- $y_{2\boldsymbol{e}_i} = \int x_i^2\, \mu(d\boldsymbol{x}) = \mathbb{E}[Y_i^2] = \mathbb{V}[Y_i] + (\mathbb{E}[Y_i])^2.$

Moment Sequences

- $\mathcal{M}^+(\mathbb{R}^n) := \{$nonnegative measures μ over $\mathbb{R}^n\}$
- The (infinite) sequence $(y_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{Z}^n_{\geq 0}}$ of moments of μ is given by:

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

For example, if μ is the probability measure corresponding to a random vector $Y \in \mathbb{R}^n$, then

- $y_{\mathbf{0}} = \displaystyle\int \mu(d\boldsymbol{x}) = 1$
- $y_{\boldsymbol{e}_i} = \displaystyle\int x_i\, \mu(d\boldsymbol{x}) = \mathbb{E}[Y_i]$
- $y_{2\boldsymbol{e}_i} = \displaystyle\int x_i^2\, \mu(d\boldsymbol{x}) = \mathbb{E}[Y_i^2] = \mathbb{V}[Y_i] + (\mathbb{E}[Y_i])^2.$
- More generally, $y_{\boldsymbol{\alpha}} = \mathbb{E}[Y_1^{\alpha_1} \cdots Y_n^{\alpha_n}]$.

The moment cone

Given a vector $\boldsymbol{y} \in \mathbb{R}^{s(n,d)}$, does $\boldsymbol{y}$ have a *representing measure*?
$\rightarrow$ Does there exist $\mu \in \mathcal{M}^+(\mathbb{R}^n)$ such that

$$\int_{\mathbb{R}^n} \boldsymbol{x}^\alpha \, \mu(d\boldsymbol{x}) = y_{\boldsymbol{\alpha}}, \qquad \forall |\boldsymbol{\alpha}| \leq d?$$

The moment cone

Given a vector $\boldsymbol{y} \in \mathbb{R}^{s(n,d)}$, does $\boldsymbol{y}$ have a *representing measure*?
$\rightarrow$ Does there exist $\mu \in \mathcal{M}^+(\mathbb{R}^n)$ such that

$$\int_{\mathbb{R}^n} \boldsymbol{x}^\alpha \, \mu(d\boldsymbol{x}) = y_{\boldsymbol{\alpha}}, \qquad \forall |\boldsymbol{\alpha}| \leq d?$$

This is the *moment problem*.

The moment cone

Given a vector $\boldsymbol{y} \in \mathbb{R}^{s(n,d)}$, does $\boldsymbol{y}$ have a *representing measure*?
$\to$ Does there exist $\mu \in \mathcal{M}^+(\mathbb{R}^n)$ such that

$$\int_{\mathbb{R}^n} \boldsymbol{x}^\alpha \, \mu(d\boldsymbol{x}) = y_{\boldsymbol{\alpha}}, \qquad \forall |\boldsymbol{\alpha}| \leq d?$$

This is the *moment problem* : $\quad \boldsymbol{y} \in \mathcal{M}_d^+(\mathbb{R}^n)$?

The vectors $\boldsymbol{y}$ with a representing measure constitute the *moment cone*

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

Nonnegative polynomials and moments

Proposition

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

Nonnegative polynomials and moments

Proposition

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

- Let $\boldsymbol{p} \in \mathcal{P}_{n,2d}^{+}$, and $\boldsymbol{y} \in \mathcal{M}_{2d}^{+}(\mathbb{R}^n)$.

Nonnegative polynomials and moments

Proposition

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

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

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

Nonnegative polynomials and moments

Proposition

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

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

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

- This already proves $\mathcal{P}^+_{n,2d} \subseteq (\mathcal{M}^+_{2d}(\mathbb{R}^n))^*$.

Nonnegative polynomials and moments

Proposition

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

- Conversely, assume $\boldsymbol{p} \notin \mathcal{P}_{n,2d}^{+}$, that is, $\exists \boldsymbol{z} \in \mathbb{R}^{n} : p(\boldsymbol{z}) < 0$.

Nonnegative polynomials and moments

Proposition

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

- Conversely, assume $\boldsymbol{p} \notin \mathcal{P}_{n,2d}^{+}$, that is, $\exists \boldsymbol{z} \in \mathbb{R}^n : p(\boldsymbol{z}) < 0$.
- Let $\boldsymbol{y}$ denote the moments of the dirac measure $\mu = \delta_{\boldsymbol{z}}$ at $\boldsymbol{z}$:

$$y_{\boldsymbol{\alpha}} = \int_{\mathbb{R}^n} \boldsymbol{x}^{\boldsymbol{\alpha}} \delta_{\boldsymbol{z}}(d\boldsymbol{x}) = \boldsymbol{z}^{\boldsymbol{\alpha}}$$

Nonnegative polynomials and moments

Proposition

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

- Conversely, assume $\boldsymbol{p} \notin \mathcal{P}^+_{n,2d}$, that is, $\exists \boldsymbol{z} \in \mathbb{R}^n : p(\boldsymbol{z}) < 0$.
- Let $\boldsymbol{y}$ denote the moments of the dirac measure $\mu = \delta_{\boldsymbol{z}}$ at $\boldsymbol{z}$:

$$y_{\boldsymbol{\alpha}} = \int_{\mathbb{R}^n} \boldsymbol{x}^{\boldsymbol{\alpha}} \delta_{\boldsymbol{z}}(d\boldsymbol{x}) = \boldsymbol{z}^{\boldsymbol{\alpha}}$$

- It holds $\langle \boldsymbol{p}, \boldsymbol{y} \rangle = \sum p_{\boldsymbol{\alpha}} \boldsymbol{z}^{\boldsymbol{\alpha}} = p(\boldsymbol{z}) < 0$, hence $\boldsymbol{p} \notin (\mathcal{M}^+_{2d}(\mathbb{R}^n))^*$.

Nonnegative polynomials and moments

Proposition

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

- Conversely, assume $\boldsymbol{p} \notin \mathcal{P}_{n,2d}^+$, that is, $\exists \boldsymbol{z} \in \mathbb{R}^n : p(\boldsymbol{z}) < 0$.
- Let $\boldsymbol{y}$ denote the moments of the dirac measure $\mu = \delta_{\boldsymbol{z}}$ at $\boldsymbol{z}$:

$$y_{\boldsymbol{\alpha}} = \int_{\mathbb{R}^n} \boldsymbol{x}^{\boldsymbol{\alpha}} \delta_{\boldsymbol{z}}(d\boldsymbol{x}) = \boldsymbol{z}^{\boldsymbol{\alpha}}$$

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

Nonnegative polynomials and moments

Proposition

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

- Conversely, assume $\boldsymbol{p} \notin \mathcal{P}_{n,2d}^{+}$, that is, $\exists \boldsymbol{z} \in \mathbb{R}^n : p(\boldsymbol{z}) < 0$.
- Let $\boldsymbol{y}$ denote the moments of the dirac measure $\mu = \delta_{\boldsymbol{z}}$ at $\boldsymbol{z}$:

$$y_{\boldsymbol{\alpha}} = \int_{\mathbb{R}^n} \boldsymbol{x}^{\boldsymbol{\alpha}} \delta_{\boldsymbol{z}}(d\boldsymbol{x}) = \boldsymbol{z}^{\boldsymbol{\alpha}}$$

- 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 $(\mathcal{M}_{2d}^{+}(\mathbb{R}^n))^{*} \subseteq \mathcal{P}_{n,2d}^{+}$.
- Proposition is proved, because cones are proper (dual=bi-dual).

A necessary LMI

- Let $r \in \mathbb{N}$ and let (y_{α}) be a (possibly infinite) sequence

A necessary LMI

- Let $r \in \mathbb{N}$ and let $(y_{\boldsymbol{\alpha}})$ be a (possibly infinite) sequence
- y is of sufficient length (it contains all elements indexed by some $\boldsymbol{\alpha}$ s.t. $|\boldsymbol{\alpha}| := \sum_i \alpha_i \leq 2r$)

A necessary LMI

- Let $r \in \mathbb{N}$ and let $(\boldsymbol{y}_{\boldsymbol{\alpha}})$ be a (possibly infinite) sequence
- $\boldsymbol{y}$ is of sufficient length (it contains all elements indexed by some $\boldsymbol{\alpha}$ s.t. $|\boldsymbol{\alpha}| := \sum_i \alpha_i \leq 2r$)
- Define the moment matrix $M_r(\boldsymbol{y}) \in \mathbb{S}^{s(n,r)}$ of order r of $\boldsymbol{y}$:

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

A necessary LMI

- Let $r \in \mathbb{N}$ and let $(\boldsymbol{y}_{\boldsymbol{\alpha}})$ be a (possibly infinite) sequence
- $\boldsymbol{y}$ is of sufficient length (it contains all elements indexed by some $\boldsymbol{\alpha}$ s.t. $|\boldsymbol{\alpha}| := \sum_i \alpha_i \leq 2r$)
- Define the moment matrix $M_r(\boldsymbol{y}) \in \mathbb{S}^{s(n,r)}$ of order r of $\boldsymbol{y}$:

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

- Note: $M_r(\boldsymbol{y}) = \sum_{\gamma \in \Delta(n,2r)} y_\gamma P_\gamma$, where P_γ is the $\{0,1\}$ matrix such that $(P_\gamma)_{\boldsymbol{\alpha},\boldsymbol{\beta}} = 1 \iff \boldsymbol{\alpha} + \boldsymbol{\beta} = \boldsymbol{\gamma}$.

A necessary LMI

- Let $r \in \mathbb{N}$ and let $(y_{\boldsymbol{\alpha}})$ be a (possibly infinite) sequence
- $\boldsymbol{y}$ is of sufficient length (it contains all elements indexed by some $\boldsymbol{\alpha}$ s.t. $|\boldsymbol{\alpha}| := \sum_i \alpha_i \leq 2r$)
- Define the moment matrix $M_r(\boldsymbol{y}) \in \mathbb{S}^{s(n,r)}$ of order r of $\boldsymbol{y}$:

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

- Note: $M_r(\boldsymbol{y}) = \sum_{\gamma \in \Delta(n,2r)} y_\gamma P_\gamma$, where P_γ is the $\{0,1\}$ matrix such that $(P_\gamma)_{\boldsymbol{\alpha},\boldsymbol{\beta}} = 1 \iff \boldsymbol{\alpha} + \boldsymbol{\beta} = \boldsymbol{\gamma}$.

Proposition

Let $r \leq d$. Then,

$$\boldsymbol{y} \in \mathcal{M}^+_{2d}(\mathbb{R}^n) \implies M_r(\boldsymbol{y}) \succeq 0.$$

A necessary LMI

Proposition

Let $r \leq d$. Then,

$$\boldsymbol{y} \in \mathcal{M}_{2d}^{+}(\mathbb{R}^n) \implies M_r(\boldsymbol{y}) \succeq 0.$$

- Assume w.l.o.g. that $y_{\mathbf{0}} = 1$.
- $\boldsymbol{y}$ has a representing *probability measure*, denote by X a corresponding random variable (in $\mathbb{R}^n$)

A necessary LMI

Proposition

Let $r \leq d$. Then,

$$\boldsymbol{y} \in \mathcal{M}_{2d}^{+}(\mathbb{R}^n) \implies M_r(\boldsymbol{y}) \succeq 0.$$

- Assume w.l.o.g. that $y_{\mathbf{0}} = 1$.
- $\boldsymbol{y}$ has a representing *probability measure*, denote by X a corresponding random variable (in $\mathbb{R}^n$)
- Given $\boldsymbol{c} \in \Delta(n, r)$, define the polynomial $c(\boldsymbol{x}) = \sum_{|\boldsymbol{\alpha}| \leq r} c_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}}$

A necessary LMI

Proposition

Let $r \leq d$. Then,

$$\boldsymbol{y} \in \mathcal{M}_{2d}^{+}(\mathbb{R}^n) \implies M_r(\boldsymbol{y}) \succeq 0.$$

- Assume w.l.o.g. that $y_{\mathbf{0}} = 1$.
- $\boldsymbol{y}$ has a representing *probability measure*, denote by X a corresponding random variable (in $\mathbb{R}^n$)
- Given $\boldsymbol{c} \in \Delta(n, r)$, define the polynomial $c(\boldsymbol{x}) = \sum_{|\boldsymbol{\alpha}| \leq r} c_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}}$
- $$\begin{aligned} 0 \leq \mathbb{E}[c(X)^2] = \mathbb{E}[\sum_{\boldsymbol{\alpha}} c_{\boldsymbol{\alpha}} X^{\boldsymbol{\alpha}} \sum_{\boldsymbol{\beta}} c_{\boldsymbol{\beta}} X^{\boldsymbol{\beta}}] &= \sum_{\boldsymbol{\alpha},\boldsymbol{\beta}} c_{\boldsymbol{\alpha}} c_{\boldsymbol{\beta}} \underbrace{\mathbb{E}[X^{\boldsymbol{\alpha}+\boldsymbol{\beta}}]}_{=y_{\boldsymbol{\alpha}+\boldsymbol{\beta}}} \\ &= \sum_{\boldsymbol{\alpha},\boldsymbol{\beta}} c_{\boldsymbol{\alpha}} c_{\boldsymbol{\beta}} \big(M_r(\boldsymbol{y})\big)_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \boldsymbol{c}^T M_r(\boldsymbol{y}) \boldsymbol{c}. \end{aligned}$$

A necessary LMI

Proposition

Let $r \leq d$. Then,

$$\boldsymbol{y} \in \mathcal{M}_{2d}^{+}(\mathbb{R}^n) \implies M_r(\boldsymbol{y}) \succeq 0.$$

- Assume w.l.o.g. that $y_{\boldsymbol{0}} = 1$.
- $\boldsymbol{y}$ has a representing *probability measure*, denote by X a corresponding random variable (in $\mathbb{R}^n$)
- Given $\boldsymbol{c} \in \Delta(n, r)$, define the polynomial $c(\boldsymbol{x}) = \sum_{|\boldsymbol{\alpha}| \leq r} c_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}}$
- $$\begin{aligned} 0 \leq \mathbb{E}[c(X)^2] = \mathbb{E}[\sum_{\boldsymbol{\alpha}} c_{\boldsymbol{\alpha}} X^{\boldsymbol{\alpha}} \sum_{\boldsymbol{\beta}} c_{\boldsymbol{\beta}} X^{\boldsymbol{\beta}}] \quad &= \sum_{\boldsymbol{\alpha},\boldsymbol{\beta}} c_{\boldsymbol{\alpha}} c_{\boldsymbol{\beta}} \underbrace{\mathbb{E}[X^{\boldsymbol{\alpha}+\boldsymbol{\beta}}]}_{=y_{\boldsymbol{\alpha}+\boldsymbol{\beta}}} \\ &= \sum_{\boldsymbol{\alpha},\boldsymbol{\beta}} c_{\boldsymbol{\alpha}} c_{\boldsymbol{\beta}} \big(M_r(\boldsymbol{y})\big)_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \boldsymbol{c}^T M_r(\boldsymbol{y}) \boldsymbol{c}. \end{aligned}$$
- $\boldsymbol{c}$ is arbitrary, so $M_r(\boldsymbol{y}) \succeq 0$.

An SDP representacable approximation

Define

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

An SDP representacable approximation

Define

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

- The necessary LMI condition for $\boldsymbol{y} \in \mathcal{M}_{2d}^{+}(\mathbb{R}^n)$ rewrites

$$\mathcal{M}_{2d}^{+}(\mathbb{R}^n) \subseteq \mathcal{M}_{2d}^{\mathrm{SDP}}(\mathbb{R}^n).$$

An SDP representacable approximation

Define

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

- The necessary LMI condition for $\boldsymbol{y} \in \mathcal{M}_{2d}^{+}(\mathbb{R}^n)$ rewrites

$$\mathcal{M}_{2d}^{+}(\mathbb{R}^n) \subseteq \mathcal{M}_{2d}^{\mathsf{SDP}}(\mathbb{R}^n).$$

- It can also be shown that

Proposition

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

Outline

Polynomial optimization as a moment problem

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

- Minimize p over semi-algebraic set

$$K := \{\boldsymbol{x} \in \mathbb{R}^n : g_i(\boldsymbol{x}) \geq 0, \ \forall i \in [m]\}.$$

Polynomial optimization as a moment problem

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

- Minimize p over semi-algebraic set

$$K := \{\boldsymbol{x} \in \mathbb{R}^n : g_i(\boldsymbol{x}) \geq 0,\ \forall i \in [m]\}.$$

- Reformulate (P) so the variable is a measure over K:

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

where $\mathcal{M}^+(K)$ is the set of nonnegative measures s.t. $\mu(\mathbb{R}^n \setminus K) = 0$.

Polynomial optimization as a moment problem

- Measure formulation of (P):

$$p^* = \inf_{\boldsymbol{\mu} \in \mathcal{M}^+(K)} \quad \int_K p(x)\,\mu(dx)$$
$$s.t. \quad \mu(K) = 1,$$

Polynomial optimization as a moment problem

- Measure formulation of (P):

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

- Define the set of *truncated moment sequences*

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

Polynomial optimization as a moment problem

- Measure formulation of (P):

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

- Define the set of *truncated moment sequences*

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

- If $\boldsymbol{y}$ is the truncated moment sequence of $\mu \in \mathcal{M}^+(K)$, we have $\mu(K) = \mu(\mathbb{R}^n) = y_0$ and $\int_K p(\boldsymbol{x})\, \mu(d\boldsymbol{x}) = \langle \boldsymbol{p}, \boldsymbol{y} \rangle$. Hence,

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

The moment problem over K

- We have $p^* = \inf\{\langle \boldsymbol{p}, \boldsymbol{y}\rangle : \ \boldsymbol{y} \in \mathcal{M}_d^+(K), \quad y_0 = 1\}.$

The moment problem over K

- We have $p^* = \inf\{\langle \boldsymbol{p}, \boldsymbol{y}\rangle : \ \boldsymbol{y} \in \mathcal{M}_d^+(K), \quad y_0 = 1\}$.
- So we look for tractable approximations of $\mathcal{M}_d^+(K)$.

The moment problem over K

- We have $p^* = \inf\{\langle \boldsymbol{p}, \boldsymbol{y}\rangle : \ \boldsymbol{y} \in \mathcal{M}_d^+(K), \quad y_0 = 1\}$.
- So we look for tractable approximations of $\mathcal{M}_d^+(K)$.
- As in the unconstrained case, we first obtain some necessary conditions

The moment problem over K

- We have $p^* = \inf\{\langle \boldsymbol{p}, \boldsymbol{y}\rangle : \ \boldsymbol{y} \in \mathcal{M}_d^+(K), \quad y_0 = 1\}$.
- So we look for tractable approximations of $\mathcal{M}_d^+(K)$.
- As in the unconstrained case, we first obtain some necessary conditions
- Given $r \in \mathbb{N}$, a polynomial g of degree $\leq 2u$ and $\boldsymbol{y}$ of sufficient length ($y_{\boldsymbol{\alpha}}$ is defined for all $|\boldsymbol{\alpha}| \leq 2(u+r)$), define the *localizing matrix* $M_r(g\boldsymbol{y}) \in \mathbb{S}^{s(n,r)}$ with elements

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

The moment problem over K

- We have $p^* = \inf\{\langle \boldsymbol{p}, \boldsymbol{y} \rangle : \ \boldsymbol{y} \in \mathcal{M}_d^+(K), \quad y_0 = 1\}$.
- So we look for tractable approximations of $\mathcal{M}_d^+(K)$.
- As in the unconstrained case, we first obtain some necessary conditions
- Given $r \in \mathbb{N}$, a polynomial g of degree $\leq 2u$ and $\boldsymbol{y}$ of sufficient length ($y_{\boldsymbol{\alpha}}$ is defined for all $|\boldsymbol{\alpha}| \leq 2(u+r)$), 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

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

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

A hierarchy of LMI-based approximations

Proposition

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

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

- Let $u_i := \lceil \deg(g_i)/2 \rceil$, and define

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

Then, $\mathcal{M}_{2r}^{+}(K) \subseteq \mathcal{M}_{2r}^{\mathsf{SDP}}(K)$.

A hierarchy of LMI-based approximations

Proposition

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

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

- Let $u_i := \lceil \deg(g_i)/2 \rceil$, and define

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

  Then, $\mathcal{M}_{2r}^+(K) \subseteq \mathcal{M}_{2r}^{\mathrm{SDP}}(K)$.
- We obtain tighter and tighter approximations by truncating y at a higher degree, say $2(r+\delta)$.

A hierarchy of LMI-based approximations

- Let $u_i := \lceil \deg(g_i)/2 \rceil$, and define

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

  Then, $\mathcal{M}_{2r}^{+}(K) \subseteq \mathcal{M}_{2r}^{\mathrm{SDP}}(K)$.
- We obtain tighter and tighter approximations by truncating $\boldsymbol{y}$ at a higher degree, say $2(r+\delta)$:

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

A hierarchy of LMI-based approximations

- Let $u_i := \lceil \deg(g_i)/2 \rceil$, and define

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

  Then, $\mathcal{M}_{2r}^{+}(K) \subseteq \mathcal{M}_{2r}^{\mathrm{SDP}}(K)$.
- We obtain tighter and tighter approximations by truncating $\boldsymbol{y}$ at a higher degree, say $2(r+\delta)$:

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

- Definition 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 longer sequence $\in \mathbb{R}^{s(n,2(r+\delta))}$ for which the moment and localizing matrices are $\succeq 0$.

A hierarchy of LMI-based approximations

- Let $u_i := \lceil \deg(g_i)/2 \rceil$, and define

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

  Then, $\mathcal{M}_{2r}^{+}(K) \subseteq \mathcal{M}_{2r}^{\mathrm{SDP}}(K)$.
- We obtain tighter and tighter approximations by truncating $\boldsymbol{y}$ at a higher degree, say $2(r+\delta)$:

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

Corollary

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

The Lasserre Hierarchy

We have shown:

- $p^* = \inf\{\langle \boldsymbol{p}, \boldsymbol{y}\rangle : \ \boldsymbol{y} \in \mathcal{M}_d^+(K), \quad y_0 = 1\}.$
- $\mathcal{M}_{2r}^+(K) \subseteq \cdots \subseteq \mathcal{M}_{2r,2}^{\text{SDP}}(K) \subseteq \mathcal{M}_{2r,1}^{\text{SDP}}(K) \subseteq \mathcal{M}_{2r,0}^{\text{SDP}}(K).$

The Lasserre Hierarchy

We have shown:

- $p^* = \inf\{\langle \boldsymbol{p}, \boldsymbol{y} \rangle : \; \boldsymbol{y} \in \mathcal{M}_d^+(K), \quad y_0 = 1\}$.
- $\mathcal{M}_{2r}^+(K) \subseteq \cdots \subseteq \mathcal{M}_{2r,2}^{\mathrm{SDP}}(K) \subseteq \mathcal{M}_{2r,1}^{\mathrm{SDP}}(K) \subseteq \mathcal{M}_{2r,0}^{\mathrm{SDP}}(K)$.

$\rightarrow$ We obtain tighter and tighter bounds by replacing the constraint $\boldsymbol{y} \in \mathcal{M}_d^+(K)$ by $\mathcal{M}_{2r,\delta}^{\mathrm{SDP}}(K)$ (for r large enough) and increasing values of $\delta \geq 0$.

The Lasserre Hierarchy

We have shown:

- $p^* = \inf\{\langle \boldsymbol{p}, \boldsymbol{y}\rangle : \ \boldsymbol{y} \in \mathcal{M}_d^+(K), \quad y_0 = 1\}.$
- $\mathcal{M}_{2r}^+(K) \subseteq \cdots \subseteq \mathcal{M}_{2r,2}^{\text{SDP}}(K) \subseteq \mathcal{M}_{2r,1}^{\text{SDP}}(K) \subseteq \mathcal{M}_{2r,0}^{\text{SDP}}(K).$

$\rightarrow$ We obtain tighter and tighter bounds by replacing the constraint $\boldsymbol{y} \in \mathcal{M}_d^+(K)$ by $\mathcal{M}_{2r,\delta}^{\text{SDP}}(K)$ (for r large enough) and increasing values of $\delta \geq 0$.

- We need $2r \geq d$ and $r - u_i \geq 0, \forall i$, where $u_i = \lceil \deg(g_i)/2 \rceil$ so we set $u := \max_{i \in [m]} u_i$ and $r := \max\big(\lceil \deg(p)/2 \rceil, u\big)$.

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

Convergence of Lasserre Hierarchy

- By construction, the optimal values p_δ^* of (Las$_\delta$) satisfy

$$p_0^* \leq p_1^* \leq p_2^* \leq \ldots \leq p^*.$$

Convergence of Lasserre Hierarchy

- By construction, the optimal values p^*_δ of (Las$_\delta$) satisfy

$$p^*_0 \leq p^*_1 \leq p^*_2 \leq \ldots \leq p^*.$$

- But do we have $\lim_{\delta \to \infty} p^*_\delta = p^*$?

Convergence of Lasserre Hierarchy

- By construction, the optimal values p_δ^* of (Las$_\delta$) satisfy

$$p_0^* \leq p_1^* \leq p_2^* \leq \ldots \leq p^*.$$

- But do we have $\lim_{\delta\to\infty} p_\delta^* = p^*$?
- A sufficient condition is given by:

Definition (Archimedean Property)

We say that K satisfies the Archimedean Property if $\exists R > 0$ and SOS polynomials $\sigma_1, \ldots, \sigma_m$ such that

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

$\Longrightarrow$ algebraic certificate of compactness for K.

Convergence of Lasserre Hierarchy

- By construction, the optimal values p_δ^* of (Las_δ) satisfy

$$p_0^* \leq p_1^* \leq p_2^* \leq \ldots \leq p^*.$$

- But do we have $\lim_{\delta\to\infty} p_\delta^* = p^*$?
- A sufficient condition is that K has the Archimedean property:

Theorem

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 converges:

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

Finite convergence of the hierarchy

Under some conditions, we can prove convergence after a finite number of steps: $\exists \delta \in \mathbb{N} : p_\delta^* = p^*$:

- When there are only equality constraints
- And, according to a result of [Nie,2014], this happens *generically*
- Moreover, we can check if convergence took place after δ rounds thanks to the following sufficient condition:

Theorem

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

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

where $u = \max_i u_i$, then $p_\delta^* = p^*$, and $\boldsymbol{y}^*$ has a representing measure $\mu^* \in \mathcal{M}^+(K)$. Moreover, μ^* has r_δ support points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{r_\delta}$, all of them are global minimizers of (P).

Outline

Dual of (Las$_\delta$)

- Recall: $M_\rho(\boldsymbol{y}) = \sum_{\boldsymbol{\gamma} \in \Delta(n,2r)} y_{\boldsymbol{\gamma}} P_{\boldsymbol{\gamma}}$, where $(P_{\boldsymbol{\gamma}})_{\boldsymbol{\alpha},\boldsymbol{\beta}} = 1 \Leftrightarrow \boldsymbol{\alpha} + \boldsymbol{\beta} = \boldsymbol{\gamma}$.
- Similarly, $M_\rho(g\boldsymbol{y}) = \sum y_\gamma Q_\gamma^g$, where $Q_\gamma^g \in \mathbb{S}^{s(n,\rho)}$ satisfies

$$(Q_\gamma^g)_{\boldsymbol{\alpha},\boldsymbol{\beta}} = g_{\boldsymbol{\tau}} \quad \text{whenever} \quad \boldsymbol{\alpha} + \boldsymbol{\beta} + \boldsymbol{\tau} = \boldsymbol{\gamma}.$$

Dual of (Las$_\delta$)

- Recall: $M_\rho(\boldsymbol{y}) = \sum_{\boldsymbol{\gamma}\in\Delta(n,2r)} y_{\boldsymbol{\gamma}} P_{\boldsymbol{\gamma}}$, where $(P_{\boldsymbol{\gamma}})_{\boldsymbol{\alpha},\boldsymbol{\beta}} = 1 \Leftrightarrow \boldsymbol{\alpha}+\boldsymbol{\beta}=\boldsymbol{\gamma}$.
- Similarly, $M_\rho(g\boldsymbol{y}) = \sum y_\gamma Q_\gamma^g$, where $Q_\gamma^g \in \mathbb{S}^{s(n,\rho)}$ satisfies

$$(Q_\gamma^g)_{\boldsymbol{\alpha},\boldsymbol{\beta}} = g_{\boldsymbol{\tau}} \quad \text{whenever} \quad \boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\tau}=\boldsymbol{\gamma}.$$

The dual SDP of (Las$_\delta$) is:

$$d_\delta^* = \sup_{\lambda,\Lambda,(\Omega_i)_{i\in[m]}} \lambda \tag{Sos$_\delta$}$$

$$\begin{aligned} s.t. \quad & (\boldsymbol{p}-\lambda\boldsymbol{e}_0)_\gamma = \langle P_\gamma, \Lambda\rangle + \sum_{i\in[m]} \langle Q_\gamma^{g_i}, \Omega_i\rangle, \quad \forall|\boldsymbol{\gamma}| \leq 2(r+\delta) \\ & \Lambda \succeq 0 \\ & \Omega_i \succeq 0, \ \forall i \in [m] \end{aligned}$$

Interpretation of (Sos_δ)

The constraint

$$(\boldsymbol{p} - \lambda \boldsymbol{e}_0)_\gamma = \langle P_\gamma, \Lambda \rangle + \sum_{i \in [m]} \langle Q_\gamma^{g_i}, \Omega_i \rangle, \qquad \forall |\boldsymbol{\gamma}| \leq 2(r + \delta)$$

means that the polynomial $p - \lambda$ 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 $\langle Q_\gamma^{g_i}, \Omega_i \rangle, \forall i \in [m]$

Interpretation of (Sos_δ)

The constraint

$$(\boldsymbol{p} - \lambda \boldsymbol{e}_0)_\gamma = \langle P_\gamma, \Lambda \rangle + \sum_{i \in [m]} \langle Q_\gamma^{g_i}, \Omega_i \rangle, \qquad \forall |\boldsymbol{\gamma}| \leq 2(r+\delta)$$

means that the polynomial $p - \lambda$ is the sum of

- a polynomial σ_0 with coefficients $\langle P_\gamma, \Lambda \rangle \implies \sigma_0$ is SOS
- and some polynomials $q_1, \ldots, q_m$ with coefficients $\langle Q_\gamma^{g_i}, \Omega_i \rangle, \forall i \in [m]$

Interpretation of (Sos_δ)

The constraint

$$(\boldsymbol{p} - \lambda \boldsymbol{e}_0)_\gamma = \langle P_\gamma, \Lambda \rangle + \sum_{i \in [m]} \langle Q_\gamma^{g_i}, \Omega_i \rangle, \qquad \forall |\boldsymbol{\gamma}| \leq 2(r+\delta)$$

means that the polynomial $p - \lambda$ is the sum of

- a polynomial σ_0 with coefficients $\langle P_\gamma, \Lambda \rangle \implies \sigma_0$ is SOS
- and some polynomials $q_1, \ldots, q_m$ with coefficients $\langle Q_\gamma^{g_i}, \Omega_i \rangle, \forall i \in [m] \implies q_i = g_i \sigma_i$ for SOS polynomial σ_i

Interpretation of (Sos_δ)

The constraint

$$(\boldsymbol{p} - \lambda \boldsymbol{e}_0)_\gamma = \langle P_\gamma, \Lambda \rangle + \sum_{i \in [m]} \langle Q_\gamma^{g_i}, \Omega_i \rangle, \qquad \forall |\boldsymbol{\gamma}| \leq 2(r+\delta)$$

means that the polynomial $p - \lambda$ is the sum of

- a polynomial σ_0 with coefficients $\langle P_\gamma, \Lambda \rangle \implies \sigma_0$ is SOS
- and some polynomials $q_1, \ldots, q_m$ with coefficients $\langle Q_\gamma^{g_i}, \Omega_i \rangle, \forall i \in [m] \implies q_i = g_i \sigma_i$ for SOS polynomial σ_i
- $\deg(\sigma_0) \leq 2(r+\delta)$ and $\deg(\sigma_i) \leq 2(r+\delta-u_i),\ \forall i \in [m]$.

Interpretation of (Sos_δ)

The constraint

$$(\boldsymbol{p} - \lambda \boldsymbol{e_0})_\gamma = \langle P_\gamma, \Lambda \rangle + \sum_{i \in [m]} \langle Q_\gamma^{g_i}, \Omega_i \rangle, \qquad \forall |\boldsymbol{\gamma}| \leq 2(r+\delta)$$

means that the polynomial $p - \lambda$ is the sum of

- a polynomial σ_0 with coefficients $\langle P_\gamma, \Lambda \rangle \implies \sigma_0$ is SOS
- and some polynomials $q_1, \ldots, q_m$ with coefficients $\langle Q_\gamma^{g_i}, \Omega_i \rangle, \forall i \in [m] \implies q_i = g_i \sigma_i$ for SOS polynomial σ_i
- $\deg(\sigma_0) \leq 2(r+\delta)$ and $\deg(\sigma_i) \leq 2(r+\delta-u_i), \ \forall i \in [m]$.

$$d_\delta^* = \sup_{\lambda, \sigma_0, (\sigma_i)_{i \in [m]}} \lambda \qquad (\text{Sos}_\delta)$$

$$s.t. \quad p(x) - \lambda = \sigma_0(x) + \sum_{i \in [m]} g_i(x) \cdot \sigma_i(x)$$

σ_0 is a SOS polynomial of degree $\leq 2(r+\delta)$

σ_i is a SOS polynomial of degree $\leq 2(r+\delta-u_i), \ \forall i \in [m]$

Putinar's Positivstellensatz

Note that $p - \lambda = \sigma_0 + \sum_{i \in [m]} g_i \cdot \sigma_i$ for some SOS polynomials $\sigma_0, \ldots, \sigma_m$ is an algebraic certificate that $p \geq \lambda$ over K.
Therefore, $\lambda^* = d_\delta^* \leq p^*$. (Note: We already knew this from weak duality: $d_\delta^* \leq p_\delta^* \leq p^*$).

Putinar's Positivstellensatz

Note that $p - \lambda = \sigma_0 + \sum_{i \in [m]} g_i \cdot \sigma_i$ for some SOS polynomials $\sigma_0, \ldots, \sigma_m$ is an algebraic certificate that $p \geq \lambda$ over K.
Therefore, $\lambda^* = d_\delta^* \leq p^*$. (Note: We already knew this from weak duality: $d_\delta^* \leq p_\delta^* \leq p^*$).

The following theorem gives a partial converse:

Theorem (Putinar's Positivstellensatz)

Let p be a *positive* polynomial over the Archimedean set $K = \{\boldsymbol{x} \in \mathbb{R}^n : g_i(\boldsymbol{x}) \geq 0,\ \forall i \in [m]\}$. 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: $p > 0$ over K, nothing known about required degrees.

Positivstellensatz $\Rightarrow$ convergence of hierarchy

Theorem (Putinar's Positivstellensatz)

Let p be a *positive* polynomial over the Archimedean set $K = \{\boldsymbol{x} \in \mathbb{R}^n : g_i(\boldsymbol{x}) \geq 0, \ \forall i \in [m]\}$. Then, p can be written as

$p = \sigma_0 + \sum_{i \in [m]} g_i \cdot \sigma_i$ for some SOS polynomials $\sigma_0, \ldots, \sigma_m$.

Positivstellensatz $\Rightarrow$ convergence of hierarchy

Theorem (Putinar's Positivstellensatz)

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

This result can be used to prove that $d_\delta^* \underset{\delta\to\infty}{\longrightarrow} p^*$: (Sos$_\delta$) converges

And hence, $p_\delta^* \underset{\delta\to\infty}{\longrightarrow} p^*$ by weak duality: (Las$_\delta$) converges

Positivstellensatz $\Rightarrow$ convergence of hierarchy

Theorem (Putinar's Positivstellensatz)

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

This result can be used to prove that $d_\delta^* \underset{\delta\to\infty}{\longrightarrow} p^*$: (Sos$_\delta$) converges

And hence, $p_\delta^* \underset{\delta\to\infty}{\longrightarrow} p^*$ by weak duality: (Las$_\delta$) converges

Proof:

- $p - p^* + \epsilon$ is positive over K for all $\epsilon > 0$.

Positivstellensatz $\Rightarrow$ convergence of hierarchy

Theorem (Putinar's Positivstellensatz)

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

This result can be used to prove that $d_\delta^* \underset{\delta \to \infty}{\longrightarrow} p^*$: ($\text{Sos}_\delta$) converges

And hence, $p_\delta^* \underset{\delta \to \infty}{\longrightarrow} p^*$ by weak duality: (Las_δ) converges

Proof:

- $p - p^* + \epsilon$ 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 σ_i's

Positivstellensatz $\Rightarrow$ convergence of hierarchy

Theorem (Putinar's Positivstellensatz)

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

This result can be used to prove that $d_\delta^* \underset{\delta\to\infty}{\to} p^*$: (Sos$_\delta$) converges

And hence, $p_\delta^* \underset{\delta\to\infty}{\to} p^*$ by weak duality: (Las$_\delta$) converges

Proof:

- $p - p^* + \epsilon$ 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 σ_i's
- This means that there exists $\delta \in \mathbb{N}$ such that $p^* - \epsilon$ is feasible for (Sos$_\delta$)

Outline

Lasserre Hierarchy for Integer Programming

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

Lasserre Hierarchy for Integer Programming

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

Given a measure μ supported on $K = \{\boldsymbol{x} \in \{0,1\}^n : A\boldsymbol{x} \geq \boldsymbol{b}\}$,

- $\int x_i^2 \, \mu(d\boldsymbol{x}) = \int x_i \, \mu(d\boldsymbol{x})$

Lasserre Hierarchy for Integer Programming

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

Given a measure μ supported on $K = \{\boldsymbol{x} \in \{0,1\}^n : A\boldsymbol{x} \geq \boldsymbol{b}\}$,

- $\displaystyle\int x_i^2 \, \mu(d\boldsymbol{x}) = \int x_i \, \mu(d\boldsymbol{x})$
- $\displaystyle\int_K \boldsymbol{x}^{\boldsymbol{\alpha}} \, \mu(d\boldsymbol{x}) = \int_K \left(\prod_{i \in I} x_i\right) \mu(d\boldsymbol{x}),$ where $I = \{i \in [n] : \alpha_i \geq 1\}$

Lasserre Hierarchy for Integer Programming

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

Given a measure μ supported on $K = \{\boldsymbol{x} \in \{0,1\}^n : A\boldsymbol{x} \geq \boldsymbol{b}\}$,

- $\displaystyle\int x_i^2 \, \mu(d\boldsymbol{x}) = \int x_i \, \mu(d\boldsymbol{x})$
- $\displaystyle\int_K \boldsymbol{x}^{\boldsymbol{\alpha}} \, \mu(d\boldsymbol{x}) = \int_K \left(\prod_{i \in I} x_i \right) \mu(d\boldsymbol{x}),$ where $I = \{i \in [n] : \alpha_i \geq 1\}$

So moment $y_{\boldsymbol{\alpha}}$ depends only on *sparsity pattern* $I \subseteq [n]$ of $\boldsymbol{\alpha}$

Lasserre Hierarchy for Integer Programming

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

Given a measure μ supported on $K = \{\boldsymbol{x} \in \{0,1\}^n : A\boldsymbol{x} \geq \boldsymbol{b}\}$,

- $\int x_i^2 \, \mu(d\boldsymbol{x}) = \int x_i \, \mu(d\boldsymbol{x})$
- $\int_K \boldsymbol{x}^{\boldsymbol{\alpha}} \, \mu(d\boldsymbol{x}) = \int_K \left(\prod_{i \in I} x_i \right) \mu(d\boldsymbol{x})$, where $I = \{i \in [n] : \alpha_i \geq 1\}$

So moment $y_{\boldsymbol{\alpha}}$ depends only on *sparsity pattern* $I \subseteq [n]$ of $\boldsymbol{\alpha}$

We construct a hierarchy based on variables (y_I), for sets I of bounded cardinality $|I| \leq O(\delta)$.

Lasserre Hierarchy for Integer Programming

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

Given a measure μ supported on $K = \{\boldsymbol{x} \in \{0,1\}^n : A\boldsymbol{x} \geq \boldsymbol{b}\}$,

- $\int x_i^2 \, \mu(d\boldsymbol{x}) = \int x_i \, \mu(d\boldsymbol{x})$
- $\int_K \boldsymbol{x}^{\boldsymbol{\alpha}} \, \mu(d\boldsymbol{x}) = \int_K \left(\prod_{i \in I} x_i \right) \mu(d\boldsymbol{x})$, where $I = \{i \in [n] : \alpha_i \geq 1\}$

So moment $y_{\boldsymbol{\alpha}}$ depends only on *sparsity pattern* $I \subseteq [n]$ of $\boldsymbol{\alpha}$

We construct a hierarchy based on variables (y_I), for sets I of bounded cardinality $|I| \leq O(\delta)$.

- In particular: $M_\rho(\boldsymbol{y}) = (y_{I \cup J})_{|I| \leq \rho, |J| \leq \rho}$

Lasserre Hierarchy for Integer Programming

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

- Hierarchy at level δ depends on (y_I) for $|I| \leq O(\delta)$.

Lasserre Hierarchy for Integer Programming

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

- Hierarchy at level δ depends on (y_I) for $|I| \leq O(\delta)$.
- Let X be a random vector associated with a representing probability measure $\mu \in \mathcal{M}^+(K)$ for $\boldsymbol{y}$:

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

Lasserre Hierarchy for Integer Programming

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

- Hierarchy at level δ depends on (y_I) for $|I| \leq O(\delta)$.
- Let X be a random vector associated with a representing probability measure $\mu \in \mathcal{M}^+(K)$ for $\boldsymbol{y}$:

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

- Standard interpretation of LP-relaxation of *IP*: round x_i to $X_i = 1$ with probability x_i^*, so $\mathbb{E}[\boldsymbol{c}^T X] = \boldsymbol{c}^T \boldsymbol{x}^*$.

Lasserre Hierarchy for Integer Programming

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

- Hierarchy at level δ depends on (y_I) for $|I| \leq O(\delta)$.
- Let X be a random vector associated with a representing probability measure $\mu \in \mathcal{M}^+(K)$ for $\boldsymbol{y}$:

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

- Standard interpretation of LP-relaxation of *IP*: round x_i to $X_i = 1$ with probability x_i^*, so $\mathbb{E}[\boldsymbol{c}^T X] = \boldsymbol{c}^T \boldsymbol{x}^*$.
- But rounded solution X might violate the constraints: $AX \not\geq \boldsymbol{b}$: We need information on joint events, e.g. $\mathbb{P}[X_i = X_j = 1]$

Lasserre Hierarchy for Integer Programming

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

- Hierarchy at level δ depends on (y_I) for $|I| \leq O(\delta)$.
- Let X be a random vector associated with a representing probability measure $\mu \in \mathcal{M}^+(K)$ for $\boldsymbol{y}$:

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

- Standard interpretation of LP-relaxation of *IP*: round x_i to $X_i = 1$ with probability x_i^*, so $\mathbb{E}[\boldsymbol{c}^T X] = \boldsymbol{c}^T \boldsymbol{x}^*$.
- But rounded solution X might violate the constraints: $AX \not\geq \boldsymbol{b}$: We need information on joint events, e.g. $\mathbb{P}[X_i = X_j = 1]$
- Lasserre hierarchy brings *correlation structure* for joint events of bounded cardinality.

Lasserre Hierarchy for Integer Programming

The feasible set of the Lasserre hierarchy (Las$_\delta$) for Problem (IP) takes the following form:

Definition

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

$$y_\emptyset = 1$$

$$M_{\delta+1}(\boldsymbol{y}) := (y_{I \cup J})_{|I|,|J| \leq \delta+1} \succeq 0$$

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

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

First round of the hierarchy: $\mathcal{L}_0$

$$\mathcal{L}_0 = \{(\boldsymbol{y}_I)_{|I|\leq 2} :\ y_\emptyset = 1,\ M_1(\boldsymbol{y}) = 0,\ M_0(g_i \boldsymbol{y}) \succeq 0, \forall i \in [m]\},$$

where $g_i(\boldsymbol{x}) = \boldsymbol{a}_i^T \boldsymbol{x} - b_i$.

First round of the hierarchy: $\mathcal{L}_0$

$$\mathcal{L}_0 = \{(\boldsymbol{y}_I)_{|I|\leq 2} : \ y_\emptyset = 1,\ M_1(\boldsymbol{y}) = 0,\ M_0(g_i \boldsymbol{y}) \succeq 0, \forall i \in [m]\},$$

where $g_i(\boldsymbol{x}) = \boldsymbol{a}_i^T \boldsymbol{x} - b_i$.

- Denote by $\boldsymbol{z} \in \mathbb{R}^n$ the vector with elements $z_i = y_{\{i\}}$
- Denote by $Z \in \mathbb{S}^n$ the matrix with coordinates $Z_{ij} = y_{\{i,j\}}$

First round of the hierarchy: $\mathcal{L}_0$

$$\mathcal{L}_0 = \{(\boldsymbol{y}_I)_{|I|\leq 2} : \ y_\emptyset = 1,\ M_1(\boldsymbol{y}) = 0,\ M_0(g_i\boldsymbol{y}) \succeq 0, \forall i \in [m]\},$$

where $g_i(\boldsymbol{x}) = \boldsymbol{a}_i^T \boldsymbol{x} - b_i$.

- Denote by $\boldsymbol{z} \in \mathbb{R}^n$ the vector with elements $z_i = y_{\{i\}}$
- Denote by $Z \in \mathbb{S}^n$ the matrix with coordinates $Z_{ij} = y_{\{i,j\}}$

We have:

- $M_1(\boldsymbol{y}) = \begin{pmatrix} y_\emptyset & \boldsymbol{z}^T \\ \boldsymbol{z} & Z \end{pmatrix} = \begin{pmatrix} 1 & \boldsymbol{z}^T \\ \boldsymbol{z} & Z \end{pmatrix},$
- $M_0(g_i\boldsymbol{y}) = \sum_{j\in[n]} a_{ij} y_{\{j\}} - b_i y_\emptyset = \boldsymbol{a}_i^T \boldsymbol{z} - b_i$
- $Z_{ii} = y_{\{i,i\}} = y_{\{i\}} = z_i$

So, the first level of the Lasserre Hierarchy coinicides with the *recipe* to construct the SDP relaxation of a problem with binary variables:

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

Finite convergence of hierarchy

So (Las$_\delta$) rewrites $\min\{\boldsymbol{c}^T\boldsymbol{x} : \boldsymbol{x} \in \mathcal{L}_\delta^{proj}\}$.
By construction, the $\mathcal{L}_\delta^{proj}$ are nested and contain K:

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

Finite convergence of hierarchy

So (Las$_\delta$) rewrites $\min\{\boldsymbol{c}^T\boldsymbol{x} : \boldsymbol{x} \in \mathcal{L}_\delta^{proj}\}$.
By construction, the $\mathcal{L}_\delta^{proj}$ are nested and contain K:

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

By convexity of $\mathcal{L}_\delta^{proj}$, these sets also contain the *integer hull* H of Problem (IP):

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

Finite convergence of hierarchy

So (Las$_\delta$) rewrites $\min\{\boldsymbol{c}^T\boldsymbol{x} : \boldsymbol{x} \in \mathcal{L}_\delta^{proj}\}$.
By construction, the $\mathcal{L}_\delta^{proj}$ are nested and contain K:

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

By convexity of $\mathcal{L}_\delta^{proj}$, these sets also contain the *integer hull* H of Problem (IP):

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

In fact, the hierarchy converges after at most n rounds to H:

Proposition

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

Finite convergence of hierarchy

Proposition

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

- This means: $\exists \delta \leq n : p_\delta^* = p^*$.
  (because solving the IP is equivalent to minimizing $\boldsymbol{c}^T x$ over integer hull H of K)
- So, the IP is equivalent to a big SDP (but it might have exponentially many variables)
- Proof relies on several lemmas (cf. next slide)

Finite convergence of hierarchy

Lemma

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

Finite convergence of hierarchy

Lemma

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

Lemma

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

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

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

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

Finite convergence of the hierarchy

Corollary

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

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

Finite convergence of the hierarchy

Corollary

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

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

Iterating the above result, we get: $\forall S \subseteq [n]$,

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

Finite convergence of the hierarchy

Corollary

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

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

Iterating the above result, we get: $\forall S \subseteq [n]$,

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

And in particular for $S = [n]$ and $\delta = n$, we obtain:

$$\mathcal{L}_{n|n} \subseteq \textbf{conv}\,\{\boldsymbol{z} \in \mathcal{L}_0 : \boldsymbol{z} \in \{0,1\}^n\}.$$

$$\implies \mathcal{L}_n^{proj} \subseteq H := \textbf{conv}\,\{\boldsymbol{x} \in \{0,1\}^n : A\boldsymbol{x} \geq \boldsymbol{b}\}.$$