The moment-LP and moment-SOS approaches in optimization

Jean B. Lasserre

LAAS-CNRS and Institute of Mathematics, Toulouse, France

Master 2 Optimization: Paris-Orsay, 2016

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

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 $(c_{\alpha\beta})$*.*

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

Given a real sequence $\mathbf{y} = (y_\alpha)$, $\alpha \in \mathbb{N}^n$, determine where there exists some finite Borel measure μ on $\mathbf{K}$ such that

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

Theorem

If $\mathbf{K} = \{\mathbf{x} : g_j(\mathbf{x}) \geq 0, \; j = 1, \ldots, m\}$ is compact (+Archimedean assumption) then $\dagger$ holds if and only if

$$(\star) \qquad L_{\mathbf{y}}\left(h^2 \, g_j\right) \geq 0, \qquad \forall h \in \mathbb{R}[\mathbf{x}]_d.$$

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

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

Given a real sequence $\mathbf{y} = (y_\alpha)$, $\alpha \in \mathbb{N}^n$, determine where there exists some finite Borel measure μ on $\mathbf{K}$ such that

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

Theorem

If $\mathbf{K} = \{\mathbf{x} : g_j(\mathbf{x}) \geq 0, \ j = 1, \ldots, m\}$ is compact (+Archimedean assumption) then $\dagger$ holds if and only if

$$(\star) \qquad L_y\left(h^2 g_j\right) \geq 0, \qquad \forall h \in \mathbb{R}[\mathbf{x}]_d.$$

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

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

Given a real sequence $\mathbf{y} = (y_\alpha)$, $\alpha \in \mathbb{N}^n$, determine where there exists some finite Borel measure μ on $\mathbf{K}$ such that

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

Theorem

If $\mathbf{K} = \{\mathbf{x} : g_j(\mathbf{x}) \geq 0, \; j = 1, \ldots, m\}$ *is compact (+Archimedean assumption) then* $\dagger$ *holds if and only if*

$$(\star) \qquad L_y\left(h^2 \, g_j\right) \geq 0, \qquad \forall h \in \mathbb{R}[\mathbf{x}]_d.$$

The condition $(\star)$ is equivalent to countably many LINEAR MATRIX INEQUALITIES on the y_α'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* $\dagger$ *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_α'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* $\dagger$ *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_α's

The Generalized problem of Moments

- 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.

The Generalized Moment Problem (GMP) is the infinite-dimensional linear program

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

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

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

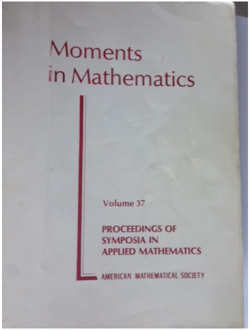
$$\sup_{\lambda_j} \quad \{ \sum_{j \in J}^{s} \lambda_j\, b_j : \quad f_i - \sum_{j \in J} \lambda_j\, h_{ij} \geq 0 \text{ on } \mathbf{K}_i, \quad i = 1, \ldots, s \}$$

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 $\mathbf{K}_i$, $i = 1, \ldots, s$.



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

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

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 μ 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 $\quad \rightarrow f^* = \inf_{\mathbf{x}} \{ f(\mathbf{x}) : \; \mathbf{x} \in \mathbf{K} \}$

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 μ 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

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

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

• 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_α.

• 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 $f \in \mathbb{R}[x,y]$, $\mathbf{B}$ 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.)

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 $f^* = \sup_\lambda \{ \lambda : f(\mathbf{x}) - \lambda \geq 0 \quad \forall \mathbf{x} \in \mathbf{K}\}$ 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 $f^* = \sup_\lambda \{ \lambda : \; f(\mathbf{x}) - \lambda \geq 0 \quad \forall \mathbf{x} \in \mathbf{K} \}$ 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. (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 ($\rightarrow$ links with quantum computing)

There has been also recents attempts to use other types of algebraic certificates of positivity that try to avoid the size explosion due to the semidefinite matrices associated with the SOS weights in Putinar's positivity certificate

Recent work by Ahmadi et al. and Lasserre, Toh and Zhang