# LECTURE 6

In the first 5 lectures, we have seen that:

- $\mathcal{P}(\mathbb{R}^n)^*_{2d}$: The dual of $\mathcal{P}(\mathbb{R}^n)$ is given by functionals $L_y$ such that $y$ has a representing measure
- $\Sigma^*_{n,2d}$: The dual of $\Sigma_n$ contains those functional $\ell$ such that the quadratic form $q \mapsto \ell(q^2)$ is positive semidefinite (moment matrix $\succcurlyeq 0$)
- $\mathcal{P}(\mathbb{R}^n)_{2d}$: It can be approximated as close as possible by projected spectrahedra (from inside) and from spectrahedra (from outside).

FOGLIO

Why is this useful? $\longrightarrow$ **RECAP**

## Univariate global polys optimization

$$f \in \mathbb{R}[x]_{2d} \quad (n=1), \quad f = f_0 + f_1 x + \dots + f_{2d} x^{2d}.$$

$$f^* = \inf_{x \in \mathbb{R}^n} f(x) = \begin{array}{l} \sup \lambda \\ \text{s.t. } f - \lambda \geq 0 \text{ on } \mathbb{R}^n \\ \quad\; f - \lambda \in \mathcal{P}(\mathbb{R}^n)_{2d} \end{array}$$

but by Hilbert's theorem $\mathcal{P}(\mathbb{R})_{2d} = \Sigma_{1,2d}$

$$\rightarrow \quad f^* = \begin{array}{l} \sup \lambda \\ \text{s.t. } f - \lambda \text{ is a sum of squares} \end{array}$$

We deduce that computing $f^*$ can be done

# LECTURE 6

In the first 5 lectures, we have seen that:

- $\mathcal{P}(\mathbb{R}^n)^*_{2d}$: The dual of $\mathcal{P}(\mathbb{R}^n)$ is given by functionals $\mathcal{L}_y$ such that $y$ has a representing measure
- $\Sigma^*_{n,2d}$: The dual of $\Sigma_n$ contains those functional $\ell$ such that the quadratic form $q \mapsto \ell(q^2)$ is positive semidefinite (moment matrix $\succeq 0$)
- $\mathcal{P}(\mathbb{R}^n)_{2d}$: It can be approximated as close as possible by projected spectrahedra (from inside) and from spectrahedra (from outside).

FOGLIO

Why is this useful? $\longrightarrow$ RECAP

Univariate global polys optimization

$$f \in \mathbb{R}[x]_{2d} \quad (n=1), \quad f = f_0 + f_1 x + \dots + f_{2d} x^{2d}.$$

$$f^* = \inf_{x \in \mathbb{R}^n} f(x) = \sup \lambda$$
$$\text{s.t. } f - \lambda \geq 0 \text{ on } \mathbb{R}^n$$
$$f - \lambda \in \mathcal{P}(\mathbb{R}^n)_{2d}$$

but by Hilbert's theorem $\mathcal{P}(\mathbb{R})_{2d} = \Sigma_{1,2d}$

$$\rightarrow \quad f^* = \sup \lambda \quad \text{s.t. } f - \lambda \text{ is a sum of squares}$$

We deduce that computing $f^*$ can be done solving just one SDP problem! Hence Convex!

__Example__ $f = x^4 + x^3 - 7x^2 - x + 6 = (x+3)(x+1)(x-1)(x-2)$

We look for $\lambda$ such that $f - \lambda$ is a sum of squares:

$$f - \lambda = (1\ x\ x^2)\begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{12} & x_{22} & x_{23} \\ x_{13} & x_{23} & x_{33} \end{pmatrix}\begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix}$$

We get the linear system + SD constraint:

$$X = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{12} & x_{22} & x_{23} \\ x_{13} & x_{23} & x_{33} \end{pmatrix} \succcurlyeq 0$$

5 linear equations
7 variables

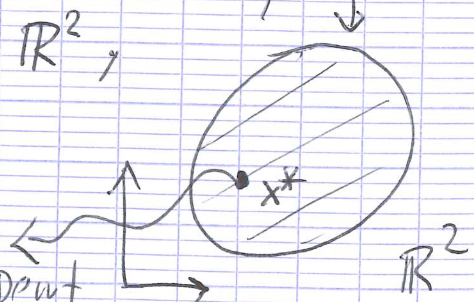
$$x_{33} = 1 \qquad 2x_{23} = 1 \qquad 2x_{13} + x_{22} = -7$$

$$2x_{12} = -1 \qquad x_{11} = 6 - \lambda$$

So we get the LMI

$$\begin{pmatrix} 6-\lambda & -\frac{1}{2} & x_{13} \\ -\frac{1}{2} & -7-2x_{13} & \frac{1}{2} \\ x_{13} & \frac{1}{2} & 1 \end{pmatrix} \succcurlyeq 0$$

which defines a convex semialgebraic set of $\mathbb{R}^2$, which is a spectrahedron.




every point $x^*$ corresponds to a SOS certificate for $f - \lambda$.

Using Matlab we get $f^* \approx$

**Exercise** (TO DO ~~know~~ NOW) Minimize $f(x) = x^2+1$ with the method of sums of squares / SDP.

**Exercise** Write the constraints of the previous example in the classical format of SDP:

$$X \succeq 0 \quad \langle A_i, X \rangle = b_i \quad .$$

What if $\Sigma_{n,2d} \subsetneq \mathcal{P}(\mathbb{R}^n)_{2d}$ ?

For instance, the Motzkin polynomial

$$M(x,y) \in \mathcal{P}(\mathbb{R}^2)_6 \setminus \Sigma_{2,6}$$

has also the property that

$$M(x,y) - \lambda \notin \Sigma_{2,6} \quad \forall \lambda \in \mathbb{R} \quad (\text{exercise})$$

**Multivariate**

In general we can define the "relaxation"

$$f^*_{sos} = \sup \lambda \quad \text{s.t. } f - \lambda \in \Sigma_{n,2d}$$

when $2d = \deg(f)$ (must be even, otherwise $f^* = -\infty$).

**Proposition** $f^*_{sos} \leq f^*$

Proof

Indeed, $\mathcal{P}(\mathbb{R}^n)_{2d} \supset \Sigma_{n,2d} \to \sup_{\mathcal{P}(\mathbb{R}^n)_{2d}} \geq \sup_{\Sigma_{n,2d}}$ $\square$

Hence we can compute lower bounds for (SAO) problem via sums of squares / SDP.

But $M(x,y) - \lambda \notin \Sigma_{2,6}$ $\forall \lambda$ !

In this case one could compute a Artin-representation, still via SDP:

$$M(x,y) - \lambda = \frac{f_1^2}{g_1^2} + \dots + \frac{f_r^2}{g_r^2}$$

but this is costly (size of SDP huge, no degree bounds for $g_i$).

<u>Univariate local case</u>

Using representation theorems (Polya-Szego) if $f - \lambda \in \mathcal{P}([0, +\infty))$ then

$$f - \lambda = \sigma_0 + \sigma_1 \cdot x \qquad \sigma_i \in \Sigma_{1,d}.$$

<u>Exercise</u> (TO DO NOW) Use POLYA-SZEGO representation theorem to compute the infimum of $f(x) = x^2 - 1$ over $[0, +\infty)$.

## Multivariate SAO ($n \geq 2$)

We have seen

$$f^* = \begin{array}{c} \text{Inf } f(x) \\ \text{s.t. } x \in S(g) \end{array} = \begin{array}{l} \text{Inf } \langle f, \mu \rangle \\ \text{s.t. } \langle 1, \mu \rangle = 1 \\ \quad \mu \geq 0 \\ \quad \text{supp}\,\mu \subset S(g) \end{array}$$

LP over a $\infty$-dimensional space

where $\langle f, \mu \rangle = \int_{\mathbb{R}^n} f \, d\mu$.

So (SAO) is an $\infty$-dimensional LP.

Its LP-dual is

$$\lambda^* = \begin{array}{l} \text{Sup } \lambda \\ \text{s.t. } f - \lambda \in \mathcal{P}(S(g)) \end{array}$$

We have also characterized (SAO) as

$$f^* = \begin{array}{l} \text{Inf } \mathcal{L}_y(f) = \sum f_\alpha y_\alpha \\ \text{s.t. } y \in \mathcal{P}(S(g))^*_d \\ \quad y_0 = 1 \end{array}$$

$d = \deg f$

sequences $y$ having a representing measure supported on $S(g)$

and the dual as well

$$\lambda^* = \begin{array}{l} \text{Sup } \lambda \\ \text{s.t. } f - \lambda \in \mathcal{P}(S(g))_d \end{array}$$

convex and finite-dimensional (only 1 variable!)

# Lasserre's LMI hierarchy

From now on, we suppose:

- $S(g) := \{ x \in \mathbb{R}^n : g_i(x) \geq 0, \ i = 1..m \}$, $g_0 := 1$
- $S(g)$ compact and $Q(g)$ archimedean

Remember by Putinar's theorem:

$y = (y_\alpha)$ has a representing measure with support in $S(g)$ if and only if

$$M(y) \succeq 0 \quad \text{and} \quad M(g_i y) \succeq 0 \quad i = 1 \ldots m$$

Hence the idea is the following:

In order to ~~xxxx~~ characterize the set $\mathcal{P}(S(g))$ we add infinitely many redundant constraints:

$$p \in \mathbb{R}[x]_d, \quad p = \underline{p}^T v_d(x) \to p^2 \underset{\text{over } S(g)}{\geq 0}$$

and also $p^2 \cdot g_i \underset{\text{over } S(g)}{\geq 0}$ (because $g_i \geq 0$ over $S(g)$)

Now remark:

$$p^2 = (\underline{p}^T v_d(x))(\underline{p}^T v_d(x)) = \underline{p}^T \left( v_d(x) v_d(x)^T \right) \underline{p}$$

$$\forall \underline{p} \in \mathbb{R}^{\binom{n+d}{d}} \quad \forall d$$

Similarly

$$p^2 g_i = \underline{p}^T \big( \overbrace{g_i(x)\, v_{d-d_j}(x)\, v_{d-d_j}(x)^T}^{\succcurlyeq 0} \big) \underline{p} \quad \text{with } d_j = \left\lceil \frac{\deg g_j}{2} \right\rceil$$

$$2 \deg p + \underbrace{\deg g_i}_{= 2d_j} \le d \qquad \forall \underline{p} \in \mathbb{R}$$

$$\deg p \le \frac{d - 2d_j}{2} = \frac{d}{2} - d_j$$

Hence we find the constraints $M_d(y) \succcurlyeq 0$

$$M_{d-d_j}(g_j y) \succcurlyeq 0.$$

__Definition__ The __Lasserre LMI hierarchy__ of semidefinite relaxations of

$$(SAO) \qquad f^* = \operatorname{Inf}_{\text{s.t. } x \in S(g)} f(x)$$

is given by

$$(L_d) \qquad \begin{aligned} \rho_d = \operatorname{Inf} \ & \sum f_\alpha y_\alpha \\ \text{s.t.} \ & y_0 = 1 \\ & M_d(y) \succcurlyeq 0 \\ & M_{d-d_j}(g_j y) \succcurlyeq 0, \ j = 1 \ldots m. \end{aligned}$$

$d \in \mathbb{N}$

__Proposition__ For all $d \in \mathbb{N}$, $\rho_d \le \rho_{d+1} \le f^*$.

**Proof**

If $\tilde{x} \in S(g)$ is a feasible solution, then $\tilde{y} = (\tilde{x}^{\alpha})_{\alpha \in \mathbb{N}^n_{2d}}$ is the vector of moments of the Dirac measure $\delta_{\tilde{x}}$, which is feasible for $(L_d)$. Hence

$$\mathcal{L}_{\tilde{y}}(f) = \int f \, d\delta_{\tilde{x}} = f(\tilde{x}) = f^*$$

hence $\quad \ell_d \leq \mathcal{L}_{\tilde{y}}(f) = f^*$.

Moreover, by the properties of positive semi-definite matrices, we have

$$M_{d+1}(y) \succeq 0 \implies M_d(y) \succeq 0$$

$$M_{d+1}(g_i y) \succeq 0 \implies M_d(gy) \succeq 0$$

$\rightarrow$ Feasible set$(L_{d+1}) \subset$ Feasible set$(L_d)$

$\rightarrow \quad \ell_d \leq \ell_{d+1} \ \forall d$ $\square$

**Remark** The $d$-th Lasserre LMI hierarchy $(L_d)$ is a SDP problem.

**Q** What is his dual SDP?

We come back to the notation:

$$M_d(y) = \sum_{|\alpha| \le 2d} y_\alpha C_\alpha^0$$

$$M_{d-d_j}(g_j y) = \sum_{\alpha \in \mathbb{N}^n_{2d}} y_\alpha C_\alpha^j \qquad j = 1..m$$

$$\left( ex: \begin{pmatrix} 1 & x \\ x & x^2 \end{pmatrix} = \underbrace{1}_{= y_0} \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}}_{= C_0^0} + \underbrace{x}_{= y_1} \underbrace{\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}}_{= C_1^0} + \underbrace{x^2}_{= y_2} \underbrace{\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}}_{= C_2^0} \right)$$

We apply the definition of dual SDP:

$$(L_d)^\vee \quad \begin{aligned} \rho_d^\vee = \; & \sup_{\lambda, X_j} \lambda \\ & \text{s.t. } f_\alpha - \cancel{\lambda}\,\lambda \cdot \mathbb{1}_{\alpha=0} = \sum_{j=0}^m \langle X_j, C_\alpha^j \rangle \quad \alpha \in \mathbb{N}^n_{2d} \\ & X_j \succeq 0 \quad \forall j = 1..m. \end{aligned}$$

$$\underline{f} - \begin{pmatrix} \lambda \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \vdots \\ \sum_{j=0}^m \langle X_j, C_\alpha^j \rangle \\ \vdots \end{pmatrix}_\alpha$$

$\underline{f}$ = vector of coefficients of $f$ in the basis $\langle x^\alpha \rangle$.

$$= \begin{pmatrix} \vdots \\ f_\alpha \\ \vdots \end{pmatrix}$$

Multiplying the constraints in $\rho_d^\vee$ by $x^\alpha$ and adding up:

$$\sum_\alpha f_\alpha x^\alpha - \lambda = \sum_{j=0}^m \langle X_j, \sum_\alpha x^\alpha C_\alpha^j \rangle$$

$$f - \lambda = \sum_{j=0}^m \langle X_j, g_j(x) \cdot v_{d-d_j}(x) \cdot v_{d-d_j}(x)^T \rangle$$

$$= \sum_{j=0}^m g_j(x) \langle X_j, v_{d-d_j}(x) v_{d-d_j}(x)^T \rangle$$

$$\langle X_j, v_{d-d_j} v_{d-d_j}^T \rangle = \mathrm{Trace}\,( X_j v_{d-d_j} v_{d-d_j}^T ) =$$

$$= \mathrm{Trace}\,( v_{d-d_j}^T X_j v_{d-d_j} ) =$$

$$= v_{d-d_j}^T(x) X_j v_{d-d_j}(x) = \sigma_j$$

$$\rightarrow \quad f - \lambda = \sigma_0 + \sigma_1 g_1 + \cdots + \sigma_m g_m$$

Putinar's representation.

If $X_j \succcurlyeq 0 \Rightarrow \sigma_j$ is SOS.

<u>Remark</u> The dual $(Ld)^\vee$ replaces the condition "$f - \lambda \in \mathcal{P}(S(g))$" of the LP-dual, with the stronger condition

$$f - \lambda = \sigma_0 + \sigma_1 g_1 + \cdots + \sigma_m g_m \, .$$

Convergence We know $\ell_d \leq \ell_{d+1} \leq f^*$

Q Under which assumptions $\ell_d \xrightarrow{d\to\infty} f^*$ ?

Theorem Suppose $Q(g)$ is archimedean :

(1) $\ell_d \xrightarrow{d\to\infty} f^*$

(2) If (SAO) has a unique global minimizer $x^* \in S(g)$, and if $y^{(d)}$ is feasible for (Ld), with $\mathcal{L}_{y^{(d)}}(f) \leq \ell_d + \frac{1}{d}$, then

$$\underbrace{\mathcal{L}_{y^{(d)}}(x_j)}_{\substack{\| \\ \text{moment of order } \alpha = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \leftarrow j}} \to x_j^* .$$

Proof

(1) $S(g)$ is compact (because $Q(g)$ arch.)

hence $f^* > -\infty$ and $f - (f^* - \varepsilon) > 0$ on $S(g)$

$$\to \quad f - (f^* - \varepsilon) = \sigma_0^{(\varepsilon)} + \sum \underbrace{\sigma_j^{(\varepsilon)}}_{\text{SOS}} g_j$$

So for $d$ large enough

$$(\underbrace{f^* - \varepsilon}_{\substack{\| \\ \lambda}}, \sigma_0^{(\varepsilon)}, \sigma_1^{(\varepsilon)}, \ldots, \sigma_m^{(\varepsilon)})$$

is a feasible point of $(L_d)^\vee$, hence $f^* - \varepsilon \leq \rho_d^\vee$.

By weak duality, one has $\rho_d^\vee \leq \rho_d$, and $\rho_d \leq f^*$.

Hence

$$f^* - \varepsilon \leq \rho_d^\vee \leq \rho_d \leq f^*$$

and $\varepsilon > 0$ is arbitrary $\Rightarrow$ $\rho_d \to f^*$.

(2) There exists a subsequence $y^{(d_n)}$ converging point-wise to an infinite sequence $y \in \mathbb{R}^\infty$, that is $\lim_{n\to\infty} y_\alpha^{(d_n)} = y_\alpha$ $\forall \alpha \in \mathbb{N}^n$. Then, $y$ has a representing measure $\mu$ with support $S(g)$, and value

$$L_y(f) = \lim_n L_{y^{(d_n)}}(f) \leq f^*.$$