Moments, Sums of Squares and Semidefinite Programming

Jean B. LASSERRE

LAAS-CNRS, and Institute of Mathematics, Toulouse, France

Semidefinite Programming and its Applications, IMS Singapore, January 2006

- the Generalized Problem of Moments (GPM)
- Some applications
- Duality between moments and nonnegative polynomials
- SDP-relaxations for the basic GPM.
- s.o.s. vs nonnegative polynomials. Alternative SDP-relaxations
- How to handle sparsity

The generalized problem of moments (GPM)

$$\min_{\mu \in M(\mathbf{K})} \{ \int f_0 \, d\mu \quad | \quad \int f_j \, d\mu \overset{=}{\geq} b_j, \quad j = 1, \ldots, p \}$$

with $\mathbf{K} \subseteq \mathbf{R}^n$ and $M(\mathbf{K})$ a convex set of finite Borel measures on $\mathbf{K}$. We even consider the more general GPM

$$\min_{\mu_i \in M(\mathbf{K}_i)} \{ \sum_{i \in I} \int f_{oi} \, d\mu_i \quad | \quad \sum_{i \in I} \int f_{ji} \, d\mu_i \overset{=}{\geq} b_j, \quad j = 1, 2, \ldots \}$$

where for all $i \in I$, $\mathbf{K}_i \subseteq \mathbf{R}^{n_i}$ and $M(\mathbf{K}_i)$ is a convex set of finite Borel measures on $\mathbf{K}_i$. The index set I may be countable.

• GPM has great modelling power, in various fields. Global Optimization (continuous, discrete), Control (Robust and optimal control), Nonlinear Equations, Probability and Statistics, Performance Evaluation (in e.g. Mathematical finance, Markov chains), Inverse Problems (cristallography, tomography), Numerical multivariate Integration, etc ...

• GPM is a useful theoretical tool to prove existence and characterization of optimal solutions.

• **BUT** ... in full generality .... GPM is unsolvable numerically.

HOWEVER ... if the $\mathbf{K}_i\,, (\subset \mathbf{R}^{n_i})$ are basic semi-algebraic sets and the f_{ij} are polynomials (or even piecewise polynomials), then ... by using results of real algebraic geometry and on the problem of moments, one may now define efficient numerical approximation chemes, based on Semidefinite Programming (SDP).

- the Generalized Problem of Moments (GPM)
- Some applications
- Duality between moments and nonnegative polynomials
- SDP-relaxations for the basic GPM
- s.o.s. vs nonnegative polynomials. Alternative SDP-relaxations
- How to handle sparsity

A few examples:

PROBLEM 1: Probability:
Let $\mathbf{K} \subseteq \mathbb{R}^n$, $S \subset \mathbf{K}$ be Borel subsets, and $\Gamma \subset \mathbb{N}^n$,

Finding an upper bound (if possible optimal) on Prob $(\mathbf{X} \in S)$, the probability that a $\mathbf{K}$-valued random variable $\mathbf{X} \in S$, given some of its moments $\gamma = \{\gamma_\alpha\}$, $\alpha \in \Gamma \subset \mathbb{N}^n$ ....

.... is equivalent to solving:

$$\sup_{\mu \in M(\mathbf{K})} \left\{ \mu(S) \quad | \quad \int x^\alpha \, d\mu = \gamma_\alpha, \quad \alpha \in \Gamma \right\}$$

- $M(\mathbf{K})$ is the (convex) set of **probability measures** on $\mathbf{K} \subseteq \mathbb{R}^n$.
- $f_\alpha \equiv x^\alpha$, $\alpha \in \Gamma$ (polynomial); $f_0 = \mathrm{I}_S$ (piecewise polynomial)

PROBLEM 2: Moments problems in financial economics:

Under no arbitrage, the price of an European Call Option with strike k, is given by $E[(\mathbf{X}-k)^+]$ where E is the expectation operator w.r.t. the distribution of the underlying asset $\mathbf{X}$.

Hence, finding an (**optimal**) upper bound on the price of a European Call Option with strike k, given the first p moments $\{\gamma_j\}$, reduces to solving:

$$\sup_{\mu \in M(\mathbf{K})} \left\{ \int (x-k)^+ \, d\mu \quad \Big| \quad \int x^j \, d\mu = \gamma_j, \quad j = 1, \ldots, p \right\}$$

with $\mathbf{K} = \mathbf{R}_+$, and $M(\mathbf{K})$ the set of probability measures on $\mathbf{K}$.

$f_j \equiv x^j$ (polynomials), and $f_0 \equiv (x-k)^+$ (piecewise polynomial)

PROBLEM 3: Global Optimization:

Let $\mathbf{K} \subseteq \mathbf{R}^n$, $f : \mathbf{R}^n \to \mathbf{R}$, and consider the optimization problem

$$f^* := \inf_x \{ f(x) \quad | \quad x \in \mathbf{K} \}$$

with f^* being the global minimum.

Finding f^* is equivalent to solving

$$\inf_{\mu \in M(\mathbf{K})} \int f \, d\mu$$

with $M(\mathbf{K})$ being the set of probability measures on $\mathbf{K}$.

PROBLEM 4: Measures with given marginals:

Let $\mathbf{K}_j \subset \mathbf{R}^{n_j}$, $j = 1, \ldots, p$, and $\mathbf{K} := \mathbf{K}_1 \times \mathbf{K}_2 \cdots \times \mathbf{K}_p \subset \mathbf{R}^n$, and with natural projections $\pi_j : \mathbf{K} :\to \mathbf{K}_j$, $j = 1, \ldots, p$.
Let ν_j be a given Borel measure on $\mathbf{K}_j$, $j = 1, \ldots, p$,

For a measure μ on $\mathbf{K}$, denote $\pi_j \, \mu$ its marginal on $\mathbf{K}_j$, i.e.

$$\pi_j \, \mu(B) := \mu\,(\pi_j^{-1}(B)) = \mu\,(\{x \in \mathbf{K} \,:\, \pi_j \, x \in B\}), \quad B \in \mathcal{B}(\mathbf{K}_j)$$

$$\inf_{\mu \in M(\mathbf{K})} \left\{ \int f \, d\mu \quad | \quad \pi_j \, \mu = \nu_j, \quad j = 1, \ldots, p \right\}$$

with $M(\mathbf{K})$ being the set of finite Borel measures on $\mathbf{K}$.

Generalization of the famous Monge-Kantorovich transportation problem, with many other interesting applications, particularly in Probability.

•• If $\mathbf{K}_j$ is compact then the constraint on marginal

$$\pi_j \, \mu = \nu_j$$

is equivalent to the countably many linear equalities

$$\int x^\alpha \, d\mu \quad = \quad \int x^\alpha \, d\nu_j, \quad \forall \alpha \in \mathbf{N}^{n_j}$$

between moments of μ and ν_j ...

because the space of polynomials is dense (for the sup-norm) in the space $C(\mathbf{K}_j)$ of continuous functions on $\mathbf{K}_j$.

PROBLEM 5: Deterministic Optimal Control:

$$
\begin{array}{rcl}
j^* := & \min_{\mathbf{u}} & \int_0^T h(s, x(s), u(s))\, ds + H(x(T)) \\
\dot{x}(s) & = & f(s,\, x(s), u(s)), \quad s \in [0, T) \\
(x(s), u(s)) & \in & X \times U, \quad s \in [0, T) \\
x(T) & \in & X_T,
\end{array}
\tag{1}
$$

and with initial condition $x(0) = x_0 \in X$, and

- $X, X_T \subset \mathbf{R}^n$ and $U \subset \mathbf{R}^m$ are basic semi-algebraic sets.

- $h, f \in \mathbf{R}[t, x, u]$, $H \in \mathbf{R}[x]$

Let $\mathbf{u} = \{u(t),\ 0 \le t < T\}$ be an admissible control.

Introduce the probability measure $\nu^{\mathbf{u}}$ on $\mathbf{R}^n$, and the measure $\mu^{\mathbf{u}}$ on $[0, T] \times \mathbf{R}^n \times \mathbf{R}^m$, defined by

$$\begin{aligned} \nu^{\mathbf{u}}(B) &:= I_B\,[x(T)], \qquad B \in \mathcal{B}_n \\ \mu^{\mathbf{u}}(A \times B \times C) &:= \int_{[0,T]\cap A} I_{B\times C}\,[(x(s), u(s))]\,ds, \end{aligned}$$

for all hyper-rectangles (A, B, C).

The measure $\mu^{\mathbf{u}}$ is called the *occupation measure* of the state-action (deterministic) process $(s, x(s), u(s))$ *up to* time T, whereas $\nu^{\mathbf{u}}$ is the *occupation measure* of the state $x(T)$ *at* time T.

- Observe that for an admissible trajectory $(s, x(s), u(s))$

$$\dot{x}(t) = f(t, x(t), u(t)), \quad t \in [0, T)$$

implies that for suitable $g : [0, T] \times X \to \mathbf{R}$, the *time* integration

$$g(x(T)) = g(0, x(0)) + \int_0^T \frac{\partial g(s, x(s))}{\partial t} + \frac{\partial g(s, x(s))}{\partial x} f(s, x(s), u(s))\, ds$$

is equivalent to the *spatial* integration

$$\int_{X_T} g_T\, d\nu^{\mathbf{u}} = g(0, x_0) + \int_{[0,T]\times X \times U} \left[\frac{\partial g}{\partial t} + \frac{\partial g}{\partial x} f\right] d\mu^{\mathbf{u}}$$

with $g_T(x) := g(T, x)$ for all x.

• Similarly, the criterion $\int_0^T h(s, x(s), u(s))\, ds + H(x(T))$ reads

$$\int_{X_T} H\, d\nu^{\mathbf{u}} + \int_{[0,T]\times X\times U} h\, d\mu^{\mathbf{u}} = L_y(H) + L_z(h).$$

The so-called weak formulation is the infinite-dimensional LP

$$\left\{ \begin{array}{ll} \rho^* = \min_{\mu,\nu} \int H\, d\nu + \int h\, d\mu & \\ \text{s.t.} \quad \int g_T\, d\nu - \int \frac{\partial g}{\partial t} + \frac{\partial g}{\partial x} f\, d\mu = g(0, x_0), & \forall\, g \in \mathbf{R}[t, x] \\ & \\ \mu : \quad \text{measure supported on } [0, T] \times X \times U & \\ \nu : \quad \text{prob. measure supported on } X_T & \end{array} \right.$$

• **Theorem:** [R. Vinter]. If X, X_T, U are compact, $f(s, x, U)$ is convex for all $(s, x) \in [0, T] \times X$, and h, H are convex, then $\rho^* = j^*$.

- the Generalized Problem of Moments (GPM)

- Some applications

- Duality between moments and nonnegative polynomials

- SDP-relaxations for the basic GPM

- s.o.s. vs nonnegative polynomials. Alternative SDP-relaxations

- How to handle sparsity

Duality

With $M(\mathbf{K})$ the space of Borel prob. measures on $\mathbf{K}$, the GPM

$$\min_{\mu\,\in\,M(\mathbf{K})} \{ \int f_0\,d\mu \quad \vert \quad \int f_j\,d\mu = b_j, \quad j=1,\ldots,p\,\}$$

is the **infinite-dimensional** **LP**

$$\min_{\mu\,\in\,\mathcal{M}} \{\,\langle f_0,\mu\rangle \mid \quad \langle f_j,\mu\rangle = b_j,\ j=1,\ldots,p; \quad \langle 1,\mu\rangle = 1;\ \mu\geq 0\,\}$$

where $\mathcal{M}$ is the **vector space** of **finite** *signed* **Borel measures** on $\mathbf{K}$. The **dual** LP reads:

$$\max_{\lambda\,\in\,\mathbf{R}^p,\gamma\in\mathbf{R}} \{ \quad \gamma \quad \vert \quad f_0 - \sum_{j=1}^p \lambda_j\,(f_j-b_j)\ \geq\ \gamma \quad \text{on } \mathbf{K}\,\}$$

To solve (or at least approximate) either LP, one needs :

- to handle $\int f_j \, d\mu$, and

relatively simple and tractable characterizations of :

- measures μ with support contained in $\mathbf{K}$, ... or

- $f_0 - \sum_{j=1}^{p} \lambda_j \, (f_j - b_j)$ nonnegative on $\mathbf{K}$.

A first good news ...

When $\mathbf{K} \subset \mathbf{R}^n$ is the basic compact semi-algebraic set

$$\mathbf{K} := \{ x \in \mathbf{R}^n \mid \quad g_j(x) \geq 0, \quad j = 1, \ldots, m \}$$

with $\{g_j\} \subset \mathbf{R}[x] \,(= \mathbf{R}[x_1, \ldots, x_n])$ ...

Powerful results of real algebraic geometry and on the moment problem, provide necessary and sufficient conditions for :

• a finite Borel measure μ to be supported on $\mathbf{K}$ (i.e., $\mu(\mathbf{K}^c) = 0$)

• a polynomial f to be > 0 on $\mathbf{K}$.

As one may expect, the conditions are *dual* to each other ....

A second good news ... (continued)

In both cases ... these conditions can translate into **Linear Matrix Inequalities** (**LMI**) on :

- the moments $y_\alpha := \int x^\alpha \, d\mu$, $\alpha \in \mathbb{N}^n$, of μ (with support in $\mathbf{K}$)

- the coefficients $\{q_{j\alpha}\}$ of sum of squares (s.o.s.) polynomials $\{q_j\}_{j=0}^m \subset \mathbf{R}[x]$, in e.g. Putinar's s.o.s. representation

$$f = q_0 + \sum_{j=1}^m q_j \, g_j, \qquad \text{if } f > 0 \text{ on } \mathbf{K}.$$

† Linear Inequalities instead of LMIs are also available .. but less efficient and ill-behaved ... despite so far, LP software packages are more powerful than SDP packages!!

Putinar-Jacobi-Prestel's Positivstellensatz

Let $Q(g_1,\ldots,g_m)$ be the quadratic module generated by the g_j's.

$$f \in Q(g_1,\ldots,g_m) \Rightarrow f \;=\; f_0 + \sum_{j=1}^{m} f_j\, g_j,$$

for some (finite) family $\{f_j\}_{j=0}^{m}$ of s.o.s. polynomials. It is an obvious **certificate of nonnegativity** on $\mathbf{K}$.

Assumption 1: There exists some $u \in Q(g_1,\ldots,g_m)$ such that the level set $\{x \in \mathbf{R}^n \,|\, u(x) \geq 0\}$ is compact.

Theorem (Putinar): Let $\mathbf{K}$ compact and Assumption 1 hold.
Then $[\,\mathbf{f} \in \mathbf{R}[\mathbf{x}]$ and $\mathbf{f} > \mathbf{0}$ on $\mathbf{K}\,] \Rightarrow \mathbf{f} \in \mathbf{Q}(\mathbf{g_1},\ldots,\mathbf{g_m})$.

If one fixes an apriori bound on the degree of the s.o.s. polynomials $\{f_j\}$, checking $f \in Q(g_1, \ldots, g_m)$ reduces to solving a SDP!!

Moreover, Assumption 1 holds true if e.g.

- all the g_j's are linear (hence $\mathbf{K}$ is a polytope), or if

- the set $\{\, x \mid g_j(x) \geq 0\}$ is compact for some $j \in \{1, \ldots, m\}$.

If $x \in \mathbf{K} \Rightarrow \|x\| \leq M$ for some (known) M, then it suffices to add the redundant quadratic constraint $M^2 - \|x\|^2 \geq 0$, in the definition of $\mathbf{K}$.

Putinar's dual condition: **The $\mathbf{K}$-moment problem**

Let $v(x) = \{x^\alpha\} := [1, x_1, \ldots, x_n, x_1^2, x_1x_2, \ldots]$ be a basis for $\mathbf{R}[x]$, and let $y := \{y_\alpha\}$ be a given sequence indexed in the basis $v(x)$.

Given $\mathbf{K} \subset \mathbf{R}^n$, does there exist a measure μ on $\mathbf{K}$, such that

$$y_\alpha = \int_{\mathbf{K}} x^\alpha \, d\mu, \qquad \forall \alpha \in \mathbf{N}^n$$

Given $y = \{y_\alpha\}$, let $L_y : \mathbf{R}[x] \to \mathbf{R}$, be the linear functional

$$f \; (= \sum_\alpha f_\alpha \, x^\alpha) \; \mapsto \; L_y(f) \; := \sum_{\alpha \in \mathbf{N}^n} f_\alpha \, y_\alpha.$$

Recall that $\mathbf{K} \subset \mathbf{R}^n$ is the semi-algebraic set

$$\mathbf{K} := \{\, \mathbf{x} \in \mathbf{R^n} \mid \quad \mathbf{g_j(x)} \geq \mathbf{0},\, \mathbf{j} = \mathbf{1}, \ldots, \mathbf{m}\}.$$

Assumption 1: There exists some $u \in Q(g_1, \ldots, g_m)$ such that the level set $\{x \in \mathbf{R}^n \mid u(x) \geq 0\}$ is compact.

Theorem (Putinar): Let $\mathbf{K}$ compact, and Assumption 1 hold. Then $y = \{y_\alpha\}$ has a representing measure μ on $\mathbf{K}$ if and only if

$$(**) \quad L_y(f^2); \quad L_y(f^2\, g_j) \geq 0, \quad \forall j = 1, \ldots, m; \quad \forall\, f \in \mathbf{R}[x]$$

Checking (**) for all $f \in \mathbf{R}[x]$ with degree less than r, reduces to solving an SDP ... to check!!

Given $y = \{y_\alpha\}$ indexed in the basis $v(x)$, introduce the **moment matrix** $M_r(y)$ with rows and columns also indexed in the basis $v(x)$, and defined as follows:

$$M_r(y)(\alpha, \beta) := L_y(x^{\alpha+\beta}) = y_{\alpha+\beta}, \quad \alpha, \beta \in \mathbf{N}^n, \quad |\alpha|, |\beta| \leq r.$$

For instance, and for illustration purposes, in $\mathbf{R}^2$,

$$M_1(y) = \left[\begin{array}{c|cc} y_{00} & y_{10} & y_{01} \\ - & - & - \\ y_{10} & y_{20} & y_{11} \\ y_{01} & y_{11} & y_{02} \end{array}\right]$$

Then

$$\left[L_y(f^2) \geq 0, \quad \forall f, \deg(f) \leq r \right] \quad \Leftrightarrow \quad M_r(y) \succeq 0$$

Similarly, given $\theta \in \mathbf{R}[x]$, $x \mapsto \theta(x) = \sum_\gamma \theta_\gamma \, x^\gamma$, one defines the localizing matrix $M_r(\theta \, y)$, with respect to y, θ, also indexed in the basis $v(x)$, by

$$M_r(\theta \, y)(\alpha,\beta) = L_y(\theta \, x^{\alpha+\beta}) = \sum_{\gamma \in \mathbf{N}^n} \theta_\gamma \, y_{\alpha+\beta+\gamma}, \quad \begin{cases} \alpha, \beta \in \mathbf{N}^n \\ |\alpha|, |\beta| \leq r. \end{cases}$$

For instance, in $\mathbf{R}^2$, and with $x \mapsto \theta(x) := 1 - x_1^2 - x_2^2$,

$$M_1(\theta \, y) = \begin{bmatrix} y_{00} - y_{20} - y_{02}, & y_{10} - y_{30} - y_{12}, & y_{01} - y_{21} - y_{03} \\ y_{10} - y_{30} - y_{12}, & y_{20} - y_{40} - y_{22}, & y_{11} - y_{21} - y_{12} \\ y_{01} - y_{21} - y_{03}, & y_{11} - y_{21} - y_{12}, & y_{02} - y_{22} - y_{04} \end{bmatrix}.$$

Then

$$\left[L_y(f^2 \, \theta) \geq 0, \quad \forall f, \deg(f) \leq r \right] \quad \Leftrightarrow \quad M_r(\theta \, y) \succeq 0$$

- the Generalized Problem of Moments (GPM)
- Some applications
- Duality between moments and nonnegative polynomials
- SDP-relaxations for the basic GPM
- s.o.s. vs nonnegative polynomials. Alternative SDP-relaxations
- How to handle sparsity

SDP-relaxations for solving the basic GPM

$$\min_{\mu \in M(\mathbf{K})} \{ \int f_0 \, d\mu \quad | \quad \int f_j \, d\mu = b_j, \quad j = 1, \ldots, p \}$$

($M(\mathbf{K})$ space of Borel prob. measures on $\mathbf{K}$, and $\{f_j\} \subset \mathbf{R}[x]$)

Let $\deg g_i = 2v_i$ or $2v_i - 1$. SDP-relaxation $\mathbf{Q}_r$ reads:

$$\mathbf{Q}_r \left\{ \begin{array}{lll} \min_{y} & L_y(f) & \\ & & \\ \text{s.t.} & M_r(y) & \succeq 0 \\ & M_{r-v_i}(g_i \, y) & \succeq 0 \quad i = 1, \ldots m. \\ & L_y(1) & = 1 \\ & & \\ & L_y(f_j - b_j) & = 0 \quad j = 1, \ldots p. \end{array} \right.$$

... whose dual is the SDP

$$\mathbf{Q}_r^* \begin{cases} \max\limits_{\lambda,\gamma,\{q_j\}} & \gamma \\ \text{s.t.} & f_0 - \sum_{j=1}^{p} \lambda_j \, (f_j - b_j) - \gamma \; = \; q_0 + \sum_{j=1}^{m} q_j \, g_j \\ & \{q_j\} \text{ are s.o.s.}; \quad \deg q_0, \, \deg q_j g_j \; \leq 2r \end{cases}$$

Recall that $\mathbf{K} \subset \mathbb{R}^n$ is the semi-algebraic set

$$\mathbf{K} := \{\ \mathbf{x} \in \mathbf{R}^{\mathbf{n}} \mid \quad \mathbf{g_j(x)} \geq \mathbf{0},\ \mathbf{j} = \mathbf{1}, \ldots, \mathbf{m}\}.$$

Assumption 1: There exists some $u \in Q(g_1, \ldots, g_m)$ such that the level set $\{x \in \mathbb{R}^n \mid u(x) \geq 0\}$ is compact.

Theorem: Let $\mathbf{K}$ be compact, and let Assumption 1 hold, and consider the basic **GPM** with optimal value ρ^*. Then :

- $\sup \mathbf{Q}_r^* \leq \inf \mathbf{Q}_r$ and $\inf \mathbf{Q}_r \uparrow \rho^*$ as $r \to \infty$

- If $\operatorname{int} \mathbf{K} \neq \emptyset$ and the GPM has a feasible solution with a density $\sup \mathbf{Q}_r^* = \max \mathbf{Q}_r^* = \inf \mathbf{Q}_r \uparrow \rho^*$.

Detecting global optimality and extracting solutions

• When $\mathbf{K}$ is compact, then the basic GPM has an optimal solution μ^*, with optimal value ρ^*.

• By Caratheodory theorem there exists an at most $(p+2)$-atomic probability measure φ on $\mathbf{K}$ such that

$$\int f_j \, d\varphi = \int f_j \, d\mu, \quad j = 1, \ldots, p; \qquad \int f_0 \, d\varphi = \rho^*$$

• Let y be an optimal solution of $\mathbf{Q}_r$ and let $2v \geq \max_j \deg g_j$. If

$$\operatorname{rank} M_r(y) = \operatorname{rank} M_{r-v}(g_j \, y) = s$$

$\min \mathbf{Q}_r = \rho^*$ and one may extract a s-atomic optimal solution φ.

GloptiPoly is a software package initially devoted to solving global optimization problems with polynomials.

`http://www.laas.fr/~henrion/software`

with detection of optimaility and extraction of solutions.

... New version to be realeased will solve GPM problems

- the Generalized Problem of Moments (GPM)

- Some applications

- Duality between moments and nonnegative polynomials

- SDP-relaxations for the basic GPM

- s.o.s. vs nonnegative polynomials. Alternative SDP-relaxations .

- How to handle sparsity

Nonnegative versus SOS polynomials.

Theorem (Blekherman): For fixed degree, the cone of nonnegative polynomials is much larger than that of s.o.s.

... BUT ... let $\|f\|_1 := \sum_\alpha |f_\alpha|$, for all $f \in \mathbf{R}[x]$. Then

Theorem (Berg): The cone of s.o.s. polynomials is dense (for the norm $\|.\|_1$) in the space of polynomials nonnegative on $[-1,1]^n$.

The next question is: *Given $f \geq 0$ on $[-1,1]^n$, can we find an explicit sequence of s.o.s. polynomials $\{f_\epsilon\}$ converging to f as $\epsilon \downarrow 0$?* That is $\|f_\epsilon - f\|_1 \to 0$.

Let $f \in \mathbf{R}[x] = \mathbf{R}[x_1, \ldots, x_n]$, and let $x \mapsto \Theta_r(x) := \sum_{i=1}^n x_i^{2r}$.

Theorem 1: Let $f \in \mathbf{R}[x]$ be a polynomial nonnegative on $[-1,1]^n$. Then for every $\epsilon > 0$, there exists $r(\epsilon) \in \mathbf{N}$ such that, for all $r \geq r(\epsilon)$, $f_{\epsilon r} := f + \epsilon\,\Theta_r$ is s.o.s., and for all $r \geq r(\epsilon)$, $\|f - f_{\epsilon r}\|_1 \to 0$ as $\epsilon \downarrow 0$.

- So one may approximate as closely as desired, any polynomial f nonnegative on $[-1,1]^n$, by a sequence $\{f_{\epsilon r}\}$ of s.o.s., by just adding essential monomials $\{x_i^{2r}\}$, with small coefficient ϵ.

- The s.o.s. approximation $\{f_{\epsilon r}\}$ is also uniform on $[-1,1]^n$.

- In addition, the s.o.s. $f_{\epsilon r} := f + \epsilon\,\Theta_r$ provides a certificate of nonnegativity of f on $[-1,1]^n$.

I. Polynomials nonnegative on the whole $\mathbb{R}^n$

Let $f \in \mathbf{R}[x] = \mathbf{R}[x_1, \ldots, x_n]$, and let $x \mapsto \theta_r(x) := \sum_{k=0}^{r} \sum_{i=1}^{n} \frac{x_i^{2k}}{k!}$.

Theorem 1: Let $f \in \mathbf{R}[x]$ be a nonnegative polynomial. Then for every $\epsilon > 0$, there exists $r(\epsilon) \in \mathbf{N}$ such that, for all $r \geq r(\epsilon)$, $f_{\epsilon r} := f + \epsilon\,\theta_r$ is s.o.s., and for all $r \geq r(\epsilon)$, $\|f - f_{\epsilon r}\|_1 \to 0$ as $\epsilon \downarrow 0$.

- So perturbating any nonnegative polynomial f to f_ϵ, by adding essential monomials $\{\frac{x_i^{2k}}{k!}\}$, with small associated coefficients ϵ, makes f_ϵ s.o.s., and close to f!

- The s.o.s. approximation $\{f_{\epsilon r}\}$ is uniform on compact sets!

- The s.o.s. $f_{\epsilon r}$ provides a certificate of nonnegativity of f.

II. SOS approximations of polynomials nonnegative on a real variety

Let $V \subset \mathbf{R}^n$ be the real algebraic set

$$V := \{x \in \mathbf{R}^n \mid \quad g_j(x) = 0, \quad j = 1, \ldots, m\},$$

for some family of real polynomials $\{g_j\} \subset \mathbf{R}[x]$.

Motivation: Provide a *certificate of positivity* for polynomials $f \in \mathbf{R}[x]$, nonnegative on V. In addition, and in view of the many potential applications, obtain if possible a representation that is also *useful* from a computational point of view.

Theorem 2: Let $f \in \mathbf{R}[x]$ be nonnegative on V, and let

$$f_{\epsilon r} = f + \epsilon\,\theta_r = f + \epsilon \sum_{k=0}^{r} \sum_{i=1}^{n} \frac{x_i^{2k}}{k!}, \qquad \epsilon \geq 0, \quad r \in \mathbf{N}.$$

(So, for every $r \in \mathbf{N}$, $\|f - f_{\epsilon r}\|_1 \to 0$ as $\epsilon \downarrow 0$.)

Then, for every $\epsilon > 0$, there exist nonnegative *scalars* $\{\lambda_j(\epsilon)\}_{j=1}^{m}$,

such that for all r sufficiently large (say $r \geq r(\epsilon)$),

$$f_{\epsilon r} = f + \epsilon\theta_r = q_\epsilon - \sum_{j=1}^{m} \lambda_j(\epsilon)\, g_j^2,$$

for some s.o.s. polynomial $q_\epsilon \in \mathbf{R}[x]$. In other words,

$$f + \epsilon\theta_r + \sum_{j=1}^{m} \lambda_j(\epsilon)\, g_j^2 \quad \text{is s.o.s.}$$

• The representation

$$f_{\epsilon r} = f + \epsilon\theta_r = q_\epsilon - \sum_{j=1}^{m} \lambda_j(\epsilon)\, g_j^2,$$

is an obvious certificate of nonnegativity of f on V as

$$f_{\epsilon r} \equiv q_\epsilon \text{ (s.o.s.)}, \quad \text{everywhere on } V \text{ and } \theta_r \text{ is bounded}$$

• Instead of a **certificate** for the approximation $f_{\epsilon o} = f + \epsilon\theta_o = f + \epsilon$ of f, as in Schmüdgen, Putinar, Jacobi and Prestel, Krivine, Vasilescu, ... one has a certificate for the approximation $f_{\epsilon r} = f + \epsilon\,\theta_r$.

• Notice that $\|f_{\epsilon o} - f\|_\infty \to 0$, whereas $\|f_{\epsilon r} - f\|_1 \to 0$ only. On the other hand, this latter s.o.s. representation holds with *no* assumption on the variety V, and the s.o.s. approximation is uniform on compact subsets of V.

Consequences: Simplified SDP-relaxations

Theorem: Let $V := \{x \in \mathbf{R}^n \mid \quad g_j(x) = 0, \quad j = 1, \ldots, m\}$, for some $\{g_j\} \subset \mathbf{R}[x]$. Assume that $\inf_{x \in V} f =: f^* > -\infty$, with $f^* = f(x^*)$ for some $x^* \in V$.

(i) Let $M > \|x^*\|_\infty$, and consider the SDP problem

$$\mathbf{Q}_r \begin{cases} \min_y & L_y(f) \\ \text{s.t.} & M_r(y) \succeq 0 \\ & \\ & L_y(\sum_{j=1}^m g_j^2) \leq 0; \quad L_y(\theta_r) \leq n e^{M^2}; \quad y_0 = 1. \end{cases}$$

- Then: $\inf \mathbf{Q}_r = \min \mathbf{Q}_r \uparrow f^*$, as $r \to \infty$.

- If $y^{(r)}$ is an optimal solution of $\mathbf{Q}_r$ then $y_1^{(r)} \to x^*$ if x^* is unique.

(ii) Given $\epsilon > 0$ fixed, let $f_{\epsilon r} := f + \epsilon\theta_r$, and consider the SDP problem

$$\mathbf{Q}_{\epsilon r} \left\{ \begin{array}{ll} \min\limits_{y} & L_y(f_{\epsilon r}) \\ \text{s.t.} & M_r(y) \succeq 0 \\ & \\ & L_y(\sum_{j=1}^m g_j^2) \leq 0; \quad y_0 = 1. \end{array} \right.$$

and its associated *dual* SDP problem $\mathbf{Q}^*_{\epsilon r}$. Then:

$$f^* \;\leq\; \sup \mathbf{Q}^*_{\epsilon r} \;\leq\; \inf \mathbf{Q}_{\epsilon r} \;\leq\; f(x^*) + \epsilon\theta_r(x^*) \;\leq\; f^* + \epsilon \sum_{i=1}^n \mathrm{e}^{(x^*)_i^2}$$

provided that r is sufficiently large.

(*) In all cases, each SDP-relaxation has a single LMI-constraint $M_r(y) \succeq 0$, and at most two linear equality/inequality.

(**) The LMI-constraint $M_r(y) \succeq 0$ does not depend on the problem data, and has a lot of structure, which could be exploited in a specialized SDP-solver

- the Generalized Problem of Moments (GPM)

- Some applications

- Duality between moments and nonnegative polynomials

- SDP-relaxations for the basic GPM

- s.o.s. vs nonnegative polynomials. Alternative SDP-relaxations

- How to handle sparsity

The no-free lunch rule ..... size of SDP-relaxations grows rapidly with the original problem size ... In particular:

- $O(n^{2r})$ variables for the r^{th} SDP-relaxation in the hierarchy
- $O(n^r)$ matrix size for the LMIs

$\longrightarrow$ In view of the present status of SDP-solvers ... only small to medium size problems can be solved by "standard" SDP-relaxations ...

$\longrightarrow$ .... How to handle larger size problems ?

• develop more efficient general purpose SDP-solvers ... (limited impact) ... or perhaps dedicated solvers ....?

• exploit symmetries when present ... Recent promising works by De Klerk, Gaterman, Gvozdenovic, Laurent, Pasechnick, Parrilo, Schrijver .. in particular for combinatorial optimization problems. Algebraic techniques permit to define an equivalent SDP of much smaller size.

• exploit sparsity in the data. In general, each constraint involves a small number of variables, and the cost criterion is a sum of polynomials involving also a small number of variables. Recent works by Kim, Kojima, Lasserre, Maramatsu and Waki

Basic idea : Let $I = \{1, 2, \ldots, n\}$ be the index set of the n variables.

Then $I = \bigcup_{j=1}^{p} I_j$ and each constraint $g_k(x) \geq 0$ only involves variables $\{x_i\}$ with $i \in I_l$ for some l.

Similarly, the cost function can be written $f = \sum_{j=1}^{p} f_j$ where each f_j involves variables $\{x_i\}$ with $i \in I_j$

A typical example: discrete-time dynamical systems

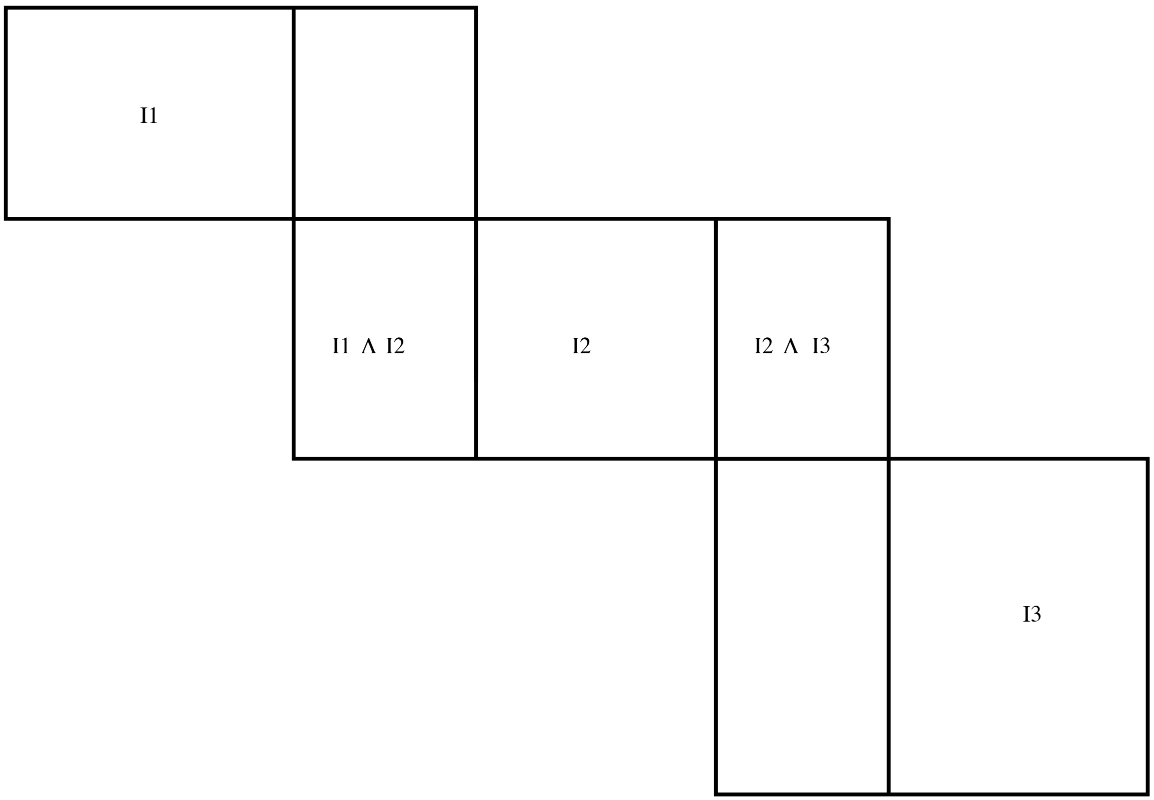
$$X_t = f(X_{t-1}, U_t), \quad t = 1, 2, \ldots T$$

with T blocks of variables (X_{t-1}, X_t, U_t), $t = 1, 2, \ldots T$.

- The coupling variables are the state-variables $\{X_t\}$.

- One usually has additional local constraints $g_t(X_{t-1}, U_t) \geq 0$.

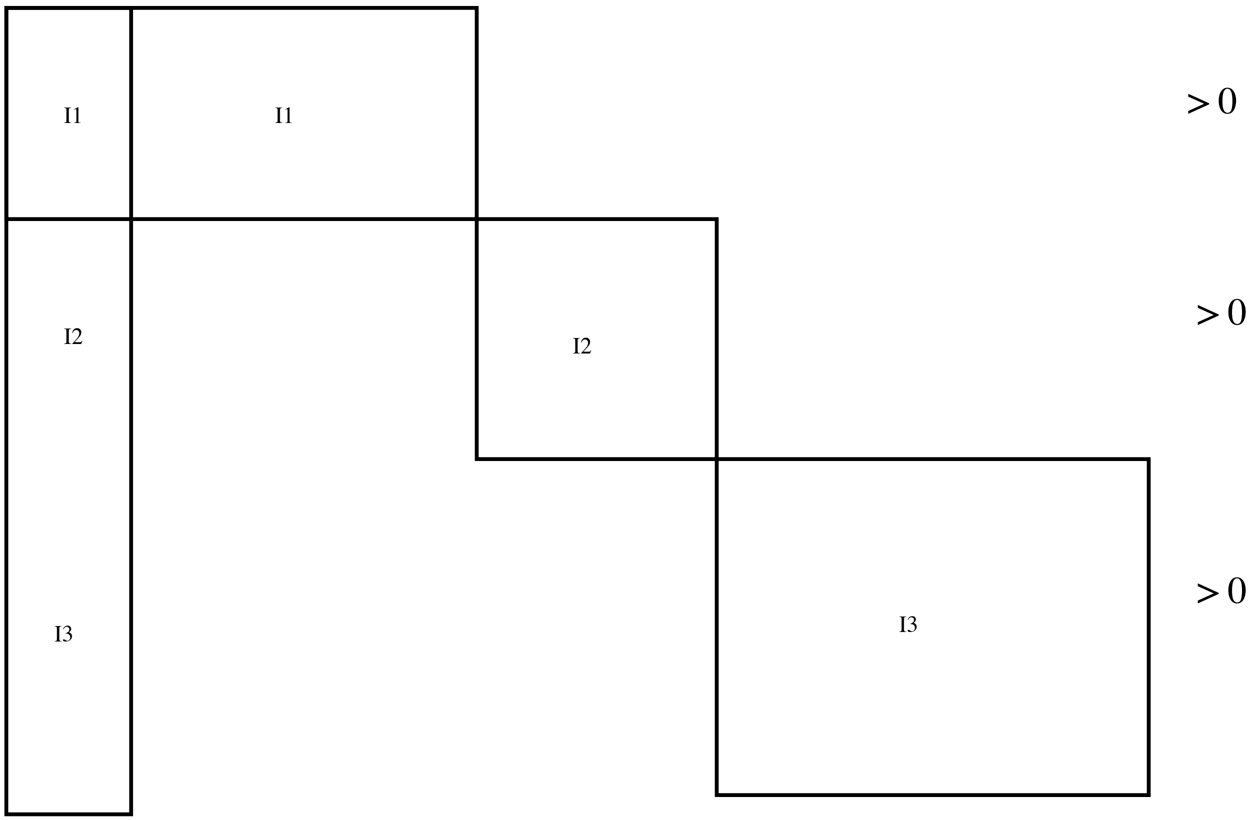
- The cost functional $f = \sum_{t=0}^{T} f_t(X_{t-1}, U_t) + H(X_T)$.



> 0

> 0

> 0


I1
I1
> 0
I2
I2
> 0
I3
I3
> 0

In recent works, Koijma's group has developed a systematic procedure to discover sparsity patterns $I = \bigcup_{j=1}^{p} I_j$.

Essentially one looks for maximal cliques $\{I_j\}$ in some chordal graph extension of a graph associated with the problem . Then:

1. One defines a set of moment variables, and a moment matrix for each set of variables I_j.

2. If constraint $g_k(x) \geq 0$ contains only variables $x_i \in I_j$ for some j, then the resulting localizing matrix w.r.t. g_k is defined only via the moments variables associated with I_j.

3. All moments associated with the vector of variables $\{x_i\}$ with $i \in I_j \cap I_k$, and expressed w.r.t. to I_j and I_k, are constrained to be equal.

The resulting r^{th} (sparse) SDP-relaxation has

- at most $p\ O(\kappa^{2r})$ variables and

- m LMIs of matrix size at most $O(\kappa^r)$

where $\kappa := \max_{j=1,\ldots,p} |I_j|$. So if $\kappa \approx n/p$ one has approximately

- $p\ (\frac{n}{p})^{2r}$ variables and m LMIs of matrix size at most $(\frac{n}{p})^r$

instead of n^{2r} and n^r respectively.

Theorem (Lasserre): if for every $k = 2, \ldots p$

$$\dagger \qquad I_k \quad \bigcap \quad (\bigcup_{j=1}^{k-1} I_j) \quad \subseteq \quad I_l \quad \text{for some } l \leq k-1,$$

then the sparse SDP-relaxations defined above converge.

• Interestingly, $\dagger$ is called the running intersection property in chordal graphs.

Examples with n large (say $n = 500$) and small κ (e.g. $\kappa = 3, 4$) are easily solved with Kojima's group software SparsePOP