The moment-LP and moment-SOS approaches

Jean B. Lasserre

LAAS-CNRS and Institute of Mathematics, Toulouse, France

CIRM, November 2013

- Semidefinite Programming
- Why polynomial optimization?
- LP- and SDP- CERTIFICATES of POSITIVITY
- The moment-LP and moment-SOS approaches
- Two examples outside optimization:
  - Approximating sets defined with quantifiers
  - Convex polynomial underestimators

- Semidefinite Programming
- Why polynomial optimization?
- LP- and SDP- CERTIFICATES of POSITIVITY
- The moment-LP and moment-SOS approaches
- Two examples outside optimization:
  - Approximating sets defined with quantifiers
  - Convex polynomial underestimators

- Semidefinite Programming
- Why polynomial optimization?
- LP- and SDP- CERTIFICATES of POSITIVITY
- The moment-LP and moment-SOS approaches
- Two examples outside optimization:
  - Approximating sets defined with quantifiers
  - Convex polynomial underestimators

- Semidefinite Programming
- Why polynomial optimization?
- LP- and SDP- CERTIFICATES of POSITIVITY
- The moment-LP and moment-SOS approaches
- Two examples outside optimization:
  - Approximating sets defined with quantifiers
  - Convex polynomial underestimators

- Semidefinite Programming
- Why polynomial optimization?
- LP- and SDP- CERTIFICATES of POSITIVITY
- The moment-LP and moment-SOS approaches
- Two examples outside optimization:
    - Approximating sets defined with quantifiers
    - Convex polynomial underestimators

- Semidefinite Programming
- Why polynomial optimization?
- LP- and SDP- CERTIFICATES of POSITIVITY
- The moment-LP and moment-SOS approaches
- Two examples outside optimization:
  - Approximating sets defined with quantifiers
  - Convex polynomial underestimators

The CONVEX optimization problem:

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

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

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

or, equivalently

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

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,...

Why Polynomial Optimization?

After all ...

the polynomial optimization problem:

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

is just a particular case of Non Linear Programming (NLP)!

True!

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

Why Polynomial Optimization?

After all ...

the polynomial optimization problem:

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

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!

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!

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^*:

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

... and so to compute f^* one needs

TRACTABLE CERTIFICATES of POSITIVITY on $\mathbf{K}$!

BUT for GLOBAL Optimization

... the picture is different!

Remember that for the GLOBAL minimum f^*:

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

... and so to compute f^* one needs

TRACTABLE CERTIFICATES of POSITIVITY on $\mathbf{K}$!

BUT for GLOBAL Optimization

... the picture is different!

Remember that for the GLOBAL minimum f^*:

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

... and so to compute f^* one needs

TRACTABLE CERTIFICATES of POSITIVITY on $\mathbf{K}$!

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 $(\sigma_j) \subset \mathbb{R}[\mathbf{x}]$.

Testing whether † holds for some

SOS $(\sigma_j) \subset \mathbb{R}[\mathbf{x}]$ with a degree bound, is SOLVING an SDP!

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 $(\sigma_j) \subset \mathbb{R}[\mathbf{x}]$.

Testing whether † holds for some

SOS $(\sigma_j) \subset \mathbb{R}[\mathbf{x}]$ with a degree bound, is SOLVING an SDP!

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* $\{g_j, (1 - g_j)\}$ *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 $(c_{\alpha\beta})$.

Testing whether $\star$ holds for some NONNEGATIVE $(c_{\alpha\beta})$ with $|\alpha + \beta| \leq M$, is SOLVING an LP!

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* $\{g_j, (1 - g_j)\}$ *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 $(c_{\alpha\beta})$.

Testing whether $\star$ holds for some

NONNEGATIVE $(c_{\alpha\beta})$ with $|\alpha + \beta| \leq M$, is SOLVING an LP!

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!

- Farkas Lemma $\rightarrow$ Krivine-Stengle
- KKT-Optimality conditions $\rightarrow$ Schmüdgen-Putinar

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!

- Farkas Lemma $\rightarrow$ Krivine-Stengle
- KKT-Optimality conditions $\rightarrow$ Schmüdgen-Putinar

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!

- Farkas Lemma $\to$ Krivine-Stengle
- KKT-Optimality conditions $\to$ Schmüdgen-Putinar

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!

- Farkas Lemma $\rightarrow$ Krivine-Stengle
- KKT-Optimality conditions $\rightarrow$ Schmüdgen-Putinar

• In addition, polynomials NONNEGATIVE ON A SET $\mathbf{K} \subset \mathbb{R}^n$ are ubiquitous. They appear in many important applications, and not only in global optimization!

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.)

• In addition, polynomials NONNEGATIVE ON A SET $\mathbf{K} \subset \mathbb{R}^n$ are ubiquitous. They appear in many important applications, and not only in global optimization!

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-Stengle's or Putinar's certificate) in potentially any application where such a characterization is needed. (Global optimization is only one example.)

In may 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-Stengle's or Putinar's certificate) in potentially any application where such a characterization is needed. (Global optimization is only one example.)

In may 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-Stengle's or Putinar's certificate) in potentially any application where such a characterization is needed. (Global optimization is only one example.)

In may 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-Stengle's or Putinar's certificate) in potentially any application where such a characterization is needed. (Global optimization is only one example.)

In may 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 (θ_d), $d \in \mathbb{N}$, are MONOTONE NON DECREASING and when $\mathbf{K}$ 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.

• 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)

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.

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 SPECIALIZES 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.)
  $\rightarrow$ (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!

Even though the moment-SOS approach DOES NOT SPECIALIZES 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.)
  $\rightarrow$ (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!

Even though the moment-SOS approach DOES NOT SPECIALIZES 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.)
  $\rightarrow$ (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!

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!

In summary:

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

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_{α} 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_α 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

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

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:

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

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

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

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

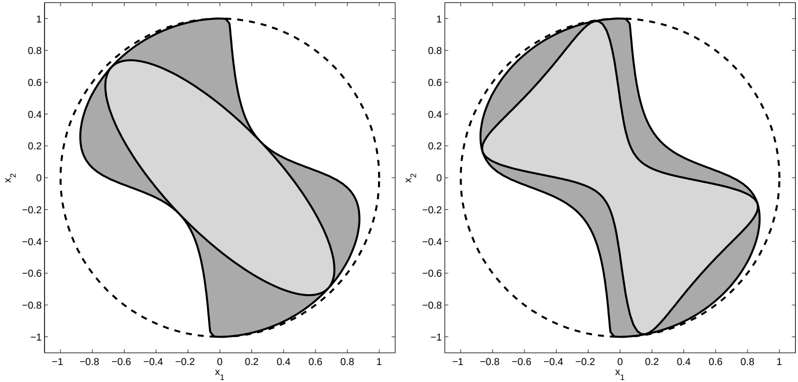
$$R_{\mathbf{A}} := \left\{ \mathbf{x} \in \mathbf{B} \;:\; \mathbf{A}(\mathbf{x}) \left(= \begin{bmatrix} 1 - 16x_1x_2 & x_1 \\ x_1 & 1 - x_1^2 - x_2^2 \end{bmatrix} \right) \succeq 0 \right\}$$

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

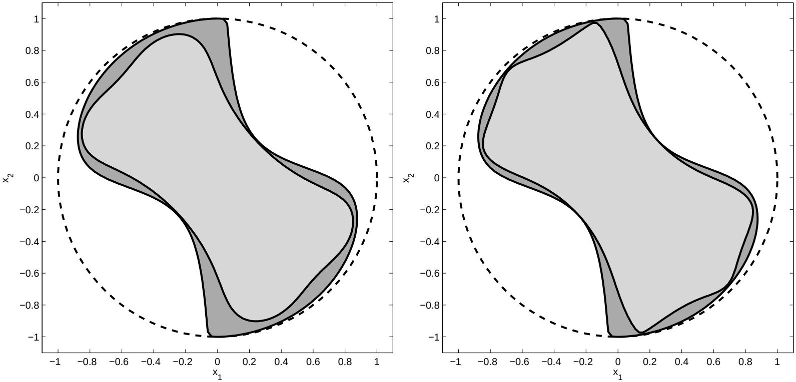
$$\Theta_k \;:=\; \{ x \in \mathbf{B} :\; J_k^*(x) \geq 0 \}$$

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

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



Θ_2 (left) and Θ_4 (right) inner approximations (light gray) of (dark gray) embedded in unit disk **B** (dashed).



Θ_6 (left) and Θ_8 (right) inner approximations (light gray) of (dark gray) embedded in unit disk **B** (dashed).

Similarly, suppose that one wants to approximate the set:

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

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.

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

for every integer k the optimization problem:

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

IS AN SDP with an optimal solution $J_k^* \in \mathbb{R}[\mathbf{x}]_k$.

From the definition of J_k^*, the sublevel sets

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

provide a nested sequence of OUTER approximations of D_f.

Theorem (Lass)

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

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

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

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

Theorem (Lass)

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

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

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

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

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 efficienty!

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

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

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

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 $\mathbf{B}$ and p convex on $\mathbf{B}$

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 $\mathbf{B}$ and p convex on $\mathbf{B}$

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 problem with sparsity and/or symmetries)

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!

An alternative for larger size problems ?

Conclusion

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

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!

An alternative for larger size problems ?

THANK YOU!!