# The moment-LP and moment-SOS approaches in optimization

Jean B. Lasserre

LAAS-CNRS and Institute of Mathematics, Toulouse, France

Workshop "Sequential Learning and Applications", Toulouse, November 2015

- Why polynomial optimization?
- LP- and SDP- CERTIFICATES of POSITIVITY
- The moment-LP and moment-SOS approaches

- Why polynomial optimization?
- LP- and SDP- CERTIFICATES of POSITIVITY
- The moment-LP and moment-SOS approaches

- Why polynomial optimization?
- LP- and SDP- CERTIFICATES of POSITIVITY
- The moment-LP and moment-SOS approaches

Consider the polynomial optimization problem:

$$\mathbf{P}: \quad f^* = \min\{ f(\mathbf{x}) : \quad g_j(\mathbf{x}) \geq 0, \; j = 1, \ldots, m \}$$

for some polynomials $f, g_j \in \mathbb{R}[\mathbf{x}]$.

**Why Polynomial Optimization?**

After all ... **P** is just a particular case of Non Linear Programming (NLP)!

Consider the polynomial optimization problem:

$$\mathbf{P}: \quad f^* = \min\{\, f(\mathbf{x}) : \quad g_j(\mathbf{x}) \geq 0,\ j = 1, \ldots, m \,\}$$

for some polynomials $f, g_j \in \mathbb{R}[\mathbf{x}]$.

**Why Polynomial Optimization?**

After all ... **P** is just a particular case of Non Linear Programming (NLP)!

## True!

... if one is interested with a LOCAL optimum only!!

## When searching for a local minimum ...

Optimality conditions and descent algorithms use basic tools from REAL and CONVEX analysis and linear algebra

The focus is on how to improve $f$ by looking at a NEIGHBORHOOD of a nominal point $\mathbf{x} \in \mathbf{K}$, i.e., LOCALLY AROUND $\mathbf{x} \in \mathbf{K}$, and in general, no GLOBAL property of $\mathbf{x} \in \mathbf{K}$ can be inferred.

The fact that $f$ and $g_j$ are POLYNOMIALS does not help much!

## True!

... if one is interested with a LOCAL optimum only!!

## When searching for a local minimum ...

Optimality conditions and descent algorithms use basic tools from REAL and CONVEX analysis and linear algebra

The focus is on how to improve $f$ by looking at a NEIGHBORHOOD of a nominal point $\mathbf{x} \in \mathbf{K}$, i.e., LOCALLY AROUND $\mathbf{x} \in \mathbf{K}$, and in general, no GLOBAL property of $\mathbf{x} \in \mathbf{K}$ can be inferred.

The fact that $f$ and $g_j$ are POLYNOMIALS does not help much!

## True!

... if one is interested with a LOCAL optimum only!!

## When searching for a local minimum ...

Optimality conditions and descent algorithms use basic tools from REAL and CONVEX analysis and linear algebra

The focus is on how to improve $f$ by looking at a NEIGHBORHOOD of a nominal point $\mathbf{x} \in \mathbf{K}$, i.e., LOCALLY AROUND $\mathbf{x} \in \mathbf{K}$, and in general, no GLOBAL property of $\mathbf{x} \in \mathbf{K}$ can be inferred.

The fact that $f$ and $g_j$ are POLYNOMIALS does not help much!

## True!

... if one is interested with a LOCAL optimum only!!

## When searching for a local minimum ...

Optimality conditions and descent algorithms use basic tools from REAL and CONVEX analysis and linear algebra

The focus is on how to improve $f$ by looking at a NEIGHBORHOOD of a nominal point $\mathbf{x} \in \mathbf{K}$, i.e., LOCALLY AROUND $\mathbf{x} \in \mathbf{K}$, and in general, no GLOBAL property of $\mathbf{x} \in \mathbf{K}$ can be inferred.

The fact that $f$ and $g_j$ are POLYNOMIALS does not help much!

# BUT for GLOBAL Optimization

## ... the picture is different!

**Remember that for the GLOBAL minimum $f^*$:**

$$f^* = \sup \{ \lambda : f(\mathbf{x}) - \lambda \geq 0 \quad \forall \mathbf{x} \in \mathbf{K} \}.$$

(Not true for a global minimum!))

and so to compute $f^*$ ...
one needs to handle EFFICIENTLY the difficult constraint

$$f(\mathbf{x}) - \lambda \geq 0 \quad \forall \mathbf{x} \in \mathbf{K},$$

i.e. one needs
TRACTABLE CERTIFICATES of POSITIVITY on $\mathbf{K}$
for the polynomial $\mathbf{x} \mapsto f(\mathbf{x}) - \lambda$!

## BUT for GLOBAL Optimization

... the picture is different!

**Remember that for the GLOBAL minimum $f^*$:**

$$f^* = \sup \{ \lambda : f(\mathbf{x}) - \lambda \geq 0 \quad \forall \mathbf{x} \in \mathbf{K} \}.$$

(Not true for a global minimum!))

and so to compute $f^*$ ...
one needs to handle EFFICIENTLY the difficult constraint

$$f(\mathbf{x}) - \lambda \geq 0 \quad \forall \mathbf{x} \in \mathbf{K},$$

i.e. one needs
TRACTABLE CERTIFICATES of POSITIVITY on $\mathbf{K}$
for the polynomial $\mathbf{x} \mapsto f(\mathbf{x}) - \lambda$!

## BUT for GLOBAL Optimization

... the picture is different!

## Remember that for the GLOBAL minimum $f^*$:

$$f^* = \sup\{\,\lambda : \; f(\mathbf{x}) - \lambda \geq 0 \quad \forall \mathbf{x} \in \mathbf{K}\}.$$

(Not true for a global minimum!))

and so to compute $f^*$ ...
one needs to handle EFFICIENTLY the difficult constraint

$$f(\mathbf{x}) - \lambda \geq 0 \quad \forall \mathbf{x} \in \mathbf{K},$$

i.e. one needs
TRACTABLE CERTIFICATES of POSITIVITY on $\mathbf{K}$
for the polynomial $\mathbf{x} \mapsto f(\mathbf{x}) - \lambda$!

**REAL ALGEBRAIC GEOMETRY helps!!!!**

Indeed, POWERFUL CERTIFICATES OF POSITIVITY EXIST!

**Moreover .... and importantly,**

Such certificates are amenable to PRACTICAL COMPUTATION!

($\star$ Stronger Positivstellensatzë exist for analytic functions but are useless from a computational viewpoint.)

## REAL ALGEBRAIC GEOMETRY helps!!!!

Indeed, POWERFUL CERTIFICATES OF POSITIVITY EXIST!

## Moreover .... and importantly,

Such certificates are amenable to PRACTICAL COMPUTATION!

($\star$ Stronger Positivstellensatzë exist for analytic functions but are useless from a computational viewpoint.)

### REAL ALGEBRAIC GEOMETRY helps!!!!

Indeed, POWERFUL CERTIFICATES OF POSITIVITY EXIST!

### Moreover .... and importantly,

Such certificates are amenable to PRACTICAL COMPUTATION!

($\star$ Stronger Positivstellensatzë exist for analytic functions but are useless from a computational viewpoint.)

# SOS-based certificate

$$\mathbf{K} = \{ \mathbf{x} : \; g_j(\mathbf{x}) \geq 0, \quad j = 1, \ldots, m \}$$

**Theorem (Putinar's Positivstellensatz)**

*If $\mathbf{K}$ is compact (+ a technical Archimedean assumption) and $f > 0$ on $\mathbf{K}$ then:*

$$\dagger \quad f(\mathbf{x}) = \sigma_0(\mathbf{x}) + \sum_{j=1}^{m} \sigma_j(\mathbf{x})\, g_j(\mathbf{x}), \qquad \forall \mathbf{x} \in \mathbb{R}^n,$$

*for some SOS polynomials $(\sigma_j) \subset \mathbb{R}[\mathbf{x}]$.*

## However ... In Putinar's theorem

... nothing is said on the DEGREE of the SOS polynomials $(\sigma_j)$!

## BUT ... GOOD news ..!!

Testing whether † holds
for some SOS $(\sigma_j) \subset \mathbb{R}[\mathbf{x}]$ with a degree bound,

is SOLVING an SDP!

## However ... In Putinar's theorem

... nothing is said on the DEGREE of the SOS polynomials $(\sigma_j)$!

## BUT ... GOOD news ..!!

Testing whether † holds
for some SOS $(\sigma_j) \subset \mathbb{R}[\mathbf{x}]$ with a degree bound,

is SOLVING an SDP!

The CONVEX optimization problem:

$$\mathbf{P} \quad \rightarrow \quad \min_{x \in \mathbb{R}^n} \{ c' x \mid \sum_{i=1}^{n} A_i x_i \succeq b \},$$

where $c \in \mathbb{R}^n$ and $b, A_i \in \mathcal{S}_m$ ($m \times m$ symmetric matrices), is called a semidefinite program.

The notation "$\cdot \succeq 0$" means the real symmetric matrix "$\cdot$" is positive semidefinite, i.e., all its (real) EIGENVALUES are nonnegative.

# Example

$$\mathbf{P}: \quad \min_{x} \quad \{x_1 + x_2 \ : \\ \text{s.t.} \quad \begin{bmatrix} 3 + 2x_1 + x_2 & x_1 - 5 \\ x_1 - 5 & x_1 - 2x_2 \end{bmatrix} \succeq 0 \Big\} ,$$

or, equivalently

$$\mathbf{P}: \min_{x} \quad \{x_1 + x_2 \ : \\ \text{s.t.} \quad \begin{bmatrix} 3 & -5 \\ -5 & 0 \end{bmatrix} + x_1 \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} + x_2 \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix} \succeq 0 \Big\}$$

**P** and its dual **P**$^*$ are **convex** problems that are solvable in polynomial time to arbitrary precision $\epsilon > 0$.

= generalization to the convex cone $\mathcal{S}_m^+$ ($X \succeq 0$) of Linear Programming on the convex polyhedral cone $\mathbb{R}_+^m$ ($x \geq 0$).

**Indeed, with DIAGONAL matrices**

Semidefinite programming = Linear Programming!

Several academic SDP software packages exist, (e.g. MATLAB "LMI toolbox", SeduMi, SDPT3, ...). However, so far, size limitation is more severe than for LP software packages. Pioneer contributions by A. Nemirovsky, Y. Nesterov, N.Z. Shor, B.D. Yudin,...

**P** and its dual **P**$^*$ are **convex** problems that are solvable in polynomial time to arbitrary precision $\epsilon > 0$.
= generalization to the convex cone $\mathcal{S}_m^+$ ($X \succeq 0$) of Linear Programming on the convex polyhedral cone $\mathbb{R}_+^m$ ($x \geq 0$).

**Indeed, with DIAGONAL matrices**

Semidefinite programming = Linear Programming!

Several academic SDP software packages exist, (e.g. MATLAB "LMI toolbox", SeduMi, SDPT3, ...). However, so far, size limitation is more severe than for LP software packages.
Pioneer contributions by A. Nemirovsky, Y. Nesterov, N.Z. Shor, B.D. Yudin,...

**P** and its dual **P**$^*$ are **convex** problems that are solvable in polynomial time to arbitrary precision $\epsilon > 0$.
= generalization to the convex cone $\mathcal{S}_m^+$ ($X \succeq 0$) of Linear Programming on the convex polyhedral cone $\mathbb{R}_+^m$ ($x \geq 0$).

**Indeed, with DIAGONAL matrices**

Semidefinite programming = Linear Programming!

Several academic SDP software packages exist, (e.g. MATLAB "LMI toolbox", SeduMi, SDPT3, ...). However, so far, size limitation is more severe than for LP software packages.
Pioneer contributions by A. Nemirovsky, Y. Nesterov, N.Z. Shor, B.D. Yudin,...

# Dual side of Putinar's theorem: The $K$-moment problem

Given a real sequence $\mathbf{y} = (y_\alpha)$, $\alpha \in \mathbb{N}^n$, does there exist a Borel measure $\mu$ on $\mathbf{K}$ such that

$$\dagger \quad y_\alpha = \int_{\mathbf{K}} x_1^{\alpha_1} \cdots x_n^{\alpha_n} \, d\mu, \qquad \forall \alpha \in \mathbb{N}^n.$$

Introduce the so-called Riesz linear functional $L_{\mathbf{y}} : \mathbb{R}[\mathbf{x}] \to \mathbb{R}$ :

$$f \left( = \sum_\alpha f_\alpha \, \mathbf{x}^\alpha \right) \mapsto L_{\mathbf{y}}(f) = \sum_{\alpha \in \mathbb{N}^n} f_\alpha \, y_\alpha$$

# Dual side of Putinar's theorem: The $K$-moment problem

Given a real sequence $\mathbf{y} = (y_\alpha)$, $\alpha \in \mathbb{N}^n$, does there exist a Borel measure $\mu$ on $\mathbf{K}$ such that

$$\dagger \quad y_\alpha = \int_{\mathbf{K}} x_1^{\alpha_1} \cdots x_n^{\alpha_n} \, d\mu, \qquad \forall \alpha \in \mathbb{N}^n.$$

Introduce the so-called Riesz linear functional $L_{\mathbf{y}} : \mathbb{R}[\mathbf{x}] \to \mathbb{R}$ :

$$f \left( = \sum_\alpha f_\alpha \, \mathbf{x}^\alpha \right) \mapsto L_{\mathbf{y}}(f) = \sum_{\alpha \in \mathbb{N}^n} f_\alpha \, y_\alpha$$

**Theorem**

*If $\mathbf{K} = \{\mathbf{x} : g_j(\mathbf{x}) \geq 0,\ j = 1, \ldots, m\}$ is compact and satisfies an Archimedean assumption then † holds if and only if for every $h \in \mathbb{R}[\mathbf{x}]^2$:*

$$(\star) \qquad L_y(h^2) \,\geq\, 0; \quad L_y(h^2\, g_j) \,\geq\, 0, \quad j = 1, \ldots, m.$$

The condition $(\star)$ is equivalent to $m+1$ positive semidefiniteness of some moment and localizing matrices, i.e.,

$$\mathbf{M}(\mathbf{y}) \,\succeq\, 0; \quad \mathbf{M}(g_j\,\mathbf{y}) \,\succeq\, 0, \quad j = 1, \ldots, m.$$

whose rows & columns are indexed by $\mathbb{N}^n$, and entries are LINEAR in the $y_\alpha$'s

## Theorem

*If* $\mathbf{K} = \{\mathbf{x} : g_j(\mathbf{x}) \geq 0,\ j = 1, \ldots, m\}$ *is compact and satisfies an Archimedean assumption then* † *holds if and only if for every* $h \in \mathbb{R}[\mathbf{x}]^2$*:*

$$(\star) \qquad L_y(h^2) \geq 0; \quad L_y(h^2\, g_j) \geq 0, \quad j = 1, \ldots, m.$$

The condition $(\star)$ is equivalent to $m+1$ positive semidefiniteness of some moment and localizing matrices, i.e.,

$$\mathbf{M}(\mathbf{y}) \succeq 0; \quad \mathbf{M}(g_j\, \mathbf{y}) \succeq 0, \quad j = 1, \ldots, m.$$

whose rows & columns are indexed by $\mathbb{N}^n$, and entries are LINEAR in the $y_\alpha$'s

# LP-based certificate

$$\mathbf{K} = \{\mathbf{x} : \; g_j(\mathbf{x}) \;\geq\; 0; \; (1 - g_j(\mathbf{x})) \geq 0, \quad j = 1, \ldots, m\}$$

**Theorem (Krivine-Vasilescu-Handelman's Positivstellensatz)**

*Let* $\mathbf{K}$ *be compact and the family* $\{1, g_j\}$ *generate* $\mathbb{R}[\mathbf{x}]$*. If* $f > 0$ *on* $\mathbf{K}$ *then:*

$$(\star) \quad f(\mathbf{x}) \;=\; \sum_{\alpha,\beta} c_{\alpha\beta} \prod_{j=1}^{m} g_j(\mathbf{x})^{\alpha_j} \, (1 - g_j(\mathbf{x}))^{\beta_j}, , \qquad \forall \mathbf{x} \in \mathbb{R}^n,$$

*for some NONNEGATIVE scalars* $(c_{\alpha\beta})$*.*

## However ... Again in Krivine's theorem

In $(\star)$ ... nothing is said on
how many nonnegative scalars $c_{\alpha\beta}$ are needed!

## BUT ... GOOD news ... again!!

Testing whether $(\star)$ holds
for some nonnegative scalars $(c_{\alpha\beta})$

is SOLVING an LP!

## However ... Again in Krivine's theorem

In $(\star)$ ... nothing is said on
how many nonnegative scalars $c_{\alpha\beta}$ are needed!

## BUT ... GOOD news ... again!!

Testing whether $(\star)$ holds
for some nonnegative scalars $(c_{\alpha\beta})$

is SOLVING an LP!

# Dual side of Krivine's theorem: The $K$-moment problem

**Theorem**

*If* $\mathbf{K} = \{\mathbf{x} : g_j(\mathbf{x}) \geq 0,\ j = 1, \ldots, m\}$ *is compact,* $0 \leq g_j \leq 1$ *on* $\mathbf{K}$*, and* $\{1, g_j\}$ *generates* $\mathbb{R}[\mathbf{x}]$*, then* $\dagger$ *holds if and only if*

$$(\star\star) \qquad L_y\left(\prod_{j=1}^{m} g_j^{\alpha_j}\,(1 - g_j^{\beta_j})\right) \geq 0, \qquad \forall \alpha, \beta \in \mathbb{N}^m.$$

The condition $(\star\star)$ is equivalent to countably many LINEAR INEQUALITIES on the $y_\alpha$'s

# Dual side of Krivine's theorem: The $K$-moment problem

**Theorem**

*If* $\mathbf{K} = \{\mathbf{x} : g_j(\mathbf{x}) \geq 0,\ j = 1, \ldots, m\}$ *is compact,* $0 \leq g_j \leq 1$ *on* $\mathbf{K}$*, and* $\{1, g_j\}$ *generates* $\mathbb{R}[\mathbf{x}]$*, then* † *holds if and only if*

$$(\star\star) \qquad L_y\left(\prod_{j=1}^{m} g_j^{\alpha_j}\,(1 - g_j)^{\beta_j}\right) \geq 0, \qquad \forall \alpha, \beta \in \mathbb{N}^m.$$

The condition $(\star\star)$ is equivalent to countably many LINEAR INEQUALITIES on the $y_\alpha$'s

## HENCE ... SUCH POSITIVITY CERTIFICATES

allow to infer GLOBAL Properties of

FEASIBILITY and OPTIMALITY,

## ... the analogue of (well-known) previous ones

valid in the CONVEX CASE ONLY!

## HENCE ... SUCH POSITIVITY CERTIFICATES

allow to infer GLOBAL Properties of

FEASIBILITY and OPTIMALITY,

## ... the analogue of (well-known) previous ones

valid in the CONVEX CASE ONLY!

• In addition, polynomials NONNEGATIVE ON A SET $\mathbf{K} \subset \mathbb{R}^n$ are ubiquitous. They also appear in many important applications (outside optimization),

**. . . modeled as**

particular instances of the so called
Generalized Moment Problem, among which:
Probability, Optimal and Robust Control, Game theory, Signal processing, multivariate integration, etc.

$$(GMP): \quad \inf_{\mu_i \in M(\mathbf{K}_i)} \{ \sum_{i=1}^{s} \int_{\mathbf{K}_i} f_i \, d\mu_i : \sum_{i=1}^{s} \int_{\mathbf{K}_i} h_{ij} \, d\mu_i \gtreqqless b_j, \quad j \in J \}$$

with $M(\mathbf{K}_i)$ space of Borel measures on $\mathbf{K}_i \subset \mathbb{R}^{n_i}$, $i = 1, \ldots, s$.

• In addition, polynomials NONNEGATIVE ON A SET $\mathbf{K} \subset \mathbb{R}^n$ are ubiquitous. They also appear in many important applications (outside optimization),

**. . . modeled as**

particular instances of the so called Generalized Moment Problem, among which: Probability, Optimal and Robust Control, Game theory, Signal processing, multivariate integration, etc.

$$(GMP): \quad \inf_{\mu_i \in M(\mathbf{K}_i)} \left\{ \sum_{i=1}^{s} \int_{\mathbf{K}_i} f_i \, d\mu_i : \sum_{i=1}^{s} \int_{\mathbf{K}_i} h_{ij} \, d\mu_i \gtreqqless b_j, \quad j \in J \right\}$$

with $M(\mathbf{K}_i)$ space of Borel measures on $\mathbf{K}_i \subset \mathbb{R}^{n_i}$, $i = 1, \ldots, s$.

The DUAL of the GMP is the linear program $GMP^*$:

$$\sup_{\lambda_j} \left\{ \sum_{j\in J}^{s} \lambda_j\, b_j : \quad f_i - \sum_{j\in J} \lambda_j\, h_{ij} \geq 0 \text{ on } \mathbf{K}_i, \quad i = 1, \ldots, s \right\}$$

**And one can see that ...**

the constraints of $GMP^*$ state that

some functions $f_i - \sum_{j\in J} \lambda_j\, h_{ij}$

must be nonnegative on a certain set $\mathbf{K}_i$, $i = 1, \ldots, s$.

# A couple of examples

I: Global OPTIM $\quad \rightarrow f^* = \inf_{\mathbf{x}} \{ f(\mathbf{x}) : \mathbf{x} \in \mathbf{K} \}$

is the SIMPLEST example of the GMP

**because ...**

$$f^* = \inf_{\mu \in M(\mathbf{K})} \{ \int_{\mathbf{K}} f \, d\mu : \int_{\mathbf{K}} 1 \, d\mu = 1 \}$$

• Indeed if $f(\mathbf{x}) \geq f^*$ for all $\mathbf{x} \in \mathbf{K}$ and $\mu$ is a probability measure on $\mathbf{K}$, then $\int_{\mathbf{K}} f \, d\mu \geq \int f^* \, d\mu = f^*$.

• On the other hand, for every $\mathbf{x} \in \mathbf{K}$ the probability measure $\mu := \delta_{\mathbf{x}}$ is such that $\int f \, d\mu = f(\mathbf{x})$.

# A couple of examples

I: Global OPTIM $\rightarrow f^* = \inf_{\mathbf{x}} \{ f(\mathbf{x}) : \mathbf{x} \in \mathbf{K} \}$

is the SIMPLEST example of the GMP

**because ...**

$$f^* = \inf_{\mu \in M(\mathbf{K})} \{ \int_{\mathbf{K}} f \, d\mu : \int_{\mathbf{K}} 1 \, d\mu = 1 \}$$

- Indeed if $f(\mathbf{x}) \geq f^*$ for all $\mathbf{x} \in \mathbf{K}$ and $\mu$ is a probability measure on $\mathbf{K}$, then $\int_{\mathbf{K}} f \, d\mu \geq \int f^* \, d\mu = f^*$.
- On the other hand, for every $\mathbf{x} \in \mathbf{K}$ the probability measure $\mu := \delta_{\mathbf{x}}$ is such that $\int f \, d\mu = f(\mathbf{x})$.

# A couple of examples

I: Global OPTIM $\rightarrow f^* = \inf_{\mathbf{x}} \{ f(\mathbf{x}) : \mathbf{x} \in \mathbf{K} \}$

is the SIMPLEST example of the GMP

**because ...**

$$f^* = \inf_{\mu \in M(\mathbf{K})} \{ \int_{\mathbf{K}} f \, d\mu : \int_{\mathbf{K}} 1 \, d\mu = 1 \}$$

- Indeed if $f(\mathbf{x}) \geq f^*$ for all $\mathbf{x} \in \mathbf{K}$ and $\mu$ is a probability measure on $\mathbf{K}$, then $\int_{\mathbf{K}} f \, d\mu \geq \int f^* \, d\mu = f^*$.
- On the other hand, for every $\mathbf{x} \in \mathbf{K}$ the probability measure $\mu := \delta_{\mathbf{x}}$ is such that $\int f \, d\mu = f(\mathbf{x})$.

# A couple of examples

I: Global OPTIM $\rightarrow f^* = \inf_{\mathbf{x}} \{ f(\mathbf{x}) : \mathbf{x} \in \mathbf{K} \}$

is the SIMPLEST example of the GMP

**because ...**

$$f^* = \inf_{\mu \in M(\mathbf{K})} \{ \int_{\mathbf{K}} f \, d\mu : \int_{\mathbf{K}} 1 \, d\mu = 1 \}$$

• Indeed if $f(\mathbf{x}) \geq f^*$ for all $\mathbf{x} \in \mathbf{K}$ and $\mu$ is a probability measure on $\mathbf{K}$, then $\int_{\mathbf{K}} f \, d\mu \geq \int f^* \, d\mu = f^*$.

• On the other hand, for every $\mathbf{x} \in \mathbf{K}$ the probability measure $\mu := \delta_{\mathbf{x}}$ is such that $\int f \, d\mu = f(\mathbf{x})$.

II. Let $\mathbf{K} \subset \mathbb{R}^n$ and $S \subset \mathbf{K}$ be given, and let $\Gamma \subset \mathbb{N}^n$ be also given.

**BOUNDS on measures with moment conditions**

$$\max_{\mu \in M(\mathbf{K})} \left\{ \langle 1_S, \mu \rangle \; : \; \int_{\mathbf{K}} x^\alpha \, d\mu \; = \; m_\alpha, \quad \alpha \in \Gamma \right\}$$

to compute an upper bound on $\mu(S)$ over all distributions $\mu \in M(\mathbf{K})$ with a certain fixed number of moments $m_\alpha$.

• If $\Gamma = \mathbb{N}^n$ then one may use this to compute the Lebesgue volume of a compact basic semi-algebraic set $S \subset \mathbf{K} := [-1, 1]^n$.

$$\text{Take } m_\alpha := \int_{[-1,1]^n} \mathbf{x}^\alpha \, d\mathbf{x}, \qquad \alpha \in \mathbb{N}^n.$$

II. Let $\mathbf{K} \subset \mathbb{R}^n$ and $S \subset \mathbf{K}$ be given, and let $\Gamma \subset \mathbb{N}^n$ be also given.

**BOUNDS on measures with moment conditions**

$$\max_{\mu \in M(\mathbf{K})} \left\{ \langle 1_S, \mu \rangle \; : \; \int_{\mathbf{K}} x^\alpha \, d\mu \; = \; m_\alpha, \quad \alpha \in \Gamma \right\}$$

to compute an upper bound on $\mu(S)$ over all distributions $\mu \in M(\mathbf{K})$ with a certain fixed number of moments $m_\alpha$.

• If $\Gamma = \mathbb{N}^n$ then one may use this to compute the Lebesgue volume of a compact basic semi-algebraic set $S \subset \mathbf{K} := [-1, 1]^n$.

$$\text{Take } m_\alpha := \int_{[-1,1]^n} \mathbf{x}^\alpha \, d\mathbf{x}, \qquad \alpha \in \mathbb{N}^n.$$

III. For instance, one may also want:

- To approximate sets defined with QUANTIFIERS, like .e.g.,

$$R_f := \{x \in \mathbf{B} \quad : \quad f(x,y) \leq 0 \text{ for all } y \text{ such that } (x,y) \in \mathbf{K}\}$$

$$D_f := \{x \in \mathbf{B} \quad : \quad f(x,y) \leq 0 \text{ for some } y \text{ such that } (x,y) \in \mathbf{K}\}$$

where $f \in \mathbb{R}[x,y]$, **B** is a simple set (box, ellipsoid).

- To compute convex polynomial underestimators $p \leq f$ of a polynomial $f$ on a box $\mathbf{B} \subset \mathbb{R}^n$. (Very useful in MINLP.)

III. For instance, one may also want:

• To approximate sets defined with QUANTIFIERS, like .e.g.,

$$R_f := \{x \in \mathbf{B} \quad : \quad f(x,y) \leq 0 \text{ for all } y \text{ such that } (x,y) \in \mathbf{K}\}$$

$$D_f := \{x \in \mathbf{B} \quad : \quad f(x,y) \leq 0 \text{ for some } y \text{ such that } (x,y) \in \mathbf{K}\}$$

where $f \in \mathbb{R}[x,y]$, $\mathbf{B}$ is a simple set (box, ellipsoid).

• To compute convex polynomial underestimators $p \leq f$ of a polynomial $f$ on a box $\mathbf{B} \subset \mathbb{R}^n$. (Very useful in MINLP.)

## The moment-LP and moment-SOS approaches

consist of using a certain type of positivity certificate (Krivine-Vasilescu-Handelman's or Putinar's certificate) in potentially any application where such a characterization is needed. (Global optimization is only one example.)

## In many situations this amounts to

solving a HIERARCHY of :

- LINEAR PROGRAMS, or
- SEMIDEFINITE PROGRAMS

... of increasing size!.

## The moment-LP and moment-SOS approaches

consist of using a certain type of positivity certificate (Krivine-Vasilescu-Handelman's or Putinar's certificate) in potentially any application where such a characterization is needed. (Global optimization is only one example.)

## In many situations this amounts to

solving a HIERARCHY of :

- LINEAR PROGRAMS, or
- SEMIDEFINITE PROGRAMS

... of increasing size!.

## The moment-LP and moment-SOS approaches

consist of using a certain type of positivity certificate (Krivine-Vasilescu-Handelman's or Putinar's certificate) in potentially any application where such a characterization is needed. (Global optimization is only one example.)

## In many situations this amounts to

solving a HIERARCHY of :

- LINEAR PROGRAMS, or
- SEMIDEFINITE PROGRAMS

... of increasing size!.

## The moment-LP and moment-SOS approaches

consist of using a certain type of positivity certificate (Krivine-Vasilescu-Handelman's or Putinar's certificate) in potentially any application where such a characterization is needed. (Global optimization is only one example.)

## In many situations this amounts to

solving a HIERARCHY of :

- LINEAR PROGRAMS, or
- SEMIDEFINITE PROGRAMS

... of increasing size!.

# LP- and SDP-hierarchies for optimization

Replace $f^* = \sup_{\lambda,\sigma_j} \{ \lambda : f(\mathbf{x}) - \lambda \geq 0 \quad \forall \mathbf{x} \in \mathbf{K} \}$ with:

**The SDP-hierarchy indexed by $d \in \mathbb{N}$:**

$$f_d^* = \sup \{ \lambda : f - \lambda = \underbrace{\sigma_0}_{SOS} + \sum_{j=1}^m \underbrace{\sigma_j}_{SOS} g_j; \quad \deg(\sigma_j g_j) \leq 2d \}$$

**or, the LP-hierarchy indexed by $d \in \mathbb{N}$:**

$$\theta_d = \sup \{ \lambda : f - \lambda = \sum_{\alpha,\beta} \underbrace{c_{\alpha\beta}}_{\geq 0} \prod_{j=1}^m g_j^{\alpha_j} (1 - g_j)^{\beta_j}; \quad |\alpha + \beta| \leq 2d \}$$

# LP- and SDP-hierarchies for optimization

Replace $f^* = \sup_{\lambda,\sigma_j} \{ \lambda : f(\mathbf{x}) - \lambda \geq 0 \quad \forall \mathbf{x} \in \mathbf{K} \}$ with:

**The SDP-hierarchy indexed by $d \in \mathbb{N}$:**

$$f^*_d = \sup \{ \lambda : f - \lambda = \underbrace{\sigma_0}_{SOS} + \sum_{j=1}^m \underbrace{\sigma_j}_{SOS} g_j; \quad \deg(\sigma_j \, g_j) \leq 2d \}$$

**or, the LP-hierarchy indexed by $d \in \mathbb{N}$:**

$$\theta_d = \sup \{ \lambda : f - \lambda = \sum_{\alpha,\beta} \underbrace{c_{\alpha\beta}}_{\geq 0} \prod_{j=1}^m g_j^{\alpha_j} (1 - g_j)^{\beta_j}; \quad |\alpha + \beta| \leq 2d \}$$

## Theorem

*Both sequence* $(f^*_d)$, *and* $(\theta_d)$, $d \in \mathbb{N}$, *are MONOTONE NON DECREASING and when* **K** *is compact (and satisfies a technical Archimedean assumption) then:*

$$f^* = \lim_{d\to\infty} f^*_d = \lim_{d\to\infty} \theta_d.$$

• What makes this approach exciting is that it is at the crossroads of several disciplines/applications:

- Commutative, Non-commutative, and Non-linear ALGEBRA
- Real algebraic geometry, and Functional Analysis
- Optimization, Convex Analysis
- Computational Complexity in Computer Science,

which BENEFIT from interactions!

• As mentioned ... potential applications are ENDLESS!

• What makes this approach exciting is that it is at the crossroads of several disciplines/applications:

- Commutative, Non-commutative, and Non-linear ALGEBRA
- Real algebraic geometry, and Functional Analysis
- Optimization, Convex Analysis
- Computational Complexity in Computer Science,

which BENEFIT from interactions!

• As mentioned ... potential applications are ENDLESS!

• Has already been proved useful and successful in applications with modest problem size, notably in optimization, control, robust control, optimal control, estimation, computer vision, etc. (If sparsity then problems of larger size can be addressed)

• HAS initiated and stimulated new research issues:

- in Convex Algebraic Geometry (e.g. semidefinite representation of convex sets, algebraic degree of semidefinite programming and polynomial optimization)
- in Computational algebra (e.g., for solving polynomial equations via SDP and Border bases)
- Computational Complexity where LP- and SDP-HIERARCHIES have become an important tool to analyze Hardness of Approximation for 0/1 combinatorial problems ($\rightarrow$ links with quantum computing)

• Has already been proved useful and successful in applications with modest problem size, notably in optimization, control, robust control, optimal control, estimation, computer vision, etc. (If sparsity then problems of larger size can be addressed)

• HAS initiated and stimulated new research issues:

- in Convex Algebraic Geometry (e.g. semidefinite representation of convex sets, algebraic degree of semidefinite programming and polynomial optimization)
- in Computational algebra (e.g., for solving polynomial equations via SDP and Border bases)
- Computational Complexity where LP- and SDP-HIERARCHIES have become an important tool to analyze Hardness of Approximation for 0/1 combinatorial problems ($\to$ links with quantum computing)

Recall that both LP- and SDP- hierarchies are

GENERAL PURPOSE METHODS ....

NOT TAILORED to solving specific hard problems!!

Recall that both LP- and SDP- hierarchies are

GENERAL PURPOSE METHODS ....

NOT TAILORED to solving specific hard problems!!

When solving the optimization problem

$$\mathbf{P}: \quad f^* = \min \{ f(\mathbf{x}) : \; g_j(\mathbf{x}) \geq 0, \; j = 1, \ldots, m \}$$

one does NOT distinguish between CONVEX, CONTINUOUS NON CONVEX, and 0/1 (and DISCRETE) problems! A boolean variable $x_i$ is modelled via the equality constraint "$x_i^2 - x_i = 0$".

**In Non Linear Programming (NLP),**

modeling a 0/1 variable with the polynomial equality constraint "$x_i^2 - x_i = 0$" and applying a standard descent algorithm would be considered "stupid"!

Each class of problems has its own *ad hoc* tailored algorithms.

# A remarkable property of the SOS hierarchy: I

When solving the optimization problem

$$\mathbf{P}: \quad f^* = \min \{ f(\mathbf{x}) : \; g_j(\mathbf{x}) \geq 0, \; j = 1, \ldots, m \}$$

one does NOT distinguish between CONVEX, CONTINUOUS NON CONVEX, and 0/1 (and DISCRETE) problems! A boolean variable $x_i$ is modelled via the equality constraint "$x_i^2 - x_i = 0$".

**In Non Linear Programming (NLP),**

modeling a 0/1 variable with the polynomial equality constraint
"$x_i^2 - x_i = 0$"
and applying a standard descent algorithm would be considered "stupid"!

Each class of problems has its own *ad hoc* tailored algorithms.

# A remarkable property of the SOS hierarchy: I

When solving the optimization problem

$$\mathbf{P}: \quad f^* = \min\{f(\mathbf{x}) : g_j(\mathbf{x}) \geq 0, \; j = 1, \ldots, m\}$$

one does NOT distinguish between CONVEX, CONTINUOUS NON CONVEX, and 0/1 (and DISCRETE) problems! A boolean variable $x_i$ is modelled via the equality constraint "$x_i^2 - x_i = 0$".

**In Non Linear Programming (NLP),**

modeling a 0/1 variable with the polynomial equality constraint "$x_i^2 - x_i = 0$" and applying a standard descent algorithm would be considered "stupid"!

Each class of problems has its own *ad hoc* tailored algorithms.

Even though the moment-SOS approach DOES NOT SPECIALIZE to each class of problems:

- It recognizes the class of (easy) SOS-convex problems as FINITE CONVERGENCE occurs at the FIRST relaxation in the hierarchy.
- Finite convergence also occurs for general convex problems and generically for non convex problems
- $\rightarrow$ (NOT true for the LP-hierarchy.)
- The SOS-hierarchy dominates other lift-and-project hierarchies (i.e. provides the best lower bounds) for hard 0/1 combinatorial optimization problems! The Computer Science community talks about a META-Algorithm.

Even though the moment-SOS approach DOES NOT SPECIALIZE to each class of problems:

- It recognizes the class of (easy) SOS-convex problems as FINITE CONVERGENCE occurs at the FIRST relaxation in the hierarchy.
- Finite convergence also occurs for general convex problems and generically for non convex problems
- $\rightarrow$ (NOT true for the LP-hierarchy.)
- The SOS-hierarchy dominates other lift-and-project hierarchies (i.e. provides the best lower bounds) for hard 0/1 combinatorial optimization problems! The Computer Science community talks about a META-Algorithm.

Even though the moment-SOS approach DOES NOT SPECIALIZE to each class of problems:

- It recognizes the class of (easy) SOS-convex problems as FINITE CONVERGENCE occurs at the FIRST relaxation in the hierarchy.
- Finite convergence also occurs for general convex problems and generically for non convex problems
- $\rightarrow$ (NOT true for the LP-hierarchy.)
- The SOS-hierarchy dominates other lift-and-project hierarchies (i.e. provides the best lower bounds) for hard 0/1 combinatorial optimization problems! The Computer Science community talks about a META-Algorithm.

Even though the moment-SOS approach DOES NOT SPECIALIZE to each class of problems:

- It recognizes the class of (easy) SOS-convex problems as FINITE CONVERGENCE occurs at the FIRST relaxation in the hierarchy.
- Finite convergence also occurs for general convex problems and generically for non convex problems
- $\rightarrow$ (NOT true for the LP-hierarchy.)
- The SOS-hierarchy dominates other lift-and-project hierarchies (i.e. provides the best lower bounds) for hard 0/1 combinatorial optimization problems! The Computer Science community talks about a META-Algorithm.

Even though the moment-SOS approach DOES NOT SPECIALIZE to each class of problems:

- It recognizes the class of (easy) SOS-convex problems as FINITE CONVERGENCE occurs at the FIRST relaxation in the hierarchy.
- Finite convergence also occurs for general convex problems and generically for non convex problems
- $\rightarrow$ (NOT true for the LP-hierarchy.)
- The SOS-hierarchy dominates other lift-and-project hierarchies (i.e. provides the best lower bounds) for hard 0/1 combinatorial optimization problems! The Computer Science community talks about a META-Algorithm.

# A remarkable property: II

FINITE CONVERGENCE of the SOS-hierarchy is GENERIC!

... and provides a GLOBAL OPTIMALITY CERTIFICATE,

the analogue for the NON CONVEX CASE of the

KKT-OPTIMALITY conditions in the CONVEX CASE!

### Theorem (Marshall, Nie)

*Let* $\mathbf{x}^* \in \mathbf{K}$ *be a global minimizer of*

$$\mathbf{P}: \quad f^* = \min\{f(\mathbf{x}) : g_j(\mathbf{x}) \geq 0, j = 1, \ldots, m\}.$$

*and assume that:*

(i) *The gradients* $\{\nabla g_j(\mathbf{x}^*)\}$ *are linearly independent,*

(ii) *Strict complementarity holds* ($\lambda_j^* \, g_j(\mathbf{x}^*) = 0$ *for all* $j$.)

(iii) *Second-order sufficiency conditions hold at* $(\mathbf{x}^*, \lambda^*) \in \mathbf{K} \times \mathbb{R}_+^m$.

*Then* $f(\mathbf{x}) - f^* = \sigma_0^*(\mathbf{x}) + \sum_{j=1}^{m} \sigma_j^*(\mathbf{x}) g_j(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n$, *for some SOS polynomials* $\{\sigma_j^*\}$.

*Moreover, the conditions (i)-(ii)-(iii) HOLD GENERICALLY!*

## Theorem (Marshall, Nie)

*Let* $\mathbf{x}^* \in \mathbf{K}$ *be a global minimizer of*

$$\mathbf{P}: \quad f^* = \min\{f(\mathbf{x}) : g_j(\mathbf{x}) \geq 0, \, j = 1, \ldots, m\}.$$

*and assume that:*

(i) *The gradients* $\{\nabla g_j(\mathbf{x}^*)\}$ *are linearly independent,*

(ii) *Strict complementarity holds* ($\lambda_j^* \, g_j(\mathbf{x}^*) = 0$ *for all* $j$.)

(iii) *Second-order sufficiency conditions hold at* $(\mathbf{x}^*, \lambda^*) \in \mathbf{K} \times \mathbb{R}_+^m$.

*Then* $f(\mathbf{x}) - f^* = \sigma_0^*(\mathbf{x}) + \sum_{j=1}^m \sigma_j^*(\mathbf{x}) g_j(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n$, *for some SOS polynomials* $\{\sigma_j^*\}$.

*Moreover, the conditions (i)-(ii)-(iii) HOLD GENERICALLY!*

Certificates of positivity already exist in convex optimization

$$f^* = f(\mathbf{x}^*) = \min \{ f(\mathbf{x}) : g_j(\mathbf{x}) \geq 0, \quad j = 1, \ldots, m \}$$

when $f$ and $-g_j$ are CONVEX. Indeed if Slater's condition holds there exist nonnegative KKT-multipliers $\lambda_j^* \in \mathbb{R}_+^m$ such that:

$$\nabla f(\mathbf{x}^*) - \sum_{j=1}^{m} \lambda_j{}^* \, g_j(\mathbf{x}^*) = 0; \quad \lambda_j{}^* \, g_j(\mathbf{x}^*) = 0, \; j = 1, \ldots, m.$$

... and so ... the Lagrangian

$$L_{\lambda^*}(\mathbf{x}) := f(\mathbf{x}) - f^* - \sum_{j=1} \lambda_j{}^* \, g_j(\mathbf{x}),$$

**satisfies**

$L_{\lambda^*}(\mathbf{x}^*) = 0$ and $L_{\lambda^*}(\mathbf{x}) \geq 0$ for all $\mathbf{x}$. Therefore:

$$L_{\lambda^*}(\mathbf{x}) \geq 0 \Rightarrow f(\mathbf{x}) \geq f^* \quad \forall \mathbf{x} \in \mathbf{K}!$$

## Certificates of positivity already exist in convex optimization

$$f^* = f(\mathbf{x}^*) = \min \{ f(\mathbf{x}) : \; g_j(\mathbf{x}) \geq 0, \quad j = 1, \ldots, m \}$$

when $f$ and $-g_j$ are CONVEX. Indeed if Slater's condition holds there exist nonnegative KKT-multipliers $\lambda_j^* \in \mathbb{R}_+^m$ such that:

$$\nabla f(\mathbf{x}^*) - \sum_{j=1}^{m} \lambda_j{}^* \, g_j(\mathbf{x}^*) \; = \; 0; \quad \lambda_j{}^* \, g_j(\mathbf{x}^*) = 0, \; j = 1, \ldots, m.$$

... and so ... the Lagrangian

$$L_{\lambda^*}(\mathbf{x}) := f(\mathbf{x}) - f^* - \sum_{j=1} \lambda_j{}^* \, g_j(\mathbf{x}),$$

**satisfies**

$L_{\lambda^*}(\mathbf{x}^*) = 0$ and $L_{\lambda^*}(\mathbf{x}) \geq 0$ for all $\mathbf{x}$. Therefore:

$$L_{\lambda^*}(\mathbf{x}) \geq 0 \Rightarrow f(\mathbf{x}) \geq f^* \quad \forall \mathbf{x} \in \mathbf{K}!$$

In summary:

| KKT-OPTIMALITY when $f$ and $-g_j$ are CONVEX | PUTINAR's CERTIFICATE in the non CONVEX CASE |
|---|---|
| $\nabla f(\mathbf{x}^*) - \sum_{j=1}^m \lambda_j^* \nabla g_j(\mathbf{x}^*) = 0$ | $\nabla f(\mathbf{x}^*) - \sum_{j=1}^m \sigma_j(\mathbf{x}^*) \nabla g_j(\mathbf{x}^*) = 0$ |
| $f(\mathbf{x}) - f^* - \sum_{j=1}^m \lambda_j^* g_j(\mathbf{x})$ | $f(\mathbf{x}) - f^* - \sum_{j=1}^m \sigma_j^*(\mathbf{x}) g_j(\mathbf{x})$ |
| $\geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$ | $(= \sigma_0^*(\mathbf{x})) \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$. |
| | for some SOS $\{\sigma_j^*\}$, and $\sigma_j^*(\mathbf{x}^*) = \lambda_j^*$. |

In summary:

| KKT-OPTIMALITY when $f$ and $-g_j$ are CONVEX | PUTINAR's CERTIFICATE in the non CONVEX CASE |
| --- | --- |
| $\nabla f(\mathbf{x}^*) - \sum_{j=1}^m \lambda_j^* \nabla g_j(\mathbf{x}^*) = 0$ | $\nabla f(\mathbf{x}^*) - \sum_{j=1}^m \sigma_j(\mathbf{x}^*) \nabla g_j(\mathbf{x}^*) = 0$ |
| $f(\mathbf{x}) - f^* - \sum_{j=1}^m \lambda_j^* g_j(\mathbf{x})$ | $f(\mathbf{x}) - f^* - \sum_{j=1}^m \sigma_j^*(\mathbf{x}) g_j(\mathbf{x})$ |
| $\geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$ | $(= \sigma_0^*(\mathbf{x})) \ \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$. |
| | for some SOS $\{\sigma_j^*\}$, and $\sigma_j^*(\mathbf{x}^*) = \lambda_j^*$. |

# II. Approximation of sets with quantifiers

Let $f \in \mathbb{R}[x, y]$ and let $\mathbf{K} \subset \mathbb{R}^n \times \mathbb{R}^p$ be the semi-algebraic set:

$$\mathbf{K} := \{(x, y) : \quad g_j(x, y) \geq 0, \quad j = 1, \ldots, m\},$$

and let $\mathbf{B} \subset \mathbb{R}^n$ be the unit ball or the box $[-1, 1]^n$.

**Suppose that one wants to approximate the set:**

$$R_f := \{x \in \mathbf{B} : \ f(x, y) \leq 0 \text{ for all } y \text{ such that } (x, y) \in \mathbf{K}\}$$

as closely as desired by a sequence of sets of the form:

$$\Theta_k := \{\mathbf{x} \in \mathbf{B} : \ J_k(x) \leq 0 \}$$

for some polynomials $J_k$.

With $g_0 = 1$ and with $\mathbf{K} \subset \mathbb{R}^n \times \mathbb{R}^p$ and $k \in \mathbb{N}$, let

$$Q_k(g) := \left\{ \sum_{j=0}^{m} \sigma_j(x,y)\, g_j(x,y) : \quad \sigma_j \in \Sigma[x,y], \deg \sigma_j\, g_j \leq 2k \right\}$$

Let $x \mapsto F(x) := \max\{f(x,y) : (x,y) \in \mathbf{K}\}$, and

**for every integer $k$ consider the optimization problem:**

$$\rho_k = \min_{J \in \mathbb{R}[x]_k} \left\{ \int_{\mathbf{B}} (J - F)\, dx \; : \; J(x) - f(x,y) \in Q_k(g) \right\}$$

## 1. The criterion

$$\int_{\mathbf{B}} (J - F)\, dx = \underbrace{\int_{\mathbf{B}} -F\, dx}_{\text{unknown but constant}} + \sum_{\alpha} J_{\alpha} \underbrace{\int_{\mathbf{B}} \mathbf{x}^{\alpha}\, dx}_{\text{easy to compute}}$$

is LINEAR in the coefficients $J_{\alpha}$ of the unknown polynomial $J \in \mathbb{R}[\mathbf{x}]_k$!

## 2. The constraint

$$J(x) - f(x, y) = \sum_{j=0}^{m} \sigma_j(x, y)\, g_j(x, y)$$

is just LINEAR CONSTRAINTS + LMIs!

## 1. The criterion

$$\int_{\mathbf{B}} (J - F)\, dx = \underbrace{\int_{\mathbf{B}} -F\, dx}_{\text{unknown but constant}} + \sum_{\alpha} J_{\alpha} \underbrace{\int_{\mathbf{B}} \mathbf{x}^{\alpha}\, dx}_{\text{easy to compute}}$$

is LINEAR in the coefficients $J_{\alpha}$ of the unknown polynomial $J \in \mathbb{R}[\mathbf{x}]_k$!

## 2. The constraint

$$J(x) - f(x, y) = \sum_{j=0}^{m} \sigma_j(x, y)\, g_j(x, y)$$

is just LINEAR CONSTRAINTS + LMIs!

## Hence, the optimization problem

$$\rho_k = \min_{J \in \mathbb{R}[x]_k} \left\{ \int_{\mathbf{B}} (J - F)\, dx \;:\; J(\mathbf{x}) - f(\mathbf{x}, y) \;\in\; Q_k(g) \right\}$$

IS AN SDP! Moreover, it has an optimal solution $J_k^* \in \mathbb{R}[x]_k$!

- Alternatively, if one uses LP-based positivity certificates for $J(\mathbf{x}) - f(\mathbf{x}, y)$, one ends up with solving an LP!

## From the definition of $J_k^*$, the sublevel sets

$$\Theta_k := \{ x \in \mathbf{B} : \; J_k^*(x) \leq 0 \} \subset R_f, \quad k \in \mathbb{N},$$

provide a nested sequence of INNNER approximations of $R_f$.

### Hence, the optimization problem

$$\rho_k = \min_{J\in\mathbb{R}[x]_k} \left\{ \int_{\mathbf{B}} (J - F)\, dx \;:\; J(\mathbf{x}) - f(\mathbf{x}, y) \in Q_k(g) \right\}$$

IS AN SDP! Moreover, it has an optimal solution $J_k^* \in \mathbb{R}[x]_k$!

• Alternatively, if one uses LP-based positivity certificates for $J(\mathbf{x}) - f(\mathbf{x}, y)$, one ends up with solving an LP!

### From the definition of $J_k^*$, the sublevel sets

$$\Theta_k := \{x \in \mathbf{B} : J_k^*(x) \leq 0\} \subset R_f, \quad k \in \mathbb{N},$$

provide a nested sequence of INNNER approximations of $R_f$.

### Hence, the optimization problem

$$\rho_k = \min_{J \in \mathbb{R}[x]_k} \left\{ \int_{\mathbf{B}} (J - F)\, dx \;:\; J(\mathbf{x}) - f(\mathbf{x}, y) \in Q_k(g) \right\}$$

IS AN SDP! Moreover, it has an optimal solution $J_k^* \in \mathbb{R}[x]_k$!

- Alternatively, if one uses LP-based positivity certificates for $J(\mathbf{x}) - f(\mathbf{x}, y)$, one ends up with solving an LP!

### From the definition of $J_k^*$, the sublevel sets

$$\Theta_k := \{x \in \mathbf{B} : J_k^*(x) \leq 0\} \subset R_f, \quad k \in \mathbb{N},$$

provide a nested sequence of INNNER approximations of $R_f$.

### Theorem (Lass)

*(Strong) convergence in $L_1(\mathbf{B})$-norm takes place, that is:*

$$\lim_{k\to\infty} \int_{\mathbf{B}} |\, J_k^* - F \,|\, dx \;=\; 0$$

*and, if in addition the set $\{x \in \mathbf{B} : \; F(x) = 0\}$ has Lebesgue measure zero, then*

$$\lim_{k\to\infty} \mathrm{VOL}(R_f \setminus \Theta_k) \;=\; 0$$

## Theorem (Lass)

*(Strong) convergence in $L_1(\mathbf{B})$-norm takes place, that is:*

$$\lim_{k\to\infty} \int_{\mathbf{B}} |\, J_k^* - F \,|\, dx \;=\; 0$$

*and, if in addition the set $\{x \in \mathbf{B} :\; F(x) = 0\}$ has Lebesgue measure zero, then*

$$\lim_{k\to\infty} \; \mathrm{VOL}(R_f \setminus \Theta_k) \;=\; 0$$

# Ex: Polynomial Matrix Inequalities: (with D. Henrion)

Let $x \mapsto \mathbf{A}(x) \in \mathbb{R}^{p\times p}$ where $\mathbf{A}(x)$ is the matrix-polynomial

$$x \mapsto \mathbf{A}(x) = \sum_{\alpha\in\mathbb{N}^n} \mathbf{A}_\alpha \, x^\alpha \quad \left( = \sum_{\alpha\in\mathbb{N}^n} \mathbf{A}_\alpha \, x_1^{\alpha_1} \cdots x_n^{\alpha_n} \right).$$

for finitely many real symmetric matrices $(\mathbf{A}_\alpha)$, $\alpha \in \mathbb{N}^n$.

**... and suppose one wants to approximate the set**

$$R_{\mathbf{A}} := \{x \in \mathbf{B} : \mathbf{A}(x) \succeq 0\} = \{x : \lambda_{\min}(\mathbf{A}(x)) \geq 0\}.$$

**Then:**

$$R_{\mathbf{A}} = \left\{ x \in \mathbf{B} : \underbrace{y^T\mathbf{A}(x)y}_{f(x,y)} \geq 0, \quad \forall\, y \text{ s.t. } \|y\|^2 = 1 \right\}$$

# Ex: Polynomial Matrix Inequalities: (with D. Henrion)

Let $x \mapsto \mathbf{A}(x) \in \mathbb{R}^{p\times p}$ where $\mathbf{A}(x)$ is the matrix-polynomial

$$x \mapsto \mathbf{A}(x) = \sum_{\alpha\in\mathbb{N}^n} \mathbf{A}_\alpha\, x^\alpha \quad \left( = \sum_{\alpha\in\mathbb{N}^n} \mathbf{A}_\alpha\, x_1^{\alpha_1}\cdots x_n^{\alpha_n}\right).$$

for finitely many real symmetric matrices $(\mathbf{A}_\alpha)$, $\alpha \in \mathbb{N}^n$.

**... and suppose one wants to approximate the set**

$$R_\mathbf{A} := \{x \in \mathbf{B} \,:\, \mathbf{A}(x) \succeq 0\} = \{x \,:\, \lambda_{\min}(\mathbf{A}(x)) \geq 0\}.$$

**Then:**

$$R_\mathbf{A} = \left\{ x \in \mathbf{B} \,:\, \underbrace{y^T\mathbf{A}(x)y}_{f(x,y)} \geq 0, \quad \forall\, y \text{ s.t. } \|y\|^2 = 1 \right\}$$

# Illustrative example (continued)

Let **B** be the unit disk $\{\mathbf{x} : \|\mathbf{x}\| \leq 1\}$ and let:

$$R_{\mathbf{A}} := \left\{ \mathbf{x} \in \mathbf{B} \; : \; \mathbf{A}(\mathbf{x}) \left( = \begin{bmatrix} 1 - 16x_1x_2 & x_1 \\ x_1 & 1 - x_1^2 - x_2^2 \end{bmatrix} \right) \succeq 0 \right\}$$

Then by solving relatively simple semidefinite programs, one may approximate $R_{\mathbf{A}}$ with sublevel sets of the form:

$$\Theta_k := \{x \in \mathbf{B} : J_k^*(x) \geq 0\}$$

for some polynomial $J_k^*$ of degree $k = 2, 4, \ldots$ and with

$$\mathrm{VOL}\,(R_A \setminus \Theta_k) \to 0 \qquad \text{as } k \to \infty.$$

# Illustrative example (continued)

Let **B** be the unit disk $\{\mathbf{x} : \|\mathbf{x}\| \leq 1\}$ and let:

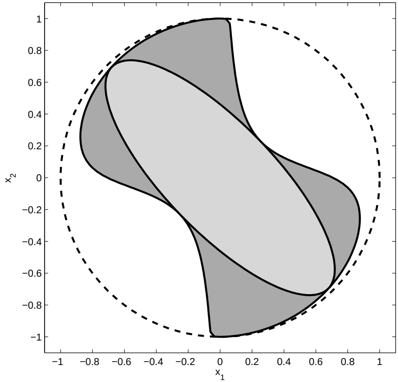
$$R_{\mathbf{A}} := \left\{ \mathbf{x} \in \mathbf{B} \ : \ \mathbf{A}(\mathbf{x}) \left( = \begin{bmatrix} 1 - 16x_1x_2 & x_1 \\ x_1 & 1 - x_1^2 - x_2^2 \end{bmatrix} \right) \succeq 0 \right\}$$

Then by solving relatively simple semidefinite programs, one may approximate $R_{\mathbf{A}}$ with sublevel sets of the form:

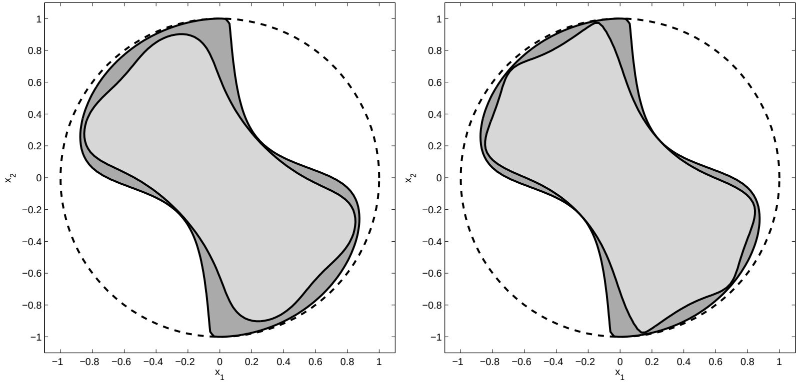
$$\Theta_k := \{ x \in \mathbf{B} : \ J_k^*(x) \geq 0 \}$$

for some polynomial $J_k^*$ of degree $k = 2, 4, \ldots$ and with

$$\mathrm{VOL}\,(R_A \setminus \Theta_k) \to 0 \qquad \text{as } k \to \infty.$$




$\Theta_2$ (left) and $\Theta_4$ (right) inner approximations (light gray) of (dark gray) embedded in unit disk **B** (dashed).




$\Theta_6$ (left) and $\Theta_8$ (right) inner approximations (light gray) of (dark gray) embedded in unit disk **B** (dashed).

# III. Convex underestimators of polynomials

In large scale Mixed Integer Nonlinear Programming (MINLP), a popular method is to use B & B where LOWER BOUNDS at each node of the search tree must be computed EFFICIENTLY!

**In such a case ... one needs**

CONVEX UNDERESTIMATORS

of the objective function, say on a BOX $\mathbf{B} \subset \mathbb{R}^n$!

**Message:**

"Good" CONVEX POLYNOMIAL UNDERESTIMATORS can be computed efficiently!

# III. Convex underestimators of polynomials

In large scale Mixed Integer Nonlinear Programming (MINLP), a popular method is to use B & B where LOWER BOUNDS at each node of the search tree must be computed EFFICIENTLY!

**In such a case ... one needs**

CONVEX UNDERESTIMATORS

of the objective function, say on a BOX $\mathbf{B} \subset \mathbb{R}^n$!

**Message:**

"Good" CONVEX POLYNOMIAL UNDERESTIMATORS can be computed efficiently!

## Solving

$$\inf_{p\in\mathbb{R}[x]_d} \left\{ \int_{\mathbf{B}} (f(x) - p(x))\,dx : \right.$$

$$\text{s.t.} \quad f - p \geq 0 \text{ on } \mathbf{B} \text{ and } p \text{ convex on } \mathbf{B} \}$$

will provide a degree-$d$ POLYNOMIAL CONVEX UNDERESTIMATOR $p^*$ of $f$ on $\mathbf{B}$ that minimizes the $L_1(\mathbf{B})$-norm $\|f - p\|_1$ !

Notice that:

- $\int_{\mathbf{B}} (f(x) - p(x))\,dx$ is LINEAR in the coefficients of $p$!
- $p$ convex on $\mathbf{B}$ $\Leftrightarrow \underbrace{\mathbf{y}^T \nabla^2 p(x)\,\mathbf{y}}_{\in\mathbb{R}[x\mathbf{y}]_d} \geq 0$ on $\mathbf{B} \times \{\mathbf{y} : \|\mathbf{y}\|^2 = 1\}$!

## Hence replace the positivity and convexity constraints

$f - p \geq 0$ on **B** and $p$ convex on **B**

## with the positivity certificates

$$f(x) - p(x) = \sum_{k=0}^{m} \underbrace{\sigma_j(x)}_{SOS} g_j(x)$$

$$y^T \nabla^2 p(x)\, y = \sum_{k=0}^{m} \underbrace{\psi(x,y)}_{SOS} g_j(x) + \psi_{m+1}(x,y)\,(1 - \|y\|^2)$$

## Hence replace the positivity and convexity constraints

$f - p \geq 0$ on **B** and $p$ convex on **B**

## with the positivity certificates

$$f(x) - p(x) \;=\; \sum_{k=0}^{m} \underbrace{\sigma_j(x)}_{SOS}\, g_j(x)$$

$$y^T \nabla^2 p(x)\, y \;=\; \sum_{k=0}^{m} \underbrace{\psi(x,y)}_{SOS}\, g_j(x) + \psi_{m+1}(x,y)\,(1 - \|y\|^2)$$

## and apply the moment-SOS approach

to obtain a sequence of polynomials $p_k^* \in R[x]_d$, $k \in \mathbb{N}$, of degree $d$ which converges to the BEST convex polynomial underestimator of degree $d$.

# Conclusion

- The moment-SOS hierarchy is a powerful general methodology.
- Works for problems of modest size (or larger size problems with sparsity and/or symmetries)

# An alternative for larger size problems ?

**Mixed LP-SOS positivity certificate**

$$f(\mathbf{x}) = \sum_{\alpha,\beta} \underbrace{c_{\alpha\beta}}_{\geq 0} \prod_j g_j(\mathbf{x})^{\alpha_j} \prod_j (1 - g_j(\mathbf{x}))^{\beta_j} + \underbrace{\sigma_0(\mathbf{x})}_{\text{sos of degree } k}$$

where $k$ IS FIXED!

→ A bounded degree SOS hierarchy for polynomial optimization, Eur. J. Comput. Optimization, with K. Toh & S. Yang

THANK YOU!!