Positive polynomials and convergence of LP and SDP relaxations

Markus Schweighofer

Universität Konstanz

ETH

Zürich, May 31, 2005

Notation for the whole talk

Notation for the whole talk

- $X_1, \ldots, X_n$ variables

Notation for the whole talk

- $X_1, \ldots, X_n$ variables
- $X := X_1$ when $n = 1$

Notation for the whole talk

- $X_1, \ldots, X_n$ variables
- $X := X_1$ when $n = 1$, $(X, Y) := (X_1, X_2)$ when $n = 2$, $\ldots$

Notation for the whole talk

- $X_1, \ldots, X_n$ variables
- $X := X_1$ when $n = 1$, $(X, Y) := (X_1, X_2)$ when $n = 2$, $\ldots$
- $\mathbb{R}[X_1, \ldots, X_n]$ polynomial ring

Notation for the whole talk

- $X_1, \dots, X_n$ variables
- $X := X_1$ when $n = 1$, $(X, Y) := (X_1, X_2)$ when $n = 2$, $\dots$
- $\mathbb{R}[X_1, \dots, X_n]$ polynomial ring
- $f \in \mathbb{R}[X_1, \dots, X_n]$ an arbitrary polynomial

Notation for the whole talk

- $X_1, \ldots, X_n$ variables
- $X := X_1$ when $n = 1$, $(X, Y) := (X_1, X_2)$ when $n = 2$, ...
- $\mathbb{R}[X_1, \ldots, X_n]$ polynomial ring
- $f \in \mathbb{R}[X_1, \ldots, X_n]$ an arbitrary polynomial
- $g_1, \ldots, g_m \in \mathbb{R}[X_1, \ldots, X_n]$ polynomials defining...

Notation for the whole talk

- $X_1, \dots, X_n$ variables
- $X := X_1$ when $n = 1$, $(X, Y) := (X_1, X_2)$ when $n = 2$, ...
- $\mathbb{R}[X_1, \dots, X_n]$ polynomial ring
- $f \in \mathbb{R}[X_1, \dots, X_n]$ an arbitrary polynomial
- $g_1, \dots, g_m \in \mathbb{R}[X_1, \dots, X_n]$ polynomials defining...
- ... the set $S := \{x \in \mathbb{R}^n \mid g_1(x) \geq 0, \dots, g_m(x) \geq 0\}$

n

f

$g_1, \ldots, g_m$

S

Optimization

We consider the problem of minimizing f on S.

Optimization

We consider the problem of minimizing f on S. So we want to compute numerically the infimum

$$f^* := \inf\{f(x) \mid x \in S\} \in \mathbb{R} \cup \{\pm\infty\}$$

Optimization

We consider the problem of minimizing f on S. So we want to compute numerically the infimum

$$f^* := \inf\{f(x) \mid x \in S\} \in \mathbb{R} \cup \{\pm\infty\}$$

and, if possible, a minimizer, i.e., an element of the set

$$S^* := \{x^* \in S \mid f(x^*) \leq f(x) \text{ for all } x \in S\}.$$

L **P**

Linear Programming

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & x \in \mathbb{R}^n \\ & g_1(x) \geq 0 \\ & \vdots \\ & g_m(x) \geq 0 \end{aligned}$$

where all polynomials f and g_i are linear, i.e., their degree is ≤ 1. In particular, $S \subseteq \mathbb{R}^n$ is a polyhedron.

Linear Programming

$$\begin{array}{rl} \text{minimize} & f(x) \\ \\ \text{subject to} & x \in \mathbb{R}^n \\ & \begin{pmatrix} g_1(x) & & \\ & \ddots & \\ & & g_m(x) \end{pmatrix} \quad \text{is psd} \end{array}$$

where all polynomials f and g_i are linear, i.e.,
their degree is ≤ 1. In particular, $S \subseteq \mathbb{R}^n$ is a polyhedron.

S **D** **P**

$$
\begin{array}{rl}
\text{minimize} & f(x) \\
\\
\text{subject to} & x \in \mathbb{R}^n \\
& \begin{pmatrix} g_{11}(x) & \dots & g_{1m}(x) \\ \vdots & \ddots & \vdots \\ & \dots & g_{mm}(x) \end{pmatrix} \text{ is psd}
\end{array}
$$

where all polynomials f and g_{ij} are linear, i.e., their degree is ≤ 1.

Semidefinite Programming

$$\begin{array}{ll} \text{minimize} & f(x) \\ \\ \text{subject to} & x \in \mathbb{R}^n \\ & \begin{pmatrix} g_{11}(x) & \dots & g_{1m}(x) \\ \vdots & \ddots & \vdots \\ & \dots & g_{mm}(x) \end{pmatrix} \text{is psd} \end{array}$$

where all polynomials f and g_{ij} are linear, i.e., their degree is ≤ 1.

Positive semidefinite matrices and families of vectors

Proposition. A real symmetric $k \times k$ matrix is psd if and only if there are vectors $v_1, \dots, v_k \in \mathbb{R}^k$ such that

$$M = \begin{pmatrix} \langle v_1, v_1 \rangle & \dots & \langle v_1, v_k \rangle \\ \vdots & & \vdots \\ \langle v_k, v_1 \rangle & \dots & \langle v_k, v_k \rangle \end{pmatrix}.$$

Duality

- Every linear program (P) has an optimal value P^*.

Duality

- Every linear program (P) has an optimal value P^*.
- To every linear program (P), one can define a dual program (D) which is again a linear program.

Duality

- Every linear program (P) has an optimal value P^*.
- To every linear program (P), one can define a dual program (D) which is again a linear program.
- If (P) is a minimization problem, then (D) is a maximization problem and weak duality holds:

$$D^* \leq P^*$$

Duality

- Every linear program (P) has an optimal value P^*.
- To every linear program (P), one can define a dual program (D) which is again a linear program.
- If (P) is a minimization problem, then (D) is a maximization problem and weak duality holds:

$$D^* \leq P^*$$

- Strong duality is desired and often holds:

$$D^* = P^*$$

Duality

- Every semidefinite program (P) has an optimal value P^*.
- To every semidefinite program (P), one can define a dual program (D) which is again a semidefinite program.
- If (P) is a minimization problem, then (D) is a maximization problem and weak duality holds:

$$D^* \leq P^*$$

- Strong duality is desired and often holds:

$$D^* = P^*$$

$$\begin{aligned} \text{minimize} \quad & \sum_{i=0}^{2d} a_i x^i \\ \text{subject to} \quad & x \in \mathbb{R} \end{aligned}$$

where $a_0, \ldots, a_{2d} \in \mathbb{R}$.

$$\text{minimize} \quad \sum_{i=0}^{2d} a_i x^i$$

$$\text{subject to} \quad x \in \mathbb{R}$$

Note that $\begin{pmatrix} 1 & x & x^2 & \dots & x^d \\ x & x^2 & \ddots & \ddots & \\ x^2 & \ddots & \ddots & & \\ \vdots & \ddots & & & \\ x^d & & & & x^{2d} \end{pmatrix}$ is psd

where $a_0, \dots, a_{2d} \in \mathbb{R}$.

$$\text{minimize} \quad \sum_{i=0}^{2d} a_i x^i$$

$$\text{subject to} \quad x \in \mathbb{R}$$

Note that

$$\begin{array}{c} \\ 1 \\ X \\ X^2 \\ \vdots \\ X^d \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & X & X^2 & \dots & X^d \end{array} \\ \begin{pmatrix} 1 & x & x^2 & \dots & x^d \\ x & x^2 & \ddots & \ddots & \\ x^2 & \ddots & \ddots & & \\ \vdots & \ddots & & & \\ x^d & & & & x^{2d} \end{pmatrix} \end{array} \text{is psd}$$

where $a_0, \dots, a_{2d} \in \mathbb{R}$.

$$(P) \qquad \text{minimize} \quad \sum_{i=1}^{2d} a_i y_i + a_0$$

$$\text{subject to} \quad y \in \mathbb{R}^{2d}$$

$$\begin{array}{c} \\ 1 \\ X \\ X^2 \\ \vdots \\ X^d \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & X & X^2 & \dots & X^d \end{array} \\ \begin{pmatrix} 1 & y_1 & y_2 & & y_d \\ y_1 & y_2 & \ddots & \ddots & \\ y_2 & \ddots & \ddots & & \\ \vdots & \ddots & & & \\ y_d & & & & y_{2d} \end{pmatrix} \end{array} \text{ is psd}$$

where $a_0, \dots, a_{2d} \in \mathbb{R}$.

Set $f := \sum_{i=0}^{2d} a_i X^i$ and denote by (D) the semidefinite program dual to (P).

Set $f := \sum_{i=0}^{2d} a_i X^i$ and denote by (D) the semidefinite program dual to (P). Then it is clear that

$$D^* \leq P^* \leq f^*.$$

Set $f := \sum_{i=0}^{2d} a_i X^i$ and denote by (D) the semidefinite program dual to (P). Then it is clear that

$$D^* \leq P^* \leq f^*.$$

It turns out that (D) can be interpreted as:

$$(D) \qquad \begin{array}{rl} \text{maximize} & \mu \\ \text{subject to} & f - \mu \text{ is sos} \end{array}$$

Set $f := \sum_{i=0}^{2d} a_i X^i$ and denote by (D) the semidefinite program dual to (P). Then it is clear that

$$D^* \leq P^* \leq f^*.$$

It turns out that (D) can be interpreted as:

$$(D) \qquad \begin{array}{rl} \text{maximize} & \mu \\ \text{subject to} & f - \mu \text{ is sos} \end{array}$$

Proposition. For every $p \in \mathbb{R}[X]$,

$$p \geq 0 \text{ on } \mathbb{R} \implies p \text{ is a sum of two squares in } \mathbb{R}[X].$$

Set $f := \sum_{i=0}^{2d} a_i X^i$ and denote by (D) the semidefinite program dual to (P). Then it is clear that

$$D^* \leq P^* \leq f^*.$$

It turns out that (D) can be interpreted as:

$$(D) \qquad \begin{array}{rl} \text{maximize} & \mu \\ \text{subject to} & f - \mu \text{ is sos} \end{array}$$

Proposition. For every $p \in \mathbb{R}[X]$,

$$p \geq 0 \text{ on } \mathbb{R} \implies p \text{ is a sum of two squares in } \mathbb{R}[X].$$

Corollary.

$$D^* = P^* = f^*$$

$$\text{minimize} \quad \sum_{i+j \leq 4} a_{ij} x^i y^j$$

$$\text{subject to} \quad x, y \in \mathbb{R}$$

where $a_{ij} \in \mathbb{R}$ $(i + j \leq 4)$.

$$\text{minimize} \quad \sum_{i+j \le 4} a_{ij} x^i y^j$$

$$\text{subject to} \quad x, y \in \mathbb{R}$$

Note that

$$\begin{pmatrix} 1 & x & y & x^2 & xy & y^2 \\ x & x^2 & xy & x^3 & x^2y & xy^2 \\ y & xy & y^2 & x^2y & xy^2 & y^3 \\ x^2 & x^3 & x^2y & x^4 & x^3y & x^2y^2 \\ xy & x^2y & xy^2 & x^3y & x^2y^2 & xy^3 \\ y^2 & xy^2 & y^3 & x^2y^2 & xy^3 & y^4 \end{pmatrix} \text{ is psd}$$

where $a_{ij} \in \mathbb{R}$ $(i + j \le 4)$.

$$\text{minimize} \quad \sum_{i+j\leq 4} a_{ij} x^i y^j$$

$$\text{subject to} \quad x, y \in \mathbb{R}$$

Note that

$$\begin{array}{c} \\ 1 \\ X \\ Y \\ X^2 \\ XY \\ Y^2 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & X & Y & X^2 & XY & Y^2 \end{array} \\ \begin{pmatrix} 1 & x & y & x^2 & xy & y^2 \\ x & x^2 & xy & x^3 & x^2y & xy^2 \\ y & xy & y^2 & x^2y & xy^2 & y^3 \\ x^2 & x^3 & x^2y & x^4 & x^3y & x^2y^2 \\ xy & x^2y & xy^2 & x^3y & x^2y^2 & xy^3 \\ y^2 & xy^2 & y^3 & x^2y^2 & xy^3 & y^4 \end{pmatrix} \end{array} \text{is psd}$$

where $a_{ij} \in \mathbb{R}$ $(i + j \leq 4)$.

$$(P) \qquad \text{minimize} \quad \sum_{1 \le i+j \le 4} a_{ij} y_{ij} + a_{00}$$

$$\text{subject to} \quad y_{ij} \in \mathbb{R} \; (1 \le i+j \le 4)$$

$$\begin{array}{c} \\ 1 \\ X \\ Y \\ X^2 \\ XY \\ Y^2 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & X & Y & X^2 & XY & Y^2 \end{array} \\ \begin{pmatrix} 1 & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} \\ y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} \\ y_{01} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} \\ y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22} \\ y_{11} & y_{21} & y_{12} & y_{31} & y_{22} & y_{13} \\ y_{02} & y_{12} & y_{03} & y_{22} & y_{13} & y_{04} \end{pmatrix} \end{array} \text{ is psd}$$

where $a_{ij} \in \mathbb{R}$ $(i + j \le 4)$.

Set $f := \sum_{i+j \leq 4} a_{ij} X^{ij}$ and denote by (D) the semidefinite program dual to (P).

Set $f := \sum_{i+j\le 4} a_{ij} X^{ij}$ and denote by (D) the semidefinite program dual to (P). Then it is clear that

$$D^* \le P^* \le f^*.$$

Set $f := \sum_{i+j \le 4} a_{ij} X^{ij}$ and denote by (D) the semidefinite program dual to (P). Then it is clear that

$$D^* \le P^* \le f^*.$$

It turns out that (D) can be interpreted as:

$$(D) \qquad \begin{array}{rl} \text{maximize} & \mu \\ \text{subject to} & f - \mu \text{ is sos} \end{array}$$

Set $f := \sum_{i+j \leq 4} a_{ij} X^{ij}$ and denote by (D) the semidefinite program dual to (P). Then it is clear that

$$D^* \leq P^* \leq f^*.$$

It turns out that (D) can be interpreted as:

$$(D) \qquad \begin{array}{rl} \text{maximize} & \mu \\ \text{subject to} & f - \mu \text{ is sos} \end{array}$$

Theorem (Hilbert). For every $p \in \mathbb{R}[X, Y]$ of degree ≤ 4,

$$p \geq 0 \text{ on } \mathbb{R}^2 \implies p \text{ is a sum of three squares in } \mathbb{R}[X, Y].$$

David Hilbert: Ueber die Darstellung definiter Formen als Summe von Formenquadraten
Math. Ann. XXXII 342-350 (1888)

http://www-gdz.sub.uni-goettingen.de/cgi-bin/digbib.cgi?PPN235181684_0032

Set $f := \sum_{i+j \leq 4} a_{ij} X^{ij}$ and denote by (D) the semidefinite program dual to (P). Then it is clear that

$$D^* \leq P^* \leq f^*.$$

It turns out that (D) can be interpreted as:

$$(D) \qquad \begin{array}{rl} \text{maximize} & \mu \\ \text{subject to} & f - \mu \text{ is sos} \end{array}$$

Theorem (Hilbert). For every $p \in \mathbb{R}[X, Y]$ of degree ≤ 4,

$$p \geq 0 \text{ on } \mathbb{R}^2 \implies p \text{ is a sum of three squares in } \mathbb{R}[X, Y].$$

Corollary. $D^* = P^* = f^*$

David Hilbert: Ueber die Darstellung definiter Formen als Summe von Formenquadraten
Math. Ann. XXXII 342-350 (1888)

http://www-gdz.sub.uni-goettingen.de/cgi-bin/digbib.cgi?PPN235181684_0032

The Motzkin polynomial

- Unfortunately, not every polynomial $p \in \mathbb{R}[X_1, \ldots, X_n]$ with $p \geq 0$ on $\mathbb{R}^n$ is a sum of squares of polynomials.

The Motzkin polynomial

- Unfortunately, not every polynomial $p \in \mathbb{R}[X_1, \ldots, X_n]$ with $p \geq 0$ on $\mathbb{R}^n$ is a sum of squares of polynomials.
- The first explicit example was found in 1967 by Motzkin:

$$p := X^4Y^2 + X^2Y^4 - 3X^2Y^2 + 1$$

The Motzkin polynomial

- Unfortunately, not every polynomial $p \in \mathbb{R}[X_1, \dots, X_n]$ with $p \geq 0$ on $\mathbb{R}^n$ is a sum of squares of polynomials.
- The first explicit example was found in 1967 by Motzkin:

$$p := X^4Y^2 + X^2Y^4 - 3X^2Y^2 + 1$$

- In fact, there is even no $N \in \mathbb{N}$ such that $p + N$ is a sum of squares in $\mathbb{R}[X, Y, Z]$.

The Motzkin polynomial

- Unfortunately, not every polynomial $p \in \mathbb{R}[X_1, \ldots, X_n]$ with $p \geq 0$ on $\mathbb{R}^n$ is a sum of squares of polynomials.
- The first explicit example was found in 1967 by Motzkin:

$$p := X^4Y^2 + X^2Y^4 - 3X^2Y^2 + 1$$

- In fact, there is even no $N \in \mathbb{N}$ such that $p + N$ is a sum of squares in $\mathbb{R}[X, Y, Z]$.
- Described method always yields certified lower bounds, but they might by $-\infty$:

$$-\infty \leq D^* = P^* \leq f^*$$

The Motzkin polynomial

- Unfortunately, not every polynomial $p \in \mathbb{R}[X_1, \ldots, X_n]$ with $p \geq 0$ on $\mathbb{R}^n$ is a sum of squares of polynomials.
- The first explicit example was found in 1967 by Motzkin:

$$p := X^4Y^2 + X^2Y^4 - 3X^2Y^2 + 1$$

- In fact, there is even no $N \in \mathbb{N}$ such that $p + N$ is a sum of squares in $\mathbb{R}[X, Y, Z]$.
- Described method always yields certified lower bounds, but they might by $-\infty$:

$$-\infty \leq D^* = P^* \leq f^*$$

- But there are a lot of remedies...

Case where S is compact.

For simplicity, we suppose $m = 1$ and write $g := g_1$ (technical difficulties which are however not very serious otherwise), i.e.

$$S = \{x \in \mathbb{R}^n \mid g(x) \geq 0\}.$$

Case where S is compact.

For simplicity, we suppose $m = 1$ and write $g := g_1$ (technical difficulties which are however not very serious otherwise), i.e.

$$S = \{x \in \mathbb{R}^n \mid g(x) \geq 0\}.$$

Now we get a sequence $(P_k)_{2k \geq d}$ of relaxations such that

$$D_k^* \leq P_k^* \leq f^* \qquad \text{and} \qquad \lim_{k \to \infty} D_k^* = \lim_{k \to \infty} P_k^* = f^*.$$

Jean Lasserre: Global optimization with polynomials and the problem of moments
SIAM J. Optim. **11**, No. 3, 796–817 (2001)

$$\text{minimize} \quad \sum_{|\alpha| \le d} a_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n}$$

$$\text{subject to} \quad x \in S$$

where $k \in \mathbb{N}$, $2k \ge d$, $a_\alpha \in \mathbb{R}$ $(|\alpha| \le k)$.

$$\text{minimize} \quad \sum_{|\alpha| \le d} a_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n}$$

$$\text{subject to} \quad x \in S$$

Note that

$$\left(\begin{pmatrix} 1 & x_1 & \cdots & x_n^k \\ x_1 & & & \vdots \\ \vdots & & & \\ x_n^k & \cdots\cdots\cdots & & x_n^{2k} \end{pmatrix} \qquad \begin{pmatrix} \text{``localization} \\ \text{matrix''} \end{pmatrix} \right) \text{is psd}$$

where $k \in \mathbb{N}$, $2k \ge d$, $a_\alpha \in \mathbb{R}$ $(|\alpha| \le k)$.

minimize $\displaystyle\sum_{|\alpha|\leq d} a_\alpha x_1^{\alpha_1}\cdots x_n^{\alpha_n}$

subject to $x \in S$

Note that

$$\begin{matrix} & \\ & \begin{matrix} 1 & X_1 & \ldots & X_n^k \end{matrix} \\ \begin{matrix} 1 \\ X_1 \\ \vdots \\ X_n^k \end{matrix} & \left(\begin{pmatrix} 1 & x_1 & \ldots & x_n^k \\ x_1 & & & \vdots \\ \vdots & & & \\ x_n^k & & \ldots\ldots\ldots & x_n^{2k} \end{pmatrix} \quad \begin{pmatrix} \text{“localization} \\ \text{matrix”} \end{pmatrix} \right) \end{matrix} \text{ is psd}$$

where $k \in \mathbb{N}$, $2k \geq d$, $a_\alpha \in \mathbb{R}$ $(|\alpha| \leq k)$.

$$(P_k) \quad \text{minimize} \quad \sum_{1 \le |\alpha| \le d} a_\alpha y_\alpha + a_0$$

$$\text{subject to} \quad y_\alpha \in \mathbb{R} \qquad (|\alpha| \le k)$$

$$\begin{matrix} & & 1 & X_1 & \dots & X_n^k & & \\ 1 & \left(\left(\vphantom{\begin{matrix}1\\1\\1\\1\end{matrix}} \right. \right. & 1 & y_{10\dots0} & \cdots & & \left. \vphantom{\begin{matrix}1\\1\\1\\1\end{matrix}} \right) & \\ X_1 & & y_{10\dots0} & & & & & \\ \vdots & & \vdots & & & & & \\ X_n^k & & & & & & & \begin{pmatrix} \text{“localization} \\ \text{matrix”} \end{pmatrix} \Bigg) \end{matrix} \text{ is psd}$$

where $k \in \mathbb{N}$, $2k \ge d$, $a_\alpha \in \mathbb{R}$ $(|\alpha| \le k)$.

Case where S is compact.

Theorem (Schmüdgen, Putinar, ...) If $f > 0$ on S, then $f = s + gt$ for sums of squares s, t in $\mathbb{R}[X_1, \ldots, X_n]$.

Case where S is compact.

Theorem (Schmüdgen, Putinar, ...) If $f > 0$ on S, then $f = s + gt$ for sums of squares s, t in $\mathbb{R}[X_1, \ldots, X_n]$.

Corollary (Lasserre). $(D_k^*)_{k \in \mathcal{N}}$ and $(P_k^*)_{k \in \mathcal{N}}$ are increasing sequences that converge to f^* and satisfy $D_k^* \leq P_k^* \leq f^*$. How fast?

Case where S is compact.

Theorem (Schmüdgen, Putinar, ...) If $f > 0$ on S, then $f = s + gt$ for sums of squares s, t in $\mathbb{R}[X_1, \ldots, X_n]$.

Corollary (Lasserre). $(D_k^*)_{k\in\mathcal{N}}$ and $(P_k^*)_{k\in\mathcal{N}}$ are increasing sequences that converge to f^* and satisfy $D_k^* \leq P_k^* \leq f^*$. How fast?

Theorem. There exists $C \in \mathbb{N}$ depending on f and g and $c \in \mathbb{N}$ depending on g such that

$$f^* - D_k^* \leq \frac{C}{\sqrt[c]{k}} \qquad \text{for big } k.$$

On the complexity of Schmüdgen's Positivstellensatz
Journal of Complexity **20**, No. 4, 529—543 (2004)

Optimization of polynomials on compact semialgebraic sets
SIAM Journal on Optimization 15, No. 3, 805-825 (2005)

Further properties of the method for compact S

- Feasible solutions of (D_k) are certified lower bounds of f^*.

Further properties of the method for compact S

- Feasible solutions of (D_k) are certified lower bounds of f^*.
- Method converges from below to f^*.

Further properties of the method for compact S

- Feasible solutions of (D_k) are certified lower bounds of f^*.
- Method converges from below to f^*.
- Method converges to unique minimizers.

Optimization of polynomials on compact semialgebraic sets
SIAM Journal on Optimization 15, No. 3, 805-825 (2005)

Further properties of the method for compact S

- Feasible solutions of (D_k) are certified lower bounds of f^*.
- Method converges from below to f^*.
- Method converges to unique minimizers. Disadvantage: Possibly from outside the set S.

Optimization of polynomials on compact semialgebraic sets
SIAM Journal on Optimization 15, No. 3, 805-825 (2005)

Further properties of the method for compact S

- Feasible solutions of (D_k) are certified lower bounds of f^*.
- Method converges from below to f^*.
- Method converges to unique minimizers. Disadvantage: Possibly from outside the set S.
- If there is a unique minimizer and it lies in the interior of S,

Optimization of polynomials on compact semialgebraic sets
SIAM Journal on Optimization 15, No. 3, 805-825 (2005)

Further properties of the method for compact S

- Feasible solutions of (D_k) are certified lower bounds of f^*.
- Method converges from below to f^*.
- Method converges to unique minimizers. Disadvantage: Possibly from outside the set S.
- If there is a unique minimizer and it lies in the interior of S, then the method produces a sequence of intervals containing f^* whose endpoints converge to f^*.

Optimization of polynomials on compact semialgebraic sets
SIAM Journal on Optimization 15, No. 3, 805-825 (2005)

Implementations

- Henrion and Lasserre: GloptiPoly
  `http://www.laas.fr/~henrion/software/gloptipoly/`
- Prajna, Papachristodoulou, Parrilo: SOSTOOLS
  `http://control.ee.ethz.ch/~parrilo/sostools/`
- Both use the free SeDuMi solver by Jos Sturm
- But they need MATLAB and the MATLAB Symbolic Toolbox

Example: The maximum cut problem

Given a graph, i.e., an $n \in \mathbb{N}$ (number of nodes) and a set

$$E \subseteq \{(i, j) \in \{1, \ldots, n\}^2 \mid i < j\}$$

(of edges),

Example: The maximum cut problem

Given a graph, i.e., an $n \in \mathbb{N}$ (number of nodes) and a set

$$E \subseteq \{(i,j) \in \{1,\ldots,n\}^2 \mid i < j\}$$

(of edges), find the maximum cut value, i.e., the maximal possible number of edges that connect nodes with different signs when each node is assigned a sign $+$ or $-$.

Example: The maximum cut problem

Given a graph, i.e., an $n \in \mathbb{N}$ (number of nodes) and a set

$$E \subseteq \{(i,j) \in \{1,\ldots,n\}^2 \mid i < j\}$$

(of edges), find the maximum cut value, i.e., the maximal possible number of edges that connect nodes with different signs when each node is assigned a sign $+$ or $-$.

$$\begin{array}{ll} \text{maximize} & \displaystyle\sum_{(i,j)\in E} \frac{1}{2}(1 - x_i x_j) \\ \text{subject to} & x_i^2 = 1 \text{ for all } i \in \{1,\ldots,n\} \end{array}$$

MAXCUT

$$\text{maximize} \quad \sum_{(i,j)\in E} \frac{1}{2}(1 - x_i x_j)$$

$$\text{subject to} \quad x \in \{-1, 1\}^n$$

MAXCUT

$$\text{maximize} \quad \sum_{(i,j)\in E} \frac{1}{2}(1 - x_i x_j)$$

$$\text{subject to} \quad x \in \{-1, 1\}^n$$

Note that $\begin{pmatrix} 1 & x_1x_2 & \dots & x_1x_n \\ x_2x_1 & 1 & & x_2x_n \\ \vdots & & \ddots & \vdots \\ x_nx_1 & \dots\dots\dots & & 1 \end{pmatrix}$ is psd

MAXCUT

$$\text{maximize} \quad \sum_{(i,j)\in E} \frac{1}{2}(1 - x_i x_j)$$

$$\text{subject to} \quad x \in \{-1, 1\}^n$$

Note that
$$\begin{array}{c} \\ X_1 \\ \vdots \\ \vdots \\ X_n \end{array} \begin{array}{c} \begin{array}{ccccc} X_1 & \dots & \dots & \dots & X_n \end{array} \\ \begin{pmatrix} 1 & x_1x_2 & \dots & x_1x_n \\ x_2x_1 & 1 & & x_2x_n \\ \vdots & & \ddots & \vdots \\ x_nx_1 & \dots\dots\dots & & 1 \end{pmatrix} \end{array} \text{is psd}$$

First MAXCUT relaxation

$$(P_1) \qquad \text{maximize} \quad \sum_{(i,j)\in E} \frac{1}{2}(1 - y_{ij})$$

$$\text{subject to} \quad y_{ij} \in \mathbb{R} \qquad (1 \le i < j \le n)$$

$$\begin{array}{c} \\ X_1 \\ \vdots \\ \vdots \\ X_n \end{array} \begin{array}{c} \begin{array}{ccccc} X_1 & \dots & \dots & \dots & X_n \end{array} \\ \begin{pmatrix} 1 & y_{12} & \dots & y_{1n} \\ y_{12} & 1 & & y_{2n} \\ \vdots & & \ddots & \vdots \\ y_{1n} & \dots\dots\dots & & 1 \end{pmatrix} \end{array} \text{is psd}$$

MAXCUT

$$\text{maximize} \quad \sum_{(i,j)\in E} \frac{1}{2}(1 - x_i x_j)$$

$$\text{subject to} \quad x \in \{-1, 1\}^n$$

MAXCUT

$$\text{maximize} \quad \sum_{(i,j)\in E} \frac{1}{2}(1 - x_i x_j)$$

$$\text{subject to} \quad x \in \{-1, 1\}^n$$

Note that

$$\begin{pmatrix} 1 & x_1 x_2 & \dots & \dots & \dots & \dots \\ x_2 x_1 & 1 & & & & \\ \vdots & & \ddots & & & \\ \vdots & & & \ddots & & \\ & & & & & 1 \end{pmatrix} \text{is psd}$$

MAXCUT

$$\text{maximize} \quad \sum_{(i,j)\in E} \frac{1}{2}(1 - x_i x_j)$$

$$\text{subject to} \quad x \in \{-1, 1\}^n$$

Note that

$$\begin{array}{c} \\ 1 \\ X_1X_2 \\ X_1X_3 \\ \vdots \\ X_{n-1}X_n \end{array} \begin{array}{c} \begin{array}{cccc} 1 & X_1X_2 & X_1X_3 \ldots & X_{n-1}X_n \end{array} \\ \begin{pmatrix} 1 & x_1x_2 & \ldots & \ldots & \ldots & \ldots \\ x_2x_1 & 1 & & & & \\ \vdots & & \ddots & & & \\ \vdots & & & \ddots & & \\ & & & & & 1 \end{pmatrix} \end{array} \text{is psd}$$

Second MAXCUT relaxation

$$(P_2) \qquad \text{maximize} \quad \sum_{(i,j)\in E} \frac{1}{2}(1 - y_{ij})$$

$$\text{subject to} \quad y_{ij} \in \mathbb{R} \qquad (1 \leq i < j \leq n)$$

$$\begin{array}{c} \\ 1 \\ X_1X_2 \\ X_1X_3 \\ \vdots \\ X_{n-1}X_n \end{array} \begin{array}{c} \begin{array}{cccc} 1 & X_1X_2 & X_1X_3 \dots & X_{n-1}X_n \end{array} \\ \begin{pmatrix} 1 & y_{12} & \cdots & \cdots & \cdots & \cdots \\ y_{12} & 1 & & & & \\ \vdots & & \ddots & & & \\ \vdots & & & \ddots & & \\ & & & & & 1 \end{pmatrix} \end{array} \text{is psd}$$

- The maximum cut problem is NP–complete

- The maximum cut problem is NP–complete
- The first relaxation gives a polynomial time algorithm which overestimates the maximum cut value at most by a factor of ≈ 1.1382.

- The maximum cut problem is NP–complete
- The first relaxation gives a polynomial time algorithm which overestimates the maximum cut value at most by a factor of ≈ 1.1382.
- The first relaxation is the famous algorithm of Goemans and Williamson.

- The maximum cut problem is NP–complete
- The first relaxation gives a polynomial time algorithm which overestimates the maximum cut value at most by a factor of ≈ 1.1382.
- The first relaxation is the famous algorithm of Goemans and Williamson. From no polynomial algorithm it is known that it has a better approximation ratio.

- The maximum cut problem is NP–complete
- The first relaxation gives a polynomial time algorithm which overestimates the maximum cut value at most by a factor of ≈ 1.1382.
- The first relaxation is the famous algorithm of Goemans and Williamson. From no polynomial algorithm it is known that it has a better approximation ratio. Existence of such an algorithm with ratio < 1.0625 implies $P = NP$ (Hastad).

- The maximum cut problem is NP–complete
- The first relaxation gives a polynomial time algorithm which overestimates the maximum cut value at most by a factor of ≈ 1.1382.
- The first relaxation is the famous algorithm of Goemans and Williamson. From no polynomial algorithm it is known that it has a better approximation ratio. Existence of such an algorithm with ratio < 1.0625 implies $P = NP$ (Hastad).
- Solving the second relaxation is a polynomial time algorithm which yields the exact value for all planar graphs (consequence of results of Seymour, Barahona, Mahjoub),

- The maximum cut problem is NP–complete
- The first relaxation gives a polynomial time algorithm which overestimates the maximum cut value at most by a factor of ≈ 1.1382.
- The first relaxation is the famous algorithm of Goemans and Williamson. From no polynomial algorithm it is known that it has a better approximation ratio. Existence of such an algorithm with ratio < 1.0625 implies $P = NP$ (Hastad).
- Solving the second relaxation is a polynomial time algorithm which yields the exact value for all planar graphs (consequence of results of Seymour, Barahona, Mahjoub), and is conjectured to improve over the GW–algorithm.

- The maximum cut problem is NP–complete
- The first relaxation gives a polynomial time algorithm which overestimates the maximum cut value at most by a factor of ≈ 1.1382.
- The first relaxation is the famous algorithm of Goemans and Williamson. From no polynomial algorithm it is known that it has a better approximation ratio. Existence of such an algorithm with ratio < 1.0625 implies $P = NP$ (Hastad).
- Solving the second relaxation is a polynomial time algorithm which yields the exact value for all planar graphs (consequence of results of Seymour, Barahona, Mahjoub), and is conjectured to improve over the GW–algorithm.
- The n–th relaxation yields the exact maximum cut value.

Exactness of the n-th MAXCUT relaxation

Proposition. Suppose $p \in \mathbb{R}[X_1, \ldots, X_n]$ such that

$$p \geq 0 \text{ on } \{-1, 1\}^n.$$

Then f is a square modulo the ideal

$$I := (X_1^2 - 1, \ldots, X_n^2 - 1) \subseteq \mathbb{R}[X_1, \ldots, X_n].$$

Exactness of the n-th MAXCUT relaxation

Proposition. Suppose $p \in \mathbb{R}[X_1, \dots, X_n]$ such that

$$p \geq 0 \text{ on } \{-1, 1\}^n.$$

Then f is a square modulo the ideal

$$I := (X_1^2 - 1, \dots, X_n^2 - 1) \subseteq \mathbb{R}[X_1, \dots, X_n].$$

Proof by algebra. By chinese remainder theorem

$$\mathbb{R}[X_1, \dots, X_n]/I \cong \mathbb{R}^{\{-1,1\}^n} \cong \mathbb{R}^{2^n}.$$

Exactness of the n-th MAXCUT relaxation

Proposition. Suppose $p \in \mathbb{R}[X_1, \dots, X_n]$ such that

$$p \geq 0 \text{ on } \{-1, 1\}^n.$$

Then f is a square modulo the ideal

$$I := (X_1^2 - 1, \dots, X_n^2 - 1) \subseteq \mathbb{R}[X_1, \dots, X_n].$$

Proof by algebra. By chinese remainder theorem

$$\mathbb{R}[X_1, \dots, X_n]/I \cong \mathbb{R}^{\{-1,1\}^n} \cong \mathbb{R}^{2^n}.$$

Proof by algebraic geometry. I is a zero-dimensional radical ideal.

Exactness of the n-th MAXCUT relaxation

Proposition. Suppose $p \in \mathbb{R}[X_1, \dots, X_n]$ such that

$$p \geq 0 \text{ on } \{-1, 1\}^n.$$

Then f is a square modulo the ideal

$$I := (X_1^2 - 1, \dots, X_n^2 - 1) \subseteq \mathbb{R}[X_1, \dots, X_n].$$

Proof by algebra. By chinese remainder theorem

$$\mathbb{R}[X_1, \dots, X_n]/I \cong \mathbb{R}^{\{-1,1\}^n} \cong \mathbb{R}^{2^n}.$$

Proof by algebraic geometry. I is a zero-dimensional radical ideal.

Corollary. $D_n^* = P_n^* = f^*$

The story goes on...

Theorem (Lasserre). For every $p \in \mathbb{R}[X_1, \ldots, X_n]$, the following are equivalent:

(i) $p \geq 0$ on $\mathbb{R}^n$

The story goes on...

Theorem (Lasserre). For every $p \in \mathbb{R}[X_1, \ldots, X_n]$, the following are equivalent:

(i) $p \geq 0$ on $\mathbb{R}^n$

(ii) For every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$p + \varepsilon \sum_{i=1}^{n} \sum_{k=0}^{N} \frac{X_i^{2k}}{k!} \text{ is sos.}$$

Jean Lasserre: A sum of squares approximation of nonnegative polynomials

`http://front.math.ucdavis.edu/math.AG/0412398`

The story goes on...

Theorem (Nie, Demmel, Sturmfels). If $p > 0$ on $\mathbb{R}^n$, then p is sos modulo its own gradient ideal

$$I := \left(\frac{\partial f}{\partial X_1}, \ldots, \frac{\partial f}{\partial X_n} \right).$$

Nie, Demmel, Sturmfels: Minimizing Polynomials via Sum of Squares over the Gradient Ideal
`http://front.math.ucdavis.edu/math.OC/0411342`