From Convex Programming to Optimization over Polynomials: An Introduction to Current Research Activities

Wu-Sheng Lu
Dept. of Electrical and Computer Engineering
University of Victoria

Part I: Convex Programming (Jan. 23, 2:30 – 3:30 pm)

Part II: Polynomial Optimization Problems (Jan. 30, 2:30 – 3:30 pm)

Outline of Part I

1. Optimization Problems
2. A Duality Theory
3. Semidefinite Programming
4. Software

1. Optimization Problems

- Unconstrained Optimization

$$\underset{x \in R^n}{\text{minimize}}\, f(x)$$

  - Example: Solve polynomial system

$$p_1(x) = 0,\ p_2(x) = 0,\ \ldots,\ p_m(x) = 0 \quad (1)$$

$$(1) \Leftrightarrow \quad p_1^2(x) = 0,\ p_2^2(x) = 0,\ \ldots,\ p_m^2(x) = 0 \quad (2)$$

$$(2) \Leftrightarrow \quad f(x) = \sum_{i=1}^{m} p_i^2(x) = 0$$

    - x^* is a solution of (1) iff the global minimum of $f(x)$ is zero and x^* is a global minimizer.

- Constrained Optimization

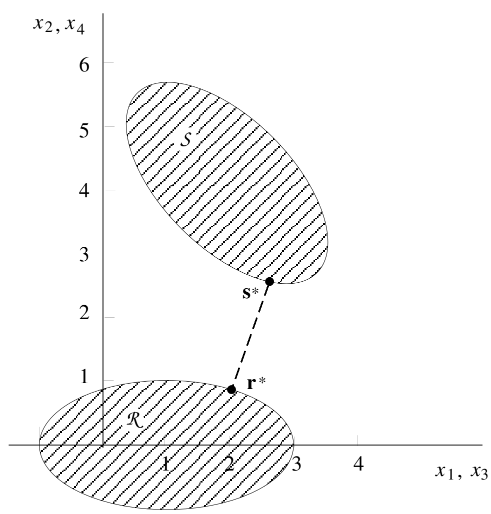
$$\begin{aligned} \text{minimize} \quad & f(x) && (3a) \\ \text{subject to:} \quad & a_i(x) = 0,\ i = 1,\ \ldots,\ p && (3b) \\ & c_j(x) \geq 0,\ j = 1,\ \ldots,\ q && (3c) \end{aligned}$$

  - Feasible region and feasible points

$$K = \{x:\ a_i(x) = 0 \text{ for } 1 \leq i \leq p,\ c_j(x) \geq 0 \text{ for } 1 \leq j \leq q\}$$

$$(3) \Leftrightarrow \underset{x \in K}{\text{minimize}}\, f(x) \quad (4)$$

- Example: Minimum distance between two ellipses



$$\begin{array}{ll} \text{minimize} & f(x) = (x_1 - x_3)^2 + (x_2 - x_4)^2 \\ \text{subject to:} & \frac{1}{4}x_1^2 + x_2^2 - \frac{1}{2}x_1 \le \frac{3}{4} \\ & \frac{5}{8}(x_3^2 + x_4^2) + \frac{3}{4}x_3x_4 - \frac{11}{2}x_3 - \frac{13}{2}x_4 \le -\frac{35}{2} \end{array}$$

- Example: Optimum data detection in wireless communications

$$\begin{array}{ll} \text{minimize} & x^T Q x + p^T x \qquad (Q \succeq 0) \\ \text{subject to:} & x_i \in \{0,\ 1\},\ i = 1,\ \ldots,\ n \\ & \Updownarrow \\ \text{minimize} & x^T Q x + p^T x \\ \text{subject to:} & x_i^2 - x_i = 0,\ i = 1,\ \ldots,\ n \end{array}$$

- Convex Programming (CP)

$$\text{minimize} \quad f(x) \tag{5a}$$
$$\text{subject to:} \quad a_i(x) = 0 \quad i = 1,\ \ldots,\ p \tag{5b}$$
$$c_j(x) \ge 0 \quad j = 1,\ \ldots,\ q \tag{5c}$$

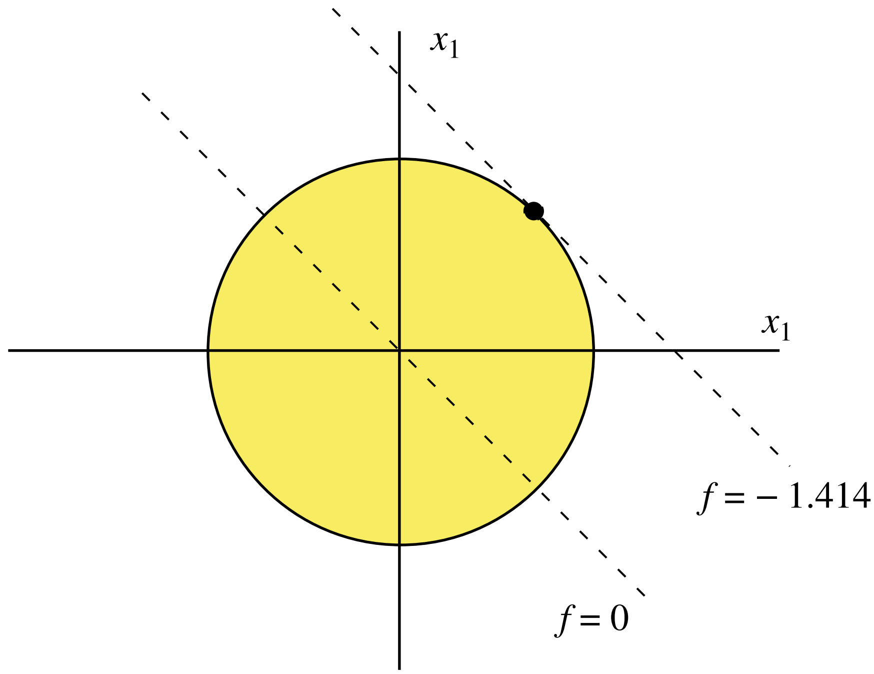
where $f(x)$ is convex
$a_i(x)$ are linear, $1 \le i \le p$
$-c_j(x)$ are convex, $1 \le j \le q$

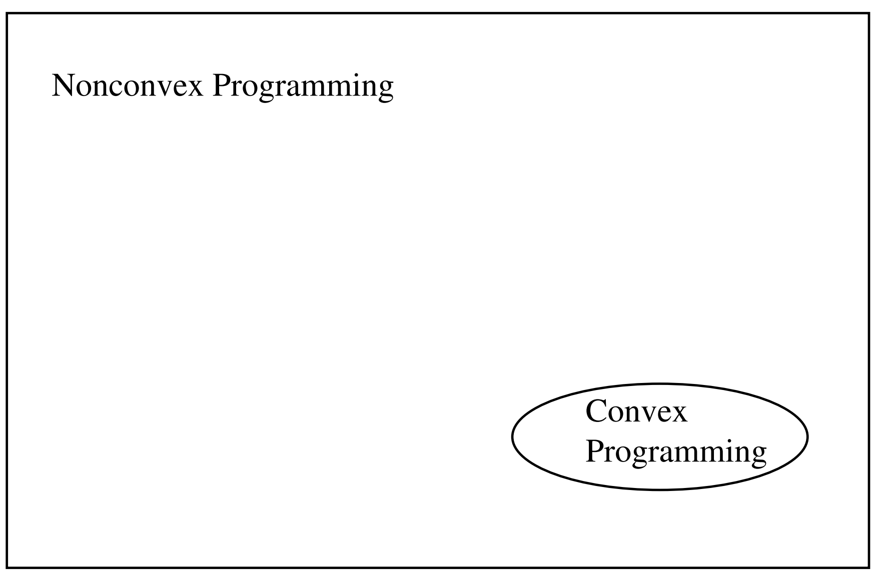
- A CP problem minimizes a convex objective function over a convex feasible region.
  - Example:

$$\begin{array}{ll} \text{minimize} & -x_1 - x_2 \\ \text{subject to:} & x_1^2 + x_2^2 \le 1 \end{array}$$

The objective function is convex and the feasible region is convex ($-c_1(x) = x_1^2 + x_2^2 - 1$ convex), hence a CP problem.

- Why convex programming?
  - CP has several desirable properties:
    Globlaness and uniqueness of solution; convexity of solution set; Karush-Kuhn-Tucker (KKT) conditions.
    For CP problem (5), the KKT conditions are both necessary and sufficient:
    Suppose x^* is a minimizer of (5) that is regular for the constraints active at x^*, then
    (a) $a_i(x^*) = 0$ for $1 \le i \le p$


x_1
x_1
$f = -1.414$
$f = 0$


Nonconvex Programming
Convex
Programming

(b) $c_j(x^*) \geq 0$ for $1 \leq j \leq q$
(c) $\exists \lambda_i^*$ and μ_j^* such that

$$\nabla f(x^*) = \sum_{i=1}^{p} \lambda_i^* \nabla a_i(x^*) + \sum_{j=1}^{q} \mu_j^* \nabla c_j(x^*)$$

(d)$\mu_j^* c_j(x^*) = 0$ for $1 \leq j \leq q$ (complementarity conditions)
(e) $\mu_j^* \geq 0$ for $1 \leq j \leq q$

- There exists a nice duality theory

- There exist efficient solvers.

- Classification of CP problems

  - Linear programming (LP)

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to:} \quad & Ax \geq b \end{aligned}$$

  - Convex quadratic programming (QP)

$$\begin{aligned} \text{minimize} \quad & x^T Q x + q^T x + K \quad (Q \succeq 0) \\ \text{subject to:} \quad & Ax \geq b \end{aligned}$$

  - Second-order cone programming (SOCP)

$$\begin{aligned} \text{minimize} \quad & b^T x \\ \text{subject to:} \quad & \|A_i x + b_i\| \leq c_i^T x + d_i, \ i = 1, \ \ldots, \ q \end{aligned}$$

  - Semidefinite programming (SDP)

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to:} \quad & F(x) = F_0 + \sum_{i=1}^{q} x_i F_i \succeq 0 \end{aligned}$$

  $(F_0, \ F_i$ are symmetric matrices)

- Relations of SDP with LP, QP, and SOCP

  - LP:

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to:} \quad & Ax \geq b \end{aligned}$$

  Write $Ax \geq b$ as

$$-b + Ax = -b + [a_1 \ \ldots \ a_n]x = -b + \sum_{i=1}^{n} a_i x_i \geq 0$$

$$\iff \underbrace{\text{diag}\{-b\}}_{F_0} + \sum_{i=1}^{n} x_i \underbrace{\text{diag}\{a_i\}}_{F_i} \succeq 0$$

$$\iff F_0 + \sum_{i=1}^{n} x_i F_i \succeq 0$$

- SOCP

$$\begin{array}{ll} \text{minimize} & b^T x \\ \text{subject to:} & \|A_i x + b_i\| \leq c_i^T x + d_i \quad 1 \leq i \leq q \end{array}$$

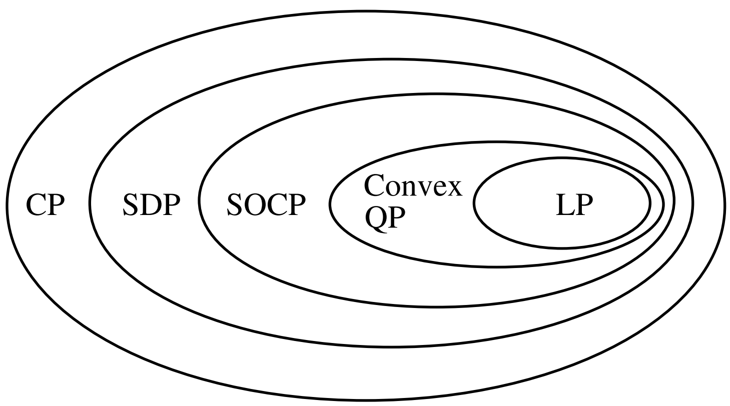
Note:

$$||u|| \leq t \iff \begin{bmatrix} tI & u \\ u^T & t \end{bmatrix} \succeq 0$$

Hence $\|A_i x + b_i\| \leq c_i^T x + d_i$

$$\iff \begin{bmatrix} (c_i^T x + d_i)I & A_i x + b_i \\ (A_i x + b_i)^T & c_i^T x + d_i \end{bmatrix} \succeq 0$$

- LP $\subset$ QP $\subset$ SOCP $\subset$ SDP



2. Wolfe's Theorem on Duality
   Consider the general CP problem (as the primal problem)

$$\begin{array}{lll} \text{(P)} & \text{minimize} & f(x) \\ & \text{subject to:} & a_i(x) = a_i^T x - b_i = 0, \ i = 1, \ \ldots, \ p \\ & & c_j(x) \geq 0, \ j = 1, \ \ldots, \ q \end{array}$$

   Define its Lagrangian

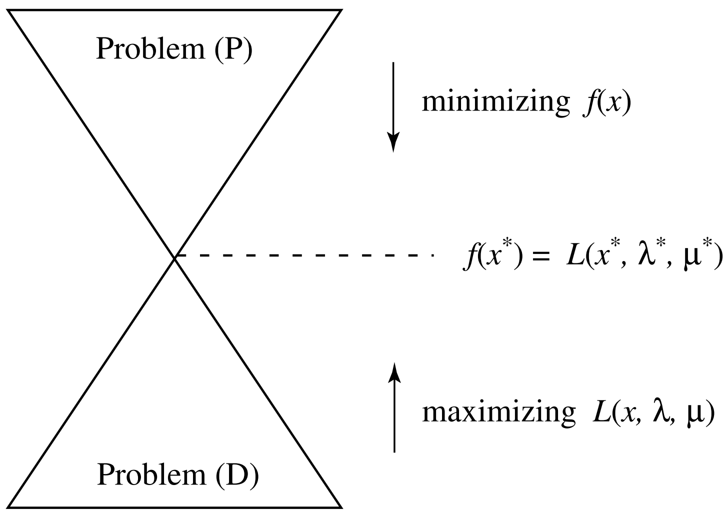
$$L(x, \ \lambda, \ \mu) = f(x) - \sum_{i=1}^{p} \lambda_i a_i(x) - \sum_{j=1}^{q} \mu_j c_j(x)$$

- The dual of problem (P) can be created by the following theorem due to P. Wolfe (1961):

  Let x^* be a minimizer of (P) and $\lambda^*,\ \mu^*$ be the associated Lagrange multipliers. Assume x^* is a regular point of the constraints. Then (x^*, λ^*, μ^*) also solves the dual problem

$$\text{(D)} \qquad \underset{x,\lambda,\mu}{\text{maximize}} \quad L(x, \lambda, \mu)$$
$$\text{subject to:} \quad \nabla L(x,\ \lambda,\ \mu) = 0$$
$$\mu \geq 0$$

In addition, $f(x^*) = L(x^*,\ \lambda^*,\ \mu^*)$



- Define *duality gap* $\delta(x,\ \lambda,\ \mu) = f(x) - L(x,\ \lambda,\ \mu)$
  Wolfe's theorem says $\delta \geq 0$ for feasible $x,\ \lambda,\ \mu$,
  and $\delta = 0$ at $(x^*,\ \lambda^*,\ \mu^*)$.
  In fact, for a feasible set $(x,\ \lambda,\ \mu)$

$$\delta = f(x) - L(x,\ \lambda,\ \mu) = \sum_{i=1}^{p} \lambda_i a_i(x) + \sum_{j=1}^{q} \mu_j c_j(x)$$
$$= \sum_{j=1}^{q} \mu_j c_j(x) \geq 0$$

  and δ reduces to zero at $(x^*,\ \lambda^*,\ \mu^*)$ because of the complementarity conditions

$$\mu_j^* c_j(x^*) = 0, \quad j = 1,\ \ldots,\ q$$

- Example: Linear programming

$$\text{(P)} \qquad \text{minimize} \quad c^T x \qquad\qquad \text{(Engineering)}$$
$$\text{subject to:} \quad Ax \geq b$$

$L(x,\ \lambda,\ \mu) = c^T x - (Ax-b)^T \mu$

$$
\begin{array}{llll}
\text{(D)} & \text{maximize} & c^T x - (Ax-b)^T \mu & \\
 & \text{subject to:} & c - A^T\mu = 0,\ \mu \geq 0 & \\
 & & \Updownarrow & \\
\text{(D)} & \text{maximize} & b^T \mu & \text{(Math)} \\
 & \text{subject to:} & A^T \mu = c,\ \mu \geq 0 &
\end{array}
$$

3. Semidefinite Programming (SDP)

- Why is SDP popular?
  - SDP is a class of CP problems (theoretical tractability)
  - LP, QP, SOCP are subclasses of SDP
  - SDP arises in a number of important applications in science and engineering
  - Efficient SDP solvers are available

- Examples

  Let $A(x) = A_0 + x_1 A_1 + \cdots + x_n A_n$

  - Find $x^* = [x_1^* \ \ldots \ x_n^*]^T$ that minimizes the largest eigenvalue of $A(x)$, i.e., find x^* to solve

    $$\underset{x}{\text{minimize}} \max \lambda[A(x)]$$

    ■ $A(x)$ is symmetric

    $$\Longrightarrow \quad A(x) = U^T \begin{bmatrix} \lambda_{max} & & 0 \\ & \ddots & \\ 0 & & \lambda_{min} \end{bmatrix} U$$

    with U orthogonal $\Rightarrow$ $tI - A(x) = t \cdot U^T I U - A(x)$

    $$= U^T \begin{bmatrix} t - \lambda_{max} & & 0 \\ & \ddots & \\ 0 & & t - \lambda_{min} \end{bmatrix} U$$

    ■ Hence $tI - A(x) \succeq 0$ iff $t \geq \lambda_{max}$

    ■ The value of t satisfying $tI - A(x) \succeq 0$ provides at right upper bound for λ_{max}
    This is an SDP problem:

    $$\begin{array}{ll} \text{minimize} & t \\ \text{subject to:} & tI - A(x) \succeq 0 \end{array}$$

– Find x^* that minimizes the 2-norm of $A(x)$, i.e., find x^* to solve

$$\underset{x}{\text{minimize}} \max \lambda^{1/2}[A^T(x)A(x)]$$

We need to solve

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to:} & t^2 I - A^T(x)A(x) \succeq 0 \end{array}$$

which can be converted into the SDP problem

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to:} & \begin{bmatrix} tI & A(x) \\ A^T(x) & tI \end{bmatrix} \succeq 0 \end{array}$$

- Duality

  – the primal SDP assumes the form

$$(P) \qquad \begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to:} & F(x) = F_0 + \sum_{i=1}^{n} x_i F_i \succeq 0 \end{array}$$

  – the Lagrangian:

$$L(x,\ Y) = c^T x - Y \cdot (F_0 + \sum_{i=1}^{n} x_i F_i)$$

  where the inner product $A \cdot B = \text{trace}(AB) = \sum_i \sum_j a_{ij} b_{ij}$

  – the Wolfe dual:

$$\begin{array}{ll} \text{maximize} & c^T x - Y \cdot F_0 - \sum_{i=1}^{n} x_i (Y \cdot F_i) \\ \text{subject to:} & \nabla_x L(x,\ Y) = 0 \\ & Y \succeq 0 \end{array}$$

$$\Longrightarrow \qquad \text{(D)} \qquad \begin{array}{ll} \text{maximize} & -Y \cdot F_0 \\ \text{subject to:} & Y \cdot F_i = c_i, \ \ 1 \le i \le n \\ & Y \succeq 0 \end{array}$$

  – Duality gap:

$$\delta = c^T x + Y \cdot F_0 = \sum_{i=1}^{n} c_i x_i + Y \cdot F_0 = \sum_{i=1}^{n} x_i (Y \cdot F_i) + Y \cdot F_0$$

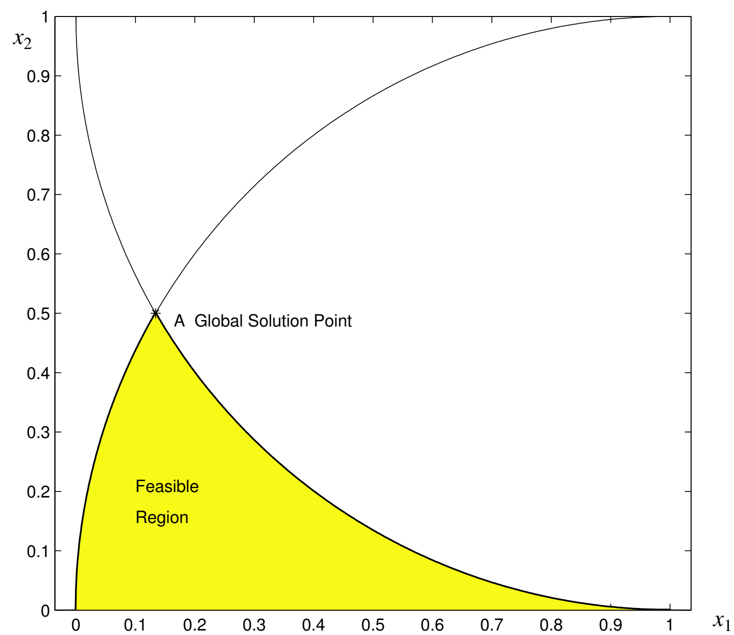
$$= Y \cdot (F_0 + \sum_{i=1}^{n} x_i F_i) = Y \cdot F(x) \ge 0$$

– Example

$$\begin{array}{rl} \text{minimize} & x_1 - 2x_2 \\ \text{subject to:} & x_1 \geq 0,\ x_2 \geq 0 \\ & (x_1-1)^2 + x_2^2 \leq 1 \\ & (x_1-1)^2 + (x_2-1)^2 \geq 1 \end{array}$$

Number of variables: 2, Number of constraints: 4

■ This is a nonconvex problem because its feasible region is not convex:



■ The global solutin of the problem is $x^* = [0.1340\ \ 0.5]^T$.

■ Put the problem in an extended space (dimensional extension)
Let $y_{10} = x_1,\ y_{01} = x_2,\ y_{20} = x_1^2,\ y_{02} = x_2^2$ and express the problem as

$$\begin{array}{rl} \text{minimize} & y_{10} - 2y_{01} \\ \text{subject to:} & y_{10} \geq 0,\ y_{01} \geq 0 \\ & -y_{20} + 2y_{10} - y_{02} \geq 0 \\ & y_{20} - 2y_{10} + y_{02} - 2y_{01} + 1 \geq 0 \\ \text{(one might add:)} & y_{20} \geq 0,\ y_{02} \geq 0 \end{array}$$

Number of variables: 4, Number of constraints: 6
This is an LP problem whose unique global solution is

$$y^* = [0\ \ 0.5\ \ 0\ \ 0]^T$$

which gives $\tilde{x}^* = [0\ \ 0.5]^T$.

■ Now use a further dimensional extension $y_{11} = x_1 x_2$
to allow a semidefinite constraint:

$$0 \preceq \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix} [1 \;\; x_1 \;\; x_2] = \begin{bmatrix} 1 & x_1 & x_2 \\ x_1 & x_1^2 & x_1 x_2 \\ x_2 & x_1 x_2 & x_2^2 \end{bmatrix} = \begin{bmatrix} 1 & y_{10} & y_{01} \\ y_{10} & y_{20} & y_{11} \\ y_{01} & y_{11} & y_{02} \end{bmatrix}$$

This leads to a SDP problem:

$$\begin{aligned} \text{minimize} \quad & y_{10} - 2y_{01} \\ \text{subject to:} \quad & y_{10} \geq 0, \; y_{01} \geq 0, \; y_{20} \geq 0, \; y_{11} \geq 0, y_{02} \geq 0 \\ & -y_{20} + 2y_{10} - y_{02} \geq 0 \\ & y_{20} - 2y_{10} + y_{02} - 2y_{01} + 1 \geq 0 \\ & \begin{bmatrix} 1 & y_{10} & y_{01} \\ y_{10} & y_{20} & y_{11} \\ y_{01} & y_{11} & y_{22} \end{bmatrix} \succeq 0 \end{aligned}$$

Number of variables: 5, Number of constraints: 8
The unique global solution of the SDP problem is given by

$$y^* = [0.1340 \;\; 0.5 \;\; 0.0179 \;\; 0.0670 \;\; 0.25]^T$$

which gives $\tilde{x}^* = [0.1340 \;\; 0.5]^T$.

4. Software for CP (MATLAB-compatible)

- Commercial
  Optimization Toolbox (MathWorks) — LP, QP
  Robust Control Toolbox (MathWorks) — SDP

- Public-domain
  SDPT3 (Cornell, NUS, CMU) — LP, QP SOCP, SDP
  SDPA (Tokyo Inst. Tech.) — LP, QP, SOCP, SDP
  SeDuMi (J.F. Sturm; McMaster) — LP, QP, SOCP, SDP
  http://sedumi.mcmaster.ca

Part II: Polynomial Optimization Problems (POP)

1. Unconstrained and Constrained POP
2. Lasserre's Conversion
3. Moments and Moment Matrices
4. Connection of Moment Problems to SDP
5. Solving POP via SDP Relaxation
6. Software and An Example

1. Unconstrained and Constrained POP

- Unconstrained POP

$$\underset{x\in R^n}{\text{minimize}}\, p(x) \tag{1}$$

- Constrained POP

$$\underset{x\in K}{\text{minimize}}\ \ p(x) \tag{2a}$$
$$K = \{x \in R^n : h_1(x) \ge 0,\ \ldots,\ h_m(x) \ge 0\} \tag{2b}$$

K is known as a semi-algebraic set.

- LP, QP, SOCP, SDP are subclasses of POP

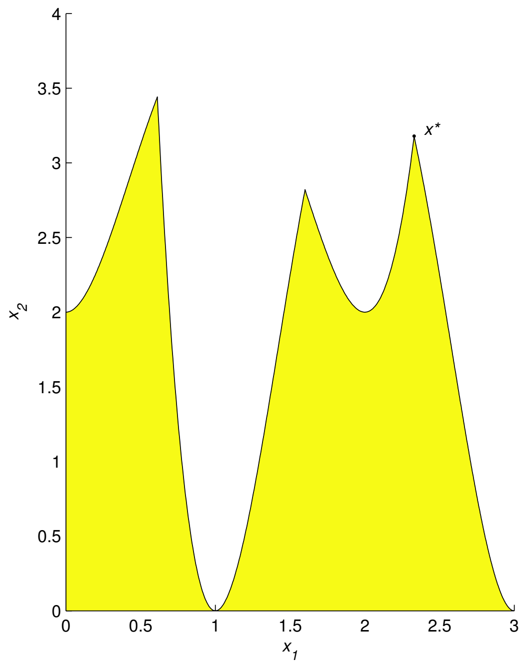
LP and QP — obvious.

$$\begin{aligned}
\text{SOCP:}\quad & \|A_i^T x + b_i\| \le c_i^T x + d_i \\
\Longleftrightarrow\quad & \|A_i^T x + b_i\|^2 \le (c_i^T x + d_i)^2,\ \ c_i^T x + d_i \ge 0 \\
\text{SDP:}\quad & F(x) = F_0 + x_1 F_1 + \cdots + x_n F_n \succeq 0 \\
\Longleftrightarrow\quad & \underbrace{\text{its principal minors}}_{\text{polynomials in } x}\ \ \text{are nonnegative}
\end{aligned}$$

- But POP also include a great many nonconvex problems.

  - Example (Laurent, 2006)

$$\begin{aligned}
\text{minimize}\quad & p(x) = -x_1 - x_2 \\
\text{subject to:}\quad & x_2 \le 2x_1^4 - 8x_1^3 + 8x_1^2 + 2 \\
& x_2 \le 4x_1^4 - 32x_1^3 + 88x_1^2 - 96x_1 + 36 \\
& 0 \le x_1 \le 3,\ 0 \le x_2 \le 4
\end{aligned}$$



– Example

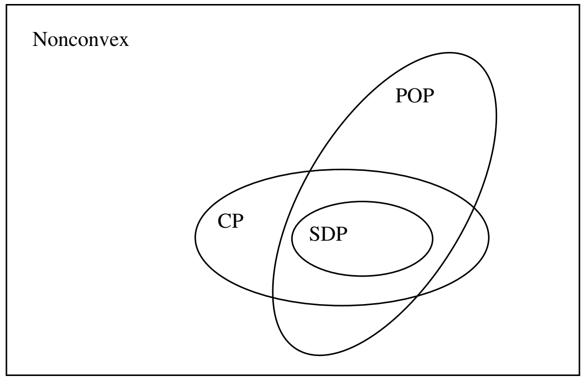
$$
\begin{array}{lll}
 & \text{minimize} & x^T Q x + q^T x \\
 & \text{subject to:} & x_i \in \{0,\ 1\} \\
\Longleftrightarrow & \text{minimize} & x^T Q x + q^T x \\
 & \text{subject to:} & x_i^2 - x_i = 0,\ 1 \le i \le n \\
\Longleftrightarrow & \text{minimize} & x^T Q x + q^T x \\
 & \text{subject to:} & x_i^2 - x_i \ge 0 \quad 1 \le i \le n \\
 & & -x_i^2 + x_i \ge 0 \quad 1 \le i \le n
\end{array}
$$

– Example (Mathematical programming with equilibrium constraints):

$$
\begin{array}{ll}
\text{minimize} & (x_1 + x_2 + y_1 - 15)^2 + (x_1 + x_2 + y_2 - 15)^2 \\
\text{subject to:} & 0 \le x_1 \le 10,\ 0 \le x_2 \le 10,\ y_1 \ge 0,\ y_2 \ge 0 \\
 & z_1 = \frac{8}{3}x_1 + 2x_2 + 2y_1 + \frac{8}{3}y_2 - 36 \ge 0 \\
 & z_2 = 2x_1 + \frac{5}{4}x_2 + \frac{5}{4}y_1 + 2y_2 - 25 \ge 0 \\
 & y_1 z_1 + y_2 z_2 = 0
\end{array}
$$

2. Lasserre's Conversion

In his SIAM 2001 paper, Jean B. Lasserre takes a fresh look at the POP problems as minimizing a linear function of sequence of moments over all probability measures, which in turn connects the problems to the theory of moment matrices and leads eventually to SDP-relaxation based solution methods.



- Probability Measures

  A probability measure is a real-valued function μ on a set $\mathcal{S}$ satisfying the following properties

  - $\mu(\phi) = 0,\ \mu(\mathcal{S}) = 1$
  - For subsets X and Y with $X \bigcap Y = \phi$, $\mu(X \bigcup Y) = \mu(X) + \mu(Y)$
  - For subsets X and Y with $X \subseteq Y$, $\mu(X) \leq \mu(Y)$

    So we see that a probability measure is a nonnegative measure, i.e., $\mu(X) \geq 0$ for any $X \subseteq S$

    - Example: Let $\mathcal{S}$ be the real line and

      $$d\mu = \tfrac{1}{\sqrt{\pi}} e^{-x^2} dx$$
      $$\mu(\mathcal{S}) = \textstyle\int_{-\infty}^{\infty} d\mu = \tfrac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-x^2} dx = 1$$

    - Example: Let $\mathcal{S}$ be the real line and

      $$d\mu = \delta(x - x^*)dx$$

      where $\delta(x)$ is Dirac's δ-function defined by

      $$\delta(x) = \begin{cases} 0 & x \neq 0 \\ \infty & x = 0 \end{cases} \qquad \int_{-\infty}^{\infty} \delta(x)dx = 1$$

      Note

      $$\mu(\mathcal{S}) = \int_{-\infty}^{\infty} d\mu = \int_{-\infty}^{\infty} \delta(x - x^*)dx = 1$$

      We call this μ the Dirac measure at x^*, having mass 1 at x^* and mass zero elsewhere.

    - Example:

      $$d\mu = \sum_{i=1}^{r} \lambda_i \delta(x - x_i)dx$$

with

$$\lambda_i > 0 \quad \text{and} \quad \sum_{i=1}^{r} \lambda_i = 1$$

This is a nonnegative measure satisfying

$$\int_{R^n} d\mu = \int_{R^n} \sum_{i=1}^{r} \lambda_i \delta(x - x_i) dx = \sum_{i=1}^{r} \lambda_i \int_{R^n} \delta(x - x_i) dx$$
$$= \sum_{i=1}^{r} \lambda_i = 1$$

The points x_i are called atoms and the measure is called r-atomic.

- Lasserre's Observations
  - If

$$p^* = \min_{x \in R^n} p(x) \quad \text{then } p^* = \min_{\mu \in \mathcal{P}(R^n)} \int_{R^n} p(x) d\mu \tag{3}$$

  - If

$$p_K^* = \min_{x \in K} p(x) \quad \text{then } p_K^* = \min_{\mu \in \mathcal{P}(K)} \int_K p(x) d\mu \tag{4}$$

where

$\mathcal{P}(R^n)$ — all probability measures over R^n
$\mathcal{P}(K)$ — all probability measures over K

Proof of (3): $p(x) \geq p^*$ for all $x \in R^n$

$$\Longrightarrow \qquad \int_{R^n} p(x) d\mu \geq p^* \int_{R^n} d\mu = p^*$$
$$\Longrightarrow \qquad \min_{\mu \in \mathcal{P}(R^n)} \int_{R^n} p(x) d\mu \geq p^* \tag{5}$$

On the other hand, suppose p^* is achieved by $p(x)$ at x^*, i.e., $p(x^*) = p^*$. We consider the Dirac measure $d\mu = \delta(x - x^*)dx$ and compute

$$\int_{R^n} p(x) d\mu = \int_{R^n} p(x) \delta(x - x^*) dx = p(x^*) = p^*$$

Hence

$$\min_{\mu \in \mathcal{P}(R^n)} \int_{R^n} p(x) d\mu \leq p^* \tag{6}$$

(5) and (6) imply (3).

– The significance of (3) and (4) is that the problem of minimizing polynomial $p(x)$ over a semi-algebraic set is now converted to the problem of minimizing the integral $\int p(x)d\mu$ over all probability measures.
– Next, the later problem is converted to the optimization over sequences of moments.

3. Moments and Moment Matrices

- Notation

$$p(x) = \sum_{\alpha} p_\alpha x^\alpha, \; x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n} \tag{7}$$

The order of $p(x)$ is equal to the largest $\sum_{i=1}^{n} \alpha_i$ among all nonzero coefficients p_α in (7).

- The terms in (7) are arranged according to a basis for d-degree real-valued polynomial $p(x)$: $1, \; x_1, \; x_2, \; \dots, \; x_n, \; x_1^2, \; x_1x_2, \; \dots, \; x_1x_n, \; \dots, \; x_n^2, \; \dots, \; x_1^d, \; \dots, \; x_n^d$
- The number of the terms in the basis is called the dimension of the basis and is denoted by $s(d)$

$$s(d) = \binom{n+d}{d} = \frac{(n+d)!}{n!d!}$$

Example:

n	d	$s(d)$
2	2	6
3	2	10
5	2	21
10	2	66
10	4	91
10	10	184756

- Writing

$$\int p(x)d\mu = \int \sum_{\alpha} p_\alpha x^\alpha d\mu = \sum_{\alpha} p_\alpha \int x^\alpha d\mu = \sum_{\alpha} p_\alpha y_\alpha$$

where

$$y_\alpha = \int x^\alpha d\mu$$

are the moments for the nonnegative measure μ, the problems in (3) and (4) become

$$\min_{\mu \in \mathcal{P}(R^n)} \int_{R^n} p(x)d\mu = \min_{\{y_\alpha\}} \sum_{\alpha} p_\alpha y_\alpha \tag{8}$$

with $\{y_\alpha\}$ a sequence of moments associated with a representing measure μ over R^n; and

$$\min_{\mu \in \mathcal{P}(K)} \int_{K} p(x)d\mu = \min_{\{y_\alpha\}} \sum_{\alpha} p_\alpha y_\alpha \tag{9}$$

with $\{y_\alpha\}$ a sequence of moments associated with a representing measure μ over set K.

- In (8) and (9), the probability measures are replaced by the sequence of moments, and the objective functions are linear functions of $\{y_\alpha\}$ so the question now is how those sequences $\{y_\alpha\}$ that are associated with nonnegative measures can be characterized. It is at this point of the development where the theory of moment matrices comes to play an important role.

  - Moment matrix $M(y)$

    - Example: $n = 2$
      $\alpha = (\alpha_1,\ \alpha_2)$: (0, 0), (1, 0), (0, 1), (2, 0), (1, 1), (0, 2), (3, 0), ...
      $x^\alpha : 1,\ x_1,\ x_2,\ x_1^2,\ x_1x_2,\ x_2^2,\ x_1^3,\ \ldots$
      $\{y_\alpha\} : y_{00},\ y_{10},\ y_{01},\ y_{20},\ y_{11},\ y_{02},\ y_{30},$ with $y_{00} = 1$

$$M(y) = \begin{bmatrix} y_{00} & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} & \cdots & y_\alpha & \cdots \\ y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} & \cdots & \vdots & \\ y_{01} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} & \cdots & \vdots & \\ y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22} & \cdots & \vdots & \\ y_{11} & y_{21} & y_{12} & y_{31} & y_{22} & y_{13} & \cdots & \vdots & \\ y_{02} & y_{12} & y_{03} & y_{22} & y_{13} & y_{04} & \cdots & \vdots & \\ \vdots & \vdots & \vdots & \cdots & \cdots & \cdots & \vdots & \vdots & \\ y_\beta & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & y_{\alpha+\beta} & \\ \vdots & & & & & & & \vdots & \end{bmatrix}$$

  - Truncated moment matrix $M_t(y)$
    It starts with a truncated sequence of moments $\{y_\alpha\} \in s(2t)$, and can be obtained from $M(y)$ as a leading principal submatrix of dimension $s(t)$.

    - Example: $n = 2,\ t = 1$:

$$M_1(y) = \begin{bmatrix} 1 & y_{10} & y_{01} \\ y_{10} & y_{20} & y_{11} \\ y_{01} & y_{11} & y_{02} \end{bmatrix}$$

      If $d\mu = \delta(x - x^*)dx$ with $x^* = [x_1^*\ \ x_2^*]^T$ then

$$M_1(y) = \begin{bmatrix} 1 & x_1^* & x_2^* \\ x_1^* & {x_1^*}^2 & x_1^*x_2^* \\ x_2^* & x_1^*x_2^* & {x_2^*}^2 \end{bmatrix} = \begin{bmatrix} 1 \\ x_1^* \\ x_2^* \end{bmatrix} [1\ \ x_1^*\ \ x_2^*] \succeq 0$$

If $d\mu = \frac{1}{\pi} e^{-(x_1^2+x_2^2)} dx_1 dx_2$ then

$$\begin{aligned} y_{10} &= y_{01} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} x e^{-x^2} dx = 0 \\ y_{20} &= y_{02} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} x^2 x e^{-x^2} dx = \frac{\pi}{2} \\ y_{11} &= 0 \end{aligned}$$

Hence

$$M_1(y) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{\pi}{2} & 0 \\ 0 & 0 & \frac{\pi}{2} \end{bmatrix} \succ 0$$

4. Connection of Moment Problems to SDP

- In general, an arbitrary vector p_α of dimension $s(t)$ can be associated with a polynomial of degree t, and we can write

$$\begin{aligned} 0 \le \int p^2(x) d\mu &= \int \left(\sum_\alpha p_\alpha x^\alpha \right) \cdot \left(\sum_\beta p_\beta x^\beta \right) d\mu \\ &= \sum_\alpha \sum_\beta p_\alpha p_\beta \int x^{\alpha+\beta} d\mu \\ &= \sum_\alpha \sum_\beta p_\alpha p_\beta y_{\alpha+\beta} = p_\alpha^T M_t(y) p_\alpha \end{aligned}$$

  Hence

$$M_t(y) \succeq 0 \implies M(y) \succeq 0$$

- The Constrained Case

  Consider one polynomial constraint: $K = \{x : h(x) \ge 0\}$

  - We need to define "shift vector" as $h * y$ whose α-component is $\sum_\beta h_\beta y_{\alpha+\beta}$
  - If $h(x)$ is a polynomial of degree $2d$ or $2d-1$, $y \in R^{s(2t)}$ is the truncated sequence of moments up to order $2t$ of a nonnegative measure supported by $K = \{x : h(x) \ge 0\}$, then $M_{t-d}(h * y) \succeq 0$.
    This is because

$$p^T M_{t-d}(h * y) p = \int_K h(x) p^2(x) d\mu \ge 0$$

  - In general, for $K = \{x : h_1(x) \ge 0, \ \ldots, \ h_m(x) \ge 0\}$ where each $h_j(x)$ is a polynomial of degree $2d_j$ or $2d_j - 1$, and $y \in R^{s(2t)}$ is truncated sequence of moments of a nonnegative measure supported by K, then

$$M_t(y) \succeq 0, \ M_{t-d_j}(h_j * y) \succeq 0 \quad 1 \le j \le m$$

- In summary,
  - $\min_{x \in R^n} p(x) = \min_{\mu \in \mathcal{P}(R^n)} \int_{R^n} p(x) d\mu = \min_{y \in \mathcal{M}} p^T y$

    where $\mathcal{M}$ is the set of sequences y, each of which admits a representing measure.
  - $\min_{x \in K} p(x) = \min_{\mu \in \mathcal{P}(K)} \int_{K} p(x) d\mu = \min_{y \in \mathcal{M}(K)} p^T y$

    where $\mathcal{M}(K)$ — the set of sequences having respresenting measures supported by set K.
  - $y \in \mathcal{M} \Rightarrow M(y) \succeq 0$
  - $y \in \mathcal{M}(K) \Rightarrow M_t(y) \succeq 0$, $M_{t-d_j}(h_j * y) \succeq 0$

    So if we let $\mathcal{M}_+ = \{y : M(y) \succeq 0\}$, then $\mathcal{M} \subseteq \mathcal{M}_+$.
  - Is $\mathcal{M} = \mathcal{M}_+$?

    If yes, then the unconstrained POP becomes an "SDP" problem:

$$\hat{p}^* = \underset{y \in \mathcal{M}_+}{\text{minimize}}\, p^T y \tag{10}$$

  - It is known that $\mathcal{M} \subset \mathcal{M}_+$ is proper, so the minimum of (10) offers only a lower bound of the global minimum p^* in (3):

$$\hat{p}^* \leq p^*$$

  - Also, technical difficulties exist for implementing (10) because $M(y)$ is of infinite dimension.
- For constrained POP:
  - Let $\mathcal{M}_+^{put} = \{y : M(y) \succeq 0,\ M(h_j * y) \succeq 0,\ 1 \leq j \leq m\}$, then

$$\mathcal{M}(K) \subset \mathcal{M}_+^{put}$$

    Thus the minimum of the SDP problem

$$\hat{p}_K^* = \underset{y \in \mathcal{M}_+^{put}}{\text{minimize}}\, p^T y \tag{11}$$

    offers a lower bound for p_K^* in (4):

$$\hat{p}_K^* \leq p_K^*$$

5. Soving POP via SDP Relaxation

- Lasserre proposes to deal with the unconstrained POP problem

$$\underset{x \in R^n}{\text{minimize}}\, p(x)$$

by solving a series of truncated SDP problems

$$p_t^* = \min p^T y \ \text{ s.t. } \ y_0 = 1,\ M_t(y) \succeq 0 \tag{12}$$

where $t \geq \deg(p(x))/2$.

- Similarly, the constrained POP $\underset{x \in K}{\text{minimize}}\, p(x)$ is treated by solving

$$\begin{aligned} &p_{tK}^* = \min p^T y \quad \text{s.t. } \ y_0 = 1, \\ &M_t(y) \succeq 0,\ M_{t-d_j}(h_j * y) \succeq 0,\ 1 \leq j \leq m \end{aligned} \tag{13}$$

where $t \geq \max(d_0,\ d_1,\ \ldots,\ d_m)$, $d_0 = \lceil \deg(p)/2 \rceil$, $d_j = \lceil \deg(h_j)/2 \rceil$, for $1 \leq j \leq m$.

- Properties of Lasserre's Solution Method
  - For a fixed t, (13) is a standard SDP problem and can be solved efficiently.
  - Under certain compactness condition on set K, $p_{tK}^* \to p_K^*$ as $t \to \infty$, and the convergence sometimes can even be achieved in finite number of steps.
  - Optimality certificate: If $M_t(y)$ satisfies certain rank condition, then $p_{tK}^* = p_K^*$.
  - Under this rank condition, the global minimizers can be constructed.

6. Software and An Example

- Public-domain

  GloptiPoly (Henrion and Lasserre)

  SparsePOP (Waki, Kim, Kojima, and Muramatsu)

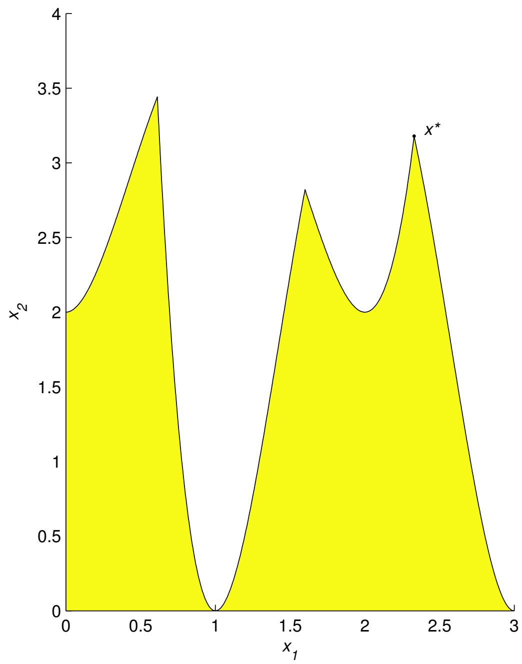
  SOSTOOLS (Prajna, Papachristodoulou, and Parrilo)

  - All three packages incorporate SeDuMi for solving the SDP problems involved.

- Example (Laurent, 2006)

$$\begin{array}{rl} \text{minimize} & p(x) = -x_1 - x_2 \\ \text{subject to:} & x_2 \leq 2x_1^4 - 8x_1^3 + 8x_1^2 + 2 \\ & x_2 \leq 4x_1^4 - 32x_1^3 + 88x_1^2 - 96x_1 + 36 \\ & 0 \leq x_1 \leq 3,\ 0 \leq x_2 \leq 4 \end{array}$$

- MATLAB code using GloptiPoly




```
function x = laurent_ex42_f(order)

P{1}.c = [0 -1 0 0 0; -1 0 0 0 0]; P{1}.t = 'min';
P{2}.c = [2 -1; 0 0; 8 0; -8 0; 2 0]; P{2}.t = '>=';
P{3}.c = [36 -1; -96 0; 88 0; -32 0; 4 0]; P{3}.t = '>=';
P{4}.c = [0; 1]; P{4}.t = '>=';
P{5}.c = [3; -1]; P{5}.t = '>=';
P{6}.c = [0 1]; P{6}.t = '>=';
P{7}.c = [4 -1]; P{7}.t = '>=';
out = gloptipoly(P,order);
x = out.sol{:};
```

- Results

Order t	Bound $\hat{p}_{tK}$	Solution
2	-7.00	none
3	-6.67	none
4	-5.51	$[2.3295\ \ 3.1785]^\dagger$

†: Global minimizer.

- MATLAB code using SparsePOP

```
function x = laurent_ex42_j(order)

% Name of the problem to be solved.
  problemName = 'laurent_ex42_j.gms';
% objPoly
  objPoly.typeCone = 1;
  objPoly.dimVar   = 2;
  objPoly.degree   = 1;
  objPoly.noTerms  = 2;
  objPoly.supports = [1 0; 0 1];
  objPoly.coef     = [-1; -1];
% ineqPolySys
  ineqPolySys{1}.typeCone = 1;
  ineqPolySys{1}.dimVar   = 2;
  ineqPolySys{1}.degree   = 4;
  ineqPolySys{1}.noTerms  = 5;
  ineqPolySys{1}.supports = [0 0; 0 1; 2 0; 3 0; 4 0];
  ineqPolySys{1}.coef     = [2; -1; 8; -8; 2];
% ineqPolySys
  ineqPolySys{2}.typeCone = 1;
  ineqPolySys{2}.dimVar   = 2;
  ineqPolySys{2}.degree   = 4;
  ineqPolySys{2}.noTerms  = 6;
  ineqPolySys{2}.supports = [0 0; 0 1; 1 0; 2 0; 3 0; 4 0];
  ineqPolySys{2}.coef     = [36; -1; -96; 88; -32; 4];
% lower bounds for variables x1 and x2
  lbd = [0,0];
% upper bounds for variables x1 and x2
  ubd = [3,4];

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Default values of parameters
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
param.dummy = 0; param.symbolicMath = 0;
param.reduceMomentMatSW = 1; param.relaxOrder = order;
[param,SDPobjValue,POP,cpuTime,SeDuMiInfo,SDPinfo] = ...
        sparsePOP(param,objPoly,ineqPolySys,lbd,ubd);
fileId = 1; printLevel=2;
printSolution(fileId,printLevel,problemName,param,...
        SDPobjValue,POP,cpuTime,SeDuMiInfo,SDPinfo);
x = POP.xVect; return
```

- Results

Order t	Bound $\hat{p}_{tK}$	Solution
2	-7.00	$[2.9907 \;\; 4.0]^*$
3	-6.64	$[2.6377 \;\; 4.0]^*$
4	-5.51	$[2.3295 \;\; 3.1785]^\dagger$

$*$: Not feasible.

†: Global minimizer.