The moment-SOS approach in & outside optimization

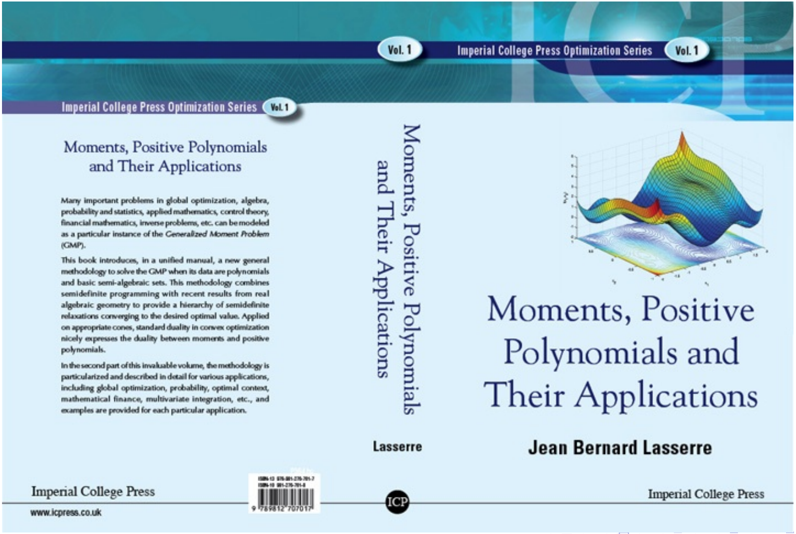
Jean B. Lasserre

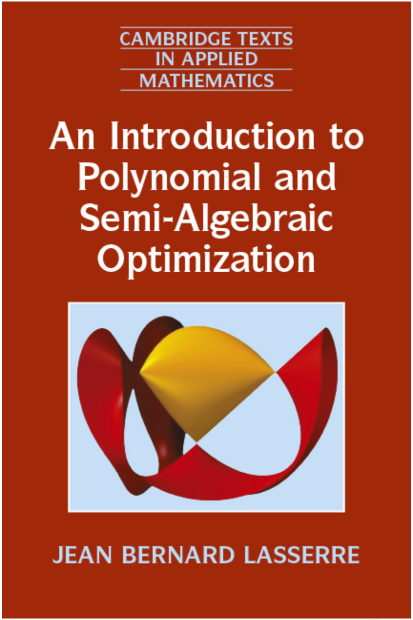
LAAS-CNRS and Institute of Mathematics, Toulouse, France

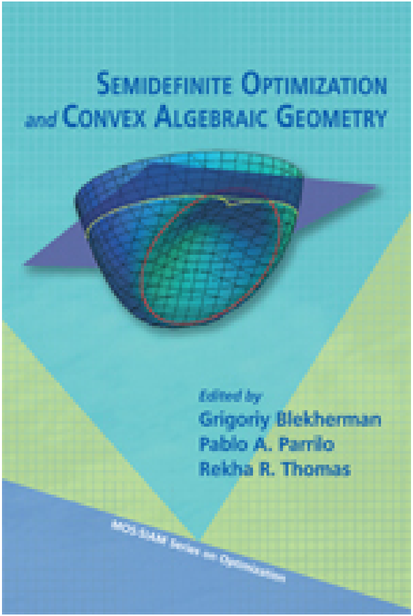
NIPS-2016 Workshop on Non-Convex Optimization for Machine Learning

Barcelona, December 2016

The author gratefully acknowledges financial support from Amazon and the European Research Council (ERC)


Vol. 1
Imperial College Press Optimization Series
Vol. 1
Imperial College Press Optimization Series
Vol. 1
Moments, Positive Polynomials
and Their Applications
Many important problems in global optimization, algebra, probability and statistics, applied mathematics, control theory, financial mathematics, inverse problems, etc. can be modeled as a particular instance of the Generalized Moment Problem (GMP).
This book introduces, in a unified manual, a new general methodology to solve the GMP when its data are polynomials and basic semi-algebraic sets. This methodology combines semidefinite programming with recent results from real algebraic geometry to provide a hierarchy of semidefinite relaxations converging to the desired optimal value. Applied on appropriate cones, standard duality in convex optimization nicely expresses the duality between moments and positive polynomials.
In the second part of this invaluable volume, the methodology is particularized and described in detail for various applications, including global optimization, probability, optimal control, mathematical finance, multivariate integration, etc., and examples are provided for each particular application.
Imperial College Press
www.icpress.co.uk
Moments, Positive Polynomials and Their Applications
Lasserre
ICP
Moments, Positive Polynomials and Their Applications
Jean Bernard Lasserre
Imperial College Press


CAMBRIDGE TEXTS
IN APPLIED
MATHEMATICS
An Introduction to
Polynomial and
Semi-Algebraic
Optimization
JEAN BERNARD LASSERRE


Semidefinite Optimization
and Convex Algebraic Geometry
Edited by
Grigoriy Blekherman
Pablo A. Parrilo
Rekha R. Thomas
MOS-SIAM Series on Optimization

- Why polynomial optimization?
- SDP- CERTIFICATE of POSITIVITY
- The moment-SOS approach
- Some applications

- Why polynomial optimization?
- SDP- CERTIFICATE of POSITIVITY
- The moment-SOS approach
- Some applications

- Why polynomial optimization?
- SDP- CERTIFICATE of POSITIVITY
- The moment-SOS approach
- Some applications

- Why polynomial optimization?
- SDP- CERTIFICATE of POSITIVITY
- The moment-SOS approach
- Some applications

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 LOCAL 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 LOCAL 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 LOCAL 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 (so far) 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 (so far) 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 (so far) are useless from a computational viewpoint.)

SOS-based certificate

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

be compact (with $g_1(\mathbf{x}) = M - \|\mathbf{x}\|^2$, so that $\mathbf{K} \subset \mathbf{B}(0, M)$).

Theorem (Putinar's Positivstellensatz)

If $f \in \mathbb{R}[\mathbf{x}]$ is strictly positive ($f > 0$) 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 $(\sigma_j) \subset \mathbb{R}[\mathbf{x}]$.

SOS-based certificate

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

be compact (with $g_1(\mathbf{x}) = M - \|\mathbf{x}\|^2$, so that $\mathbf{K} \subset \mathbf{B}(0, M)$).

Theorem (Putinar's Positivstellensatz)

If $f \in \mathbb{R}[\mathbf{x}]$ is strictly positive ($f > 0$) 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 (σ_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 (σ_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!

Dual side: The K-moment problem

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

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

If yes y is said to have a representing measure supported on $\mathbf{K}$.

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: The K-moment problem

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

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

If yes y is said to have a representing measure supported on $\mathbf{K}$.

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

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

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

be compact (with $g_1(\mathbf{x}) = M - \|\mathbf{x}\|^2$, so that $\mathbf{K} \subset \mathbf{B}(0, M)$).

Theorem

A sequence $\mathbf{y} = (y_\alpha)$, $\alpha \in \mathbb{N}^n$, has a representing measure supported on $\mathbf{K}$ if and only if for every $h \in \mathbb{R}[\mathbf{x}]$:

$$(\star) \qquad L_{\mathbf{y}}(h^2) \, \geq \, 0; \quad L_{\mathbf{y}}(h^2 \, g_j) \, \geq \, 0, \quad j = 1, \ldots, m.$$

The condition $(\star)$ for all $h \in \mathbb{R}[\mathbf{x}]_d$ is equivalent to $m + 1$ positive semidefiniteness of some moment and localizing matrices, i.e.,

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

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

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

be compact (with $g_1(\mathbf{x}) = M - \|\mathbf{x}\|^2$, so that $\mathbf{K} \subset \mathbf{B}(0, M)$).

Theorem

A sequence $\mathbf{y} = (y_\alpha)$, $\alpha \in \mathbb{N}^n$, *has a representing measure supported on* $\mathbf{K}$ *if and only if for every* $h \in \mathbb{R}[\mathbf{x}]$*:*

$$(\star) \qquad L_{\mathbf{y}}(h^2) \,\geq\, 0; \quad L_{\mathbf{y}}(h^2\, g_j) \,\geq\, 0, \quad j = 1,\ldots,m.$$

The condition $(\star)$ for all $h \in \mathbb{R}[\mathbf{x}]_d$ is equivalent to $m+1$ positive semidefiniteness of some moment and localizing matrices, i.e.,

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

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

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

be compact (with $g_1(\mathbf{x}) = M - \|\mathbf{x}\|^2$, so that $\mathbf{K} \subset \mathbf{B}(0, M)$).

Theorem

A sequence $\mathbf{y} = (y_\alpha)$, $\alpha \in \mathbb{N}^n$, has a representing measure supported on $\mathbf{K}$ if and only if for every $h \in \mathbb{R}[\mathbf{x}]$:

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

The condition $(\star)$ for all $h \in \mathbb{R}[\mathbf{x}]_d$ is equivalent to $m + 1$ positive semidefiniteness of some moment and localizing matrices, i.e.,

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

whose rows & columns are indexed by $\mathbb{N}^n_d$, and entries are LINEAR in the y_α'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: The primal view

The GMP is the infinite-dimensional LP:

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

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

GMP: The dual view

The DUAL GMP* is the infinite-dimensional LP:

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

GMP: The primal view

The GMP is the infinite-dimensional LP:

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

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

GMP: The dual view

The DUAL GMP* is the infinite-dimensional LP:

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

And one can see that ...

the constraints of GMP^* state that the functions

$$\mathbf{x} \mapsto f_i(\mathbf{x}) - \sum_{j \in J} \lambda_j \, h_{ij}(\mathbf{x})$$

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

Several examples will follow .... and

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 $\int f \, d\mu = f(\mathbf{x})$.

Several examples will follow .... and

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 $\int f \, d\mu = f(\mathbf{x})$.

Several examples will follow .... and

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 $\int f \, d\mu = f(\mathbf{x})$.

Several examples will follow .... and

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 $\int f \, d\mu = f(\mathbf{x})$.

The moment-SOS approach

consists of using Putinar's certificate in potentially any application where one has to handle a positivity constraint "$f \geq 0$ on $\mathbf{K}$" on a compact semi-algebraic set $\mathbf{K}$ (Global optimization is only one example.)

Alternatively, the moment-LP approach uses Krivine-Vasilescu-Handelman's positivity certificate (but has several drawbacks).

In many situations this amounts to

solving a HIERARCHY of :

- LINEAR PROGRAMS, or
- SEMIDEFINITE PROGRAMS

... of increasing size!.

The moment-SOS approach

consists of using Putinar's certificate in potentially any application where one has to handle a positivity constraint "$f \geq 0$ on $\mathbf{K}$" on a compact semi-algebraic set $\mathbf{K}$ (Global optimization is only one example.)

Alternatively, the moment-LP approach uses Krivine-Vasilescu-Handelman's positivity certificate (but has several drawbacks).

In many situations this amounts to

solving a HIERARCHY of :

- LINEAR PROGRAMS, or
- SEMIDEFINITE PROGRAMS

... of increasing size!.

The moment-SOS approach

consists of using Putinar's certificate in potentially any application where one has to handle a positivity constraint "$f \geq 0$ on $\mathbf{K}$" on a compact semi-algebraic set $\mathbf{K}$ (Global optimization is only one example.)

Alternatively, the moment-LP approach uses Krivine-Vasilescu-Handelman's positivity certificate (but has several drawbacks).

In many situations this amounts to

solving a HIERARCHY of :

- LINEAR PROGRAMS, or
- SEMIDEFINITE PROGRAMS

... of increasing size!.

The moment-SOS approach

consists of using Putinar's certificate in potentially any application where one has to handle a positivity constraint "$f \geq 0$ on $\mathbf{K}$" on a compact semi-algebraic set $\mathbf{K}$ (Global optimization is only one example.)

Alternatively, the moment-LP approach uses Krivine-Vasilescu-Handelman's positivity certificate (but has several drawbacks).

In many situations this amounts to

solving a HIERARCHY of :

- LINEAR PROGRAMS, or
- SEMIDEFINITE PROGRAMS

... of increasing size!.

SDP-hierarchy 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 $d \in \mathbb{N}$:

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

Theorem

The sequence (f^*_d), $d \in \mathbb{N}$, is MONOTONE NON DECREASING and when $\mathbf{K}$ is compact then:

$$f^* = \lim_{d \to \infty} f^*_d \qquad \text{and finite convergence is generic.}$$

SDP-hierarchy 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 $d \in \mathbb{N}$:

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

Theorem

The sequence (f_d^*), $d \in \mathbb{N}$, is MONOTONE NON DECREASING and when $\mathbf{K}$ is compact then:

$$f^* = \lim_{d \to \infty} f_d^* \qquad \text{and finite convergence is generic.}$$

- 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, 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)

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, 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)

which BENEFIT from interactions!

• As mentioned ... potential applications are ENDLESS!

Recall that the SDP- hierarchy is a

GENERAL PURPOSE METHOD ....

NOT TAILORED to solving specific hard problems!!

Recall that the SDP- hierarchy is a

GENERAL PURPOSE METHOD ....
NOT TAILORED to solving specific hard problems!!

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 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.

Global minimizers are obtained from an optimal solution of the SDP, via a simple linear algebra routine.

A remarkable property: II

FINITE CONVERGENCE of the SOS-hierarchy is GENERIC!

... and provides a GLOBAL OPTIMALITY CERTIFICATE.

Global minimizers are obtained from an optimal solution of the SDP, via a simple linear algebra routine.

The no-free lunch rule ...

The size of SDP-relaxations grows rapidly with the original problem size ... In particular:

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

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

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

The no-free lunch rule ...

The size of SDP-relaxations grows rapidly with the original problem size ... In particular:

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

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

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

The no-free lunch rule ...

The size of SDP-relaxations grows rapidly with the original problem size ... In particular:

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

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

→ .... How to handle larger size problems ?

The no-free lunch rule ...

The size of SDP-relaxations grows rapidly with the original problem size ... In particular:

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

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

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

The no-free lunch rule ...

The size of SDP-relaxations grows rapidly with the original problem size ... In particular:

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

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

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

Exploit SPARSITY in the data!

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

☞ Yields a SPARSE VARIANT of the SOS-hierarchy where

- Convergence to the global optimum is preserved.
- Finite Convergence for the class of SOS-convex problems.

☞ Can solve Sparse non-convex quadratic problems with more than 2000 variables.

Exploit SPARSITY in the data!

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

☞ Yields a SPARSE VARIANT of the SOS-hierarchy where

- Convergence to the global optimum is preserved.
- Finite Convergence for the class of SOS-convex problems.

☞ Can solve Sparse non-convex quadratic problems with more than 2000 variables.

Exploit SPARSITY in the data!

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

☞ Yields a SPARSE VARIANT of the SOS-hierarchy where

- Convergence to the global optimum is preserved.
- Finite Convergence for the class of SOS-convex problems.

☞ Can solve Sparse non-convex quadratic problems with more than 2000 variables.

There has been also recent 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. ☞ Hierarchy of LP or SOCP programs.
- Lasserre, Toh and Zhang ☞ Hierarchy of SDP with semidefinite constraint of fixed size

EXAMPLES

I. 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:

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

where $\mathbf{B} \subset \mathbb{R}^n$ is a box $[-a, a]^n$.

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.

Using Putinar's positivity certificate one may build up a hierarchy of SDPs whose sizes increase with d, and whose optimal solution if the vector of coefficients of a polynomial $\mathbf{x} \mapsto J_d^*(\mathbf{x})$ of degree $2d$.

Theorem (Lass)

The associated level set $\Theta_k^* := \{\mathbf{x} \in \mathbf{B} : J_k^*(x) \leq 0\}$ *satisfies:*

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

Using Putinar's positivity certificate one may build up a hierarchy of SDPs whose sizes increase with d, and whose optimal solution if the vector of coefficients of a polynomial $\mathbf{x} \mapsto J_d^*(\mathbf{x})$ of degree $2d$.

Theorem (Lass)

The associated level set $\Theta_k^* := \{\mathbf{x} \in \mathbf{B} : \; J_k^*(x) \, \leq \, 0 \,\}$ *satisfies:*

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

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 $\mathbf{B}$ 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

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

Illustrative example (continued)

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

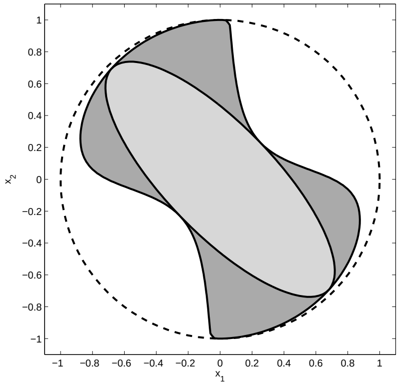
$$R_{\mathbf{A}} := \left\{ \mathbf{x} \in \mathbf{B} \;:\; \mathbf{A}(\mathbf{x}) \left(= \begin{bmatrix} 1 - 16x_1x_2 & x_1 \\ x_1 & 1 - x_1^2 - x_2^2 \end{bmatrix} \right) \succeq 0 \right\}$$

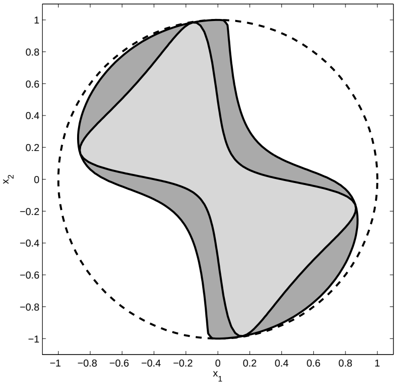
Then by solving relatively simple semidefinite programs, one may approximate $R_{\mathbf{A}}$ with sublevel sets of the form:

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

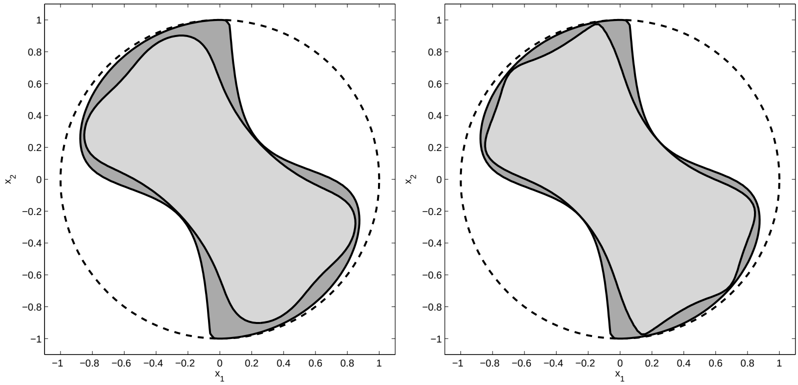
for some polynomial J_k^* of degree $k = 2, 4, \ldots$ and with

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





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).

II. Convex Underestimators of Polynomials

Consider the generic problem:

Compute a "tight" convex polynomial underestimator $p \leq f$ of a non convex polynomial f on a box $\mathbf{B} \subset \mathbb{R}^n$.

☞ Very useful in large scale MINLP to compute efficiently LOWER BOUNDS at the nodes of a BRANCH & BOUND search tree (One minimizes the convex p instead of the non-convex f).

Message:

"Good" CONVEX POLYNOMIAL UNDERESTIMATORS can be computed efficiently!

II. Convex Underestimators of Polynomials

Consider the generic problem:

Compute a "tight" convex polynomial underestimator $p \leq f$ of a non convex polynomial f on a box $\mathbf{B} \subset \mathbb{R}^n$.

☞ Very useful in large scale MINLP to compute efficiently LOWER BOUNDS at the nodes of a BRANCH & BOUND search tree (One minimizes the convex p instead of the non-convex f).

Message:

"Good" CONVEX POLYNOMIAL UNDERESTIMATORS can be computed efficiently!

I: Characterizing convex polynomial underestimators

1. $p(\mathbf{x}) \leq f(\mathbf{x})$ for every $\mathbf{x} \in \mathbf{B}$.
2. p convex on $\mathbf{B} \rightarrow \nabla^2 p(\mathbf{x}) \succeq 0$ for all $\mathbf{x} \in \mathbf{B}$,

$$\Longleftrightarrow \quad \mathbf{u}^T \nabla^2 p(\mathbf{x})\,\mathbf{u} \geq 0, \; \forall (\mathbf{x}, \mathbf{u}) \in \mathbf{B} \times \mathbf{U},$$

where $\mathbf{U} := \{\mathbf{u} : \|\mathbf{u}\|^2 \leq 1\}$.

☞ Hence with $d \in \mathbb{N}$ fixed, one would like to solve:

$$\min_{p \in \mathbb{R}[\mathbf{x}]_d} \{ \|f - p\|_{\mathbf{B}} \text{ under the two "Positivity constraints" :}$$

$$f(\mathbf{x}) - p(\mathbf{x}) \geq 0, \forall \mathbf{x} \in \mathbf{B}; \quad \mathbf{u}^T \nabla^2 p(\mathbf{x})\,\mathbf{u} \geq 0, \forall (\mathbf{x}, \mathbf{u}) \in \mathbf{B} \times \mathbf{U} \}.$$

which as an optimal solution $p^* \in \mathbb{R}[\mathbf{x}]_d$, the best convex polynomial underestimator of degree d.

I: Characterizing convex polynomial underestimators

1. $p(\mathbf{x}) \leq f(\mathbf{x})$ for every $\mathbf{x} \in \mathbf{B}$.
2. p convex on $\mathbf{B} \to \nabla^2 p(\mathbf{x}) \succeq 0$ for all $\mathbf{x} \in \mathbf{B}$,

$$\Longleftrightarrow \quad \mathbf{u}^T \nabla^2 p(\mathbf{x})\, \mathbf{u} \,\geq\, 0,\ \forall (\mathbf{x}, \mathbf{u}) \in \mathbf{B} \times \mathbf{U},$$

where $\mathbf{U} := \{\, \mathbf{u} : \|\mathbf{u}\|^2 \leq 1 \}$.

☞ Hence with $d \in \mathbb{N}$ fixed, one would like to solve:

$$\min_{p \in \mathbb{R}[\mathbf{x}]_d} \{\, \|f - p\|_{\mathbf{B}} \text{ under the two "Positivity constraints"} :$$

$$f(\mathbf{x}) - p(\mathbf{x}) \,\geq\, 0,\ \forall \mathbf{x} \in \mathbf{B}; \quad \mathbf{u}^T \nabla^2 p(\mathbf{x})\, \mathbf{u} \,\geq\, 0,\ \forall (\mathbf{x}, \mathbf{u}) \in \mathbf{B} \times \mathbf{U} \}.$$

which as an optimal solution $p^* \in \mathbb{R}[\mathbf{x}]_d$, the best convex polynomial underestimator of degree d.

I: Characterizing convex polynomial underestimators

1. $p(\mathbf{x}) \leq f(\mathbf{x})$ for every $\mathbf{x} \in \mathbf{B}$.
2. p convex on $\mathbf{B} \rightarrow \nabla^2 p(\mathbf{x}) \succeq 0$ for all $\mathbf{x} \in \mathbf{B}$,

$$\Longleftrightarrow \quad \mathbf{u}^T \nabla^2 p(\mathbf{x})\,\mathbf{u} \geq 0,\ \forall (\mathbf{x}, \mathbf{u}) \in \mathbf{B} \times \mathbf{U},$$

where $\mathbf{U} := \{\mathbf{u} : \|\mathbf{u}\|^2 \leq 1\}$.

☞ Hence with $d \in \mathbb{N}$ fixed, one would like to solve:

$$\min_{p \in \mathbb{R}[\mathbf{x}]_d} \ \{ \|f - p\|_{\mathbf{B}} \text{ under the two "Positivity constraints" :}$$

$$f(\mathbf{x}) - p(\mathbf{x}) \geq 0,\ \forall\, \mathbf{x} \in \mathbf{B}; \quad \mathbf{u}^T \nabla^2 p(\mathbf{x})\,\mathbf{u} \geq 0,\ \forall (\mathbf{x}, \mathbf{u}) \in \mathbf{B} \times \mathbf{U} \}.$$

which as an optimal solution $p^* \in \mathbb{R}[\mathbf{x}]_d$, the best convex polynomial underestimator of degree d.

Again, for fixed d, one may build up a hierarchy of SDPs whose associated sequence of optimal solutions are polynomials $(p^*_\ell)_{\ell\in\mathbb{N}}$, each of degree d, with $p_\ell \leq f$ on **B**, and p^*_ℓ is CONVEX on **B**. Moreover:

Theorem (Lass & T. Phan Thanh (JOGO 2013))

$p^*_\ell \to p^* \in \mathbb{R}[\mathbf{x}]_d$, as $\ell \to \infty$

$\rightarrow$ Provides the best results in the comparison:

Guzman, Y. A; Hasan, M. M. F.; Floudas, C. A: *Computational Comparison of Convex Underestimators for Use in a Branch-and-Bound Global Optimization Framework*, Optimization in Science and Engineering; Springer, 2014; pp 229-246.

III. Super Resolution

Suppose that an unknown SIGNED measure ϕ^* (signal) is supported on finitely many atoms $(\mathbf{x}(k))_{k=1}^p \subset \mathbf{K}$, i.e.,

$$\phi^* = \sum_{k=1}^p \gamma_k \, \delta_{\mathbf{x}(k)}, \quad \text{for some real numbers } (\gamma_k).$$

The goal is to find

the SUPPORT $(\mathbf{x}(k))_{k=1}^p \subset \mathbf{K}$ and WEIGHTS $(\gamma_k)_{k=1}^p$ from only FINITELY MANY MEASUREMENTS (moments)

$$q_\alpha = \int_{\mathbf{K}} \mathbf{x}^\alpha \, d\phi^*(\mathbf{x}), \quad \alpha \in \Gamma.$$

Solve the infinite-dimensional LP

$$\mathbf{P}: \quad \inf_{\phi} \{ \|\phi\|_{TV} : \int_{\mathbf{K}} \mathbf{x}^\alpha \, d\phi(\mathbf{x}) = q_\alpha, \quad \alpha \in \Gamma.$$

Univariate case on a bounded interval $I \subset \mathbb{R}$: If the distance between any two $\mathbf{x}(k)$'s is sufficiently large then exact recovery is obtained by solving a single SDP.

☞ Candès & Fernandez-Granda: Comm. Pure & Appl. Math. (2013)

Writing the signed measure ϕ on I as $\phi^+ - \phi^-$, **P** reads

$$\inf_{\phi^+,\phi^-} \underbrace{\int_I d(\phi^+ + \phi^-)}_{y_0+z_0} : \underbrace{\int_I \mathbf{x}^k \, d\phi^+(\mathbf{x})}_{y_k} - \underbrace{\int_I \mathbf{x}^k \, d\phi^+(\mathbf{x})}_{z_k} = q_k, \quad k = 1, \ldots, r$$

... again an instance of the GMP!

The dual **P*** reads: $\sup_{p \in \mathbb{R}[\mathbf{x}]_r} \{ \langle p, q \rangle : \sup_{\mathbf{x} \in I} |p(\mathbf{x})| \leq 1 \}$.

☞ As we are in the univariate case, both **P** and **P*** reduce to solving a single SDP.

Writing the signed measure ϕ on I as $\phi^+ - \phi^-$, **P** reads

$$\inf_{\phi^+,\phi^-} \underbrace{\int_I d(\phi^+ + \phi^-)}_{y_0+z_0} : \underbrace{\int_I \mathbf{x}^k\, d\phi^+(\mathbf{x})}_{y_k} - \underbrace{\int_I \mathbf{x}^k\, d\phi^+(\mathbf{x})}_{z_k} = q_k, \quad k = 1, \ldots, r$$

... again an instance of the GMP!

The dual $\mathbf{P}^*$ reads: $\sup_{p \in \mathbb{R}[\mathbf{x}]_r} \{ \langle p, q \rangle : \sup_{\mathbf{x} \in I} |p(\mathbf{x})| \leq 1 \}$.

☞ As we are in the univariate case, both **P** and $\mathbf{P}^*$ reduce to solving a single SDP.

Extension to compact semi-algebraic domains $\mathbf{K} \subset \mathbb{R}^n$ via the moment-SOS approach: FINITE RECOVERY is also possible, via a hierarchy of SDPs

☞ De Castro, Gamboa, Henrion & Lasserre: IEEE Trans. Info. Theory (2016).

IV. Bounds on measures with moment conditions

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

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

.... is equivalent to solving:

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

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

IV. Bounds on measures with moment conditions

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

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

.... is equivalent to solving:

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

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

IV. Bounds on measures with moment conditions

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

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

.... is equivalent to solving:

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

- $M(\mathbf{K})$ is the (convex) set of **probability measures** on $\mathbf{K} \subseteq \mathbb{R}^n$.

- $f_\alpha \equiv X^\alpha$, $\alpha \in \Gamma$ (polynomial); $f_0 = \mathrm{I}_S$ (piecewise polynomial)

Assume that $\Gamma \subset \mathbb{N}^n_d$. Then the dual of **P** reads:

$$\mathbf{P}^*: \quad \rho^* = \inf_{p_\alpha} \left\{ \sum_{\alpha\in\Gamma} p_\alpha\, \gamma_\alpha : \quad p \geq 1 \text{ on } S; \quad p \geq 0 \text{ on } \mathbf{K} \right\}$$

where $p \in \mathbb{R}[\mathbf{x}]_d$ is a polynomial

$$\mathbf{x} \mapsto p(\mathbf{x}) = \sum_{\alpha\in\mathbb{N}^n_d} p_\alpha\, \mathbf{x}^\alpha; \qquad p_\alpha = 0 \quad \forall \alpha \in \mathbb{N}^n_d \setminus \Gamma.$$

The moment-SOS approach: II. The (dual) SOS-side

☞ REPLACE the positivity constraints

$$p - 1 \geq 0 \text{ on } S; \quad p \geq 0 \text{ on } \mathbf{K}$$

with Putinar's positivity certificates of increasing degree

Assume that $\Gamma \subset \mathbb{N}^n_d$. Then the dual of **P** reads:

$$\mathbf{P}^* : \quad \rho^* = \inf_{p_\alpha} \left\{ \sum_{\alpha\in\Gamma} p_\alpha \, \gamma_\alpha : \quad p \geq 1 \text{ on } S; \quad p \geq 0 \text{ on } \mathbf{K} \right\}$$

where $p \in \mathbb{R}[\mathbf{x}]_d$ is a polynomial

$$\mathbf{x} \mapsto p(\mathbf{x}) = \sum_{\alpha\in\mathbb{N}^n_d} p_\alpha \, \mathbf{x}^\alpha; \qquad p_\alpha = 0 \quad \forall \alpha \in \mathbb{N}^n_d \setminus \Gamma.$$

The moment-SOS approach: II. The (dual) SOS-side

☞ REPLACE the positivity constraints

$$p - 1 \geq 0 \text{ on } S; \quad p \geq 0 \text{ on } \mathbf{K}$$

with Putinar's positivity certificates of increasing degree

Assume that $\Gamma \subset \mathbb{N}^n_d$. Then the dual of **P** reads:

$$\mathbf{P}^* : \quad \rho^* = \inf_{p_\alpha} \left\{ \sum_{\alpha \in \Gamma} p_\alpha \, \gamma_\alpha : \quad p \geq 1 \text{ on } S; \quad p \geq 0 \text{ on } \mathbf{K} \right\}$$

where $p \in \mathbb{R}[\mathbf{x}]_d$ is a polynomial

$$\mathbf{x} \mapsto p(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}^n_d} p_\alpha \, \mathbf{x}^\alpha; \qquad p_\alpha = 0 \quad \forall \alpha \in \mathbb{N}^n_d \setminus \Gamma.$$

The moment-SOS approach: II. The (dual) SOS-side

☞ REPLACE the positivity constraints

$$p - 1 \geq 0 \text{ on } S; \quad p \geq 0 \text{ on } \mathbf{K}$$

with Putinar's positivity certificates of increasing degree

☞ One ends up in solving the hierarchy of semidefinite programs of increasing size, indexed by $t \in \mathbb{N}$, and such that the associated sequence of optimal values $(\rho_t)_{t\in\mathbb{N}}$ converges to $\rho = \rho^*$.

V. Computing the volume of semi algebraic sets

Let $S \subset \mathbb{R}^n$ be a compact basic semi-algebraic set. Let $\mathbf{K}$ be a BOX $[0, a]^n$ containing S and let:

$$\gamma_\alpha = \int_{\mathbf{K}} X^\alpha \, dx = \frac{a^{n+|\alpha|}}{\prod_{k=1}^n (1+\alpha_k)!}, \quad \forall \alpha \in \mathbb{N}^n$$

Theorem

The (Lebesgue) volume of the set S is obtained as:

$$\sup_{\nu,\varphi} \left\{ \int_S 1 \, d\varphi : \quad \int_S X^\alpha \, d\varphi + \int_{\mathbf{K}} X^\alpha \, d\nu = \gamma_\alpha, \quad \alpha \in \mathbb{N}^n \right\}$$

☞ The unique optimal solution φ^* is the Lebesgue measure on S.

V. Computing the volume of semi algebraic sets

Let $S \subset \mathbb{R}^n$ be a compact basic semi-algebraic set. Let $\mathbf{K}$ be a BOX $[0, a]^n$ containing S and let:

$$\gamma_\alpha = \int_{\mathbf{K}} X^\alpha \, dx = \frac{a^{n+|\alpha|}}{\prod_{k=1}^n (1+\alpha_k)!}, \quad \forall \alpha \in \mathbb{N}^n$$

Theorem

The (Lebesgue) volume of the set S is obtained as:

$$\sup_{\nu,\varphi} \left\{ \int_S 1 \, d\varphi : \quad \int_S X^\alpha \, d\varphi + \int_{\mathbf{K}} X^\alpha \, d\nu = \gamma_\alpha, \quad \alpha \in \mathbb{N}^n \right\}$$

☞ The unique optimal solution φ^* is the Lebesgue measure on S.

V. Computing the volume of semi algebraic sets

Let $S \subset \mathbb{R}^n$ be a compact basic semi-algebraic set. Let $\mathbf{K}$ be a BOX $[0, a]^n$ containing S and let:

$$\gamma_\alpha = \int_{\mathbf{K}} X^\alpha \, dx = \frac{a^{n+|\alpha|}}{\prod_{k=1}^n (1+\alpha_k)!}, \quad \forall \alpha \in \mathbb{N}^n$$

Theorem

The (Lebesgue) volume of the set S is obtained as:

$$\sup_{\nu,\varphi} \left\{ \int_S 1 \, d\varphi \; : \quad \int_S X^\alpha \, d\varphi + \int_{\mathbf{K}} X^\alpha \, d\nu = \gamma_\alpha, \quad \alpha \in \mathbb{N}^n \right\}$$

☞ The unique optimal solution φ^* is the Lebesgue measure on S.

Same methodology

The only difference is that we now have COUNTABLY MANY moments constraints

☞ Henrion D., Lasserre J.B., Savorgnan C. (2009) *Approximate volume and integration for basic semi-algebraic sets*. SIAM Review 51, pp. 722–743.

VI. Gaussian measures of semi-algebraic sets

Let μ be the Gaussian measure on $\mathbb{R}^n$ with density $\mathbf{x} \mapsto \exp(-\|\mathbf{x}\|^2)$ and let $\mathbf{K} \subset \mathbb{R}^n$ be the non necessarily compact basic semi-algebraic set

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

Goal:

Approximate $\mu(\mathbf{K})$ as closely as desired

☞ Can be difficult even in small dimension $n = 2, 3$.

VI. Gaussian measures of semi-algebraic sets

Let μ be the Gaussian measure on $\mathbb{R}^n$ with density $\mathbf{x} \mapsto \exp(-\|\mathbf{x}\|^2)$ and let $\mathbf{K} \subset \mathbb{R}^n$ be the non necessarily compact basic semi-algebraic set

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

Goal:

Approximate $\mu(\mathbf{K})$ as closely as desired

☞ Can be difficult even in small dimension $n = 2, 3$.

Theorem (Lass 2015)

Let $f \in \mathbb{R}[\mathbf{x}]$ *be strictly positive* μ*-a.e. on* $\mathbf{K}$*, and let* $M(\mathbf{K})$ *(resp.* $M(\mathbb{R}^n)$*) be the space of finite Borel measures on* $\mathbf{K}$ *(resp.* $\mathbb{R}^n$*). Then the optimization problem:*

$$f_1^* = \sup_{\nu,\phi} \left\{ \int_{\mathbf{K}} f\, d\phi : \ \phi + \nu \ = \ \mu;\ \phi \in M(\mathbf{K}),\ \nu \in M(\mathbb{R}^n) \right\},$$

has a unique optimal solution $(\phi^*, \nu^*) = (\mu_{\mathbf{K}}, \mu - \mu_{\mathbf{K}})$ *where* $\mu_{\mathbf{K}}$ *is the restriction of* μ *to* $\mathbf{K}$*, that is:*

$$\phi^*(B) \ = \ \mu_{\mathbf{K}}(B) \ = \ \mu(\mathbf{K} \cap B), \qquad \forall B \in \mathcal{B}(\mathbb{R}^n).$$

In particular, $\phi^*(\mathbf{K}) = \mu(\mathbf{K})$*, and* $f^* = \mu(\mathbf{K})$ *if* $f = 1$.

Proof

From $\phi + \nu = \mu$ one deduces $\phi \leq \mu$ and therefore

$$f^* \leq \int_{\mathbf{K}} f d\mu = \int f \, d\mu_{\mathbf{K}}.$$

On the other hand the pair $(\phi^*, \nu^*) = (\mu_{\mathbf{K}}, \mu - \mu_{\mathbf{K}})$ is a feasible solution with associated cost

$$\int_{\mathbf{K}} f \, d\phi^* = \int f \, d\mu_{\mathbf{K}},$$

which proves the optimality of (ϕ^*, ν^*).

Uniqueness is more delicate. Assume there is another optimal solution $\phi, \nu)$. From $\phi \leq \mu$ one deduces $\phi \ll \mu$ and so by Radon-Nykodim

$$\phi(B \cap \mathbf{K}) = \int_{B \cap \mathbf{K}} g \, d\mu \leq \int_{B \cap \mathbf{K}} d\mu, \quad \forall B \in \mathcal{B}(\mathbb{R}^n),$$

for some nonnegative measurable function g. Hence $g \leq 1$.

Proof

From $\phi + \nu = \mu$ one deduces $\phi \leq \mu$ and therefore

$$f^* \leq \int_{\mathbf{K}} f d\mu = \int f \, d\mu_{\mathbf{K}}.$$

On the other hand the pair $(\phi^*, \nu^*) = (\mu_{\mathbf{K}}, \mu - \mu_{\mathbf{K}})$ is a feasible solution with associated cost

$$\int_{\mathbf{K}} f \, d\phi^* = \int f \, d\mu_{\mathbf{K}},$$

which proves the optimality of (ϕ^*, ν^*).

Uniqueness is more delicate. Assume there is another optimal solution ϕ, ν). From $\phi \leq \mu$ one deduces $\phi \ll \mu$ and so by Radon-Nykodim

$$\phi(B \cap \mathbf{K}) = \int_{B \cap \mathbf{K}} g \, d\mu \leq \int_{B \cap \mathbf{K}} d\mu, \quad \forall B \in \mathcal{B}(\mathbb{R}^n),$$

for some nonnegative measurable function g. Hence $g \leq 1$, On the other hand, by optimality of ϕ^* and ϕ,

$$f^* = \int_{\mathbf{K}} f\, d\mu = \int f\, d\phi^* = \int f\, d\phi = \int_{\mathbf{K}} f\, g\, d\mu$$

which implies

$$0 = \int_{\mathbf{K}} f(1-g)\, d\mu,$$

Combining this with $f > 0$ and $g \leq 1$ μ-a.e. on $\mathbf{K}$, yields $g = 1$, μ-a.e. on $\mathbf{K}$.

☞ This yields the desired result that $\phi = \phi^*$. $\square$

A dual view

A possible dual for the above LP is the LP:

$$\rho^* = \inf_{p\in\mathbb{R}[\mathbf{x}]} \{\int_{\mathbf{K}} p\, d\mu \,:\; p \geq f \text{ on } \mathbf{K};\; p \geq 0 \text{ on } \mathbb{R}^n\},$$

Indeed it trivially holds that $\rho^* \geq f^*$.

A tractable version is obtained by replacing:

- the "hard" positivity constraint $p - f \geq 0$ on $\mathbf{K}$, with the positivity-on-$\mathbf{K}$ certificate

$$p - f = \sigma_0 + \sum_{j=1}^{m} \sigma_j\, g_j; \quad \sigma_j \text{ is SOS for all } j\}.$$

- the "hard" positivity constraint $p \geq 0$ on $\mathbb{R}^n$ with p is SOS.

A dual view

A possible dual for the above LP is the LP:

$$\rho^* = \inf_{p\in\mathbb{R}[\mathbf{x}]} \{\int_{\mathbf{K}} p\, d\mu : \; p \geq f \text{ on } \mathbf{K};\; p \geq 0 \text{ on } \mathbb{R}^n\},$$

Indeed it trivially holds that $\rho^* \geq f^*$.

A tractable version is obtained by replacing:

- the "hard" positivity constraint $p - f \geq 0$ on $\mathbf{K}$, with the positivity-on-$\mathbf{K}$ certificate

$$p - f \;=\; \sigma_0 + \sum_{j=1}^{m} \sigma_j\, g_j; \quad \sigma_j \text{ is SOS for all } j\}.$$

- the "hard" positivity constraint $p \geq 0$ on $\mathbb{R}^n$ with p is SOS.

so as to obtain the hierarchy of semidefinite approximations

indexed by $d \in \mathbb{N}$:

$$\rho_d^* = \inf_{p \in \mathbb{R}[\mathbf{x}]_d} \{ \int_{\mathbb{R}^n} p \, d\mu : \; p - f \; = \; \sigma_0 + \sum_{j=1}^m \sigma_j \, g_j; \quad p, \sigma_j \text{ all SOS } \}$$

where the degree of the SOS p, σ_j is bounded by $2d$.

Theorem (Lass 2015)

For every $d \in \mathbb{N}$, $\rho_d^* \geq f^*$ and $\rho_d^* \to f^*$ as $d \to \infty$.

so as to obtain the hierarchy of semidefinite approximations

indexed by $d \in \mathbb{N}$:

$$\rho_d^* = \inf_{p \in \mathbb{R}[\mathbf{x}]_d} \left\{ \int_{\mathbb{R}^n} p \, d\mu : \; p - f \;=\; \sigma_0 + \sum_{j=1}^m \sigma_j \, g_j; \quad p, \sigma_j \text{ all SOS } \right\}$$

where the degree of the SOS p, σ_j is bounded by $2d$.

Theorem (Lass 2015)

For every $d \in \mathbb{N}$, $\rho_d^* \geq f^*$ and $\rho_d^* \; \rightarrow \; f^*$ as $d \rightarrow \infty$.

One may do the same for the complement $\mathbf{K}^c := \mathbb{R}^n \setminus \mathbf{K}$ as soon as one can write

$$\mathbf{K}^c = \bigcup_{i=1}^{p} \Omega_i; \quad \mu(\Omega_i \cap \Omega_j) = 0 \quad \forall (i,j)$$

so that $\mu(\mathbf{K}^c) = \sum_{i=1}^{p} \mu(\Omega_i)$. In doing so one obtains for each $i = 1, \ldots, p$ a sequence $(\theta_{id})_{d \in \mathbb{N}}$ such that

$$\sum_{i=1}^{p} \theta_{id} \geq \mu(\mathbf{K}^c) \quad \text{and} \quad \lim_{d\to\infty} \sum_{i=1}^{p} \theta_{id} = \mu(\mathbf{K}^c) = \mu(\mathbb{R}^n) - \mu(\mathbf{K}).$$

Theorem (Lass 2015)

With $f = 1$ *one obtains* $\mu(\mathbb{R}^n) - \underbrace{\sum_{j=1}^{p} \theta_{id}}_{\omega_d^*} \leq \mu(\mathbf{K}) \leq \rho_d^*$ *for all* d*,*

and

$$\lim \omega_d^* = \mu(\mathbf{K}) = \lim \rho_d^*$$

One may do the same for the complement $\mathbf{K}^c := \mathbb{R}^n \setminus \mathbf{K}$ as soon as one can write

$$\mathbf{K}^c \,=\, \bigcup_{i=1}^{p} \Omega_i; \quad \mu(\Omega_i \cap \Omega_j) \,=\, 0 \quad \forall (i,j)$$

so that $\mu(\mathbf{K}^c) = \sum_{i=1}^p \mu(\Omega_i)$. In doing so one obtains for each $i = 1, \ldots, p$ a sequence $(\theta_{id})_{d\in\mathbb{N}}$ such that

$$\sum_{i=1}^{p} \theta_{id} \,\geq\, \mu(\mathbf{K}^c) \quad \text{and} \quad \lim_{d\to\infty} \sum_{i=1}^{p} \theta_{id} = \mu(\mathbf{K}^c) \,=\, \mu(\mathbb{R}^n) - \mu(\mathbf{K}).$$

Theorem (Lass 2015)

With $f = 1$ *one obtains* $\underbrace{\mu(\mathbb{R}^n) - \sum_{j=1}^{p} \theta_{id}}_{\omega_d^*} \leq \mu(\mathbf{K}) \leq \rho_d^*$ *for all* d*,*

and

$$\lim \omega_d^* \,=\, \mu(\mathbf{K}) \,=\, \lim \rho_d^*$$

Examples

Let $n = 2$, and $d\mu = \exp(-\|\mathbf{x}\|^2/\sigma)\, d\mathbf{x}$ and let $\mathbf{K}$ be the non-convex quadratic

$$\mathbf{x} \mapsto \mathbf{x}^T \mathbf{A} \mathbf{x} = 0.56\, x_1^2 + 0.96\, x_1 x_2 - 1.24\, x_2^2.$$

$$\mathbf{K} = \{(x, y) : (\mathbf{x} - \mathbf{u})^T \mathbf{A}\, (\mathbf{x} - \mathbf{u}) \leq 1\} \quad \text{(non-compact)},$$

$$\text{with} \quad \mathbf{u} \; = \; (0.1, 0.5) \text{ and } (0.5, 0.1).$$

	$\mathbf{u} = (0.5, 0.1)$		
σ	ρ_9^*	ω_9^*	$100\,(\rho^*9 - \omega_9^*)/\omega_9^*$
1	2.829605	2.824718	0.17%
0.8	1.876731	1.876609	0.006%
	$\mathbf{u} = (0.1, 0.5)$		
σ	ρ_9^*	ω_9^*	$100\,(\rho_9^* - \omega_9^*)/\omega_9^*$
1	2.989832	2.986599	0.10%
0.8	1.969188	1.969103	0.004%

Examples

Let $n = 2$, and $d\mu = \exp(-\|\mathbf{x}\|^2/\sigma)\, d\mathbf{x}$ and let $\mathbf{K}$ be the non-convex quadratic

$$\mathbf{x} \mapsto \mathbf{x}^T \mathbf{A} \mathbf{x} = 0.56\, x_1^2 + 0.96\, x_1 x_2 - 1.24\, x_2^2.$$

$$\mathbf{K} = \{(x, y) : (\mathbf{x} - \mathbf{u})^T \mathbf{A}\, (\mathbf{x} - \mathbf{u}) \leq 1\} \quad \text{(non-compact)},$$

with $\mathbf{u} = (0.1, 0.5)$ and $(0.5, 0.1)$.

	$\mathbf{u} = (0.5, 0.1)$		
σ	ρ_9^*	ω_9^*	$100\,(\rho^*9 - \omega_9^*)/\omega_9^*$
1	2.829605	2.824718	0.17%
0.8	1.876731	1.876609	0.006%
	$\mathbf{u} = (0.1, 0.5)$		
σ	ρ_9^*	ω_9^*	$100\,(\rho_9^* - \omega_9^*)/\omega_9^*$
1	2.989832	2.986599	0.10%
0.8	1.969188	1.969103	0.004%

More details and (non-compact) examples in arXiv:1508.06132.

Conclusion

- Provides a sequence of converging upper and lower bounds on $\mu(\mathbf{K})$ for non necessarily compact basic semi-algebraic sets $\mathbf{K}$.
- A general methodology not set-$\mathbf{K}$-dependent.
- Also works for the exponential measure on the positive orthant $\mathbb{R}^n_+$, and in fact any measure μ provided that it satisfies Carleman's condition and one knows all its moments.

... but of course ...

With rough basic implementation and present state-of-the-art SDP solvers, one can obtain a few upper and lower bounds only and for dimension $n = 2$ or $n = 3$. For $d \geq 15$ numerical accuracy problems show up.

- ☞ Some non-trivial tricks (based on Stokes' formula) permit to improve the quality of bounds.
- ☞ Much remains to be done for a better implementation

... but of course ...

With rough basic implementation and present state-of-the-art SDP solvers, one can obtain a few upper and lower bounds only and for dimension $n = 2$ or $n = 3$. For $d \geq 15$ numerical accuracy problems show up.

- ☞ Some non-trivial tricks (based on Stokes' formula) permit to improve the quality of bounds.
- ☞ Much remains to be done for a better implementation

... but of course ...

With rough basic implementation and present state-of-the-art SDP solvers, one can obtain a few upper and lower bounds only and for dimension $n = 2$ or $n = 3$. For $d \geq 15$ numerical accuracy problems show up.

- ☞ Some non-trivial tricks (based on Stokes' formula) permit to improve the quality of bounds.
- ☞ Much remains to be done for a better implementation

VII. Lebesgue decomposition in action

Given two measures μ and ν on $\mathbb{R}^n$,

one would like to approximate the Lebesgue decomposition

$$\phi + \psi = \mu; \quad \phi \ll \nu; \quad \psi \perp \nu,$$

of μ with respect to ν.

... based on the sole knowledge of the moments

$$y_\alpha = \int_{\mathbb{R}^n} \mathbf{x}^\alpha \, d\mu, \quad z_\alpha = \int_{\mathbb{R}^n} \mathbf{x}^\alpha \, d\nu, \quad \alpha \in \mathbb{N}^n.$$

of μ and ν.

VII. Lebesgue decomposition in action

Given two measures μ and ν on $\mathbb{R}^n$,

one would like to approximate the Lebesgue decomposition

$$\phi + \psi \,=\, \mu; \quad \phi \ll \nu; \quad \psi \perp \nu,$$

of μ with respect to ν.

... based on the sole knowledge of the moments

$$y_\alpha \,=\, \int_{\mathbb{R}^n} \mathbf{x}^\alpha \, d\mu, \quad z_\alpha \,=\, \int_{\mathbb{R}^n} \mathbf{x}^\alpha \, d\nu, \quad \alpha \in \mathbb{N}^n.$$

of μ and ν.

By definition of ϕ and ψ:

ϕ has a DENSITY w.r.t. ν in $L_1(\nu)$ (called the Radon-Nikodym derivative of μ w.r.t. ν). That is, there exists a nonnegative measurable function $f \in L_1(\nu)$ such that:

$$\phi(A) = \int_A f(\mathbf{x})\, d\nu(\mathbf{x}), \qquad \forall A \in \mathcal{B}(\mathbb{R}^n).$$

CLAIM: If one assumes that :

- f is in $L_\infty(\nu)$ (instead of $L_1(\nu)$), and $\|f\|_\infty < M$ for some M,
- Both moment sequences (y_α) and (z_α), $\alpha \in \mathbb{N}^n$ satisfy Carleman's condition:

$$+\infty = \sum_{k=1}^{\infty} \left(\int X_i^{2k} \, d\mu \right)^{-1/2k} = \sum_{k=1}^{\infty} \left(\int X_i^{2k} \, d\nu \right)^{-1/2k}$$

for all $i = 1, \ldots, n$.

THEN ... one may approximate as closely as desired any fixed set of moments of ϕ and ψ.

CLAIM: If one assumes that :

- f is in $L_\infty(\nu)$ (instead of $L_1(\nu)$), and $\|f\|_\infty < M$ for some M,
- Both moment sequences (y_α) and (z_α), $\alpha \in \mathbb{N}^n$ satisfy Carleman's condition:

$$+\infty = \sum_{k=1}^{\infty} \left(\int X_i^{2k} \, d\mu \right)^{-1/2k} = \sum_{k=1}^{\infty} \left(\int X_i^{2k} \, d\nu \right)^{-1/2k}$$

for all $i = 1, \ldots, n$.

THEN ... one may approximate as closely as desired any fixed set of moments of ϕ and ψ.

CLAIM: If one assumes that :

- f is in $L_\infty(\nu)$ (instead of $L_1(\nu)$), and $\|f\|_\infty < M$ for some M,
- Both moment sequences (y_α) and (z_α), $\alpha \in \mathbb{N}^n$ satisfy Carleman's condition:

$$+\infty = \sum_{k=1}^{\infty} \left(\int X_i^{2k} \, d\mu \right)^{-1/2k} = \sum_{k=1}^{\infty} \left(\int X_i^{2k} \, d\nu \right)^{-1/2k}$$

for all $i = 1, \ldots, n$.

THEN ... one may approximate as closely as desired any fixed set of moments of ϕ and ψ.

A hierarchy of semidefinite approximations

Denote the moments of μ and ν by:

$$\mu_\alpha = \int \mathbf{x}^\alpha \, d\mu, \quad \nu_\alpha = \int \mathbf{x}^\alpha \, d\nu, \quad \alpha \in \mathbb{N}^n.$$

Let $\gamma > 0$ be fixed, and consider the hierarchy of semidefinite programs $\mathbf{P}_d$ indexed by $d \in \mathbb{N}$:

$$\mathbf{P}_d: \quad \rho_d = \sup_{y,u,v} y_0$$

$$\text{s.t.} \quad y_\alpha + u_\alpha = \mu_\alpha, \quad \forall \alpha \in \mathbb{N}^n_d$$

$$y_\alpha + v_\alpha = \gamma \, \nu_\alpha, \quad \forall \alpha \in \mathbb{N}^n_d$$

$$\mathbf{M}_d(y), \mathbf{M}_d(u), \mathbf{M}_d(v) \succeq 0$$

Let ϕ^* and ψ^* be the Lebesgue decomposition of μ w.r.t. ν, and let $f^* \in L_1(\nu)$ be the density of ϕ^* w.r.t. ν.

Theorem (Lass 2015)

(i) For each $d \in \mathbb{N}$, the semidefinite program has an optimal solution (y^d, u^d, v^d).

(ii) Moreover as $d \to \infty$, the triplet of sequences (y^d, u^d, v^d) converges to some triplet of sequences (y^*, u^*, v^*) where

$$y^*_\alpha = \int \mathbf{x}^\alpha \, (\gamma \wedge f^*) \, d\nu = \int \mathbf{x}^\alpha \, f^*_\gamma \, d\nu, \quad \forall \alpha \in \mathbb{N}^n.$$

with $\|f^*_\gamma\|_\infty \leq \gamma$.

(iii) So if $f^* \in L_\infty(\nu)$ with $\|f^*\|_\infty \leq \gamma$, then

$$y^*_\alpha = \int \mathbf{x}^\alpha \, d\phi, \quad \forall \alpha \in \mathbb{N}^n.$$

Let ϕ^* and ψ^* be the Lebesgue decomposition of μ w.r.t. ν, and let $f^* \in L_1(\nu)$ be the density of ϕ^* w.r.t. ν.

Theorem (Lass 2015)

(i) For each $d \in \mathbb{N}$, the semidefinite program has an optimal solution (y^d, u^d, v^d).

(ii) Moreover as $d \to \infty$, the triplet of sequences (y^d, u^d, v^d) converges to some triplet of sequences (y^*, u^*, v^*) where

$$y^*_\alpha = \int \mathbf{x}^\alpha \, (\gamma \wedge f^*) \, d\nu = \int \mathbf{x}^\alpha \, f^*_\gamma \, d\nu, \quad \forall \alpha \in \mathbb{N}^n.$$

with $\|f^*_\gamma\|_\infty \leq \gamma$.

(iii) So if $f^* \in L_\infty(\nu)$ with $\|f^*\|_\infty \leq \gamma$, then

$$y^*_\alpha = \int \mathbf{x}^\alpha \, d\phi, \quad \forall \alpha \in \mathbb{N}^n.$$

Let ϕ^* and ψ^* be the Lebesgue decomposition of μ w.r.t. ν, and let $f^* \in L_1(\nu)$ be the density of ϕ^* w.r.t. ν.

Theorem (Lass 2015)

(i) For each $d \in \mathbb{N}$, the semidefinite program has an optimal solution (y^d, u^d, v^d).

(ii) Moreover as $d \to \infty$, the triplet of sequences (y^d, u^d, v^d) converges to some triplet of sequences (y^*, u^*, v^*) where

$$y^*_\alpha = \int \mathbf{x}^\alpha \, (\gamma \wedge f^*) \, d\nu = \int \mathbf{x}^\alpha \, f^*_\gamma \, d\nu, \quad \forall \alpha \in \mathbb{N}^n.$$

with $\|f^*_\gamma\|_\infty \leq \gamma$.

(iii) So if $f^* \in L_\infty(\nu)$ with $\|f^*\|_\infty \leq \gamma$, then

$$y^*_\alpha = \int \mathbf{x}^\alpha \, d\phi, \quad \forall \alpha \in \mathbb{N}^n.$$

Let ϕ^* and ψ^* be the Lebesgue decomposition of μ w.r.t. ν, and let $f^* \in L_1(\nu)$ be the density of ϕ^* w.r.t. ν.

Theorem (Lass 2015)

(i) For each $d \in \mathbb{N}$, the semidefinite program has an optimal solution (y^d, u^d, v^d).

(ii) Moreover as $d \to \infty$, the triplet of sequences (y^d, u^d, v^d) converges to some triplet of sequences (y^*, u^*, v^*) where

$$y^*_\alpha = \int \mathbf{x}^\alpha \, (\gamma \wedge f^*) \, d\nu = \int \mathbf{x}^\alpha \, f^*_\gamma \, d\nu, \quad \forall \alpha \in \mathbb{N}^n.$$

with $\|f^*_\gamma\|_\infty \leq \gamma$.

(iii) So if $f^* \in L_\infty(\nu)$ with $\|f^*\|_\infty \leq \gamma$, then

$$y^*_\alpha = \int \mathbf{x}^\alpha \, d\phi, \quad \forall \alpha \in \mathbb{N}^n.$$

Examples

Let $n = 2$, $p \in (0, 1)$ and

- ν is the Gaussian with density $\mathbf{x} \mapsto \exp(-\|\mathbf{x}\|^2)$,
- θ is the measure uniformy distributed on the circle $\{\mathbf{x} : x_1^2 + x_2^2 = 1\}$.

Define the measure μ to be

$$\mu = p\,\nu + (1-p)\,\theta,$$

so that the Lebesgue decomposition of μ w.r.t. ν is $(\phi, \psi) = (p\,\nu, (1-p)\,\theta)$.

The table below show relative error between the approximate moments $\mathbf{u} = (u_\alpha)$ of degree 2 and 4, of the singular part ψ and those of $p\,\theta$ computed with moments up to order $2d = 14$.

approx. moments	x_1^2	x_1^4	$x_1^2 x_2^2$	$L_{\mathbf{u}^d}((x_1^2 + x_2^2 - 1)^2)$
p=0.1	0.19%	0.52%	0.53%	0.001
p=0.2	3.7%	8.12%	12.14%	0.16

Same thing but now with ν being uniformly supported on the unit box $[-1, 1]^n$.

approx. moments	x_1^2	x_1^4	$x_1^2 x_2^2$	$L_{\mathbf{u}^d}((x_1^2 + x_2^2 - 1)^2)$
p=0.1	0.26%	0.93%	0.61%	0.0001
p=0.2	8.8%	10.2%	7.5%	0.08

The table below show relative error between the approximate moments $\mathbf{u} = (u_\alpha)$ of degree 2 and 4, of the singular part ψ and those of $p\,\theta$ computed with moments up to order $2d = 14$.

approx. moments	x_1^2	x_1^4	$x_1^2 x_2^2$	$L_{\mathbf{u}^d}((x_1^2 + x_2^2 - 1)^2)$
p=0.1	0.19%	0.52%	0.53%	0.001
p=0.2	3.7%	8.12%	12.14%	0.16

Same thing but now with ν being uniformly supported on the unit box $[-1, 1]^n$.

approx. moments	x_1^2	x_1^4	$x_1^2 x_2^2$	$L_{\mathbf{u}^d}((x_1^2 + x_2^2 - 1)^2)$
p=0.1	0.26%	0.93%	0.61%	0.0001
p=0.2	8.8%	10.2%	7.5%	0.08