Sums of Squares, Moments and Applications in Polynomial Optimization

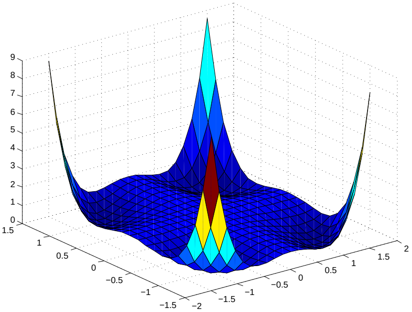
CWI

TILBURG UNIVERSITY

Monique Laurent

Fields Distinguished Lecture Series - May 10, 2021

What is polynomial optimization?



(P)

Minimize a polynomial function f over a region

$$K = \{x \in \mathbb{R}^n : g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}$$

defined by polynomial inequalities (and equations)

Some instances

Testing nonnegativity of polynomials

The unconstrained quadratic case is easy

The quadratic form $x^{\mathsf{T}}Mx$ is nonnegative over $\mathbb{R}^n$ if and only if the matrix M is positive semidefinite ($M \succeq 0$)

This can be tested in **polynomial time**, using Gaussian elimination

Testing nonnegativity of polynomials

The unconstrained quadratic case is easy

The quadratic form $x^\mathsf{T} M x$ is nonnegative over $\mathbb{R}^n$ if and only if the matrix M is positive semidefinite ($M \succeq 0$)

This can be tested in **polynomial time**, using Gaussian elimination

Constrained quadratic / unconstrained quartic is hard

Testing nonnegativity of polynomials

The unconstrained quadratic case is easy

The quadratic form $x^\mathsf{T} M x$ is nonnegative over $\mathbb{R}^n$ if and only if the matrix M is positive semidefinite $(M \succeq 0)$

This can be tested in **polynomial time**, using Gaussian elimination

Constrained quadratic / unconstrained quartic is hard

Testing matrix copositivity: co-NP complete [Kabadi-Murty 1987]

A symmetric matrix M is **copositive** if $x^\mathsf{T} M x = \sum_{i,j} M_{ij} x_i x_j \geq 0 \ \ \forall x \geq 0$

Testing nonnegativity of polynomials

The unconstrained quadratic case is easy

The quadratic form $x^\mathsf{T} Mx$ is nonnegative over $\mathbb{R}^n$ if and only if the matrix M is positive semidefinite ($M \succeq 0$)

This can be tested in **polynomial time**, using Gaussian elimination

Constrained quadratic / unconstrained quartic is hard

Testing matrix copositivity: co-NP complete [Kabadi-Murty 1987]

A symmetric matrix M is **copositive** if $x^\mathsf{T} Mx = \sum_{i,j} M_{ij} x_i x_j \geq 0 \quad \forall x \geq 0$

Equivalently, the quartic polynomial $\sum_{i,j} M_{ij} x_i^2 x_j^2$ is nonnegative over $\mathbb{R}^n$

Testing nonnegativity of polynomials

The unconstrained quadratic case is easy

The quadratic form $x^\top Mx$ is nonnegative over $\mathbb{R}^n$ if and only if the matrix M is positive semidefinite ($M \succeq 0$)

This can be tested in **polynomial time**, using Gaussian elimination

Constrained quadratic / unconstrained quartic is hard

Testing matrix copositivity: co-NP complete [Kabadi-Murty 1987]

A symmetric matrix M is **copositive** if $x^\top Mx = \sum_{i,j} M_{ij}x_ix_j \geq 0 \quad \forall x \geq 0$

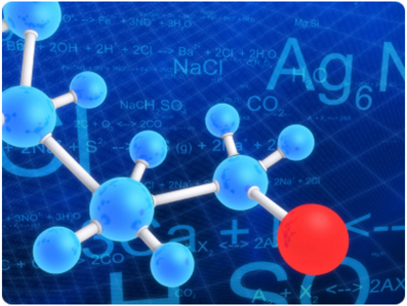
Equivalently, the quartic polynomial $\sum_{i,j} M_{ij}x_i^2x_j^2$ is nonnegative over $\mathbb{R}^n$

Testing convexity: NP-hard [Ahmadi et al. 2013]

A polynomial $f(x)$ is **convex** if and only if its Hessian matrix $H(f)(x)$ is positive semidefinite

Equivalently, $g(x, y) = y^\top H(f)(x)y$ is nonnegative on $\mathbb{R}^n \times \mathbb{R}^n$

Example from distance geometry



Reconstruct the locations of objects (say) in 3D from partial measurements of mutual distances

Given (partial) pairwise distances $d = (d_{ij})_{ij \in E}$, find (if possible) locations $u_1, \cdots, u_n \in \mathbb{R}^k$ in given dimension k ($k = 1, 2, 3, ..$) such that

$$\|u_i - u_j\|^2 = d_{ij} \qquad \text{for all } \{i, j\} \in E$$

Formulations via SDP and polynomial optimization

Find (if possible) vectors $u_1, \cdots, u_n \in \mathbb{R}^k$ ($k = 1, 2, 3, ..$) such that

$$\|u_i - u_j\|^2 = d_{ij} \quad (\{i, j\} \in E)$$

Formulations via SDP and polynomial optimization

Find (if possible) vectors $u_1, \cdots, u_n \in \mathbb{R}^k$ ($k = 1, 2, 3, ..$) such that

$$\|u_i - u_j\|^2 = d_{ij} \quad (\{i,j\} \in E)$$

$$\Updownarrow \qquad\qquad X = (\langle u_i, u_j \rangle)$$

Find (if possible) a solution X with rank $\leq k$ to the semidefinite program

$$X \succeq 0, \quad X_{ii} + X_{jj} - 2X_{ij} = d_{ij} \quad (\{i,j\} \in E)$$

Formulations via SDP and polynomial optimization

Find (if possible) vectors $u_1, \cdots, u_n \in \mathbb{R}^k$ ($k = 1, 2, 3, ..$) such that

$$\|u_i - u_j\|^2 = d_{ij} \quad (\{i,j\} \in E)$$

$$\Updownarrow \qquad X = (\langle u_i, u_j \rangle)$$

Find (if possible) a solution X with rank $\leq k$ to the semidefinite program

$$X \succeq 0, \quad X_{ii} + X_{jj} - 2X_{ij} = d_{ij} \quad (\{i,j\} \in E)$$

$$\Updownarrow$$

Decide if $p_{\min} = 0$ and find a global minimizer to the quartic polynomial

$$\min_{x \in \mathbb{R}^{kn}} p(x) = \sum_{\{i,j\} \in E} \Big(d_{ij} - \sum_{h=1}^{k} (x_{ih} - x_{jh})^2 \Big)^2$$

Formulations via SDP and polynomial optimization

Find (if possible) vectors $u_1, \cdots, u_n \in \mathbb{R}^k$ ($k = 1, 2, 3, ..$) such that

$$\|u_i - u_j\|^2 = d_{ij} \quad (\{i,j\} \in E)$$

$$\Updownarrow \qquad X = (\langle u_i, u_j \rangle)$$

Find (if possible) a solution X with rank $\leq k$ to the semidefinite program

$$X \succeq 0, \quad X_{ii} + X_{jj} - 2X_{ij} = d_{ij} \quad (\{i,j\} \in E)$$

$$\Updownarrow$$

Decide if $p_{\min} = 0$ and find a global minimizer to the quartic polynomial

$$\min_{x \in \mathbb{R}^{kn}} p(x) = \sum_{\{i,j\} \in E} \Big(d_{ij} - \sum_{h=1}^{k} (x_{ih} - x_{jh})^2 \Big)^2$$

Hard problem, already in dimension $k = 1$ when G is cycle C_n [Saxe'79]

Formulations via SDP and polynomial optimization

Find (if possible) vectors $u_1, \cdots, u_n \in \mathbb{R}^k$ ($k = 1, 2, 3, ..$) such that

$$\|u_i - u_j\|^2 = d_{ij} \quad (\{i,j\} \in E)$$

$$\Updownarrow \qquad X = (\langle u_i, u_j \rangle)$$

Find (if possible) a solution X with rank $\leq k$ to the semidefinite program

$$X \succeq 0, \quad X_{ii} + X_{jj} - 2X_{ij} = d_{ij} \quad (\{i,j\} \in E)$$

$$\Updownarrow$$

Decide if $p_{\min} = 0$ and find a global minimizer to the quartic polynomial

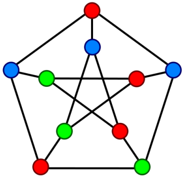
$$\min_{x \in \mathbb{R}^{kn}} p(x) = \sum_{\{i,j\} \in E} \left(d_{ij} - \sum_{h=1}^{k} (x_{ih} - x_{jh})^2 \right)^2$$

Hard problem, already in dimension $k = 1$ when G is cycle C_n [Saxe'79]
Given $a_1, \ldots, a_n \in \mathbb{N}$, assign distance $d_{i,i+1} = a_i$ to the edges of C_n. Then

$$\exists \text{ locations in } \mathbb{R} \iff \exists \epsilon \in \{\pm 1\}^n \text{ s.t. } \sum_{i=1}^{n} \epsilon_i a_i = 0$$

$\rightsquigarrow$ *hard partition problem*

Examples from combinatorial problems in graphs



$\alpha = 4 \quad \chi = 3$

- **stability number** $\alpha(G)$:

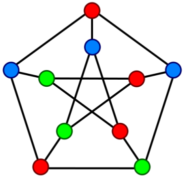
maximum cardinality of a set of pairwise non-adjacent vertices (**stable set**)

- **coloring number** $\chi(G)$:

minimum number of colors needed to properly color the vertices of G

$\alpha(G)$, $\chi(G)$ are NP-complete [Karp 1972]

Examples from combinatorial problems in graphs



$\alpha = 4 \quad \chi = 3$

- **stability number** $\alpha(G)$:

  maximum cardinality of a set of pairwise non-adjacent vertices (**stable set**)

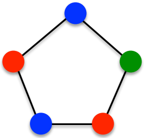
- **coloring number** $\chi(G)$:

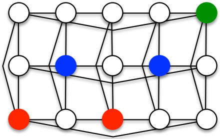
  minimum number of colors needed to properly color the vertices of G

$\alpha(G)$, $\chi(G)$ are NP-complete [Karp 1972]

Chvátal's reduction of coloring to the stability number:

$\chi(G)$ is the smallest integer c such that $\alpha(G \square K_c) = \lvert V(G)\rvert$





Polynomial optimization formulations for $\alpha(G)$

- Basic 0/1 formulation:

$$\alpha(G) = \max \sum_{i \in V} x_i \ \text{ s.t. } \ x_i x_j = 0 \ (\{i,j\} \in E), \ x_i^2 = x_i \ (i \in V)$$

Polynomial optimization formulations for $\alpha(G)$

- Basic 0/1 formulation:

$$\alpha(G) = \max \sum_{i \in V} x_i \;\; \text{s.t.} \;\; x_i x_j = 0 \; (\{i,j\} \in E), \; x_i^2 = x_i \; (i \in V)$$

- Motzkin-Straus formulation:

$$\frac{1}{\alpha(G)} = \min \; x^T (I + A_G) x \;\; \text{s.t.} \;\; \sum_{i \in V} x_i = 1, \; x_i \geq 0 \; (i \in V)$$

Polynomial optimization formulations for $\alpha(G)$

- Basic 0/1 formulation:

$$\alpha(G) = \max \sum_{i \in V} x_i \ \text{ s.t. } \ x_i x_j = 0 \ (\{i,j\} \in E), \ x_i^2 = x_i \ (i \in V)$$

- Motzkin-Straus formulation:

$$\frac{1}{\alpha(G)} = \min \ x^T (I + A_G) x \ \text{ s.t. } \ \sum_{i \in V} x_i = 1, \ x_i \geq 0 \ (i \in V)$$

$$\frac{1}{\alpha(G)} = \min \ (x^{\circ 2})^T (I + A_G) x^{\circ 2} \ \text{ s.t. } \ \sum_{i \in V} x_i^2 = 1$$

Polynomial optimization formulations for $\alpha(G)$

- Basic 0/1 formulation:

$$\alpha(G) = \max \sum_{i \in V} x_i \ \text{ s.t. } \ x_i x_j = 0 \ (\{i,j\} \in E), \ x_i^2 = x_i \ (i \in V)$$

- Motzkin-Straus formulation:

$$\frac{1}{\alpha(G)} = \min \ x^T (I + A_G) x \ \text{ s.t. } \ \sum_{i \in V} x_i = 1, \ x_i \geq 0 \ (i \in V)$$

$$\frac{1}{\alpha(G)} = \min \ (x^{\circ 2})^T (I + A_G) x^{\circ 2} \ \text{ s.t. } \ \sum_{i \in V} x_i^2 = 1$$

- Copositive formulation:

$$\alpha(G) = \min \ \lambda \ \text{ s.t. } \ \lambda(I + A_G) - J \text{ is copositive}$$

Polynomial optimization formulations for $\alpha(G)$

- Basic 0/1 formulation:

$$\alpha(G) = \max \sum_{i \in V} x_i \text{ s.t. } x_i x_j = 0 \ (\{i,j\} \in E), \ x_i^2 = x_i \ (i \in V)$$

- Motzkin-Straus formulation:

$$\frac{1}{\alpha(G)} = \min \ x^T(I + A_G)x \text{ s.t. } \sum_{i \in V} x_i = 1, \ x_i \geq 0 \ (i \in V)$$

$$\frac{1}{\alpha(G)} = \min \ (x^{\circ 2})^T(I + A_G)x^{\circ 2} \text{ s.t. } \sum_{i \in V} x_i^2 = 1$$

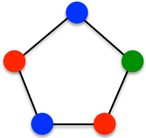
- Copositive formulation:

$$\alpha(G) = \min \ \lambda \text{ s.t. } \lambda(I + A_G) - J \text{ is copositive}$$

$\rightsquigarrow$ optimization over the **boolean cube** $\{0,1\}^n$, the **standard simplex** Δ_n, the **unit sphere** $\mathbb{S}^{n-1}$, the **copositive cone** COP_n

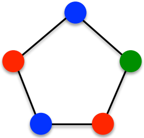
$\rightsquigarrow$ *More in Lecture 3*

Basic semidefinite bounds for $\alpha(G)$ and $\chi(G)$



S stable $\rightsquigarrow x = (1,0,0,1,0)^{\mathsf{T}} \rightsquigarrow X = \begin{pmatrix}1\\x\end{pmatrix}\begin{pmatrix}1\\x\end{pmatrix}^{\mathsf{T}}$

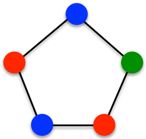
Basic semidefinite bounds for $\alpha(G)$ and $\chi(G)$



$$S \text{ stable} \rightsquigarrow x = (1,0,0,1,0)^{\mathsf{T}} \rightsquigarrow X = \binom{1}{x}\binom{1}{x}^{\mathsf{T}}$$

$X \succeq 0$ positive semidefinite

Basic semidefinite bounds for $\alpha(G)$ and $\chi(G)$

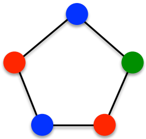


S stable $\rightsquigarrow x = (1,0,0,1,0)^{\mathsf{T}} \rightsquigarrow X = \begin{pmatrix}1\\x\end{pmatrix}\begin{pmatrix}1\\x\end{pmatrix}^{\mathsf{T}}$

$X \succeq 0$ positive semidefinite

$X \geq 0$ entry-wise nonnegative

Basic semidefinite bounds for $\alpha(G)$ and $\chi(G)$



S stable $\rightsquigarrow x = (1,0,0,1,0)^\mathsf{T} \rightsquigarrow X = \binom{1}{x}\binom{1}{x}^\mathsf{T}$

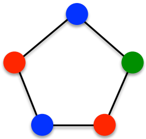
$X \succeq 0$ positive semidefinite

$X \geq 0$ entry-wise nonnegative

Theta number: [Lovász 1979]

$$\vartheta(G) = \max_{X \succeq 0} \sum_{i \in V} X_{0i} \;\; \text{s.t.} \;\; X_{00} = 1, \; X_{0i} = X_{ii} \; (i \in V), \; X_{ij} = 0 \;\; (\{i,j\} \in E)$$

Basic semidefinite bounds for $\alpha(G)$ and $\chi(G)$



$$S \text{ stable} \rightsquigarrow x = (1,0,0,1,0)^{\mathsf{T}} \rightsquigarrow X = \binom{1}{x}\binom{1}{x}^{\mathsf{T}}$$

$X \succeq 0$ positive semidefinite

$X \geq 0$ entry-wise nonnegative

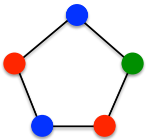
Theta number: [Lovász 1979]

$$\vartheta(G) = \max_{X \succeq 0} \sum_{i \in V} X_{0i} \ \text{ s.t. } X_{00} = 1,\ X_{0i} = X_{ii}\ (i \in V),\ X_{ij} = 0\ \ (\{i,j\} \in E)$$

Strengthen with non-negativity: [McEliece et al. 1978] [Schrijver 1979]

$$\vartheta'(G) = \max_{X \succeq 0, X \geq 0} \sum_{i \in V} X_{ii} \ \text{ s.t. } X_{00} = 1,\ X_{0i} = X_{ii}\ (i \in V),\ X_{ij} = 0\ \ (\{i,j\} \in E)$$

Basic semidefinite bounds for $\alpha(G)$ and $\chi(G)$



S stable $\rightsquigarrow x = (1,0,0,1,0)^{\mathsf{T}} \rightsquigarrow X = \binom{1}{x}\binom{1}{x}^{\mathsf{T}}$

$X \succeq 0$ positive semidefinite

$X \geq 0$ entry-wise nonnegative

Theta number: [Lovász 1979]

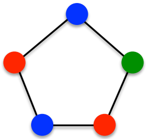
$$\vartheta(G) = \max_{X \succeq 0} \sum_{i \in V} X_{0i} \ \text{ s.t. } X_{00} = 1,\ X_{0i} = X_{ii}\ (i \in V),\ X_{ij} = 0\ \ (\{i,j\} \in E)$$

Strengthen with non-negativity: [McEliece et al. 1978] [Schrijver 1979]

$$\vartheta'(G) = \max_{X \succeq 0, X \geq 0} \sum_{i \in V} X_{ii} \ \text{ s.t. } X_{00} = 1,\ X_{0i} = X_{ii}\ (i \in V),\ X_{ij} = 0\ \ (\{i,j\} \in E)$$

'Sandwich' inequalities: $\alpha(G) \leq \vartheta'(G) \leq \vartheta(G) \leq \chi(\overline{G})$

Basic semidefinite bounds for $\alpha(G)$ and $\chi(G)$



S stable $\rightsquigarrow x = (1,0,0,1,0)^T \rightsquigarrow X = \binom{1}{x}\binom{1}{x}^T$

$X \succeq 0$ positive semidefinite

$X \geq 0$ entry-wise nonnegative

Theta number: [Lovász 1979]

$$\vartheta(G) = \max_{X \succeq 0} \sum_{i \in V} X_{0i} \text{ s.t. } X_{00} = 1,\ X_{0i} = X_{ii}\ (i \in V),\ X_{ij} = 0\ (\{i,j\} \in E)$$

Strengthen with non-negativity: [McEliece et al. 1978] [Schrijver 1979]

$$\vartheta'(G) = \max_{X \succeq 0, X \geq 0} \sum_{i \in V} X_{ii} \text{ s.t. } X_{00} = 1,\ X_{0i} = X_{ii}\ (i \in V),\ X_{ij} = 0\ (\{i,j\} \in E)$$

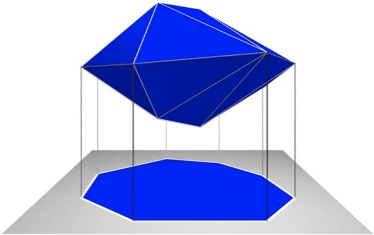
'Sandwich' inequalities: $\alpha(G) \leq \vartheta'(G) \leq \vartheta(G) \leq \chi(\overline{G})$

Stronger bounds?

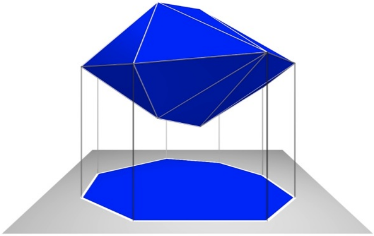
Some key ideas

to get stronger bounds

- **Lift to higher dimensional space:** add new variables modeling products of original variables, such as $x_i x_j, x_i x_j x_k, x_i x_j x_k x_l, \ldots$



- **Lift to higher dimensional space:** add new variables modeling products of original variables, such as $x_i x_j, x_i x_j x_k, x_i x_j x_k x_l, \ldots$



- Use **sums of squares of polynomials** as a 'proxy' for non-negativity of polynomials to get **tractable** relaxations

  Key fact: One can **model sums of squares of polynomials efficiently** using **semidefinite programming (SDP)**

Model sums of squares of polynomials with SDP

$$f(x) = \sum_{|\alpha| \leq 2d} f_\alpha x^\alpha \quad \text{is a } \textbf{sum of squares of polynomials}$$

$f(x) = \sum_i p_i(x)^2$

Model sums of squares of polynomials with SDP

$$f(x) = \sum_{|\alpha| \leq 2d} f_\alpha x^\alpha \quad \text{is a } \textbf{sum of squares of polynomials}$$

$f(x) = \sum_i p_i(x)^2$ [write $p_i(x) = \overline{p_i}^T [x]_d, \quad [x]_d = (x^\alpha)$]

Model sums of squares of polynomials with SDP

$$f(x) = \sum_{|\alpha| \leq 2d} f_\alpha x^\alpha$$ is a **sum of squares of polynomials**

$f(x) = \sum_i p_i(x)^2$ [write $p_i(x) = \overline{p_i}^T [x]_d, \quad [x]_d = (x^\alpha)$]

$$\Updownarrow$$

$$f(x) = \sum_i [x]_d^T \overline{p_i}\ \overline{p_i}^T [x]_d = [x]_d^T \underbrace{\left(\sum_i \overline{p_i}\ \overline{p_i}^T \right)}_{M \succeq 0} [x]_d = \sum_{\beta,\gamma} M_{\beta,\gamma} x^{\beta+\gamma}$$

Model sums of squares of polynomials with SDP

$$f(x) = \sum_{|\alpha|\le 2d} f_\alpha x^\alpha$$ is a **sum of squares of polynomials**

$f(x) = \sum_i p_i(x)^2$ [write $p_i(x) = \overline{p_i}^T [x]_d, \quad [x]_d = (x^\alpha)$]

$$\Updownarrow$$

$$f(x) = \sum_i [x]_d^T \overline{p_i}\ \overline{p_i}^T [x]_d = [x]_d^T \Big(\underbrace{\sum_i \overline{p_i}\ \overline{p_i}^T}_{M \succeq 0} \Big) [x]_d = \sum_{\beta,\gamma} M_{\beta,\gamma} x^{\beta+\gamma}$$

$$\Updownarrow$$

The SDP $\left\{ \begin{array}{l} \displaystyle\sum_{\beta,\gamma|\beta+\gamma=\alpha} M_{\beta,\gamma} = f_\alpha \quad (|\alpha| \le 2d) \\ M \succeq 0 \end{array} \right.$ **is feasible**

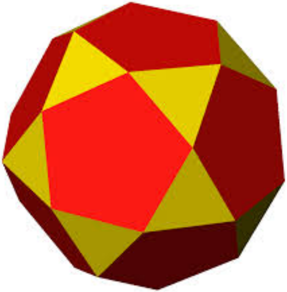
Gram-matrix method [Powers-Wörmann 1998]

Linear Programming vs Semidefinite Programming

Optimize a linear function over

a polyhedron

$a_j^T x = b_j, \; x \geq 0$



LP

a convex set (spectrahedron)

$\langle A_j, X \rangle = b_j, \; X \succeq 0$

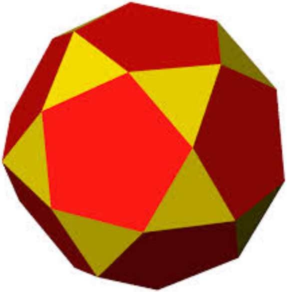
SDP

Linear Programming vs Semidefinite Programming

Optimize a linear function over

a polyhedron

$a_j^\mathsf{T} x = b_j, \; x \geq 0$



LP

a convex set (spectrahedron)

$\langle A_j, X \rangle = b_j, \; X \succeq 0$

SDP

There are efficient algorithms to solve LP and SDP (up to any precision)

About the complexity of SDP

- 1980's: There are efficient algorithms to find an **almost optimal** solution, **under some assumptions**

  Roughly: one needs a feasible point, an inscribed ball and a circumscribed ball to the feasible region

  - Grötschel-Lovász-Schrijver: based on Khachiyan *ellipsoid method*

  - Karmarkar, Nesterov-Nemirovski: *interior point algorithms*

About the complexity of SDP

- 1980's: There are efficient algorithms to find an **almost optimal** solution, **under some assumptions**

  Roughly: one needs a feasible point, an inscribed ball and a circumscribed ball to the feasible region

  - Grötschel-Lovász-Schrijver: based on Khachiyan *ellipsoid method*

  - Karmarkar, Nesterov-Nemirovski: *interior point algorithms*

- **Testing feasibility of SDP**: Given rational A_j, b_j, decide

  (F) $\exists X \succeq 0$ s.t. $\langle A_j, X\rangle = b_j \ \ (j \in [m])$?

  - Ramana (1997): (F) $\in$ NP $\Longleftrightarrow$ (F) $\in$ co-NP

About the complexity of SDP

- 1980's: There are efficient algorithms to find an **almost optimal** solution, **under some assumptions**

  Roughly: one needs a feasible point, an inscribed ball and a circumscribed ball to the feasible region

  - Grötschel-Lovász-Schrijver: based on Khachiyan *ellipsoid method*

  - Karmarkar, Nesterov-Nemirovski: *interior point algorithms*

- **Testing feasibility of SDP**: Given rational A_j, b_j, decide

  (F) $\exists X \succeq 0$ s.t. $\langle A_j, X \rangle = b_j \ \ (j \in [m])$?

  - Ramana (1997): (F) $\in$ NP $\iff$ (F) $\in$ co-NP

  - Porkolab-Khachiyan (1997): (F) $\in$ P for fixed n or m

  $mn^{O(\min\{m,n^2\})}$ arithmetic operations on $Ln^{O(\min\{m,n^2\})}$-bit length numbers

About the complexity of SDP

- 1980's: There are efficient algorithms to find an **almost optimal** solution, **under some assumptions**

  Roughly: one needs a feasible point, an inscribed ball and a circumscribed ball to the feasible region

  - Grötschel-Lovász-Schrijver: based on Khachiyan *ellipsoid method*

  - Karmarkar, Nesterov-Nemirovski: *interior point algorithms*

- **Testing feasibility of SDP**: Given rational A_j, b_j, decide

  (F) $\exists X \succeq 0$ s.t. $\langle A_j, X\rangle = b_j$ $(j \in [m])$?

  - Ramana (1997): (F) $\in$ NP $\Longleftrightarrow$ (F) $\in$ co-NP

  - Porkolab-Khachiyan (1997): (F) $\in$ P for fixed n or m

  $mn^{O(\min\{m,n^2\})}$ *arithmetic operations on* $Ln^{O(\min\{m,n^2\})}$*-bit length numbers*

- Well developed **duality theory** for LP, SDP, conic programs (with no duality gap under some strict feasibility conditions)

General approach to polynomial optimization

Strategy

(P) $$f_{\min} = \min_{x \in K} f(x)$$

Approximate **(P)** by a hierarchy of convex (semidefinite) relaxations

These relaxations can be constructed using

sums of squares of polynomials

and

the dual theory of moments

Shor (1987), Nesterov (2000), **Lasserre, Parrilo** (2000–)

Sums of squares approach

Strategy (use sums of squares)

(P) $$f_{\min} = \min_{x \in K} f(x) = \sup_{\lambda \in \mathbb{R}} \lambda \ \text{ s.t. } \ f(x) - \lambda \geq 0 \ \forall x \in K$$

Testing whether a polynomial f is nonnegative is **hard**

but one can test the *sufficient condition*:

f is a sum of squares of polynomials (SoS)

using semidefinite programming

Are all nonnegative polynomials SoS?

Are all nonnegative polynomials SoS?

Hilbert [1888]: *Every nonnegative polynomial in n variables and even degree d is a sum of squares of polynomials*

Are all nonnegative polynomials SoS?

Hilbert [1888]: *Every nonnegative polynomial in* n *variables and even degree* d *is a sum of squares of polynomials* $\iff$
$n = 1$,

Are all nonnegative polynomials SoS?

Hilbert [1888]: *Every nonnegative polynomial in n variables and even degree d is a sum of squares of polynomials* $\iff$ $n = 1$, *or* $d = 2$,

Are all nonnegative polynomials SoS?

Hilbert [1888]: *Every nonnegative polynomial in n variables and even degree d is a sum of squares of polynomials* $\iff$

$n = 1$, *or* $d = 2$, *or* ($n = 2$ *and* $d = 4$)

Are all nonnegative polynomials SoS?

Hilbert [1888]: *Every nonnegative polynomial in n variables and even degree d is a sum of squares of polynomials* $\iff$
$n = 1$, *or* $d = 2$, *or* ($n = 2$ *and* $d = 4$)

Hilbert's 17th problem [1900]: *Is every nonnegative polynomial is a sum of squares of* **rational** *functions?*

Are all nonnegative polynomials SoS?

Hilbert [1888]: *Every nonnegative polynomial in* n *variables and even degree* d *is a sum of squares of polynomials* $\iff$
$n = 1$, *or* $d = 2$, *or* ($n = 2$ *and* $d = 4$)

Hilbert's 17th problem [1900]: *Is every nonnegative polynomial is a sum of squares of* **rational** *functions?*

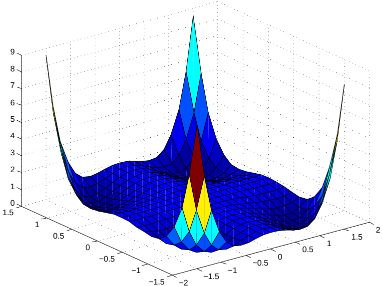
Artin [1927]: Yes

Are all nonnegative polynomials SoS?

Hilbert [1888]: *Every nonnegative polynomial in n variables and even degree d is a sum of squares of polynomials* $\iff$
$n = 1$, *or* $d = 2$, *or* ($n = 2$ and $d = 4$)

Hilbert's 17th problem [1900]: *Is every nonnegative polynomial is a sum of squares of* **rational** *functions?*

Artin [1927]: Yes



Motzkin [1967]:
$p = x^4y^2 + x^2y^4 + 1 - 3x^2y^2$
is nonnegative,
not a sum of squares,
but $(x^2 + y^2)^2 p$ is SoS

Positivity certificates over K

$$K = \{x \mid g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}$$

Positivity certificates over K

$$K = \{x \mid g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}$$

Quadratic module: $Q(g) = \{s_0 + s_1 g_1 + \ldots + s_m g_m \mid s_j \text{ SoS}\}$

Positivity certificates over K

$$K = \{x \mid g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}$$

Quadratic module: $Q(g) = \{s_0 + s_1 g_1 + \ldots + s_m g_m \mid s_j \text{ SoS}\}$

Preordering: $P(g) = \{\sum_{e \in \{0,1\}^m} s_e g_1^{e_1} \cdots g_m^{e_m} \mid s_e \text{ SoS}\} \supset Q(g)$

Positivity certificates over K

$$K = \{x \mid g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}$$

Quadratic module: $Q(g) = \{s_0 + s_1 g_1 + \ldots + s_m g_m \mid s_j \text{ SoS}\}$

Preordering: $P(g) = \{\sum_{e \in \{0,1\}^m} s_e g_1^{e_1} \cdots g_m^{e_m} \mid s_e \text{ SoS}\} \supset Q(g)$

Theorem: Assume K compact.

- [Schmüdgen 1991] $f > 0$ on $K \implies f \in P(g)$

Positivity certificates over K

$$K = \{x \mid g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}$$

Quadratic module: $Q(g) = \{s_0 + s_1 g_1 + \ldots + s_m g_m \mid s_j \text{ SoS}\}$

Preordering: $P(g) = \{\sum_{e \in \{0,1\}^m} s_e g_1^{e_1} \cdots g_m^{e_m} \mid s_e \text{ SoS}\} \supset Q(g)$

Theorem: Assume K compact.

- [Schmüdgen 1991] $f > 0$ on $K \implies f \in P(g)$
- [Putinar 1993] **Archimedean condition:** $\exists R : \; R - \sum_i x_i^2 \in Q(g)$

$$f > 0 \text{ on } K \implies f \in Q(g)$$

Positivity certificates over K

$$K = \{x \mid g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}$$

Quadratic module: $Q(g) = \{s_0 + s_1 g_1 + \ldots + s_m g_m \mid s_j \text{ SoS}\}$

Preordering: $P(g) = \{\sum_{e \in \{0,1\}^m} s_e g_1^{e_1} \cdots g_m^{e_m} \mid s_e \text{ SoS}\} \supset Q(g)$

Theorem: Assume K compact.

- [Schmüdgen 1991] $f > 0$ on $K \implies f \in P(g)$
- [Putinar 1993] **Archimedean condition:** $\exists R : \; R - \sum_i x_i^2 \in Q(g)$

$$f > 0 \text{ on } K \implies f \in Q(g)$$

Observation: If we know a ball of radius R containing K, then just add the (redundant) constraint $R^2 - \sum_i x_i^2 \geq 0$ to the description of K

SoS relaxations for (P)

Truncated quadratic module:

$$Q(g)_t := \{ \underbrace{s_0}_{\deg \leq 2t} + \underbrace{s_1 g_1}_{\deg \leq 2t} + \ldots + \underbrace{s_m g_m}_{\deg \leq 2t} \mid s_j \ \text{ SoS}\}$$

SoS relaxations for (P)

Truncated quadratic module:

$$Q(g)_t := \{ \underbrace{s_0}_{\deg \leq 2t} + \underbrace{s_1 g_1}_{\deg \leq 2t} + \ldots + \underbrace{s_m g_m}_{\deg \leq 2t} \mid s_j \ \text{ SoS} \}$$

Replace

(P) $$\boxed{f_{\min} = \ \inf_{x \in K} f(x) \ = \sup \ \lambda \ \text{ s.t. } \ f - \lambda \geq 0 \ \text{ on } K}$$

by

(SOSt) $$\boxed{f_t^{\text{sos}} = \sup \ \lambda \ \text{ s.t. } \ f - \lambda \in Q(g)_t}$$

SoS relaxations for (P)

Truncated quadratic module:

$$Q(g)_t := \{ \underbrace{s_0}_{\deg \leq 2t} + \underbrace{s_1 g_1}_{\deg \leq 2t} + \ldots + \underbrace{s_m g_m}_{\deg \leq 2t} \mid s_j \ \text{SoS} \}$$

Replace

(P) $$f_{\min} = \inf_{x \in K} f(x) = \sup \ \lambda \ \text{ s.t. } \ f - \lambda \geq 0 \ \text{ on } K$$

by

(SOSt) $$f_t^{\text{sos}} = \sup \ \lambda \ \text{ s.t. } \ f - \lambda \in Q(g)_t$$

- Each bound f_t^{sos} can be computed with SDP

SoS relaxations for (P)

Truncated quadratic module:

$$Q(g)_t := \{ \underbrace{s_0}_{\deg \leq 2t} + \underbrace{s_1 g_1}_{\deg \leq 2t} + \ldots + \underbrace{s_m g_m}_{\deg \leq 2t} \mid s_j \ \text{SoS} \}$$

Replace

(P) $f_{\min} = \inf_{x \in K} f(x) = \sup \ \lambda \ \text{ s.t. } \ f - \lambda \geq 0 \ \text{ on } K$

by

(SOSt) $f_t^{\text{sos}} = \sup \ \lambda \ \text{ s.t. } \ f - \lambda \in Q(g)_t$

- Each bound f_t^{sos} can be computed with SDP
- $f_t^{\text{sos}} \leq f_{t+1}^{\text{sos}} \leq f_{\min}$

SoS relaxations for (P)

Truncated quadratic module:

$$Q(g)_t := \{ \underbrace{s_0}_{\deg \leq 2t} + \underbrace{s_1 g_1}_{\deg \leq 2t} + \ldots + \underbrace{s_m g_m}_{\deg \leq 2t} \mid s_j \ \text{SoS} \}$$

Replace

(P) $f_{\min} = \inf_{x \in K} f(x) = \sup \ \lambda \ \text{ s.t. } \ f - \lambda \geq 0 \ \text{ on } K$

by

(SOSt) $f_t^{\text{sos}} = \sup \ \lambda \ \text{ s.t. } \ f - \lambda \in Q(g)_t$

- Each bound f_t^{sos} can be computed with SDP
- $f_t^{\text{sos}} \leq f_{t+1}^{\text{sos}} \leq f_{\min}$
- **Asymptotic convergence:** $\lim_{t \to \infty} f_t^{\text{sos}} = f_{\min}$ [Lasserre 2001]

Moment approach

$$f_{\min} = \inf_{x \in K} f(x) \quad = \inf_{\mu} \int_K f(x) d\mu \text{ s.t. } \mu \text{ is a probability measure on } K$$

$$f_{\min} = \inf_{x \in K} f(x) \quad = \inf_{\mu} \int_K f(x) d\mu \ \text{ s.t. } \mu \text{ is a probability measure on } K$$

$$= \inf_{L \in \mathbb{R}[x]^*} L(f) \ \text{ s.t. } L \text{ has a representing measure } \mu \text{ on } K$$

$$f_{\min} = \inf_{x \in K} f(x) = \inf_{\mu} \int_K f(x) d\mu \text{ s.t. } \mu \text{ is a probability measure on } K$$

$$= \inf_{L \in \mathbb{R}[x]^*} L(f) \text{ s.t. } L \text{ has a representing measure } \mu \text{ on } K$$

Deciding if a linear functional $L \in \mathbb{R}[x]^*$ has a **representing measure** μ on K

is the (difficult) **classical moment problem.**

$$f_{\min} = \inf_{x\in K} f(x) = \inf_{\mu} \int_K f(x)d\mu \text{ s.t. } \mu \text{ is a probability measure on } K$$

$$= \inf_{L\in\mathbb{R}[x]^*} L(f) \text{ s.t. } L \text{ has a representing measure } \mu \text{ on } K$$

Deciding if a linear functional $L \in \mathbb{R}[x]^*$ has a **representing measure** μ on K

is the (difficult) **classical moment problem.**

But one can use the **necessary condition**:

L is **nonnegative on the quadratic module** $Q(g) = \{s_0 + \sum_j s_j g_j : s_j \text{ SOS}\}$:

$$f_{\min} = \inf_{x \in K} f(x) = \inf_{\mu} \int_K f(x) d\mu \text{ s.t. } \mu \text{ is a probability measure on } K$$

$$= \inf_{L \in \mathbb{R}[x]^*} L(f) \text{ s.t. } L \text{ has a representing measure } \mu \text{ on } K$$

Deciding if a linear functional $L \in \mathbb{R}[x]^*$ has a **representing measure** μ on K

is the (difficult) **classical moment problem.**

But one can use the **necessary condition**:

L is **nonnegative on the quadratic module** $Q(g) = \{s_0 + \sum_j s_j g_j : s_j \text{ SOS}\}$:

$$L(p^2) \geq 0 \ \ \forall p, \quad \text{i.e.,} \quad M(L) = (L(x^{\alpha+\beta}))_{\alpha,\beta \in \mathbb{N}^n} \succeq 0$$

and $L(g_j p^2) \geq 0 \ \ \forall p, \quad$ i.e., $\quad M(g_j L) = (L(g_j x^{\alpha+\beta}))_{\alpha,\beta \in \mathbb{N}^n} \succeq 0$

$$L(p^2) = L((\textstyle\sum_\alpha p_\alpha x^\alpha)^2) = \sum_{\alpha,\beta} p_\alpha p_\beta L(x^{\alpha+\beta}) = \overline{p}^{\mathsf{T}} M(L) \overline{p}$$

$$f_{\min} = \inf_{x\in K} f(x) = \inf_{\mu} \int_K f(x)d\mu \text{ s.t. } \mu \text{ is a probability measure on } K$$

$$= \inf_{L\in\mathbb{R}[x]^*} L(f) \text{ s.t. } L \text{ has a representing measure } \mu \text{ on } K$$

Deciding if a linear functional $L \in \mathbb{R}[x]^*$ has a **representing measure** μ on K

is the (difficult) **classical moment problem.**

But one can use the **necessary condition**:

L is **nonnegative on the quadratic module** $Q(g) = \{s_0 + \sum_j s_j g_j : s_j \text{ SOS}\}$:

$$L(p^2) \geq 0 \ \ \forall p, \quad \text{i.e.,} \quad M(L) = (L(x^{\alpha+\beta}))_{\alpha,\beta\in\mathbb{N}^n} \succeq 0$$

and $$L(g_j p^2) \geq 0 \ \ \forall p, \quad \text{i.e.,} \quad M(g_j L) = (L(g_j x^{\alpha+\beta}))_{\alpha,\beta\in\mathbb{N}^n} \succeq 0$$

$$L(p^2) = L((\textstyle\sum_\alpha p_\alpha x^\alpha)^2) = \sum_{\alpha,\beta} p_\alpha p_\beta L(x^{\alpha+\beta}) = \overline{p}^{\mathsf{T}} M(L) \overline{p}$$

$M(L)$ is a **moment matrix** and $M(g_j L)$ are **localizing moment matrices**

Moment relaxations for (P)

(P) $$f_{\min} = \inf_{L \in \mathbb{R}[x]^*} L(f) \ \text{ s.t. } L \text{ has a representing measure } \mu \text{ on } K$$

Moment relaxations for (P)

(P) $$f_{\min} = \inf_{L \in \mathbb{R}[x]^*} L(f) \ \text{ s.t. } L \text{ has a representing measure } \mu \text{ on } K$$

Truncate at degree $2t$:

$$f_t^{\text{mom}} = \inf_{L \in \mathbb{R}[x]_{2t}^*} L(f) \ \text{ s.t. } \ L(1) = 1, \ L \geq 0 \text{ on } Q(g)_t$$

(MOMt)

Moment relaxations for (P)

(P) $$f_{\min} = \inf_{L\in\mathbb{R}[x]^*} L(f) \text{ s.t. } L \text{ has a representing measure } \mu \text{ on } K$$

Truncate at degree $2t$:

(MOMt) $$f_t^{\mathrm{mom}} = \inf_{L\in\mathbb{R}[x]^*_{2t}} L(f) \text{ s.t. } L(1) = 1, \; L \geq 0 \text{ on } Q(g)_t$$

$$\text{i.e., } M_t(L) \succeq 0, \;\; M_{t-d_j}(g_j L) \succeq 0 \;\; \forall j$$

Moment relaxations for (P)

(P) $$f_{\min} = \inf_{L \in \mathbb{R}[x]^*} L(f) \ \text{ s.t. } L \text{ has a representing measure } \mu \text{ on } K$$

Truncate at degree $2t$:

(MOMt) $$f_t^{\text{mom}} = \inf_{L \in \mathbb{R}[x]_{2t}^*} L(f) \ \text{ s.t. } \ L(1) = 1, \ L \geq 0 \text{ on } Q(g)_t$$

$$\text{i.e., } M_t(L) \succeq 0, \ \ M_{t-d_j}(g_j L) \succeq 0 \ \ \forall j$$

(SOSt) $$f_t^{\text{sos}} = \sup \ \lambda \ \text{ s.t. } \ f - \lambda \in Q(g)_t$$

Moment relaxations for (P)

(P)
$$f_{\min} = \inf_{L \in \mathbb{R}[x]^*} L(f) \text{ s.t. } L \text{ has a representing measure } \mu \text{ on } K$$

Truncate at degree $2t$:

(MOMt)
$$f_t^{\text{mom}} = \inf_{L \in \mathbb{R}[x]_{2t}^*} L(f) \text{ s.t. } L(1) = 1, \; L \geq 0 \text{ on } Q(g)_t$$

$$\text{i.e., } M_t(L) \succeq 0, \; M_{t-d_j}(g_j L) \succeq 0 \;\; \forall j$$

(SOSt)
$$f_t^{\text{sos}} = \sup \lambda \text{ s.t. } f - \lambda \in Q(g)_t$$

$$f_t^{\text{sos}} \leq f_t^{\text{mom}} \leq f_{\min}$$

$\rightsquigarrow$ dual sdp bounds

Some results on the full/truncated moment problem

Theorem [Putinar 1997]
Assume $L \in \mathbb{R}[x]^*$ is nonnegative on the (archimedean) quadratic module $Q(g)$.

• Then L has a **representing measure** μ supported by K: $L(f) = \int f(x)\mu(dx)$

• [Tchakaloff 1957] For any fixed degree k, the restriction of L to $\mathbb{R}[x]_k$ has a representing measure supported by K, which is **finite atomic**.

Some results on the full/truncated moment problem

Theorem [Putinar 1997]
Assume $L \in \mathbb{R}[x]^*$ is nonnegative on the (archimedean) quadratic module $Q(g)$.

• Then L has a **representing measure** μ supported by K: $L(f) = \int f(x)\mu(dx)$

• [Tchakaloff 1957] For any fixed degree k, the restriction of L to $\mathbb{R}[x]_k$ has a representing measure supported by K, which is **finite atomic**.

Theorem [Curto-Fialkow 1996 - L 2005: short algebraic proof]

Assume $L \in \mathbb{R}[x]^*_{2t}$ is nonnegative on $Q_t(g)$, i.e., $M_t(L) \succeq 0$,
and $\mathrm{rank}\ M_t(L) = \mathrm{rank}\ M_{t-1}(L)$ [**flatness condition**]

Then L has a **finite atomic representing measure** μ on K.

Main steps of proof:

- Extend L to $L \in \mathbb{R}[x]^*$ with $\mathrm{rank}\ M(L) = \mathrm{rank} M_t(L) =: r$
- $M(L) \succeq 0$ with *finite rank* $r \Longrightarrow L$ has an r-atomic measure μ

Optimality criterion for moment relaxation (MOMt)

$$K = \{x \mid g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}$$

$$d_K = \max_j \lceil \deg(g_j)/2 \rceil$$

$$f_t^{\mathrm{mom}} = \inf_{L \in \mathbb{R}[x]_{2t}^*} L(f) \ \text{ s.t. } \ L(1) = 1,\ M_t(L) \succeq 0,\ \ M_{t-d_j}(g_j L) \succeq 0 \ \ \forall j$$

Optimality criterion for moment relaxation (MOMt)

$$K = \{x \mid g_1(x) \geq 0, \ldots, g_m(x) \geq 0\} \qquad d_K = \max_j \lceil \deg(g_j)/2 \rceil$$

$$f_t^{\text{mom}} = \inf_{L \in \mathbb{R}[x]_{2t}^*} L(f) \text{ s.t. } L(1) = 1, \; M_t(L) \succeq 0, \; M_{t-d_j}(g_j L) \succeq 0 \; \forall j$$

Theorem [CF 2000 + Henrion-Lasserre 2005 + Lasserre-L-Rostalski 2008]

Assume L is an optimal solution of (MOMt) such that

$\text{rank } M_s(L) = \text{rank } M_{s-d_K}(L)$ for some $d_K \leq s \leq t$.

Optimality criterion for moment relaxation (MOMt)

$$K = \{x \mid g_1(x) \geq 0, \ldots, g_m(x) \geq 0\} \qquad d_K = \max_j \lceil \deg(g_j)/2 \rceil$$

$$f_t^{\text{mom}} = \inf_{L \in \mathbb{R}[x]_{2t}^*} L(f) \ \text{ s.t. } \ L(1) = 1,\ M_t(L) \succeq 0,\ \ M_{t-d_j}(g_j L) \succeq 0 \ \ \forall j$$

Theorem [CF 2000 + Henrion-Lasserre 2005 + Lasserre-L-Rostalski 2008]

Assume L is an optimal solution of (MOMt) such that

$$\text{rank } M_s(L) = \text{rank } M_{s-d_K}(L) \text{ for some } d_K \leq s \leq t.$$

- Then the relaxation is **exact:** $f_t^{\text{mom}} = f_{\min}$.

Optimality criterion for moment relaxation (MOMt)

$$K = \{x \mid g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}$$

$$d_K = \max_j \lceil \deg(g_j)/2 \rceil$$

$$f_t^{\text{mom}} = \inf_{L \in \mathbb{R}[x]_{2t}^*} L(f) \text{ s.t. } L(1) = 1, \; M_t(L) \succeq 0, \; M_{t-d_j}(g_j L) \succeq 0 \;\; \forall j$$

Theorem [CF 2000 + Henrion-Lasserre 2005 + Lasserre-L-Rostalski 2008]

Assume L is an optimal solution of (MOMt) such that

$$\mathrm{rank}\ M_s(L) = \mathrm{rank}\ M_{s-d_K}(L) \text{ for some } d_K \leq s \leq t.$$

- Then the relaxation is **exact:** $f_t^{\text{mom}} = f_{\min}$.
- Moreover, one can compute the **global minimizers:**

Optimality criterion for moment relaxation (MOMt)

$$K = \{x \mid g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}$$ $d_K = \max_j \lceil \deg(g_j)/2 \rceil$

$$f_t^{\text{mom}} = \inf_{L \in \mathbb{R}[x]_{2t}^*} L(f) \ \text{ s.t. } \ L(1) = 1, \ M_t(L) \succeq 0, \ \ M_{t-d_j}(g_j L) \succeq 0 \ \ \forall j$$

Theorem [CF 2000 + Henrion-Lasserre 2005 + Lasserre-L-Rostalski 2008]

Assume L is an optimal solution of (MOMt) such that

$\mathrm{rank}\ M_s(L) = \mathrm{rank}\ M_{s-d_K}(L)$ for some $d_K \leq s \leq t$.

- Then the relaxation is **exact:** $f_t^{\text{mom}} = f_{\min}$.
- Moreover, one can compute the **global minimizers:**

$V(\mathrm{Ker} M_s(L)) \subseteq \{$ global minimizers of f on $K\}$,

with **equality** if $\mathrm{rank}\ M_t(L)$ is **maximum** (rank = # minimizers).

Some properties

- Interior point algos for SDP give a **maximum rank** optimal solution

Some properties

- Interior point algos for SDP give a **maximum rank** optimal solution

- **Finite convergence** holds in **finite variety case** [L 2007, Nie 2013]
  in **convex case** [Lasserre 2009, de Klerk-L 2011]
  generically [Nie 2014]

Some properties

- Interior point algos for SDP give a **maximum rank** optimal solution

- **Finite convergence** holds in **finite variety case** [L 2007, Nie 2013]
  in **convex case** [Lasserre 2009, de Klerk-L 2011]
  generically [Nie 2014]

- Can **exploit structure** (like sparsity, symmetry, equations) to design more economical SDP relaxations

Some properties

- Interior point algos for SDP give a **maximum rank** optimal solution

- **Finite convergence** holds in **finite variety case** [L 2007, Nie 2013]
  in **convex case** [Lasserre 2009, de Klerk-L 2011]
  generically [Nie 2014]

- Can **exploit structure** (like sparsity, symmetry, equations) to design more economical SDP relaxations

- Algorithm for computing the (finitely many) **real roots** of polynomial equations (and real radical ideals)

  [Lasserre-L-Rostalski 2008,2009]
  [Lasserre-L-Mourrain-Rostalski-Trebuchet 2013]

$\rightsquigarrow$ large literature, surveys, monographs

Application for bounding matrix factorization ranks

using the moment approach

Matrix factorization ranks

- **Nonnegative factorization** of $A \in \mathbb{R}_+^{m\times n}$:
  $A = \sum_{\ell=1}^{r} a_\ell b_\ell^\mathsf{T}$, where $a_\ell \in \mathbb{R}_+^m$, $b_\ell \in \mathbb{R}_+^n$ [*atomic decomposition*]
  $A = (\langle u_i, v_j \rangle)_{i\in[m], j\in[n]}$, where $u_i, v_j \in \mathbb{R}_+^r$ [*Gram factorization*]
  Smallest such r: $\mathrm{rank}_+(A)$ $\rightsquigarrow$ **nonnegative rank**

Matrix factorization ranks

- **Nonnegative factorization** of $A \in \mathbb{R}_+^{m\times n}$:
  $A = \sum_{\ell=1}^{r} a_\ell b_\ell^\mathsf{T}$, where $a_\ell \in \mathbb{R}_+^m$, $b_\ell \in \mathbb{R}_+^n$ [*atomic decomposition*]
  $A = (\langle u_i, v_j \rangle)_{i\in[m], j\in[n]}$, where $u_i, v_j \in \mathbb{R}_+^r$ [*Gram factorization*]
  Smallest such r: $\mathrm{rank}_+(A)$ $\rightsquigarrow$ **nonnegative rank**

- **CP-factorization** of $A \in \mathcal{S}^n$: *symmetric* nonnegative factorization:
  restrict to $a_\ell = b_\ell \ \ \forall \ell$ and to $u_i = v_i \ \ \forall i$
  Smallest such r: $\mathrm{rank}_{\mathrm{cp}}(A)$ $\rightsquigarrow$ **cp-rank**
  $\mathrm{rank}_{\mathrm{cp}}(A) < \infty$ when A is **completely positive**

Matrix factorization ranks

- **Nonnegative factorization** of $A \in \mathbb{R}_+^{m\times n}$:
  $A = \sum_{\ell=1}^r a_\ell b_\ell^\mathsf{T}$, where $a_\ell \in \mathbb{R}_+^m$, $b_\ell \in \mathbb{R}_+^n$ [*atomic decomposition*]
  $A = (\langle u_i, v_j\rangle)_{i\in[m],j\in[n]}$, where $u_i, v_j \in \mathbb{R}_+^r$ [*Gram factorization*]
  Smallest such r: $\mathrm{rank}_+(A)$ $\rightsquigarrow$ **nonnegative rank**

- **CP-factorization** of $A \in \mathcal{S}^n$: *symmetric* nonnegative factorization:
  restrict to $a_\ell = b_\ell \ \ \forall \ell$ and to $u_i = v_i \ \ \forall i$
  Smallest such r: $\mathrm{rank_{cp}}(A)$ $\rightsquigarrow$ **cp-rank**
  $\mathrm{rank_{cp}}(A) < \infty$ when A is **completely positive**

- **PSD factorization** of $A \in \mathbb{R}_+^{m\times n}$:
  $A = (\langle U_i, V_j\rangle)_{i\in[m],j\in[n]}$, where $U_i, V_j \in \mathcal{S}_+^r$ [*Gram factorization*]
  Smallest such r: $\mathrm{rank_{psd}}(A)$ $\rightsquigarrow$ **psd-rank**
  Symmetric analogue: require $U_i = V_i \ \forall i$ $\rightsquigarrow$ **cpsd-rank**

Matrix factorization ranks

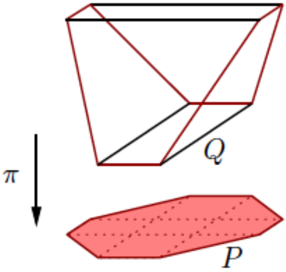
- **Nonnegative factorization** of $A \in \mathbb{R}_+^{m\times n}$:
  $A = \sum_{\ell=1}^r a_\ell b_\ell^\mathsf{T}$, where $a_\ell \in \mathbb{R}_+^m$, $b_\ell \in \mathbb{R}_+^n$ [*atomic decomposition*]
  $A = (\langle u_i, v_j \rangle)_{i\in[m], j\in[n]}$, where $u_i, v_j \in \mathbb{R}_+^r$ [*Gram factorization*]
  Smallest such r: $\mathrm{rank}_+(A)$ $\rightsquigarrow$ **nonnegative rank**

- **CP-factorization** of $A \in \mathcal{S}^n$: *symmetric* nonnegative factorization:
  restrict to $a_\ell = b_\ell \;\; \forall \ell$ and to $u_i = v_i \;\; \forall i$
  Smallest such r: $\mathrm{rank}_{\mathrm{cp}}(A)$ $\rightsquigarrow$ **cp-rank**
  $\mathrm{rank}_{\mathrm{cp}}(A) < \infty$ when A is **completely positive**

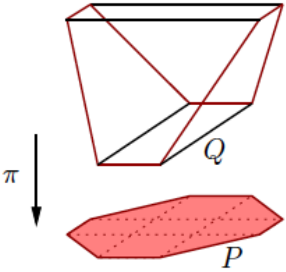
- **PSD factorization** of $A \in \mathbb{R}_+^{m\times n}$:
  $A = (\langle U_i, V_j \rangle)_{i\in[m], j\in[n]}$, where $U_i, V_j \in \mathcal{S}_+^r$ [*Gram factorization*]
  Smallest such r: $\mathrm{rank}_{\mathrm{psd}}(A)$ $\rightsquigarrow$ **psd-rank**
  Symmetric analogue: require $U_i = V_i \; \forall i$ $\rightsquigarrow$ **cpsd-rank**

Applications: extended formulations (LP/SDP) of polytopes
(quantum) communication complexity

Nonnegative/psd rank and extended formulations



Nonnegative/psd rank and extended formulations

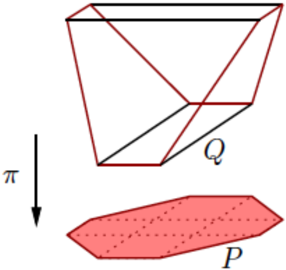


Theorem [Yannakakis 1991 - Gouveia-Parrilo-Thomas 2013]

For a polytope $P = \mathrm{conv}(V) = \{x \in \mathbb{R}^n : a_i^\mathsf{T} x \leq b_i \ \forall i \in [m]\}$

its **slack-matrix** is $S = (b_i - a_i^\mathsf{T} v)_{v \in V, i \in [m]} \in \mathbb{R}_+^{|V| \times m}$

Nonnegative/psd rank and extended formulations



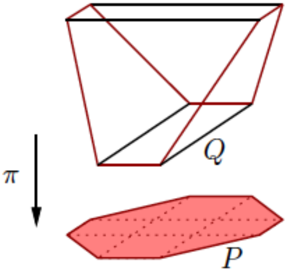
Theorem [Yannakakis 1991 - Gouveia-Parrilo-Thomas 2013]

For a polytope $P = \mathrm{conv}(V) = \{x \in \mathbb{R}^n : a_i^\top x \leq b_i \ \forall i \in [m]\}$

its **slack-matrix** is $S = (b_i - a_i^\top v)_{v \in V, i \in [m]} \ \in \mathbb{R}_+^{|V| \times m}$

- Smallest r s.t. P is **projection of affine section** of $\mathbb{R}_+^r$ is $\mathrm{rank}_+(S)$

Nonnegative/psd rank and extended formulations



Theorem [Yannakakis 1991 - Gouveia-Parrilo-Thomas 2013]

For a polytope $P = \mathrm{conv}(V) = \{x \in \mathbb{R}^n : a_i^\top x \leq b_i \ \forall i \in [m]\}$

its **slack-matrix** is $S = (b_i - a_i^\top v)_{v \in V, i \in [m]} \ \in \mathbb{R}_+^{|V| \times m}$

- Smallest r s.t. P is **projection of affine section** of $\mathbb{R}_+^r$ is $\mathrm{rank}_+(S)$
- Smallest r s.t. P is **projection of affine section** of $\mathcal{S}_+^r$ is $\mathrm{rank}_{\mathrm{psd}}(S)$

Bounds for cp-rank via polynomial optimization

Assume $A = \sum_{\ell=1}^{r} a_\ell a_\ell^\mathsf{T}$, where $a_\ell \in \mathbb{R}^n_+$, $r = \mathrm{rank}_{\mathrm{cp}}(A)$.

Bounds for cp-rank via polynomial optimization

Assume $A = \sum_{\ell=1}^{r} a_\ell a_\ell^\mathsf{T}$, where $a_\ell \in \mathbb{R}^n_+$, $r = \mathrm{rank}_{\mathrm{cp}}(A)$.

$$\rightsquigarrow \quad \frac{1}{r} A \in \mathcal{R} := \mathrm{conv}(xx^\mathsf{T} : x \in \mathbb{R}^n_+,\ A - xx^\mathsf{T} \geq 0,\ A - xx^T \succeq 0\}$$

Bounds for cp-rank via polynomial optimization

Assume $A = \sum_{\ell=1}^{r} a_\ell a_\ell^\mathsf{T}$, where $a_\ell \in \mathbb{R}^n_+$, $r = \mathrm{rank}_{\mathrm{cp}}(A)$.

$$\rightsquigarrow \quad \frac{1}{r} A \in \mathcal{R} := \mathrm{conv}(xx^\mathsf{T} : x \in \mathbb{R}^n_+,\ A - xx^\mathsf{T} \geq 0,\ A - xx^T \succeq 0\}$$

$$\tau_{\mathrm{cp}}(A) := \min \left\{ \lambda : \ \tfrac{1}{\lambda} A \in \mathcal{R} \right\} \ \leq \ \mathrm{rank}_{\mathrm{cp}}(A)$$

[Fawzi-Parrilo 2016]

Bounds for cp-rank via polynomial optimization

Assume $A = \sum_{\ell=1}^{r} a_\ell a_\ell^{\mathsf{T}}$, where $a_\ell \in \mathbb{R}^n_+$, $r = \mathrm{rank}_{\mathrm{cp}}(A)$.

$$\rightsquigarrow \quad \frac{1}{r} A \in \mathcal{R} := \mathrm{conv}(xx^{\mathsf{T}} : x \in \mathbb{R}^n_+,\ A - xx^{\mathsf{T}} \geq 0,\ A - xx^T \succeq 0\}$$

$$\tau_{\mathrm{cp}}(A) := \min\left\{\lambda :\ \tfrac{1}{\lambda} A \in \mathcal{R}\right\}\ \leq\ \mathrm{rank}_{\mathrm{cp}}(A)$$

[Fawzi-Parrilo 2016]

Moment approach: Define $L \in \mathbb{R}[x_1, \ldots, x_n]^*$ by $L(p) = \sum_{\ell=1}^{r} p(a_\ell)$.

Bounds for cp-rank via polynomial optimization

Assume $A = \sum_{\ell=1}^{r} a_\ell a_\ell^\mathsf{T}$, where $a_\ell \in \mathbb{R}^n_+$, $r = \mathrm{rank_{cp}}(A)$.

$$\rightsquigarrow \quad \frac{1}{r} A \in \mathcal{R} := \mathrm{conv}(xx^\mathsf{T} : x \in \mathbb{R}^n_+,\ A - xx^\mathsf{T} \geq 0,\ A - xx^T \succeq 0\}$$

$$\tau_{\mathrm{cp}}(A) := \min \left\{ \lambda : \ \tfrac{1}{\lambda} A \in \mathcal{R} \right\} \ \leq \ \mathrm{rank_{cp}}(A)$$

[Fawzi-Parrilo 2016]

Moment approach: Define $L \in \mathbb{R}[x_1, \ldots, x_n]^*$ by $L(p) = \sum_{\ell=1}^{r} p(a_\ell)$.

(0) $L(1) = r$ (model $\mathrm{rank_{cp}}(A)$)

Bounds for cp-rank via polynomial optimization

Assume $A = \sum_{\ell=1}^{r} a_\ell a_\ell^\mathsf{T}$, where $a_\ell \in \mathbb{R}^n_+$, $r = \mathrm{rank_{cp}}(A)$.

$$\rightsquigarrow \quad \frac{1}{r}A \in \mathcal{R} := \mathrm{conv}(xx^\mathsf{T} : x \in \mathbb{R}^n_+,\ A - xx^\mathsf{T} \geq 0,\ A - xx^T \succeq 0\}$$

$$\tau_{\mathrm{cp}}(A) := \min\left\{\lambda :\ \tfrac{1}{\lambda}A \in \mathcal{R}\right\} \ \leq\ \mathrm{rank_{cp}}(A)$$

[Fawzi-Parrilo 2016]

Moment approach: Define $L \in \mathbb{R}[x_1, \ldots, x_n]^*$ by $L(p) = \sum_{\ell=1}^{r} p(a_\ell)$.

(0) $L(1) = r$ (model $\mathrm{rank_{cp}}(A)$)

(1) $L(x_i x_j) = A_{ij}$ (recover A)

Bounds for cp-rank via polynomial optimization

Assume $A = \sum_{\ell=1}^{r} a_\ell a_\ell^{\mathsf{T}}$, where $a_\ell \in \mathbb{R}^n_+$, $r = \mathrm{rank}_{\mathrm{cp}}(A)$.

$$\rightsquigarrow \quad \frac{1}{r} A \in \mathcal{R} := \mathrm{conv}(xx^{\mathsf{T}} : x \in \mathbb{R}^n_+,\ A - xx^{\mathsf{T}} \geq 0,\ A - xx^T \succeq 0\}$$

$$\tau_{\mathrm{cp}}(A) := \min\left\{\lambda : \ \tfrac{1}{\lambda} A \in \mathcal{R}\right\} \ \leq \ \mathrm{rank}_{\mathrm{cp}}(A)$$

[Fawzi-Parrilo 2016]

Moment approach: Define $L \in \mathbb{R}[x_1, \ldots, x_n]^*$ by $L(p) = \sum_{\ell=1}^{r} p(a_\ell)$.

(0) $L(1) = r$ (model $\mathrm{rank}_{\mathrm{cp}}(A)$)

(1) $L(x_i x_j) = A_{ij}$ (recover A)

(2) $L \geq 0$ on $Q(\sqrt{A_{ii}} x_i - x_i^2,\ A_{ij} - x_i x_j)$ (model $x \geq 0,\ A - xx^T \geq 0$)

Bounds for cp-rank via polynomial optimization

Assume $A = \sum_{\ell=1}^{r} a_\ell a_\ell^\mathsf{T}$, where $a_\ell \in \mathbb{R}_+^n$, $r = \mathrm{rank_{cp}}(A)$.

$$\rightsquigarrow \quad \frac{1}{r} A \in \mathcal{R} := \mathrm{conv}(xx^\mathsf{T} : x \in \mathbb{R}_+^n,\ A - xx^\mathsf{T} \geq 0,\ A - xx^T \succeq 0\}$$

$$\tau_{\mathrm{cp}}(A) := \min \left\{ \lambda : \ \tfrac{1}{\lambda} A \in \mathcal{R} \right\} \ \leq \ \mathrm{rank_{cp}}(A)$$

[Fawzi-Parrilo 2016]

Moment approach: Define $L \in \mathbb{R}[x_1, \ldots, x_n]^*$ by $L(p) = \sum_{\ell=1}^{r} p(a_\ell)$.

(0) $L(1) = r$ (model $\mathrm{rank_{cp}}(A)$)

(1) $L(x_i x_j) = A_{ij}$ (recover A)

(2) $L \geq 0$ on $Q(\sqrt{A_{ii}} x_i - x_i^2,\ A_{ij} - x_i x_j)$ (model $x \geq 0,\ A - xx^T \geq 0$)

(3a) $L((xx^T)^{\otimes k}) \preceq A^{\otimes k}$ for $k \geq 2$ (model $A - xx^\mathsf{T} \succeq 0$)

Bounds for cp-rank via polynomial optimization

Assume $A = \sum_{\ell=1}^{r} a_\ell a_\ell^{\mathsf{T}}$, where $a_\ell \in \mathbb{R}^n_+$, $r = \mathrm{rank_{cp}}(A)$.

$$\rightsquigarrow \quad \frac{1}{r} A \in \mathcal{R} := \mathrm{conv}(xx^{\mathsf{T}} : x \in \mathbb{R}^n_+,\ A - xx^{\mathsf{T}} \geq 0,\ A - xx^{T} \succeq 0\}$$

$$\tau_{\mathrm{cp}}(A) := \min \left\{ \lambda :\ \tfrac{1}{\lambda} A \in \mathcal{R} \right\} \ \leq\ \mathrm{rank_{cp}}(A)$$

[Fawzi-Parrilo 2016]

Moment approach: Define $L \in \mathbb{R}[x_1, \ldots, x_n]^*$ by $L(p) = \sum_{\ell=1}^{r} p(a_\ell)$.

(0) $L(1) = r$ (model $\mathrm{rank_{cp}}(A)$)

(1) $L(x_i x_j) = A_{ij}$ (recover A)

(2) $L \geq 0$ on $Q(\sqrt{A_{ii}} x_i - x_i^2,\ A_{ij} - x_i x_j)$ (model $x \geq 0,\ A - xx^{T} \geq 0$)

(3b) $L((A - xx^{\mathsf{T}}) \otimes [x][x]^{\mathsf{T}}) \succeq 0$ (model $A - xx^{\mathsf{T}} \succeq 0$)

[Gribling-L-Steenkamp 2021]

Bounds for cp-rank via polynomial optimization

Assume $A = \sum_{\ell=1}^{r} a_\ell a_\ell^\mathsf{T}$, where $a_\ell \in \mathbb{R}^n_+$, $r = \mathrm{rank_{cp}}(A)$.

$$\rightsquigarrow \quad \frac{1}{r} A \in \mathcal{R} := \mathrm{conv}(xx^\mathsf{T} : x \in \mathbb{R}^n_+,\ A - xx^\mathsf{T} \geq 0,\ A - xx^T \succeq 0\}$$

$$\tau_{\mathrm{cp}}(A) := \min \left\{ \lambda :\ \tfrac{1}{\lambda} A \in \mathcal{R} \right\} \ \leq\ \mathrm{rank_{cp}}(A)$$

[Fawzi-Parrilo 2016]

Moment approach: Define $L \in \mathbb{R}[x_1, \ldots, x_n]^*$ by $L(p) = \sum_{\ell=1}^{r} p(a_\ell)$.

(0) $L(1) = r$ (model $\mathrm{rank_{cp}}(A)$)

(1) $L(x_i x_j) = A_{ij}$ (recover A)

(2) $L \geq 0$ on $Q(\sqrt{A_{ii}} x_i - x_i^2,\ A_{ij} - x_i x_j)$ (model $x \geq 0,\ A - xx^T \geq 0$)

(3b) $L((A - xx^\mathsf{T}) \otimes [x][x]^\mathsf{T}) \succeq 0$ (model $A - xx^\mathsf{T} \succeq 0$)

[Gribling-L-Steenkamp 2021]

Theorem [Gribling-de Laat-L 2019]
The bounds ξ_t^{cp}, obtained by minimizing $L(1)$ over $L \in \mathbb{R}[x]_{2t}^*$ satisfying the truncated versions of (1)-(3a), converge asymptotically to $\tau_{\mathrm{cp}}(A)$,

Bounds for cp-rank via polynomial optimization

Assume $A = \sum_{\ell=1}^{r} a_\ell a_\ell^\mathsf{T}$, where $a_\ell \in \mathbb{R}^n_+$, $r = \mathrm{rank_{cp}}(A)$.

$$\rightsquigarrow \quad \frac{1}{r} A \in \mathcal{R} := \mathrm{conv}(xx^\mathsf{T} : x \in \mathbb{R}^n_+,\ A - xx^\mathsf{T} \geq 0,\ A - xx^T \succeq 0\}$$

$$\tau_{\mathrm{cp}}(A) := \min \left\{ \lambda : \ \tfrac{1}{\lambda} A \in \mathcal{R} \right\} \ \leq \ \mathrm{rank_{cp}}(A)$$

[Fawzi-Parrilo 2016]

Moment approach: Define $L \in \mathbb{R}[x_1, \ldots, x_n]^*$ by $L(p) = \sum_{\ell=1}^{r} p(a_\ell)$.

(0) $L(1) = r$ (model $\mathrm{rank_{cp}}(A)$)

(1) $L(x_i x_j) = A_{ij}$ (recover A)

(2) $L \geq 0$ on $Q(\sqrt{A_{ii}} x_i - x_i^2,\ A_{ij} - x_i x_j)$ (model $x \geq 0,\ A - xx^T \geq 0$)

(3b) $L((A - xx^\mathsf{T}) \otimes [x][x]^\mathsf{T}) \succeq 0$ (model $A - xx^\mathsf{T} \succeq 0$)

[Gribling-L-Steenkamp 2021]

Theorem [Gribling-de Laat-L 2019]
The bounds ξ_t^{cp}, obtained by minimizing $L(1)$ over $L \in \mathbb{R}[x]_{2t}^*$ satisfying the truncated versions of (1)-(3a), converge asymptotically to $\tau_{\mathrm{cp}}(A)$, and in finitely many steps under flatness.

Extension to other factorization ranks

The moment view point for polynomial optimization offers a systematic, common approach to treat many factorization ranks

- Extension to the **nonnegative rank**, by using two sets of variables x, y; extends also to the more general tensor setting
- Extension to the **psd-rank and cpsd-rank**, by taking the Gram factorization view point and using *noncommutative* variables

Extension to other factorization ranks

The moment view point for polynomial optimization offers a systematic, common approach to treat many factorization ranks

- Extension to the **nonnegative rank**, by using two sets of variables x, y; extends also to the more general tensor setting
- Extension to the **psd-rank and cpsd-rank**, by taking the Gram factorization view point and using *noncommutative* variables

Currently working (with Gribling and Steenkamp) on bounds for the **separable rank** of a linear operator ρ acting on $\mathbb{C}^n \otimes \mathbb{C}^n$, asking for the smallest decomposition of the form

$$\rho = \sum_{\ell=1}^{r} a_\ell a_\ell^\mathsf{T} \otimes b_\ell b_\ell^\mathsf{T}$$

Understanding separable states is a fundamental question in quantum information

Concluding remarks

- The two (dual) approaches via moments and sums-of-squares provide **interesting complementary information**
- This extends to the **problem of moments** (*optimize over measures*) and to polynomial optimization in **noncommutative variables** (*optimize over matrix-valued variables*), with *many applications*
- What about the quality of the relaxations? (see Lecture 2)
- Approximation hierarchies for graph problems (see Lecture 3)

THANK YOU !

Some references

P. Parrilo: Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization, PhD thesis, CalTech, 2000.

J.B. Lasserre: Global optimization with polynomials and the problem of moments, *SIAM J. Optimization*, 2001.

J.B. Lasserre, M. Laurent, P. Rostalski: Semidefinite characterization and computation of real radical ideals, Foundations Comput. Math., 2008.

J.B. Lasserre: Moments, Positive Polynomials and their Applications, Imperial College Press, 2009.

M. Laurent: Sums of squares, moment matrices and optimization over polynomials, in IMA volume 149, 2009.

M. Anjos and J.-B. Lasserre (eds): Handbook of Semidefinite, Conic and Polynomial Optimization, Springer, 2011

G. Blekherman, P. Parrilo, R. Thomas (eds): Semidefinite Optimization and Convex Algebraic Geometry, MOS-SIAM Series Optimization, 2011