SOS and SDP Relaxation of Polynomial Optimization Problems

Masakazu Kojima
Tokyo Institute of Technology

November 2010

At National Cheng-Kung University, Tainan

Purpose of this talk —

Introduction to

Sum Of Squares (SOS) and

Semidefinite Programming (SDP)

Relaxation of

Polynomial Optimization Problems (POPs)

Contents

1. Global optimization of nonconvex problems
2. Polynomial Optimization Problems (POPs)
3. Basic idea of SOS and SDP relaxations
4. SOS relaxation of POPs
5. SDP relaxation of POPs
6. Exploiting sparsity in SOS and SDP relaxations of POPs
7. Numerical results

Purpose of this talk —

Introduction to

Sum Of Squares (SOS) and

Semidefinite Programming (SDP)

Relaxation of

Polynomial Optimization Problems (POPs)

Contents

1. Global optimization of nonconvex problems
2. Polynomial Optimization Problems (POPs)
3. Basic idea of SOS and SDP relaxations
4. SOS relaxation of POPs
5. SDP relaxation of POPs
6. Exploiting sparsity in SOS and SDP relaxations of POPs
7. Numerical results

OP : Optimization problem in the n-dim. Euclidean space $\mathbb{R}^n$

min. $f(\boldsymbol{x})$ sub.to $\boldsymbol{x} \in S \subseteq \mathbb{R}^n$, where $f : \mathbb{R}^n \rightarrow \mathbb{R}$.

We want to approximate a global optimal solution $\boldsymbol{x}^*$;

$$\boldsymbol{x}^* \in S \text{ and } f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}) \text{ for all } \boldsymbol{x} \in S$$

if it exists. But, impossible without any assumption.

Various assumptions

- continuity, differentiability, compactness, ...
- convexity $\Rightarrow$ local opt. sol. $\equiv$ global opt. sol.
  $\Rightarrow$ local improvement leads to a global opt. sol.
  - Powerful software for convex problems $\ni$ LPs, SDPs, ...

OP : Optimization problem in the n-dim. Euclidean space $\mathbb{R}^n$

min. $f(\boldsymbol{x})$ sub.to $\boldsymbol{x} \in S \subseteq \mathbb{R}^n$, where $f : \mathbb{R}^n \rightarrow \mathbb{R}$.

We want to approximate a global optimal solution $\boldsymbol{x}^*$;

$$\boldsymbol{x}^* \in S \text{ and } f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}) \text{ for all } \boldsymbol{x} \in S$$

if it exists. But, impossible without any assumption.

Various assumptions

- continuity, differentiability, compactness, ...
- convexity $\Rightarrow$ local opt. sol. $\equiv$ global opt. sol.
  $\Rightarrow$ local improvement leads to a global opt. sol.
  - Powerful software for convex problems $\ni$ LPs, SDPs, ...

What can we do beyond convexity?

- We still need proper models and assumptions
  - Polynomial Optimization Problems (POPs) — this talk
- Main tool is SDP relaxation — this talk
  Powerful in theory but expensive in practice
- Exploiting sparsity in large scale SDPs & POPs — this talk

Contents

POP: min $f_0(\boldsymbol{x})$ sub.to $f_k(\boldsymbol{x}) \geq 0$ or $= 0$ $(k = 1, \ldots, m)$

$\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$: a vector variable
$f_k(\boldsymbol{x})$: a real-valued polynomial in $x_1, \ldots, x_n$ $(k = 0, 1, \ldots, m)$

Example. $n = 3$, $\boldsymbol{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$: a vector variable

$$
\begin{array}{rl}
\text{min} & f_0(\boldsymbol{x}) \equiv x_1^3 - 2x_1x_2^2 + x_1^2x_2x_3 - 4x_3^2 \\
\text{sub.to} & f_1(\boldsymbol{x}) \equiv -x_1^2 + 5x_2x_3 + 1 \geq 0, \\
& f_2(\boldsymbol{x}) \equiv x_1^2 - 3x_1x_2x_3 + 2x_3 + 2 \geq 0, \\
& f_3(\boldsymbol{x}) \equiv -x_1^2 - x_2^2 - x_3^2 + 1 \geq 0. \\
& x_1(x_1 - 1) = 0 \text{ (0-1 integer cond.)}, \\
& x_2 \geq 0, \ x_3 \geq 0, \ x_2x_3 = 0 \text{ (comp. cond.)}.
\end{array}
$$

- Various problems (including 0-1 integer programs) $\Rightarrow$ POP
- POP serves as a unified theoretical model for global optimization in nonlinear and combinatorial optimization problems.

POP: min $f_0(\boldsymbol{x})$ sub.to $f_k(\boldsymbol{x}) \geq 0$ or $= 0$ $(k = 1, \ldots, m)$

POP: min $f_0(\boldsymbol{x})$ sub.to $f_k(\boldsymbol{x}) \geq 0$ or $= 0$ $(k = 1, \ldots, m)$

[1] Lasserre, "Global optimization with polynomials and the problems of moments", *SIAM J. on Optim.* (2001).

[2] Parrilo, "Semidefinite programming relaxations for semialgebraic problems", *Math. Prog.* (2003).

- primal approach $\Rightarrow$ a sequence of SDP relaxations.
- dual approach $\Rightarrow$ a sequence of SOS relaxations.

POP: min $f_0(\boldsymbol{x})$ sub.to $f_k(\boldsymbol{x}) \geq 0$ or $= 0$ $(k = 1, \ldots, m)$

[1] Lasserre, “Global optimization with polynomials and the problems of moments”, *SIAM J. on Optim.* (2001).

[2] Parrilo, “Semidefinite programming relaxations for semialgebraic problems”, *Math. Prog.* (2003).

- primal approach $\Rightarrow$ a sequence of SDP relaxations.
- dual approach $\Rightarrow$ a sequence of SOS relaxations.

Main features:

(a) Lower bounds for the optimal value.

(b) Convergence to global optimal solutions under assump.

(c) Each relaxed problem can be solved as an SDP; its size $\uparrow$ rapidly along “the sequence” as the size of POP $\uparrow$, the deg. of poly. $\uparrow$, and/or we require higher accuracy.

(d) Expensive to solve large scale POPs in practice.

$\Rightarrow$ Exploiting Sparsity.

Contents

1. Global optimization of nonconvex problems
2. Polynomial Optimization Problems (POPs)
3. Basic idea of SOS and SDP relaxations

Three ways of describing the SDP relaxation by Lasserre:

(a) Probabilty measure and its moments

(b) Linearization of polynomial SDPs

(c) Sum of squares of polynomials

μ : a probability measure on $\mathbb{R}^n$. We assume $n = 2$. For every $r = 0, 1, 2, \ldots$, define

$$\boldsymbol{u}_r(\boldsymbol{x}) = (1, x_1, x_2, x_1^2, x_1 x_2, x_2^2, x_1^3, \ldots, x_2^r)^T$$: a column vector

(all monomials with degree $\leq r$)

$$\boldsymbol{M}_r(\boldsymbol{y}) = \int_{\mathbb{R}^2} \boldsymbol{u}_r(\boldsymbol{x}) \boldsymbol{u}_r(\boldsymbol{x})^T d\mu$$ (moment matrix, symmetric, positive semidefinite)

$y_{\alpha\beta} = \int_{\mathbb{R}^2} x_1^\alpha x_2^\beta d\mu$ = (α, β)-element depending on μ, $y_{00} = 1$

μ : a probability measure on $\mathbb{R}^n$. We assume $n = 2$. For every $r = 0, 1, 2, \ldots$, define

$$\boldsymbol{u}_r(\boldsymbol{x}) = (1, x_1, x_2, x_1^2, x_1x_2, x_2^2, x_1^3, \ldots, x_2^r)^T$$: a column vector (all monomials with degree $\leq r$)

$$\boldsymbol{M}_r(\boldsymbol{y}) = \int_{\mathbb{R}^2} \boldsymbol{u}_r(\boldsymbol{x})\boldsymbol{u}_r(\boldsymbol{x})^T d\mu$$ (moment matrix, symmetric, positive semidefinite)

$y_{\alpha\beta} = \int_{\mathbb{R}^2} x_1^\alpha x_2^\beta d\mu$ = (α, β)-element depending on μ, $y_{00} = 1$

Example with $r = 2$:

$$y_{21} = \int_{\mathbb{R}^2} x_1^2 x_2 d\mu$$

$$\boldsymbol{M}_r(\boldsymbol{y}) = \begin{pmatrix} y_{00} & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} \\ y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} \\ y_{01} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} \\ y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22} \\ y_{11} & y_{21} & y_{12} & y_{31} & y_{22} & y_{13} \\ y_{02} & y_{12} & y_{03} & y_{22} & y_{13} & y_{04} \end{pmatrix}, \quad y_{00} = 1$$

μ : a probability measure on $\mathbb{R}^n$. We assume $n = 2$. For every $r = 0, 1, 2, \ldots$, define

$$\boldsymbol{u}_r(\boldsymbol{x}) = (1, x_1, x_2, x_1^2, x_1 x_2, x_2^2, x_1^3, \ldots, x_2^r)^T$$: a column vector (all monomials with degree $\leq r$)

$$\boldsymbol{M}_r(\boldsymbol{y}) = \int_{\mathbb{R}^2} \boldsymbol{u}_r(\boldsymbol{x}) \boldsymbol{u}_r(\boldsymbol{x})^T d\mu$$ (moment matrix, symmetric, positive semidefinite)

$y_{\alpha\beta} = \int_{\mathbb{R}^2} x_1^\alpha x_2^\beta d\mu$ = (α, β)-element depending on μ, $y_{00} = 1$

μ : a probability measure on $\mathbb{R}^n$. We assume $n = 2$. For every $r = 0, 1, 2, \ldots$, define

$$\boldsymbol{u}_r(\boldsymbol{x}) = (1, x_1, x_2, x_1^2, x_1x_2, x_2^2, x_1^3, \ldots, x_2^r)^T$$: a column vector (all monomials with degree $\leq r$)

$$\boldsymbol{M}_r(\boldsymbol{y}) = \int_{\mathbb{R}^2} \boldsymbol{u}_r(\boldsymbol{x})\boldsymbol{u}_r(\boldsymbol{x})^T d\mu$$ (moment matrix, symmetric, positive semidefinite)

$y_{\alpha\beta} = \int_{\mathbb{R}^2} x_1^\alpha x_2^\beta d\mu$ = (α, β)-element depending on μ, $y_{00} = 1$

μ : a probability measure on $\mathbb{R}^2$ $\Downarrow$ $y_{00} = 1$, $\boldsymbol{M}_r(\boldsymbol{y}) \succeq \boldsymbol{O}$ (positive semidefinite) $(r = 1, 2, \ldots)$

- We will use this necessary cond. with a finite r for μ to be a probability measure in relaxation of a POP $\Rightarrow$ next slide.

SDP relaxation (Lasserre '01) of a POP — an example

POP: min $f_0(\boldsymbol{x}) = x_1^4 - 2x_1x_2$ opt. val. ζ^* : unknown

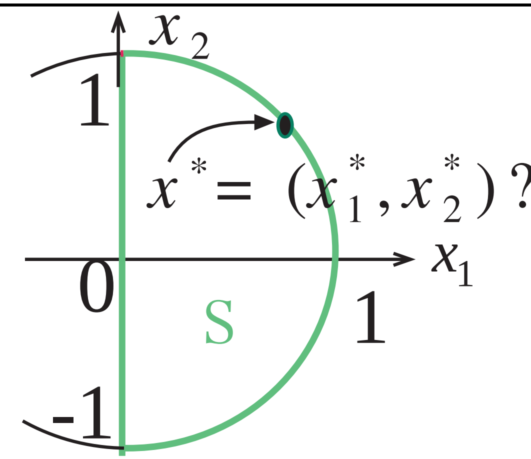
sub. to $\boldsymbol{x} \in S \equiv \left\{ \boldsymbol{x} \in \mathbb{R}^2 : \begin{array}{l} f_1(\boldsymbol{x}) = x_1 \geq 0 \\ f_2(\boldsymbol{x}) = 1 - x_1^2 - x_2^2 \geq 0 \end{array} \right\}.$

$\Updownarrow$

min $\int (x_1^4 x_2^0 - 2x_1x_2) d\mu$

sub. to μ : a prob. meas. on S.

$\Downarrow y_{\alpha\beta} = \int_{\mathbb{R}^2} x_1^\alpha x_2^\beta d\mu$



SDP relaxation (Lasserre '01) of a POP — an example

POP: min $f_0(\boldsymbol{x}) = x_1^4 - 2x_1x_2$ opt. val. ζ^* : unknown

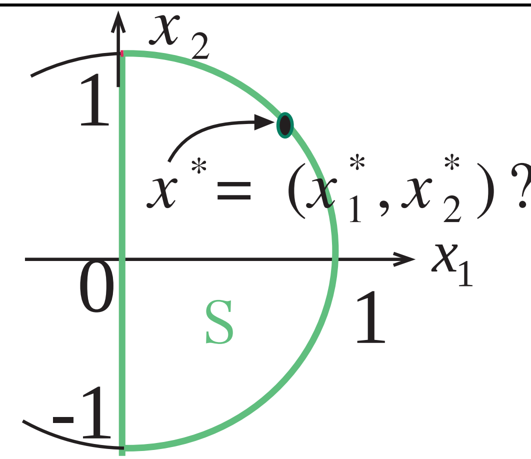
sub. to $\boldsymbol{x} \in S \equiv \left\{ \boldsymbol{x} \in \mathbb{R}^2 : \begin{array}{l} f_1(\boldsymbol{x}) = x_1 \geq 0 \\ f_2(\boldsymbol{x}) = 1 - x_1^2 - x_2^2 \geq 0 \end{array} \right\}$.

$\Updownarrow$

min $\int (x_1^4 x_2^0 - 2x_1x_2) d\mu$

sub. to μ : a prob. meas. on S.

$\Downarrow \; y_{\alpha\beta} = \int_{\mathbb{R}^2} x_1^\alpha x_2^\beta d\mu$



min $y_{40} - 2y_{11}$ $\Rightarrow$ SDP relaxation, opt. val. $\zeta_r \leq \zeta^*$

sub. to "a certain moment cond. with a parameter r

for μ to be a probability measure on S" $\Rightarrow$ next slide

- $\zeta_r \leq \zeta_{r+1} \leq \zeta^*$, and $\zeta_r \to \zeta^*$ as $r \to \infty$ under a moderate assumption that requires S is bounded (Lasserre '01).

SDP relaxation (Lasserre '01) of a POP — an example

$r = 2$

$$\min_{\mu} \; y_{40} - 2y_{11} \text{ s.t. } \int_S \begin{pmatrix} 1 \\ x_1 \\ x_1^2 \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ x_1^2 \end{pmatrix}^T x_1 d\mu \succeq \boldsymbol{O}, \Leftarrow x_1 \geq 0$$

$$1 - x_1^2 - x_2^2 \geq 0 \Rightarrow \int_S \begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix}^T (1 - x_1^2 - x_2^2) d\mu \succeq \boldsymbol{O},$$

$$\text{(moment matrix)} \quad \int_S \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1 x_2 \\ x_2^2 \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1 x_2 \\ x_2^2 \end{pmatrix}^T d\mu \succeq \boldsymbol{O}.$$

Here μ denotes a probability measure on S.

$\Downarrow$

SDP relaxation (Lasserre '01) of a POP — an example

- μ : a probability measure
- $y_{\alpha\beta} = \int_S x_1^\alpha x_2^\beta d\mu$

$$f_1(\boldsymbol{x}) = x_1 \geq 0 \Rightarrow \int_S \begin{pmatrix} 1 \\ x_1 \\ x_1^2 \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ x_1^2 \end{pmatrix}^T x_1 d\mu \succeq \boldsymbol{O}$$

$$\Leftrightarrow \int_S \begin{pmatrix} x_1 & x_1^2 & x_1^3 \\ x_1^2 & x_1^3 & x_1^4 \\ x_1^3 & x_1^4 & x_1^5 \end{pmatrix} d\mu \succeq \boldsymbol{O} \Leftrightarrow \begin{pmatrix} y_{10} & y_{20} & y_{30} \\ y_{20} & y_{30} & y_{40} \\ y_{30} & y_{40} & y_{50} \end{pmatrix} \succeq \boldsymbol{O}$$

$$f_2(\boldsymbol{x}) = 1 - x_1^2 - x_2^2 \geq 0 \Rightarrow \int_S \begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix}^T f_2(\boldsymbol{x}) d\mu \succeq \boldsymbol{O}$$

$$\Rightarrow \begin{pmatrix} 1 - y_{20} - y_{02} & y_{10} - y_{30} - y_{12} & y_{01} - y_{21} - y_{03} \\ y_{10} - y_{30} - y_{12} & y_{20} - y_{40} - y_{22} & y_{11} - y_{31} - y_{13} \\ y_{01} - y_{21} - y_{03} & y_{11} - y_{31} - y_{13} & y_{02} - y_{22} - y_{04} \end{pmatrix} \succeq \boldsymbol{O}$$

SDP relaxation (Lasserre '01) of a POP — an example

- μ : a probability measure
- $y_{\alpha\beta} = \int_S x_1^\alpha x_2^\beta d\mu$

$$\int_S \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1 x_2 \\ x_2^2 \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1 x_2 \\ x_2^2 \end{pmatrix}^T d\mu \succeq \boldsymbol{O}$$

$$\Rightarrow \begin{pmatrix} 1 & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} \\ y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} \\ y_{01} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} \\ y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22} \\ y_{11} & y_{21} & y_{12} & y_{31} & y_{22} & y_{13} \\ y_{02} & y_{12} & y_{03} & y_{22} & y_{13} & y_{04} \end{pmatrix} \succeq \boldsymbol{O}$$

Consequently, we obtain

SDP relaxation (Lasserre '01) of a POP — an example

min $y_{40} - 2y_{11}$ s.t. $\begin{pmatrix} y_{10} & y_{20} & y_{30} \\ y_{20} & y_{30} & y_{40} \\ y_{30} & y_{40} & y_{50} \end{pmatrix} \succeq \boldsymbol{O}$,

$$\begin{pmatrix} 1 - y_{20} - y_{02} & y_{10} - y_{30} - y_{12} & y_{01} - y_{21} - y_{03} \\ y_{10} - y_{30} - y_{12} & y_{20} - y_{40} - y_{22} & y_{11} - y_{31} - y_{13} \\ y_{01} - y_{21} - y_{03} & y_{11} - y_{31} - y_{13} & y_{02} - y_{22} - y_{04} \end{pmatrix} \succeq \boldsymbol{O},$$

(moment matrix) $\begin{pmatrix} 1 & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} \\ y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} \\ y_{01} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} \\ y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22} \\ y_{11} & y_{21} & y_{12} & y_{31} & y_{22} & y_{13} \\ y_{02} & y_{12} & y_{03} & y_{22} & y_{13} & y_{04} \end{pmatrix} \succeq \boldsymbol{O}$.

- We can apply SDP relaxation to general POPs in $\mathbb{R}^n$.
- Poweful in theory but very expensive in computation
  $\Rightarrow$ Exploiting sparsity is crucial in practice.

Contents

Three ways of describing the SDP relaxation by Lasserre:

(a) Probabilty measure and its moments

(b) Linearization of polynomial SDPs
— similar to (a), but easier to understand

(c) Sum of squares of polynomials

SDP relaxation (Lasserre '01) of a POP — an example

POP: min $f_0(\boldsymbol{x}) = x_1^4 - 2x_1x_2$ opt. val. ζ^* : unknown

sub. to $\boldsymbol{x} \in S \equiv \left\{ \boldsymbol{x} \in \mathbb{R}^2 : \begin{array}{l} f_2(\boldsymbol{x}) = x_1 \geq 0 \\ f_1(\boldsymbol{x}) = 1 - x_1^2 - x_2^2 \geq 0 \end{array} \right\}$.

SDP relaxation (Lasserre '01) of a POP — an example

POP: min $f_0(\boldsymbol{x}) = x_1^4 - 2x_1x_2$ opt. val. ζ^* : unknown

sub. to $\boldsymbol{x} \in S \equiv \left\{ \boldsymbol{x} \in \mathbb{R}^2 : \begin{array}{l} f_2(\boldsymbol{x}) = x_1 \geq 0 \\ f_1(\boldsymbol{x}) = 1 - x_1^2 - x_2^2 \geq 0 \end{array} \right\}$.

$\Updownarrow$

Polynomial SDP

min $x_1^4 - 2x_1x_2$

s.t. $\begin{pmatrix} 1 \\ x_1 \\ x_1^2 \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ x_1^2 \end{pmatrix}^T f_1(\boldsymbol{x}) \succeq \boldsymbol{O}$,

$\begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix}^T f_2(\boldsymbol{x}) \succeq \boldsymbol{O}$,

$$\begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1x_2 \\ x_2^2 \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1x_2 \\ x_2^2 \end{pmatrix}^T \succeq \boldsymbol{O}.$$

- Expand poly.mat.inequalities.
- Linearize the problem by replacing $x_1^\alpha x_2^\beta$ by $y_{\alpha\beta} \Rightarrow$

SDP relaxation (Lasserre '01) of a POP — an example

min $x_1^4 - 2x_1x_2 \Rightarrow$ min $y_{40} - 2y_{11}$

$$f_1(\boldsymbol{x}) = x_1 \geq 0 \Leftrightarrow \begin{pmatrix} 1 \\ x_1 \\ x_1^2 \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ x_1^2 \end{pmatrix}^T x_1 \succeq \boldsymbol{O}$$

$$\Leftrightarrow \begin{pmatrix} x_1 & x_1^2 & x_1^3 \\ x_1^2 & x_1^3 & x_1^4 \\ x_1^3 & x_1^4 & x_1^5 \end{pmatrix} \succeq \boldsymbol{O} \Rightarrow \begin{pmatrix} y_{10} & y_{20} & y_{30} \\ y_{20} & y_{30} & y_{40} \\ y_{30} & y_{40} & y_{50} \end{pmatrix} \succeq \boldsymbol{O}$$

SDP relaxation (Lasserre '01) of a POP — an example

$$f_2(\boldsymbol{x}) = 1 - x_1^2 - x_2^2 \geq 0 \Leftrightarrow \begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix}^T f_2(\boldsymbol{x}) \succeq \boldsymbol{O}$$

$$\Updownarrow$$

$$\begin{pmatrix} 1 - x_1^2x_2^0 - x_1^0x_2^2 & x_1^1x_2^0 - x_1^3x_2^0 - x_1^1x_2^2 & x_1^0x_2^1 - x_1^2x_2^1 - x_1^0x_2^3 \\ x_1^1x_2^0 - x_1^3x_2^0 - x_1^1x_2^2 & x_1^2x_2^0 - x_1^4x_2^0 - x_1^2x_2^2 & x_1^1x_2^1 - x_1^3x_2^1 - x_1^1x_2^3 \\ x_1^0x_2^1 - x_1^2x_2^1 - x_1^0x_2^3 & x_1^1x_2^1 - x_1^3x_2^1 - x_1^1x_2^3 & x_1^0x_2^2 - x_1^2x_2^2 - x_1^0x_2^4 \end{pmatrix} \succeq \boldsymbol{O}$$

$$\Downarrow$$

$$\begin{pmatrix} 1 - y_{20} - y_{02} & y_{10} - y_{30} - y_{12} & y_{01} - y_{21} - y_{03} \\ y_{10} - y_{30} - y_{12} & y_{20} - y_{40} - y_{22} & y_{11} - y_{31} - y_{13} \\ y_{01} - y_{21} - y_{03} & y_{11} - y_{31} - y_{13} & y_{02} - y_{22} - y_{04} \end{pmatrix} \succeq \boldsymbol{O}$$

SDP relaxation (Lasserre '01) of a POP — an example

Similarly,

$$\begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1 x_2 \\ x_2^2 \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1 x_2 \\ x_2^2 \end{pmatrix}^T \succeq \boldsymbol{O}$$

$$\Rightarrow \begin{pmatrix} 1 & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} \\ y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} \\ y_{01} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} \\ y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22} \\ y_{11} & y_{21} & y_{12} & y_{31} & y_{22} & y_{13} \\ y_{02} & y_{12} & y_{03} & y_{22} & y_{13} & y_{04} \end{pmatrix} \succeq \boldsymbol{O}$$

Consequently, we obtain

SDP relaxation (Lasserre '01) of a POP — an example

$$\min\ y_{40} - 2y_{11} \text{ s.t. } \begin{pmatrix} y_{10} & y_{20} & y_{30} \\ y_{20} & y_{30} & y_{40} \\ y_{30} & y_{40} & y_{50} \end{pmatrix} \succeq \boldsymbol{O},$$

$$\begin{pmatrix} 1 - y_{20} - y_{02} & y_{10} - y_{30} - y_{12} & y_{01} - y_{21} - y_{03} \\ y_{10} - y_{30} - y_{12} & y_{20} - y_{40} - y_{22} & y_{11} - y_{31} - y_{13} \\ y_{01} - y_{21} - y_{03} & y_{11} - y_{31} - y_{13} & y_{02} - y_{22} - y_{04} \end{pmatrix} \succeq \boldsymbol{O},$$

$$\begin{pmatrix} 1 & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} \\ y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} \\ y_{01} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} \\ y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22} \\ y_{11} & y_{21} & y_{12} & y_{31} & y_{22} & y_{13} \\ y_{02} & y_{12} & y_{03} & y_{22} & y_{13} & y_{04} \end{pmatrix} \succeq \boldsymbol{O}.$$

- If $\boldsymbol{x}$ is a feasible sol. of Poly. SDP (or POP), then $\boldsymbol{y} = (y_{\alpha\beta} = x_1^{\alpha} x_2^{\beta})$ is a feasible sol. of SDP with the same obj. val.; $x_1^4 - 2x_1x_2 = y_{40} - 2y_{11} \Longrightarrow$ Relaxation

Contents

1. Global optimization of nonconvex problems
2. Polynomial Optimization Problems (POPs)
3. Basic idea of SOS and SDP relaxations

Three ways of describing the SDP relaxation by Lasserre:

(a) Probabilty measure and its moments

(b) Linearization of polynomial SDPs

(c) Sum of squares of polynomials
— dual approach
— easy to understand in unconstrained case

$f(\boldsymbol{x})$: an SOS (Sum of Squares) polynomial

$$\Updownarrow$$

$\exists$ polynomials $g_1(\boldsymbol{x}), \ldots, g_q(\boldsymbol{x})$; $f(\boldsymbol{x}) = \sum_{j=1}^{q} g_j(\boldsymbol{x})^2$.

$\mathcal{N}$: the set of nonnegative polynomials in $\boldsymbol{x} \in \mathbb{R}^n$.

SOS$_*$: the set of SOS. Obviously, **SOS**$_* \subset \mathcal{N}$.

SOS$_r = \{f \in$ **SOS**$_*$: deg $f \leq 2r\}$: SOSs w. degree at most $2r$.

$n = 2$. $f(x_1, x_2) = (x_1^2 - 2x_2 + 1)^2 + (3x_1x_2 + x_2 - 4)^2 \in$ **SOS**$_2$.

$n = 2$. $f(x_1, x_2) = (x_1x_2 - 1)^2 + x_1^2 \in$ **SOS**$_2$.

- In theory, **SOS**$_*$ (SOS) $\subset \mathcal{N}$. **SOS**$_* \neq \mathcal{N}$ in general.
- If $n = 1$, **SOS**$_* = \mathcal{N}$. $\{f \in \mathcal{N}$: deg $f \leq 2\} \equiv$ **SOS**$_1$.
- In practice, $f(\boldsymbol{x}) \in \mathcal{N} \backslash$**SOS**$_*$ is rare.
- So we replace $\mathcal{N}$ by **SOS**$_* \Longrightarrow$ SOS Relaxations.

Representation of

$$\textbf{SOS}_r \equiv \left\{ \sum_{j=1}^{q} g_j(\boldsymbol{x})^2 : q \geq 1,\ g_j(\boldsymbol{x}) \text{is a poly. of deg } \leq r \right\} \subset \textbf{SOS}_*.$$

$\forall$ poly. $g(\boldsymbol{x})$ of deg $\leq r$, $\exists \boldsymbol{a} \in \mathbb{R}^{d(r)}$; $g(\boldsymbol{x}) = \boldsymbol{a}^T \boldsymbol{u}_r(\boldsymbol{x})$, where

$$\boldsymbol{u}_r(\boldsymbol{x}) = (1, x_1, \ldots, x_n, x_1^2, x_1 x_2, x_1 x_3, \ldots, x_n^2, \ldots, x_1^r, \ldots, x_n^r)^T$$

(a column vector of a basis for polynomials of degree $\leq r$),

$$d(r) = \begin{pmatrix} n+r \\ r \end{pmatrix}$$: the dimension of $\boldsymbol{u}_r(\boldsymbol{x})$.

$$\Downarrow$$

$$\begin{aligned}
\textbf{SOS}_r &= \left\{ \textstyle\sum_{j=1}^{q} \left(\boldsymbol{a}_j^T \boldsymbol{u}_r(\boldsymbol{x}) \right)^2 \ : \ q \geq 1,\ \boldsymbol{a}_j \in \mathbb{R}^{d(r)} \right\} \\
&= \left\{ \boldsymbol{u}_r(\boldsymbol{x})^T \left(\textstyle\sum_{j=1}^{q} \boldsymbol{a}_j \boldsymbol{a}_j^T \right) \boldsymbol{u}_r(\boldsymbol{x}) \ : \ q \geq 1,\ \boldsymbol{a}_j \in \mathbb{R}^{d(r)} \right\} \\
&= \left\{ \boldsymbol{u}_r(\boldsymbol{x})^T \boldsymbol{V} \boldsymbol{u}_r(\boldsymbol{x}) \ : \ \boldsymbol{V} \succeq \boldsymbol{O} \right\}.
\end{aligned}$$

Example. $n = 1$, SOS polynomials of degree ≤ 3 in $x \in \mathbb{R}$.

$$\textbf{SOS}_3 \equiv \left\{ \sum_{j=1}^{q} g_j(\boldsymbol{x})^2 \;:\; q \geq 1,\; g_j(\boldsymbol{x}) \text{ is a poly. of degree } \leq 3 \right\}$$

$$= \left\{ \begin{pmatrix} 1 \\ x \\ x^2 \\ x^3 \end{pmatrix}^T \boldsymbol{V} \begin{pmatrix} 1 \\ x \\ x^2 \\ x^3 \end{pmatrix} \;:\; \boldsymbol{V} \text{ is } 4 \times 4 \text{ psd matrix} \right\}$$

Example. $n = 2$, SOS polynomials of degree ≤ 2 in x=(x_1, x_2).

$$\textbf{SOS}_2 \equiv \left\{ \sum_{j=1}^{q} g_j(\boldsymbol{x})^2 \; : \; q \geq 1, \; g_j(\boldsymbol{x}) \text{ is a poly. of degree } \leq 2 \right\}$$

$$= \left\{ \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1 x_2 \\ x_2^2 \end{pmatrix}^T \boldsymbol{V} \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1 x_2 \\ x_2^2 \end{pmatrix} \; : \; \boldsymbol{V} \text{ is a } 6 \times 6 \text{ psd matrix} \right\}$$

$f(\boldsymbol{x}) = -x_1 + 2x_2 + 3x_1^2 - 5x_1^2x_2^2 + 7x_2^4$ $\zeta = 3.1$: fixed

$f(\boldsymbol{x}) - \zeta \geq 0 \ (\forall \boldsymbol{x}) \Longrightarrow$ SOS relaxation

$f(\boldsymbol{x}) - \zeta \in$ **SOS**$_2$ (SOS of poly. of degree ≤ 2)

$\Updownarrow$

$\exists \boldsymbol{V} \in \mathbb{S}^6; \qquad f(\boldsymbol{x}) - \zeta =$

Sum of Squares

$$\begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1x_2 \\ x_2^2 \end{pmatrix}^T \begin{pmatrix} V_{11} & V_{12} & V_{13} & V_{14} & V_{15} & V_{16} \\ V_{12} & V_{22} & V_{23} & V_{24} & V_{25} & V_{26} \\ V_{13} & V_{23} & V_{33} & V_{34} & V_{35} & V_{36} \\ V_{14} & V_{24} & V_{34} & V_{44} & V_{45} & V_{46} \\ V_{15} & V_{25} & V_{35} & V_{45} & V_{55} & V_{56} \\ V_{16} & V_{26} & V_{36} & V_{46} & V_{56} & V_{66} \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1x_2 \\ x_2^2 \end{pmatrix}$$

$(\forall (x_1, x_2)^T \in \mathbb{R}^n), \qquad \boldsymbol{V} \succeq \boldsymbol{O}$

$\Updownarrow$ Compare the coef. of $1, \ x_1, \ x_2, \ x_1^2, \ x_1x_2, \boldsymbol{x}_2^2$ on both sides of $=$

LMI: $\exists \boldsymbol{V} \in \mathbb{S}^6$?;

$$-\zeta = V_{11},\ -1 = 2V_{12},\ 2 = 2V_{13},\ 3 = 2V_{14} + V_{22},$$

$$-5 = 2V_{46} + V_{55},\ 7 = V_{66},\ \text{all others}\ \ 0 = \cdots\ ,$$

$$\boldsymbol{V} \succeq \boldsymbol{O}$$

In general, each equality constraint is a linear equation in ζ and $\boldsymbol{V}$.

Unconstrained minimization

$\mathcal{P}$: min $f(\boldsymbol{x}) = -x_1 + 2x_2 + 3x_1^2 - 5x_1^2x_2^2 + 7x_2^4$

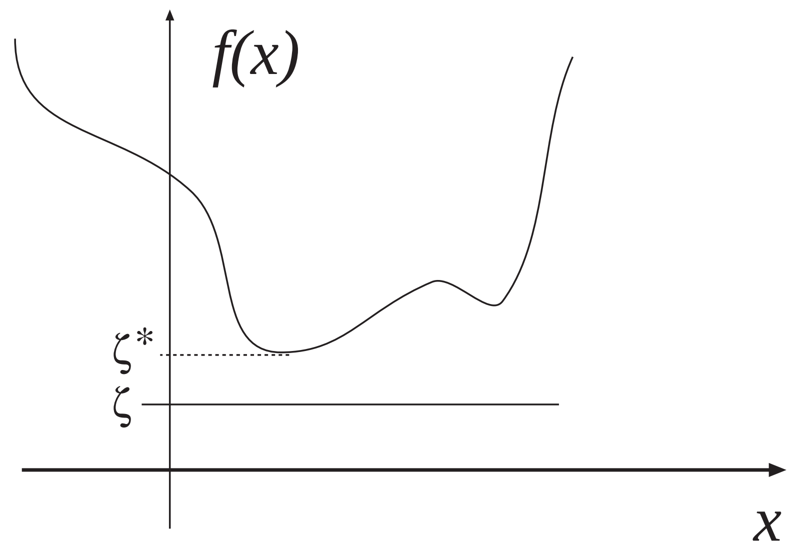
$\Updownarrow$ (a special case of Lagragian dual)

$\mathcal{P}$': $\max\ \zeta$ s.t $\quad f(\boldsymbol{x}) - \zeta \geq 0\ (\forall \boldsymbol{x} \in \mathbb{R}^n)$

$\Updownarrow$

$f(\boldsymbol{x}) - \zeta \in \mathcal{N}$ (the nonnegative polynomials)

Here $\boldsymbol{x}$ is a parameter (index) describing inequality constraints.



Unconstrained minimization

$\mathcal{P}$: min $f(\boldsymbol{x}) = -x_1 + 2x_2 + 3x_1^2 - 5x_1^2x_2^2 + 7x_2^4$

$\Updownarrow$ (a special case of Lagragian dual)

$\mathcal{P}$': $\max\ \zeta$ s.t $f(\boldsymbol{x}) - \zeta \geq 0 \ (\forall \boldsymbol{x} \in \mathbb{R}^n)$

$$\Updownarrow$$

$f(\boldsymbol{x}) - \zeta \in \mathcal{N}$ (the nonnegative polynomials)

Here $\boldsymbol{x}$ is a parameter (index) describing inequality constraints.

Unconstrained minimization

$\mathcal{P}$: min $f(\boldsymbol{x}) = -x_1 + 2x_2 + 3x_1^2 - 5x_1^2x_2^2 + 7x_2^4$

$\Updownarrow$ (a special case of Lagragian dual)

$\mathcal{P}$': $\max\ \zeta$ s.t $f(\boldsymbol{x}) - \zeta \geq 0\ (\forall \boldsymbol{x} \in \mathbb{R}^n)$

$\Updownarrow$

$f(\boldsymbol{x}) - \zeta \in \mathcal{N}$ (the nonnegative polynomials)

Here $\boldsymbol{x}$ is a parameter (index) describing inequality constraints.

$\Downarrow$ a subproblem of $\mathcal{P}$' (or a relaxation of $\mathcal{P}$)

$\mathcal{P}$": $\max\ \zeta$ sub.to $f(\boldsymbol{x}) - \zeta \in$ **SOS**$_2$ (SOS of poly. of degree ≤ 2)

$\Downarrow$

$\max \zeta$ s.t. $f(\boldsymbol{x}) - \zeta =$

Sum of Squares

$$\begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1 x_2 \\ x_2^2 \end{pmatrix}^T \begin{pmatrix} V_{11} & V_{12} & V_{13} & V_{14} & V_{15} & V_{16} \\ V_{12} & V_{22} & V_{23} & V_{24} & V_{25} & V_{26} \\ V_{13} & V_{23} & V_{33} & V_{34} & V_{35} & V_{36} \\ V_{14} & V_{24} & V_{34} & V_{44} & V_{45} & V_{46} \\ V_{15} & V_{25} & V_{35} & V_{45} & V_{55} & V_{56} \\ V_{16} & V_{26} & V_{36} & V_{46} & V_{56} & V_{66} \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1 x_2 \\ x_2^2 \end{pmatrix}$$

$(\forall (x_1, x_2)^T \in \mathbb{R}^n)$, $\quad \boldsymbol{V} \succeq \boldsymbol{O}$

$\Updownarrow$ Compare the coef. of $1,\ x_1,\ x_2,\ x_1^2,\ x_1 x_2, \boldsymbol{x}_2^2$ on both side of $=$

SDP (Semidefinite Program)

$$\max \quad \zeta \text{ s.t.} \quad -\zeta = V_{11},\ -1 = 2V_{12},\ 2 = 2V_{13},\ 3 = 2V_{14} + V_{22},$$
$$-5 = 2V_{46} + V_{55},\ 7 = V_{66},\ \text{all others } 0 = \cdots\ ,$$
$$\boldsymbol{V} \succeq \boldsymbol{O}$$

In general, each equality constraint is a linear equation in ζ and $\boldsymbol{V}$.

SDP (Semidefinite Program)

$$\max \quad \zeta \text{ s.t.} \quad -\zeta = V_{11},\ -1 = 2V_{12},\ 2 = 2V_{13},\ 3 = 2V_{14} + V_{22},$$
$$-5 = 2V_{46} + V_{55},\ 7 = V_{66},\ \text{all others } 0 = \cdots\ ,$$
$$\boldsymbol{V} \succeq \boldsymbol{O}$$

In general, each equality constraint is a linear equation in ζ and $\boldsymbol{V}$.

- SOS relaxation of general constrained POPs

  $\Longrightarrow$ Next section

- We will use a generalized Lagrangian dual.

Contents

POP: $\zeta^* = $ min. $f_0(\boldsymbol{x})$ s.t. $f_k(\boldsymbol{x}) \geq 0$ $(k = 1, \ldots, m)$

$$\Downarrow$$

A sequence of SOS relaxation problems SOS^r depending on the relaxation order $r = 1, 2, \ldots$, which determines the degree of polynomials used;

(a) Under a moderate assumption,
opt. sol. of SOS^r $\rightarrow$ opt sol. of POP as $r \rightarrow \infty$
(Lasserre 2001).

(b) $r = \lceil$"the max. deg. of poly. in POP"$/2\rceil + 0 \sim 3$ is usually large enough to attain opt sol. of POP in practice.

(c) Such an r is unknown in theory except $\exists$ special cases.

(d) SOS^r can be converted into an SDP^r and solved by the primal-dual interior-point method.

(e) The size of SDP^r increases as $r \rightarrow \infty$.

POP: $\zeta^* = $ min. $f_0(\boldsymbol{x})$ s.t. $f_k(\boldsymbol{x}) \geq 0 \ (k = 1, \ldots, m)$

POP: $\zeta^* = $ min. $f_0(\boldsymbol{x})$ s.t. $f_k(\boldsymbol{x}) \geq 0 \ (k = 1, \ldots, m)$

- Rough sketch of SOS relaxation of POP

“Generalized Lagrangian Dual”

$$+$$

“SOS relaxation of unconstrained POPs”

$$\Downarrow$$

SOS relaxation of POP

POP: $\zeta^* = \text{min. } f_0(\boldsymbol{x}) \text{ s.t. } f_k(\boldsymbol{x}) \geq 0 \ (k = 1, \ldots, m)$

POP: $\zeta^* = $ min. $f_0(\boldsymbol{x})$ s.t. $f_k(\boldsymbol{x}) \geq 0 \ (k = 1, \ldots, m)$

$\Downarrow$

Generalized Lagrangian function

$$L(\boldsymbol{x}, \lambda_1, \ldots, \lambda_m) = f_0(\boldsymbol{x}) - \sum_{k=1}^{m} \lambda_k(\boldsymbol{x}) f_k(\boldsymbol{x})$$

for $\forall \boldsymbol{x} \in \mathbb{R}^n, \ \forall \lambda_k \in \textbf{SOS}_*$

If $\mathbb{R} \ni \lambda_k \geq 0$ then L is the standard Lagrangian function.

Generalized Lagrangian Dual

$$\max_{\lambda_1 \in \textbf{SOS}_*, \ldots, \lambda_m \in \textbf{SOS}_*} \min_{\boldsymbol{x} \in \mathbb{R}^n} L(\boldsymbol{x}, \lambda_1, \ldots, \lambda_m)$$

$\Updownarrow$ $\Leftrightarrow \max \ \eta$ s.t. $L(\boldsymbol{x}, \lambda_1, \ldots, \lambda_m) - \eta \geq 0 \ (\forall \boldsymbol{x} \in \mathbb{R}^n)$

Generalized Lagrangian Dual

$$\max \quad \eta \text{ s.t. } \quad L(\boldsymbol{x}, \lambda_1, \ldots, \lambda_m) - \eta \geq 0 \ (\forall \boldsymbol{x} \in \mathbb{R}^n),$$
$$\lambda_1 \in \textbf{SOS}_*, \ldots, \lambda_m \in \textbf{SOS}_*$$

$\boldsymbol{x}$ is not a variable but the index for infinite no. of inequalities.

POP: $\zeta^* = $ min. $f_0(\boldsymbol{x})$ s.t. $f_k(\boldsymbol{x}) \geq 0 \ (k = 1, \ldots, m)$

POP: $\zeta^* = $ min. $f_0(\boldsymbol{x})$ s.t. $f_k(\boldsymbol{x}) \geq 0 \ (k = 1, \ldots, m)$

$\Downarrow$

Generalized Lagrangian function

$$L(\boldsymbol{x}, \lambda_1, \ldots, \lambda_m) = f_0(\boldsymbol{x}) - \sum_{k=1}^{m} \lambda_k(\boldsymbol{x}) f_k(\boldsymbol{x})$$

for $\forall \boldsymbol{x} \in \mathbb{R}^n, \ \forall \lambda_k \in$ **SOS**$_*$

Generalized Lagrangian Dual

$$\max \quad \eta \text{ s.t. } \quad L(\boldsymbol{x}, \lambda_1, \ldots, \lambda_m) - \eta \geq 0 \ (\forall \boldsymbol{x} \in \mathbb{R}^n),$$

$$\lambda_1 \in \textbf{SOS}_*, \ldots, \lambda_m \in \textbf{SOS}_*$$

$\Downarrow$ SOS relaxation

SOS^r: $\eta^r = \max \ \eta$ s.t. $L(\boldsymbol{x}, \lambda_1, \ldots, \lambda_m) - \eta \in \textbf{SOS}_r \ (\forall \boldsymbol{x} \in \mathbb{R}^n),$

$$\lambda_1 \in \textbf{SOS}_{r_1}, \ldots, \lambda_m \in \textbf{SOS}_{r_m}$$

- **SOS**$_{r_k}$: the set of sum of square polynomials with deg. $\leq r_k$
- $\lceil \max\{\text{deg}(f_k) : k = 0, \ldots, m\}/2 \rceil \leq r$: the relaxation order
- $r_k = r - \lceil \text{deg}(f_k)/2 \rceil \ (k = 1, \ldots, m)$; chosen to balance the degrees of all the terms $\lambda_k(\boldsymbol{x}) f_k(\boldsymbol{x}) \ (k = 1, \ldots, m)$.
- $\eta^r \leq \eta^{r+1}$, $\eta^r \to \zeta^*$ as $r \to \infty$ under a moderate assumption which requires the feasible region of POP is compact.

Contents

POP: $\zeta^* = $ min. $f_0(\boldsymbol{x})$ s.t. $f_k(\boldsymbol{x}) \geq 0 \ (k = 1, \ldots, m)$

- the relax. order $r \geq \lceil \max\{\text{deg}(f_k) : k = 0, \ldots, m\}/2 \rceil$
- $r_k = r - \lceil \text{deg}(f_k)/2 \rceil \ (k = 1, \ldots, m)$

$\Updownarrow$

$$\boldsymbol{u}_r(\boldsymbol{x}) = (1, x_1, x_2, x_1^2, x_1 x_2, x_2^2, x_1^3, \ldots, x_2^r)^T$$

(all monomials with degree $\leq r$)

Poly.SDP: min $f_0(\boldsymbol{x})$ s.t. $\boldsymbol{u}_{r_k}(\boldsymbol{x})\boldsymbol{u}_{r_k}(\boldsymbol{x})^T f_k(\boldsymbol{x}) \succeq \boldsymbol{O} \ (\forall k)$

$\boldsymbol{u}_r(\boldsymbol{x})\boldsymbol{u}_r(\boldsymbol{x})^T \succeq \boldsymbol{O}$.

Linearize:

$\Downarrow$

- Expand all the polynomial inequalities
- Replace $x_1^\alpha x_2^\beta \cdots x_n^\gamma$ by a single variable $y_{\alpha\beta\ldots\gamma}$

Linear SDP^r: $\zeta^r = $ min. a linear function in $y_{\alpha\beta\ldots\gamma}$

s.t. linear matrix inequalities in $y_{\alpha\beta\ldots\gamma}$

- $\zeta^r \leq \zeta^{r+1}$, $\zeta^r \to \zeta^*$ as $r \to \infty$ under a moderate assumption which requires the feasible region of POP is compact.

Example

POP: min. $f_0(\boldsymbol{x}) = x_1^3 - 2x_2^2$ s.t. $f_1(\boldsymbol{x}) = 1 - x_1^2 - x_2^2 \geq 0$.

$\Updownarrow$

- $r = 2 \geq \lceil \max\{\text{deg}(f_k) : k = 0, 1\}/2 \rceil = \lceil 3/2 \rceil = 2$.
- $r_1 = r - \lceil \text{deg}(f_1)/2 \rceil = 2 - \lceil 2/2 \rceil = 1$.
- $\boldsymbol{u}_1(\boldsymbol{x}) = (1, x_1, x_2)^T$, $\boldsymbol{u}_2(\boldsymbol{x}) = 1, x_1, x_2, x_1^2, x_1 x_2, x_2^2)^T$.

Poly.SDP: min $x_1^3 - 2x_2^2$

s.t. $\boldsymbol{u}_1(\boldsymbol{x})\boldsymbol{u}_1(\boldsymbol{x})^T f_1(\boldsymbol{x}) \succeq \boldsymbol{O}$

$\boldsymbol{u}_2(\boldsymbol{x})\boldsymbol{u}_2(\boldsymbol{x})^T \succeq \boldsymbol{O}$.

$\Downarrow$ Expand matrix inequalities

Poly.SDP: min $x_1^3 - 2x_2^2$ s.t.

$$\begin{pmatrix} 1 - x_1^2 - x_2^2 & x_1 - x_1^3 - x_1x_2^2 & x_2 - x_1^2x_2 - x_2^3 \\ x_1 - x_1^3 - x_1x_2^2 & x_1^2 - x_1^4 - x_1^2x_2^2 & x_1x_2 - x_1^3x_2 - x_1x_2^3 \\ x_2 - x_1^2x_2 - x_2^3 & x_1x_2 - x_1^3x_2 - x_1x_2^3 & x_2^2 - x_1^2x_2^2 - x_2^4 \end{pmatrix} \succeq \boldsymbol{O},$$

$$\begin{pmatrix} 1 & x_1 & x_2 & x_1^2 & x_1x_2 & x_2^2 \\ x_1 & x_1^2 & x_1x_2 & x_1^3 & x_1^2x_2 & x_1x_2^2 \\ x_2 & x_1x_2 & x_2^2 & x_1^2x_2 & x_1x_2^2 & x_2^3 \\ x_1^2 & x_1^3 & x_1^2x_2 & x_1^4 & x_1^3x_2 & x_1^2x_2^2 \\ x_1x_2 & x_1^2x_2 & x_1x_2^2 & x_1^3x_2 & x_1^2x_2^2 & x_1x_2^3 \\ x_2^2 & x_1x_2^2 & x_2^3 & x_1^2x_2^2 & x_1x_2^3 & x_2^4 \end{pmatrix} \succeq \boldsymbol{O}.$$

$\Downarrow$ Linearize by replacing $x_1^\alpha x_2^\beta$ by a single variable $y_{\alpha\beta}$

SDP^2: min $y_{30} - 2y_{02}$ s.t.

$$\begin{pmatrix} 1 - y_{20} - y_{02} & y_{10} - y_{30} - y_{12} & y_{01} - y_{21} - y_{03} \\ y_{10} - y_{30} - y_{12} & y_{20} - y_{40} - y_{22} & y_{11} - y_{31} - y_{13} \\ y_{01} - y_{21} - y_{03} & y_{11} - y_{31} - y_{13} & y_{02} - y_{22} - y_{04} \end{pmatrix} \succeq \boldsymbol{O},$$

$$\begin{pmatrix} 1 & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} \\ y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} \\ y_{01} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} \\ y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22} \\ y_{11} & y_{21} & y_{12} & y_{31} & y_{22} & y_{13} \\ y_{02} & y_{12} & y_{03} & y_{22} & y_{13} & y_{04} \end{pmatrix} \succeq \boldsymbol{O}.$$

Duality

POP: $\zeta^* = \text{min. } f_0(\boldsymbol{x}) \text{ s.t. } f_k(\boldsymbol{x}) \geq 0 \ (k = 1, \ldots, m)$

$\Downarrow$

- the relax. order $r \geq \lceil \max\{\text{deg}(f_k) : k = 0, \ldots, m\}/2 \rceil$
- $r_k = r - \lceil \text{deg}(f_k)/2 \rceil \ (k = 1, \ldots, m)$

SOS^r: $\eta^r = \max \ \eta \text{ s.t. } L(\boldsymbol{x}, \lambda_1, \ldots, \lambda_m) - \eta \in \textbf{SOS}_r \ (\forall \boldsymbol{x} \in \mathbb{R}^n),$
$\lambda_1 \in \textbf{SOS}_{r_1}, \ldots, \lambda_m \in \textbf{SOS}_{r_m}$

$\Downarrow$ Convert

SDP$_d^r$: $\eta^r =$ max. a linear objective function
s.t. linear matrix inequalities.

$\Updownarrow$ primal and dual to each other

SDP^r: $\zeta^r =$ min. a linear function in $y_{\alpha\beta\ldots\gamma}$
s.t. linear matrix inequalities in $y_{\alpha\beta\ldots\gamma}$

- $\eta^r \leq \zeta^r$.
- $\eta^r = \zeta^r$ if both SDP$_d^r$ and SDP^r have int. feasible solutions.

Contents

A sparse POP example, alkyl.gms from globalib

$\min f_0 = -6.3x_5x_8 + 5.04x_2 + 0.35x_3 + x_4 + 3.36x_6$ sub.to

$f_1 = -0.820x_2 + x_5 - 0.820x_6 = 0$ $f_2 = -x_2x_9 + 10x_3 + x_6 = 0$

$f_3 = 0.98x_4 - x_7(0.01x_5x_{10} + x_4) = 0,$ lbd$_i \leq x_i \leq$ ubd$_i$

$f_4 = x_5x_{12} - x_2(1.12 + 0.132x_9 - 0.0067x_9^2) = 0$

$f_5 = x_8x_{13} - 0.01x_9(1.098 - 0.038x_9) - 0.325x_7 - 0.574 = 0$

$f_6 = x_{10}x_{14} + 22.2x_{11} - 35.82 = 0$ $f_7 = x_1x_{11} - 3x_8 - 1.33 = 0$

- $n = 14$ variables. polynomials with deg ≤ 3.
- $\forall f_k$ involves at most 6 variables.

A sparse POP example, alkyl.gms from globalib

$$\min f_0 = -6.3x_5x_8 + 5.04x_2 + 0.35x_3 + x_4 + 3.36x_6 \text{ sub.to}$$

$$f_1 = -0.820x_2 + x_5 - 0.820x_6 = 0 \quad f_2 = -x_2x_9 + 10x_3 + x_6 = 0$$

$$f_3 = 0.98x_4 - x_7(0.01x_5x_{10} + x_4) = 0, \quad \text{lbd}_i \le x_i \le \text{ubd}_i$$

$$f_4 = x_5x_{12} - x_2(1.12 + 0.132x_9 - 0.0067x_9^2) = 0$$

$$f_5 = x_8x_{13} - 0.01x_9(1.098 - 0.038x_9) - 0.325x_7 - 0.574 = 0$$

$$f_6 = x_{10}x_{14} + 22.2x_{11} - 35.82 = 0 \quad f_7 = x_1x_{11} - 3x_8 - 1.33 = 0$$

- $n = 14$ variables. polynomials with deg ≤ 3.
- $\forall f_k$ involves at most 6 variables.

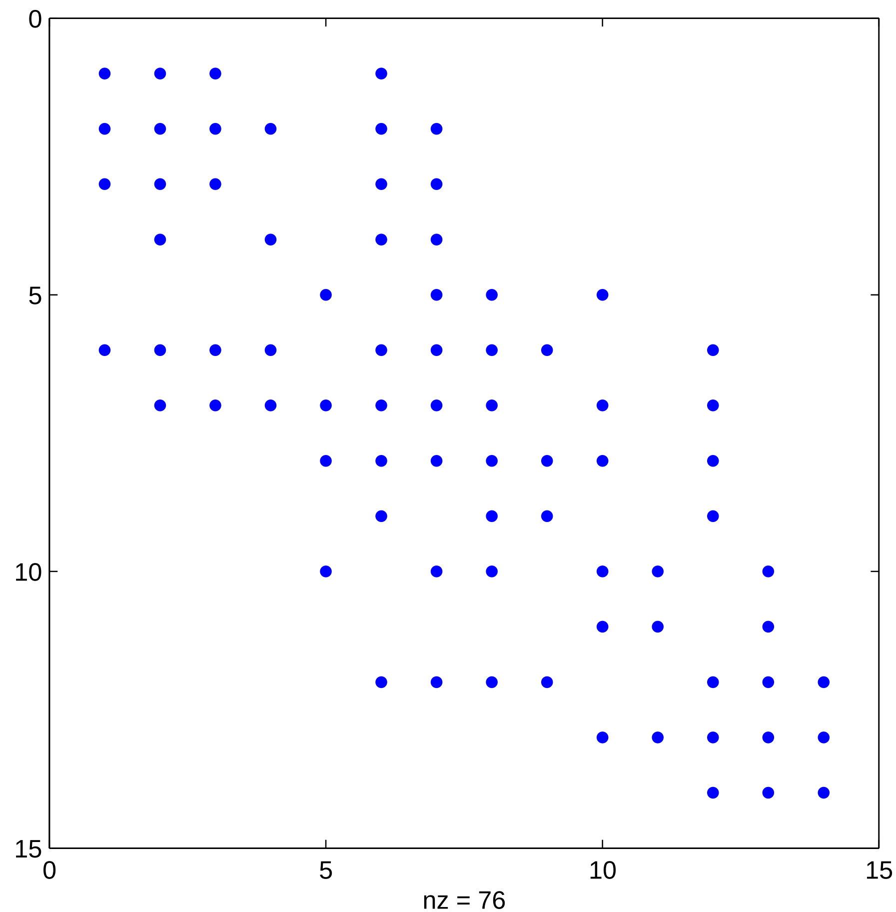
Let $F_k = \{i : x_i$ is involved $f_k\}$; $F_1 = \{2, 5, 6\}$, $F_2 = \{2, 3, 6, 9\}$.
Sparsity pattern matrix $\boldsymbol{R} \in \mathbb{S}^n$ (symbolic) & graph $G(N, E)$.

$$R_{ij} = \begin{cases} \star & \text{if } x_i, x_j \in \text{a monomial of } f_0 \text{ or } \{i, j\} \subset F_k \ \exists k \ge 1, \\ 0 & \text{otherwise .} \end{cases}$$

$$N = \{1, 2, \ldots, n\}, \ E = \{(i, j) \in N \times N : i \ne j \text{ and } R_{ij} = \star\}.$$

eg., $R_{25} = R_{52} = R_{58} = R_{85} = \star$, $R_{28} = R_{82} = 0$

The 14×14 sparsity pattern matrix $\boldsymbol{R}$ with simultaneous row and column reordering (Reverse Cuthill-McKee ordering)



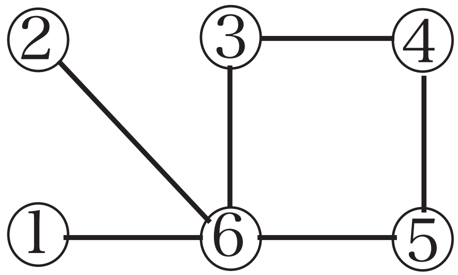
Structured sparsity

- Sparse (symbolic) Cholesky factorization
- Also, characterized by a sparse chordal graph structure

$G(N, E)$: a graph, $N = \{1, \ldots, n\}$ (nodes), $E \subset N \times N$ (edges)

chordal $\Leftrightarrow$ $\forall$ cycle with more than 3 edges has a chord

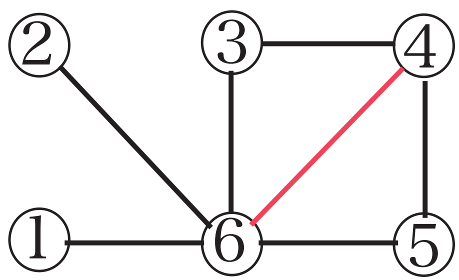
$G(N, E)$



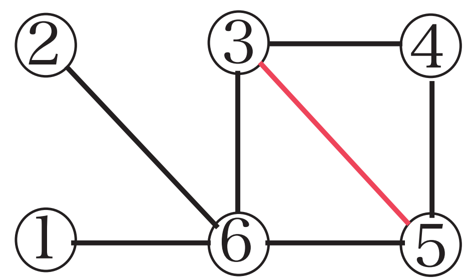
Not chordal

$\Downarrow$ chordal extension

$G(N, \overline{E})$



(a)



(b)

$\{1, 6\}$, $\{2, 6\}$, $\{3, 4, 6\}$, $\{4, 5, 6\}$

$\{1, 6\}$, $\{2, 6\}$, $\{3, 5, 6\}$, $\{3, 4, 5\}$

Maximal cliques (node sets of maximal complete subgraphs)

POP: $\zeta^* =$ min. $f_0(\boldsymbol{x})$ s.t. $f_k(\boldsymbol{x}) \geq 0$ $(k = 1, \ldots, m)$

Let $F_k = \{i : x_i$ is involved $f_k\}$.

Sparsity pattern matrix $\boldsymbol{R} \in \mathbb{S}^n$ (symbolic) & graph $G(N, E)$.

$$R_{ij} = \begin{cases} \star & \text{if } x_i, x_j \in \text{a monomial of } f_0 \text{ or } \{i, j\} \subset F_k \ \exists k \geq 1, \\ 0 & \text{otherwise .} \end{cases}$$

$$N = \{1, 2, \ldots, n\}, \ E = \{(i, j) \in N \times N : i \neq j \ \text{ and } R_{ij} = \star\}.$$

POP: $\zeta^* = $ min. $f_0(\boldsymbol{x})$ s.t. $f_k(\boldsymbol{x}) \geq 0 \ (k = 1, \ldots, m)$

Let $F_k = \{i : x_i$ is involved $f_k\}$.

Sparsity pattern matrix $\boldsymbol{R} \in \mathbb{S}^n$ (symbolic) & graph $G(N, E)$.

$$R_{ij} = \begin{cases} \star & \text{if } x_i, x_j \in \text{a monomial of } f_0 \text{ or } \{i, j\} \subset F_k \ \exists k \geq 1, \\ 0 & \text{otherwise .} \end{cases}$$

$$N = \{1, 2, \ldots, n\}, \ E = \{(i, j) \in N \times N : i \neq j \ \text{ and } R_{ij} = \star\}.$$

- $G(N, E)$ is chordal iff $\exists$ simultaneous row and column ordering of $\boldsymbol{R}$ that allows a Cholesky factorization of $\boldsymbol{R}$ with no fill-in.
- If $G(N, E)$ is not chordal, then there are fill-ins in its Cholesky factorization, which are corresponding to the edges that are added in its chordal extension $G(N, \overline{E})$.
- The chordal extension is not unique; each chordal extension is corresponding a Cholesky factorization with $\exists$ simultaneous row and column ordering of $\boldsymbol{R}$

POP: $\zeta^* = $ min. $f_0(\boldsymbol{x})$ s.t. $f_k(\boldsymbol{x}) \geq 0 \ (k = 1, \ldots, m)$

Let $F_k = \{i : x_i$ is involved $f_k\}$.

Sparsity pattern matrix $\boldsymbol{R} \in \mathbb{S}^n$ (symbolic) & graph $G(N, E)$.

$$R_{ij} = \begin{cases} \star & \text{if } x_i, x_j \in \text{a monomial of } f_0 \text{ or } \{i,j\} \subset F_k \ \exists k \geq 1, \\ 0 & \text{otherwise .} \end{cases}$$

$$N = \{1, 2, \ldots, n\}, \ E = \{(i,j) \in N \times N : i \neq j \ \text{ and } R_{ij} = \star\}.$$

POP: $\zeta^* = $ min. $f_0(\boldsymbol{x})$ s.t. $f_k(\boldsymbol{x}) \geq 0 \ (k = 1, \ldots, m)$

Let $F_k = \{i : x_i$ is involved $f_k\}$.

Sparsity pattern matrix $\boldsymbol{R} \in \mathbb{S}^n$ (symbolic) & graph $G(N, E)$.

$$R_{ij} = \begin{cases} \star & \text{if } x_i, x_j \in \text{a monomial of } f_0 \text{ or } \{i,j\} \subset F_k \ \exists k \geq 1, \\ 0 & \text{otherwise .} \end{cases}$$

$$N = \{1, 2, \ldots, n\}, \ E = \{(i,j) \in N \times N : i \neq j \text{ and } R_{ij} = \star\}.$$

Γ : the set of max. cliques of a chordal ext. $G(N, \overline{E})$ of $G(N, E)$.

Then $F_k \subset \exists C \in \Gamma$. Let $F_k \subset C_k \in \Gamma \ (k = 1, 2, \ldots, m)$. Let

the relax. order $r \geq \lceil \max\{\text{deg}(f_k) : k = 0, \ldots, m\}/2 \rceil$,

$r_k = r - \lceil \text{deg}(f_k)/2 \rceil \ (k = 1, \ldots, m)$,

$\textbf{SOS}_q(C) = $ the sum of square poly. with deg. $\leq q$ in $x_i \ (i \in C)$.

Sparse SOS relaxation SOS^r:

$$\max \ \eta \ \text{s.t.} \quad L(\boldsymbol{x}, \lambda_1, \ldots, \lambda_m) - \eta \in \textstyle\sum_{C \in \Gamma} \textbf{SOS}_r(C) \ (\forall \boldsymbol{x} \in \mathbb{R}^n),$$
$$\lambda_1 \in \textbf{SOS}_{r_1}(C_1), \ldots, \lambda_m \in \textbf{SOS}_{r_m}(C_m).$$

- Lasserre's dense SOS^r is obtained if $\Gamma = \{N\}$.

POP: $\zeta^* =$ min. $f_0(\boldsymbol{x})$ s.t. $f_k(\boldsymbol{x}) \geq 0 \ (k = 1, \ldots, m)$

Let $F_k = \{i : x_i$ is involved $f_k\}$.

Sparsity pattern matrix $\boldsymbol{R} \in \mathbb{S}^n$ (symbolic) & graph $G(N, E)$.

$$R_{ij} = \begin{cases} \star & \text{if } x_i, x_j \in \text{a monomial of } f_0 \text{ or } \{i, j\} \subset F_k \ \exists k \geq 1, \\ 0 & \text{otherwise .} \end{cases}$$

$$N = \{1, 2, \ldots, n\}, \ E = \{(i, j) \in N \times N : i \neq j \ \text{and } R_{ij} = \star\}.$$

Γ : the set of max. cliques of a chordal ext. $G(N, \overline{E})$ of $G(N, E)$.

Then $F_k \subset \exists C \in \Gamma$. Let $F_k \subset C_k \in \Gamma \ (k = 1, 2, \ldots, m)$. Let

the relax. order $r \geq \lceil \max\{\text{deg}(f_k) : k = 0, \ldots, m\}/2 \rceil$,

$r_k = r - \lceil \text{deg}(f_k)/2 \rceil \ (k = 1, \ldots, m)$,

POP: $\zeta^* = $ min. $f_0(\boldsymbol{x})$ s.t. $f_k(\boldsymbol{x}) \geq 0 \ (k = 1, \ldots, m)$

Let $F_k = \{i : x_i$ is involved $f_k\}$.

Sparsity pattern matrix $\boldsymbol{R} \in \mathbb{S}^n$ (symbolic) & graph $G(N, E)$.

$$R_{ij} = \begin{cases} \star & \text{if } x_i, x_j \in \text{a monomial of } f_0 \text{ or } \{i,j\} \subset F_k \ \exists k \geq 1, \\ 0 & \text{otherwise .} \end{cases}$$

$$N = \{1, 2, \ldots, n\}, \ E = \{(i,j) \in N \times N : i \neq j \text{ and } R_{ij} = \star\}.$$

Γ : the set of max. cliques of a chordal ext. $G(N, \overline{E})$ of $G(N, E)$.

Then $F_k \subset \exists C \in \Gamma$. Let $F_k \subset C_k \in \Gamma \ (k = 1, 2, \ldots, m)$. Let

the relax. order $r \geq \lceil \max\{\text{deg}(f_k) : k = 0, \ldots, m\}/2 \rceil$,

$r_k = r - \lceil \text{deg}(f_k)/2 \rceil \ (k = 1, \ldots, m)$,

$\boldsymbol{u}_q(\boldsymbol{x}, C) =$ the col. vect. of monomials with deg. $\leq q$ in $x_i \ (i \in C)$.

Sparse Poly.SDP $\Longrightarrow$ Sparse SDP^r by linearlization

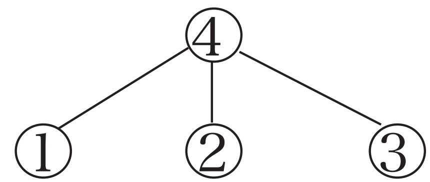
$$\min \quad f_0(\boldsymbol{x}) \quad \text{s.t.} \quad \boldsymbol{u}_{r_j}(\boldsymbol{x}, C_j)\boldsymbol{u}_{r_j}(\boldsymbol{x}, C_j)^T f_k(\boldsymbol{x}) \succeq \boldsymbol{O} \ (\forall k)$$
$$\boldsymbol{u}_r(\boldsymbol{x}, C)\boldsymbol{u}_r(\boldsymbol{x}, C)^T \succeq \boldsymbol{O} \ (\forall C \in \Gamma)$$

- Lasserre's dense SDP^r is obtained if $\Gamma = \{N\}$.

Example of a sparse POP $\Rightarrow$ a sparse SDP[2]

POP: min $\quad f_0(\boldsymbol{x}) = \sum_{i=1}^{4}(-x_i^3)$

s.t. $\quad f_k(\boldsymbol{x}) = -a_i \times x_i^2 - x_4^2 + 1 \geq 0 \ (i = 1, 2, 3).$



$$\boldsymbol{R} = \begin{pmatrix} \star & 0 & 0 & \star \\ 0 & \star & 0 & \star \\ 0 & 0 & \star & \star \\ \star & \star & \star & \star \end{pmatrix}$$

$C_1 = \{1, 4\}, \ C_2 = \{2, 3\}$, $C_3 = \{3, 4\}$ } max. cliques

$F_k = \{k, 4\} \subset C_k \ (k = 1, 2, 3)$.

the relax. order $r = 2 \geq \lceil \max\{\text{deg}(f_k) : 0 \leq k \leq 3\}/2 \rceil = 2,$

$r_k = 2 - \lceil \text{deg}(f_k)/2 \rceil = 1 \ (k = 1, 2, 3),$

$\boldsymbol{u}_1(\boldsymbol{x}, C_k) = (1, x_k, x_4)^T \ (k = 1, 2, 3),$

$\boldsymbol{u}_2(\boldsymbol{x}, C_k) = (1, x_k, x_4, x_k^2, x_k x_4, x_4^2)^T \ (k = 1, 2, 3).$

Poly.SDP: min. $f_0(\boldsymbol{x})$ s.t. $\quad \boldsymbol{u}_1(\boldsymbol{x}, C_k)\boldsymbol{u}_1(\boldsymbol{x}, C_k)^T f_k(\boldsymbol{x}) \succeq \boldsymbol{O} \ (\forall k)$

$\quad \boldsymbol{u}_2(\boldsymbol{x}, C_k)\boldsymbol{u}_2(\boldsymbol{x}, C_k)^T \succeq \boldsymbol{O} \ (\forall k)$

Contents

- SparsePOP (Waki et al.) for POPs $\Rightarrow$ sparse SDPs
- SDPA (Fujisawa et al.) for solving SDPs
- MATLAB Optimization Toolbox to refine the SDP solution
- Intel Xeon 2.66GHz with 12 GB memory

Unconstrained optimization problem

The gneralized Rosenbrock function — poly. with deg $= 4$

$$f_R(\boldsymbol{x}) \;=\; 1 + \sum_{i=2}^{n} \left(100(x_i - x_{i-1}^2)^2 + (1 - x_i^2)^2\right)$$

The chained singular function — poly. with deg $= 4$

$$\begin{aligned} f_C(\boldsymbol{x}) \;=\; & \sum_{i \in J} \Big((x_i + 10x_{i+1})^2 + 5(x_{i+2} - x_{i+3})^2 \\ & +(x_{i+1} - 2x_{i+2})^4 + 10(x_i - 10x_{i+3})^4\Big) \end{aligned}$$

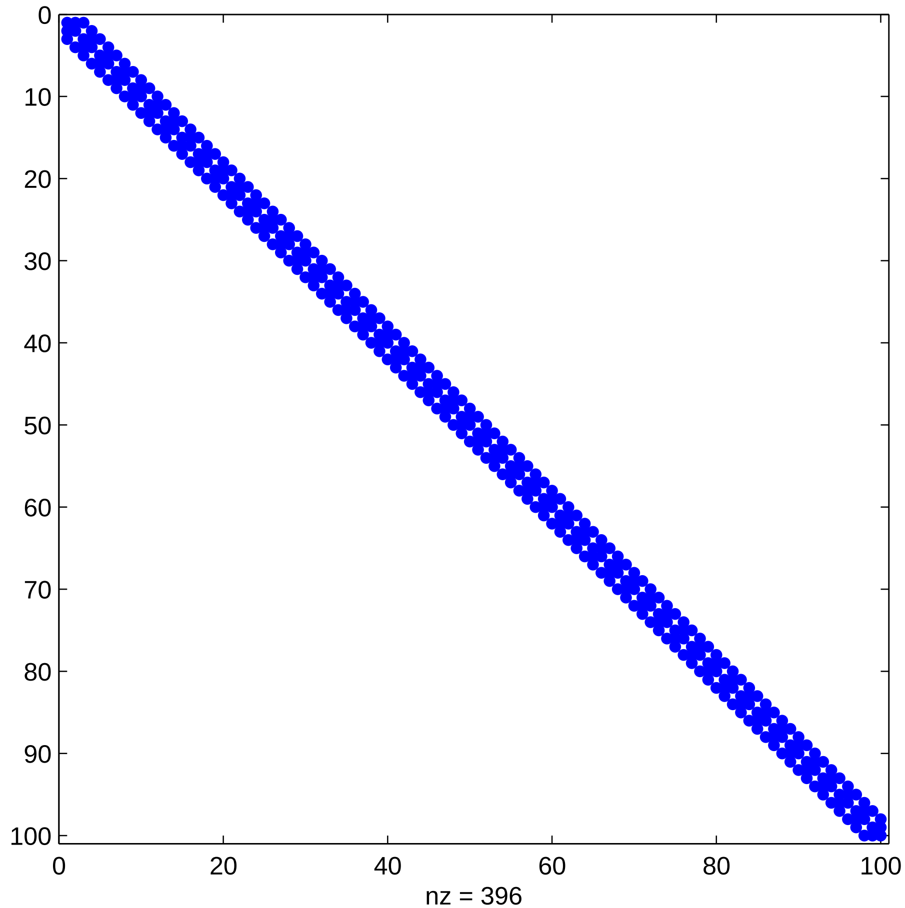
Here $J = \{1, 3, 5, \ldots, n-3\}$, n is a mutiple of 4.

minimize $f_R(\boldsymbol{x}) + f_C(\boldsymbol{x})$

— unknown global optimal value and solution

the sparsity pattern matrix $\boldsymbol{R}$
(=the sparsity pattern of the Hessian matrix)
— very sparse $\Rightarrow$ next

Sparsity pattern of the sparsity pattern matrix $\boldsymbol{R}$ with a simultaneous row and column reordering (Reverse Cuthill-McKee ordering)



Structured sparsity

- Sparse (symbolic) Cholesky factorization

min $f_R(\boldsymbol{x}) + f_C(\boldsymbol{x})$ — deg. 4, sparse, unknown opt.val.

	Sparse			Dense (Lasserre)		
n	ϵ_{obj}	$\#=$	cpu	ϵ_{obj}	$\#=$	cpu
12	6e-5	214	0.1	3e-6	1,819	6.9
16	5e-5	294	0.1	1e-9	4,844	71.3
100	7e-6	1,974	0.5	out of	mem	
1000	7e-7	19,974	6.9			
2000	7e-8	39,974	15.1			
3000	out of	mem				

$$\epsilon_{\text{obj}} = \frac{|\text{lbd. for opt.val.} - \text{approx.opt.val.}|}{\max\{1, |\text{lbd. for opt.val.}|\}}.$$

$\#=$: the number of equalities of SDP,

cpu : cpu time in second

- Global optimality is guaranteed with high accuracy.

alkyl from globalib — presented previously

$$\begin{aligned}\text{min}\quad & -6.3x_5x_8+5.04x_2+0.35x_3+x_4+3.36x_6\\ \text{sub.to}\quad & -0.820x_2+x_5-0.820x_6=0,\\ & 0.98x_4-x_7(0.01x_5x_{10}+x_4)=0,\ -x_2x_9+10x_3+x_6=0,\\ & x_5x_{12}-x_2(1.12+0.132x_9-0.0067x_9^2)=0,\\ & x_8x_{13}-0.01x_9(1.098-0.038x_9)-0.325x_7=0.574,\\ & x_{10}x_{14}+22.2x_{11}=35.82,\ x_1x_{11}-3x_8=-1.33,\\ & \text{lbd}_i\le x_i\le \text{ubd}_i\ (i=1,2,\ldots,14).\end{aligned}$$

Sparse			Dense (Lasserre)		
ϵ_{obj}	ϵ_{feas}	cpu	ϵ_{obj}	ϵ_{feas}	cpu
7.7e-8	1.8e-6	1.4	out of	memory	

$$\epsilon_{\text{obj}}=\frac{|\text{lbd. for opt.val.}-\text{approx.opt.val.}|}{\max\{1,|\text{lbd. for opt.val.}|\}}.$$

$\epsilon_{\text{feas}}=$ max. error in equalities, Elapsed time of SDPA in sec.

- Global optimality is guaranteed with high accuracy.

Some benchmark QOPs from globalib

			Sparse			Dense
Problem	n	r	ϵ_{obj}	ϵ_{feas}	E.time	E.time
ex2_1_8	24	2	1.2e-5	1.5e-6	6.7	28.7
ex3_1_1	8	2	3.5e-2	1.8e-14	0.3	0.4
ex3_1_1	8	3	2.5e-7	1.2e-13	0.6	52.3
ex5_4_2	8	2	5.2e-1	2.5e-12	0.3	0.5
ex5_4_2	8	3	8.0e-9	4.6e-16	0.6	49.0
ex5_3_2	22	2	1.5e-1	1.7e-16	1.3	47.4†
ex5_3_2	22	3	1.3e-4	2.6e-14	599.0	-

$r =$ the relax. order, E.time $=$ Elapsed time of SDPA in sec.,

$$\epsilon_{\text{obj}} = \frac{|\text{lbd. for opt.val.} - \text{approx.opt.val.}|}{\max\{1, |\text{lbd. for opt.val.}|\}}.$$

$\epsilon_{\text{feas}} =$ the max. error in equalities and inequalities,

$\dagger$: a highly accurate sol. is obtained.