The Moment-SOS Hierarchy in & outside Optimization

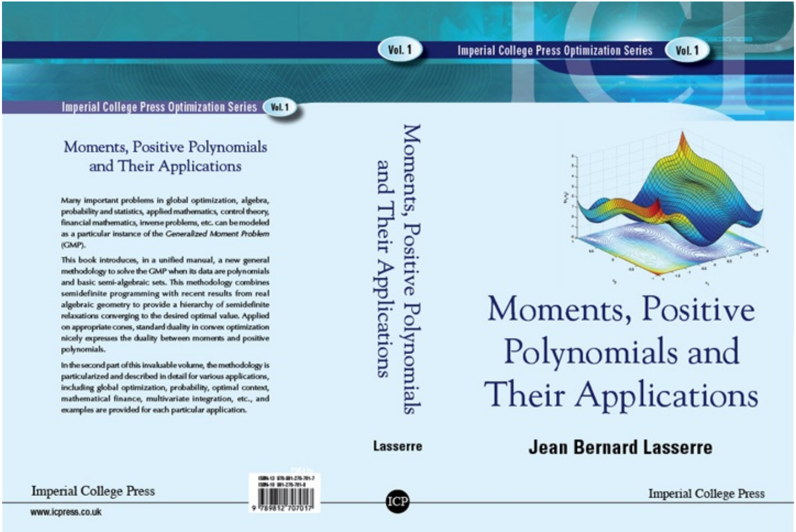
Jean B. Lasserre*

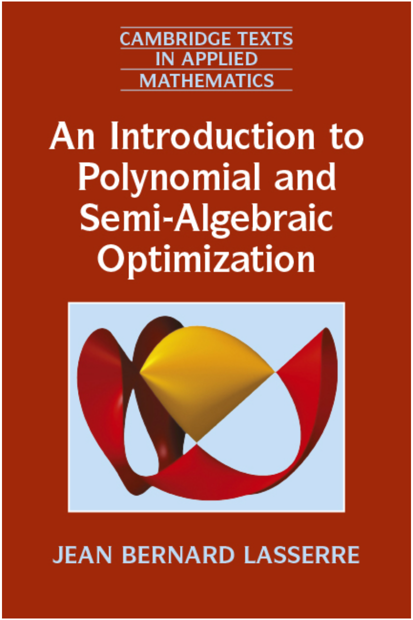
LAAS-CNRS and Institute of Mathematics, Toulouse, France

CIMI-ANITI School on Optimisation, September 2021

Research funded by ANITI institute, within the framework of the French Program “Investing for the Future C PIA3"" under grant ANR-19-PIA3-0004

ANITI
ARTIFICIAL & NATURAL INTELLIGENCE
TOULOUSE INSTITUTE


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

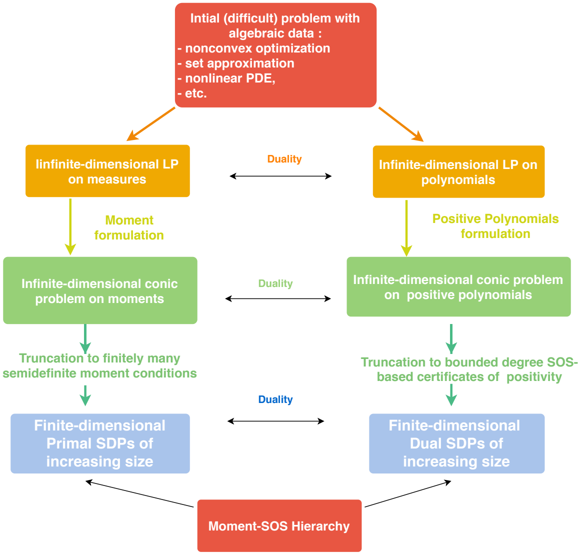

Vol. 4
Series on Optimization and its Applications – Vol. 4
The Moment-SOS Hierarchy
The Moment-SOS Hierarchy
Lectures in Probability, Statistics, Computational Geometry, Control and Nonlinear PDEs
Didier Henrion
Milan Korda
Jean B. Lasserre
Henrion
Korda
Lasserre
World Scientific
The moment-SOS hierarchy is a powerful methodology that is used to solve the Generalized Moment Problem (GMP) where the list of applications in various areas of Science and Engineering is almost endless. Initially designed for solving polynomial optimization problems (the simplest example of the GMP), it applies to solving any instance of the GMP whose description only involves semi-algebraic functions and sets. It consists of solving a sequence (a hierarchy) of convex relaxations of the initial problem, and each convex relaxation is a semidefinite program whose size increases in the hierarchy.
The goal of this book is to describe in a unified and detailed manner how this methodology applies to solving various problems in different areas ranging from Optimization, Probability, Statistics, Signal Processing, Computational Geometry, Control, Optimal Control and Analysis of a certain class of nonlinear PDEs. For each application, this unconventional methodology differs from traditional approaches and provides an unusual viewpoint. Each chapter is devoted to a particular application, where the methodology is thoroughly described and illustrated on some appropriate examples.
The exposition is kept at an appropriate level of detail to aid the different levels of readers not necessarily familiar with these tools, to better know and understand this methodology.
World Scientific
www.worldscientific.com
Q0252 hc
ISSN 2399-1593
ISBN 978-1-78634-853-1

- Illustration of the Moment-SOS hierarchy for POLYNOMIAL optimization
- LP- and SDP-based CERTIFICATES of POSITIVITY
- The moment-LP and moment-SOS hierarchies
- Some applications

- Illustration of the Moment-SOS hierarchy for POLYNOMIAL optimization
- LP- and SDP-based CERTIFICATES of POSITIVITY
- The moment-LP and moment-SOS hierarchies
- Some applications

- Illustration of the Moment-SOS hierarchy for POLYNOMIAL optimization
- LP- and SDP-based CERTIFICATES of POSITIVITY
- The moment-LP and moment-SOS hierarchies
- Some applications

- Illustration of the Moment-SOS hierarchy for POLYNOMIAL optimization
- LP- and SDP-based CERTIFICATES of POSITIVITY
- The moment-LP and moment-SOS hierarchies
- Some applications


Intial (difficult) problem with algebraic data :
- nonconvex optimization
- set approximation
- nonlinear PDE,
- etc.
Iinfinite-dimensional LP on measures
Duality
Infinite-dimensional LP on polynomials
Moment formulation
Positive Polynomials formulation
Infinite-dimensional conic problem on moments
Duality
Infinite-dimensional conic problem on positive polynomials
Truncation to finitely many semidefinite moment conditions
Truncation to bounded degree SOS-based certificates of positivity
Duality
Finite-dimensional Primal SDPs of increasing size
Finite-dimensional Dual SDPs of increasing size
Moment-SOS Hierarchy

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

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

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

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 (Dual side of Putinar's 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 $d = 0, 1, \ldots$

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

☞ The real symmetric matrix $\mathbf{M}_2(y)$ is called the MOMENT MATRIX associated with the sequence y

☞ The real symmetric matrix $\mathbf{M}_d(g_j\, \mathbf{y})$ is called the LOCALIZING MATRIX associated with the sequence y and the polynomial g_j.

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 (Dual side of Putinar's 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* $d = 0, 1, \ldots$

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

☞ The real symmetric matrix $\mathbf{M}_2(y)$ is called the MOMENT MATRIX associated with the sequence y

☞ The real symmetric matrix $\mathbf{M}_d(g_j\,\mathbf{y})$ is called the LOCALIZING MATRIX associated with the sequence y and the polynomial g_j.

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 (Dual side of Putinar's 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* $d = 0, 1, \ldots$

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

☞ The real symmetric matrix $\mathbf{M}_2(y)$ is called the MOMENT MATRIX associated with the sequence y

☞ The real symmetric matrix $\mathbf{M}_d(g_j\,\mathbf{y})$ is called the LOCALIZING MATRIX associated with the sequence y and the polynomial g_j.

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 (Dual side of Putinar's 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* $d = 0, 1, \ldots$

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

☞ The real symmetric matrix $\mathbf{M}_2(y)$ is called the MOMENT MATRIX associated with the sequence y

☞ The real symmetric matrix $\mathbf{M}_d(g_j\,\mathbf{y})$ is called the LOCALIZING MATRIX associated with the sequence y and the polynomial g_j.

Remarkably,

the Necessary & Sufficient conditions ($\star$) for existence of a representing measure are stated only in terms of countably many LMI CONDITIONS on the sequence y ! (No mention of the unknown representing measure in the conditions.)

For instance with $n = 2$, $d = 1$, the moment matrix $\mathbf{M}_2(y)$ reads

$$\mathbf{M}_2(y) = \left[\begin{array}{c|cc|ccc} \overbrace{y_{00}}^{1} & \overbrace{y_{10}}^{x_1} & \overbrace{y_{01}}^{x_2} & \overbrace{y_{20}}^{x_1^2} & \overbrace{y_{11}}^{x_1 x_2} & \overbrace{y_{02}}^{x_2^2} \\ \hline y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} \\ y_{01} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} \\ \hline y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22} \\ y_{11} & y_{12} & y_{21} & y_{31} & y_{22} & y_{13} \\ y_{02} & y_{12} & y_{03} & y_{22} & y_{13} & y_{04} \end{array}\right]$$

Remarkably,

the Necessary & Sufficient conditions $(\star)$ for existence of a representing measure are stated only in terms of countably many LMI CONDITIONS on the sequence y ! (No mention of the unknown representing measure in the conditions.)

For instance with $n = 2$, $d = 1$, the moment matrix $\mathbf{M}_2(y)$ reads

$$\mathbf{M}_2(y) = \begin{bmatrix} \overbrace{y_{00}}^{1} & | & \overbrace{y_{10}}^{X_1} & \overbrace{y_{01}}^{X_2} & | & \overbrace{y_{20}}^{X_1^2} & \overbrace{y_{11}}^{X_1X_2} & \overbrace{y_{02}}^{X_2^2} \\ - & & - & - & & - & - & - \\ y_{10} & | & y_{20} & y_{11} & | & y_{30} & y_{21} & y_{12} \\ y_{01} & | & y_{11} & y_{02} & | & y_{21} & y_{12} & y_{03} \\ - & & - & - & & - & - & - \\ y_{20} & | & y_{30} & y_{21} & | & y_{40} & y_{31} & y_{22} \\ y_{11} & | & y_{12} & y_{21} & | & y_{31} & y_{22} & y_{13} \\ y_{02} & | & y_{12} & y_{03} & | & y_{22} & y_{13} & y_{04} \end{bmatrix}$$

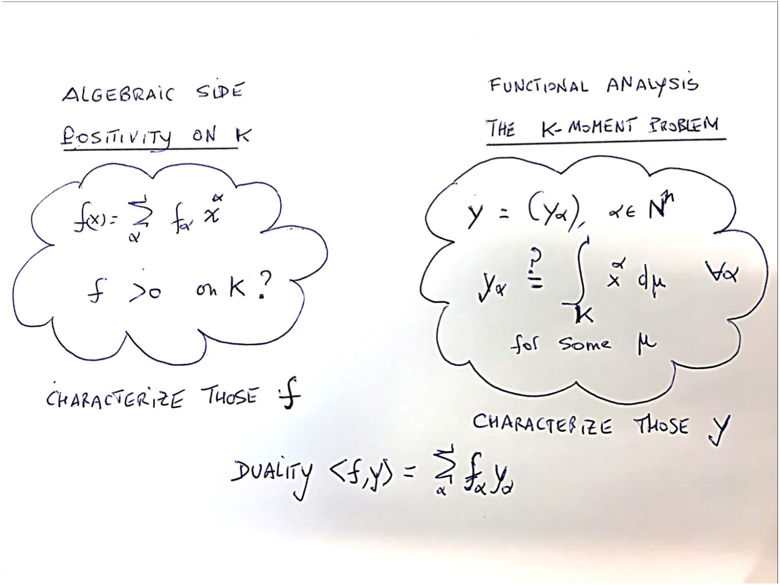
ALGEBRAIC SIDE

POSITIVITY ON K

$$f(x) = \sum_\alpha f_\alpha \, x^\alpha$$

$f > 0$ on K ?

CHARACTERIZE THOSE f


ALGEBRAIC SIDE
POSITIVITY ON K
$f(x) = \sum_\alpha f_\alpha x^\alpha$
$f > 0$ on K ?
CHARACTERIZE THOSE f
FUNCTIONAL ANALYSIS
THE K-MOMENT PROBLEM
$y = (y_\alpha),\ \alpha \in \mathbb{N}^n$
$y_\alpha \overset{?}{=} \int_K x^\alpha \, d\mu \quad \forall \alpha$
for some μ
CHARACTERIZE THOSE y
DUALITY $\langle f, y \rangle = \sum_\alpha f_\alpha y_\alpha$

There is also another ALGEBRAIC POSITIVITY CERTIFICATE due to Krivine, Vasilescu, and Handelman.

But unfortunately less powerful ... and with some drawbacks!

There is also another ALGEBRAIC POSITIVITY CERTIFICATE due to Krivine, Vasilescu, and Handelman.

But unfortunately less powerful ... and with some drawbacks!

• 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:

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

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

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

GMP: The dual view

The DUAL GMP* is the infinite-dimensional LP:

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

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

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

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

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

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

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

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 $d \in \mathbb{N}$:

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

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

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

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:

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

Moreover, and importantly,

- GENERICALLY, ... the Moment-SOS hierarchy has finite convergence, that is, $f^* = f^*_d$ for some d.
- A sufficient RANK-CONDITION on the moment matrix (which also holds GENERICALLY) permits to test whether $f^* = f^*_d$

Theorem

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

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

Moreover, and importantly,

- GENERICALLY, ... the Moment-SOS hierarchy has finite convergence, that is, $f^* = f_d^*$ for some d.
- A sufficient RANK-CONDITION on the moment matrix (which also holds GENERICALLY) permits to test whether $f^* = f_d^*$

• 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!

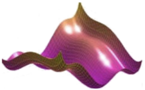
• What makes this approach exciting is that it is at the crossroads of several disciplines/applications:

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

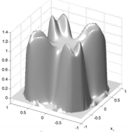
which BENEFIT from interactions!

• As mentioned ... potential applications are ENDLESS!

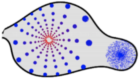
Global optimization



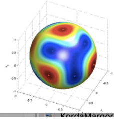
Volume of semialgebraic set



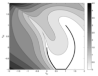
Reachable set



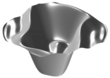
Super resolution



Optimal control

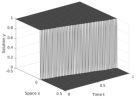


Region of attraction



Maximum invariant sets

PDE analysis & control



• 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)

The moment-SOS approach can be applied to problems defined with semi-algebraic functions via the introduction of additional variables (LIFTING)

Examples

$$\mathbf{x} \in \mathbf{K};\ |f(\mathbf{x})| \quad \Leftrightarrow \quad \mathbf{x} \in \mathbf{K};\ f(\mathbf{x})^2 - z^2 = 0; \quad z \geq 0.$$

$$f(\mathbf{x}) \geq 0 \text{ on } \mathbf{K};\ \sqrt{f(\mathbf{x})} \quad \Leftrightarrow \quad \mathbf{x} \in \mathbf{K};\ f(\mathbf{x}) - z^2 = 0; \quad z \geq 0.$$

Similarly to model the function $\mathbf{x} \mapsto g(\mathbf{x}) := \max[f_1(\mathbf{x}), f_2(\mathbf{x})]$,

$$\underbrace{(f_1(\mathbf{x}) - f_2(\mathbf{x}))^2 - z^2 = 0;\ z \geq 0}_{z=|f_1(\mathbf{x})-f_2(\mathbf{x})|} \quad \Leftrightarrow g(\mathbf{x}) = \frac{z}{2} + \frac{f_1(\mathbf{x}) + f_2(\mathbf{x})}{2}$$

(*) $\max[0, f(\mathbf{x})]$ ☞ ReLU function in ML !

etc.

Recall that both LP- and SDP- hierarchies are

GENERAL PURPOSE METHODS ....

NOT TAILORED to solving specific hard problems!!

Recall that both LP- and SDP- hierarchies are

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

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,

the analogue for the NON CONVEX CASE of the
KKT-OPTIMALITY conditions in the CONVEX CASE!

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 ?

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 ?

• develop more efficient general purpose SDP-solvers ... (limited impact) ... or perhaps dedicated solvers ....?

• exploit symmetries when present ... Recent promising works by De Klerk, Gaterman, Gvozdenovic, Laurent, Pasechnick, Parrilo, Schrijver .. in particular for combinatorial optimization problems. Algebraic techniques permit to define an equivalent SDP of much smaller size.

☞ See e.g. works in CODING and PACKING problems (Bachoc, de Laat, Oliveira de Filho, Vallentin)

• develop more efficient general purpose SDP-solvers ... (limited impact) ... or perhaps dedicated solvers ....?

• exploit symmetries when present ... Recent promising works by De Klerk, Gaterman, Gvozdenovic, Laurent, Pasechnick, Parrilo, Schrijver .. in particular for combinatorial optimization problems. Algebraic techniques permit to define an equivalent SDP of much smaller size.

☞ See e.g. works in CODING and PACKING problems (Bachoc, de Laat, Oliveira de Filho, Vallentin)

• exploit sparsity in the data. In general, each constraint involves a small number of variables and /or monomials, and the cost criterion is a sum of polynomials involving also a small number of 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

☞ Most of such extensions yield SPARSE VARIANTS 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 4000 variables.

☞ Most of such extensions yield SPARSE VARIANTS 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 4000 variables.

See e.g. works by :

- Kim, Kojima, Matamatsu ☞ Correlative-sparsity.
- Ahmadi et al. ☞ Hierarchy of LP or SOCP programs.
- Lass., Toh and Zhang ☞ Hierarchy of SDP with semidefinite constraint of fixed size
- Hoang Mai, Lass., V. Magron, J. Wang ☞ term- and/or correlative-sparsity

EXAMPLES

Suppose that we want to evaluate

the ROBUSTNESS of a GIVEN Deep Learning Network (DNN) where the activation function are polynomial or ReLU functions.

☞ à posteriori analysis since the weights of the DNN have been already determined.

Example in Classification with m labels

The input $\mathbf{x} \in \mathbb{R}^n$ and the output $F(\mathbf{x})$ of the DN is a vector $\mathbf{y} \in \mathbb{R}^m$ where

$F : \mathbb{R}^n \to \mathbb{R}^m$ is the input-output map(or score) of the DNN

and the "prediction" $\mathbf{y}(\mathbf{x})$ associated with the input $\mathbf{x}$ reads:

$$\mathbf{y}(\mathbf{x}) := \arg\max[F(\mathbf{x})_1, \ldots, F(\mathbf{x})_m] \in \{1, \ldots, m\}.$$

0. Robustness evaluation of Deep Learning Networks

Suppose that we want to evaluate

the ROBUSTNESS of a GIVEN Deep Learning Network (DNN) where the activation function are polynomial or ReLU functions.

☞ à posteriori analysis since the weights of the DNN have been already determined.

Example in Classification with m labels
The input $\mathbf{x} \in \mathbb{R}^n$ and the output $F(\mathbf{x})$ of the DN is a vector $\mathbf{y} \in \mathbb{R}^m$ where

$F : \mathbb{R}^n \to \mathbb{R}^m$ is the input-output map(or score) of the DNN

and the "prediction" $\mathbf{y}(\mathbf{x})$ associated with the input $\mathbf{x}$ reads:

$$\mathbf{y}(\mathbf{x}) := \arg\max[\, F(\mathbf{x})_1, \ldots, F(\mathbf{x})_m \,] \in \{1, \ldots, m\}.$$

Let $\varepsilon > 0$ and $\mathbf{x} \in \mathbb{R}^n$ be fixed, and let $\mathbf{B}(\mathbf{x};\varepsilon) := \{\mathbf{z} : \|\mathbf{z}\| \leq \varepsilon\}$

The DNN is ε-robust at $\mathbf{x} \in \mathbb{R}^n$ if

$$F(\mathbf{z})_i < F_{\mathbf{y}(\mathbf{x})}(\mathbf{z}), \quad \forall \mathbf{z} \in \mathbf{B}(\mathbf{x};\varepsilon), \forall i \neq \mathbf{y}(\mathbf{x}).$$

Indeed then $\mathbf{y}(\mathbf{z}) = \mathbf{y}(\mathbf{x})$ for all $\mathbf{z} \in \mathbf{B}(\mathbf{x};\varepsilon)$.

Fact

The function $x \mapsto \mathrm{ReLu}(\mathbf{x}) := \max[0, \mathbf{x}]$ is semi-algebraic, as

$$u = \mathrm{ReLU}(x) \quad \Leftrightarrow \quad u \geq x;\ u \geq 0;\ u\,(u - x) = 0.$$

Let $\varepsilon > 0$ and $\mathbf{x} \in \mathbb{R}^n$ be fixed, and let $\mathbf{B}(\mathbf{x};\varepsilon) := \{\mathbf{z} : \|\mathbf{z}\| \leq \varepsilon\}$

The DNN is ε-robust at $\mathbf{x} \in \mathbb{R}^n$ if

$$F(\mathbf{z})_i \,<\, F_{\mathbf{y}(\mathbf{x})}(\mathbf{z})\,, \quad \forall \mathbf{z} \in \mathbf{B}(\mathbf{x};\varepsilon)\,,\, \forall i \neq \mathbf{y}(\mathbf{x})\,.$$

Indeed then $\mathbf{y}(\mathbf{z}) \,=\, \mathbf{y}(\mathbf{x})$ for all $\mathbf{z} \in \mathbf{B}(\mathbf{x};\varepsilon)$.

Fact

The function $x \mapsto \mathrm{ReLu}(\mathbf{x}) := \max[0, \mathbf{x}]$ is semi-algebraic, as

$$u = \mathrm{ReLU}(x) \quad \Leftrightarrow \quad u \geq x\,;\; u \geq 0\,;\; u\,(u - x) = 0\,.$$

then checking whether

the DNN is ε-robust at $\mathbf{x} \in \mathbb{R}^n$ reduces to solving a Quadratically Constrained Quadratic Program (QCQP).

☞ Hence one may apply the Moment-SOS hierarchy to certify where the DNN is ε-robust at $\mathbf{x} \in \mathbb{R}^n$

An alternative to check robustness

is to provide a valid upper bound bound τ on the LIPSCHITZ constant of the mapping F w.r.t. a set $S \subset \mathbb{R}^n$.

Then once τ has been computed, one may check ε-robustness for all $\mathbf{x} \in S$ instead of for a single $\mathbf{x}$.

then checking whether

the DNN is ε-robust at $\mathbf{x} \in \mathbb{R}^n$ reduces to solving a Quadratically Constrained Quadratic Program (QCQP).

☞ Hence one may apply the Moment-SOS hierarchy to certify where the DNN is ε-robust at $\mathbf{x} \in \mathbb{R}^n$

An alternative to check robustness

is to provide a valid upper bound bound τ on the LIPSCHITZ constant of the mapping F w.r.t. a set $S \subset \mathbb{R}^n$.

Then once τ has been computed, one may check ε-robustness for all $\mathbf{x} \in S$ instead of for a single $\mathbf{x}$.

then checking whether

the DNN is ε-robust at $\mathbf{x} \in \mathbb{R}^n$ reduces to solving a Quadratically Constrained Quadratic Program (QCQP).

☞ Hence one may apply the Moment-SOS hierarchy to certify where the DNN is ε-robust at $\mathbf{x} \in \mathbb{R}^n$

An alternative to check robustness

is to provide a valid upper bound bound τ on the LIPSCHITZ constant of the mapping F w.r.t. a set $S \subset \mathbb{R}^n$.

Then once τ has been computed, one may check ε-robustness for all $\mathbf{x} \in S$ instead of for a single $\mathbf{x}$.

by exploiting the fact that

the sub-gradient $x \mapsto \partial \mathrm{ReLu}(x)$ is also semi-algebraic, computing τ again reduces to solving a QCQP.

☞ Hence one may also apply the Moment-SOS hierarchy to certify where the DNN is ε-robust at $\mathbf{x} \in S$.

☞ **Tong Chen, Lass J.B., Pauwels E., Magron V.** Semi-algebraic optimization for Lipschitz constants of ReLU networks, NEURIPS 2020.

☞ **Tong Chen, Lass J.B., Magron V., Pauwels E..** Semialgebraic Representation of Monotone DeepEquilibrium Models and Applications to Certification, Submitted.

I. Optimal Control

Consider the OPTIMAL CONTROL (OCP) problem:

$$\rho = \inf_{u} \int_0^T h(\mathbf{x}(t), u(t))\, dt$$

$$\begin{array}{rrcl} \text{s.t.} & \dot{\mathbf{x}}(t) & = & f(\mathbf{x}(t), u(t)), \quad t \in [0, T] \\ & \mathbf{x}(0) & = & x_0 \\ & \mathbf{x}(t) \in X \subset \mathbb{R}^n; & & u(t) \in U \subset \mathbb{R}^m, \end{array}$$

that is, the goal is now to compute a function $u : [0, T] \to \mathbb{R}^m$ (in a suitable space).

In general OCP problems are hard to solve, and particularly when STATE CONSTRAINTS $\mathbf{x}(t) \in X$ are present !

By introducing the concept of OCCUPATION MEASURE, there exists a so-called WEAK FORMULATION of the OCP which is an infinite-dimensional LINEAR PROGRAM (LP) on a suitable space of measures, and in fact an instance of the Generalized Problem of Moments.

☞ Under some conditions the optimal values of OCP and LP are the same.

☞ When the vector field f is a polynomial and the sets X and U are compact basic semi-algebraic then the MOMENT-SOS approach can be applied to approximate ρ as closely as desired.

By introducing the concept of OCCUPATION MEASURE, there exists a so-called WEAK FORMULATION of the OCP which is an infinite-dimensional LINEAR PROGRAM (LP) on a suitable space of measures, and in fact an instance of the Generalized Problem of Moments.

☞ Under some conditions the optimal values of OCP and LP are the same.

☞ When the vector field f is a polynomial and the sets X and U are compact basic semi-algebraic then the MOMENT-SOS approach can be applied to approximate ρ as closely as desired.

By introducing the concept of OCCUPATION MEASURE, there exists a so-called WEAK FORMULATION of the OCP which is an infinite-dimensional LINEAR PROGRAM (LP) on a suitable space of measures, and in fact an instance of the Generalized Problem of Moments.

☞ Under some conditions the optimal values of OCP and LP are the same.

☞ When the vector field f is a polynomial and the sets X and U are compact basic semi-algebraic then the MOMENT-SOS approach can be applied to approximate ρ as closely as desired.

☞ It yields a HIERARCHY OF SEMIDEFINITE PROGRAMS of increasing size whose associated monotone sequence of optimal values CONVERGES to the optimal value ρ of the OCP.

☞ Lass. J.B., Henrion D., Prieur C., Trelat E. (2008), Nonlinear optimal control via occupation measures and LMI-relaxations, SIAM J. Contr. Optim. 47, pp. 1649–1666.

How do we transform an OCP to a moment problem?

Illustration on an ode with no control

$$\dot{\mathbf{x}}(t) \;=\; f(\mathbf{x}(t)), \quad t \in [0, T]; \quad \mathbf{x}(0) = \mathbf{x}_0.$$

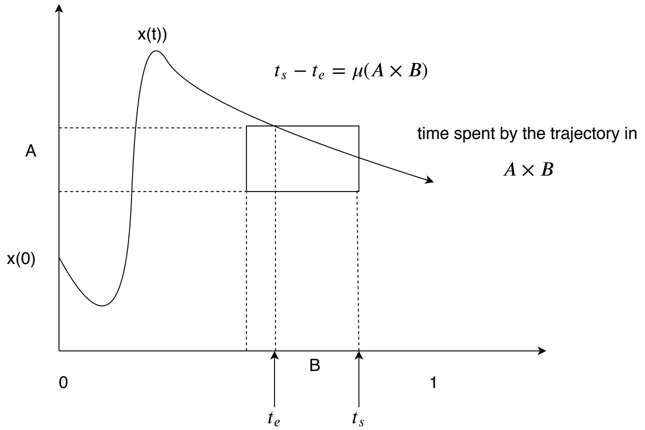
How do we characterize a feasible trajectory?

Introduce the OCCUPATION MEASURE μ up to time 1 by

$$\mu(A \times B) \;=\; \int_{B\cap[0,T]} 1_A(\mathbf{x}(t))\,dt, \quad \forall A \in \mathcal{B}(\mathbb{R}^n),\; B \in \mathcal{B}([0,T])$$

and the OCCUPATION MEASURE ν at time T by:

$$\nu(A) \;=\; 1_A(\mathbf{x}(T)) \quad \forall A \in \mathcal{B}(\mathbb{R}^n).$$


x(t))
$t_s - t_e = \mu(A \times B)$
time spent by the trajectory in
$A \times B$
A
x(0)
B
0
1
t_e
t_s

☞ Then use TEST FUNCTIONS $\phi(x,t)$!

The time-integral

$$\phi(\mathbf{x}(T),T)-\phi(\mathbf{x}(0),0) = \int_0^1 \frac{\partial\phi(\mathbf{x}(t),t)}{\partial t}+\frac{\partial\phi(\mathbf{x}(t),t)}{\partial x}f(\mathbf{x}(t),t)\,dt$$

is the same as the "spatial integral"

$$\int \phi(\mathbf{x},T)\,d\nu(\mathbf{x})-\phi(\mathbf{x}_0,0) = \int_{\mathbb{R}^n\times\mathbb{R}} \frac{\partial\phi(\mathbf{x},t)}{\partial t}+\frac{\partial\phi(\mathbf{x},t)}{\partial x}f(\mathbf{x},t)\,d\mu(\mathbf{x},t)$$

For each test function $\phi(\mathbf{x},t) := \mathbf{x}^\alpha\, t^k$

$$(\mathbf{x},t) \mapsto p(\mathbf{x},t) := \frac{\partial \phi(\mathbf{x},t)}{\partial t} + \frac{\partial \phi(\mathbf{x},t)}{\partial x} f(\mathbf{x},t)$$

is a polynomial

and therefore the constraint

$$\int \phi(\mathbf{x},T)\, d\nu(\mathbf{x}) - \phi(\mathbf{x}_0,0) = \int_{\mathbb{R}^n\times\mathbb{R}} \frac{\partial \phi(\mathbf{x},t)}{\partial t} + \frac{\partial \phi(\mathbf{x},t)}{\partial x} f(\mathbf{x},t)\, d\mu(\mathbf{x},t)$$

reads

$$T^k \int \mathbf{x}^\alpha\, d\nu(\mathbf{x}) - \mathbf{x}_0^k\, 0^k = \int_{\mathbb{R}^n\times\mathbb{R}} p(\mathbf{x},t)\, d\mu(\mathbf{x},t)$$

☞ a linear constraint on the moments of μ and ν!

The trajectory $(\mathbf{x}(t), t)$ satisfies the ODE $\dot{\mathbf{x}}(t) = f(\mathbf{x}(t))$

translates into

☞ Its occupation measures μ and ν satisfy countably many LINEAR MOMENT CONSTRAINTS

Non-linear hyperbolic PDEs

☞ The same approach works for MEASURE-VALUED SOLUTIONS for weak formulation of certain NONLINEAR HYPERBOLIC PDE's, e.g. BURGERS Equation:

$$\begin{aligned} \frac{\partial y(t,x)}{\partial t} + \frac{\partial f(y(t,x))}{\partial x} &= 0, \quad (t,x) \in \mathbb{R}_+ \times \mathbb{R} \\ y(0,x) &= y_0(x), \quad x \in \mathbb{R} \end{aligned}$$

Step-d of the Moment-SOS hierarchy aims at computing moments up to order $2d$ of the measure μ supported on the graph $\{(t,x,y(t,x)) : (t,x) \in \Omega\}$ of the solution $(t,x) \mapsto y(t,x)$.

☞ AVOIDS A DISCRETIZATION of the (t,x) domain Ω!

Non-linear hyperbolic PDEs

☞ The same approach works for MEASURE-VALUED SOLUTIONS for weak formulation of certain NONLINEAR HYPERBOLIC PDE's, e.g. BURGERS Equation:

$$\begin{aligned} \frac{\partial y(t,x)}{\partial t} + \frac{\partial f(y(t,x))}{\partial x} &= 0, \quad (t,x) \in \mathbb{R}_+ \times \mathbb{R} \\ y(0,x) &= y_0(x), \quad x \in \mathbb{R} \end{aligned}$$

Step-d of the Moment-SOS hierarchy aims at computing moments up to order $2d$ of the measure μ supported on the graph $\{(t,x,y(t,x)) : (t,x) \in \Omega\}$ of the solution $(t,x) \mapsto y(t,x)$.

☞ AVOIDS A DISCRETIZATION of the (t,x) domain Ω!

Ex: The Burgers equation

Two initial conditions: One yields a solution $y(t, x)$ with a discontinuity (shock) and the other yields a continuous solution $y(t, x)$ (rarefaction).

☞ At step 6 of the Moment-SOS hierarchy (i.e., with moments up to order 12), the moments (computed from the SDP-relaxation) match those of the measure μ supported on the graph of $y(t, x)$ (with at least 4 digits of precision).

We also have an approximation procedure (not described here) to recover the solution $(t, x) \mapsto y(t, x)$ from finitely many moments of μ.

☞ S. Marx, E. Pauwels, T. Weisser, D. Henrion and J.B. Lass (2018). Tractable semi-algebraic approximation using Christoffel-Darboux kernel. arXiv:1807.02306

Ex: The Burgers equation

Two initial conditions: One yields a solution $y(t,x)$ with a discontinuity (shock) and the other yields a continuous solution $y(t,x)$ (rarefaction).

☞ At step 6 of the Moment-SOS hierarchy (i.e., with moments up to order 12), the moments (computed from the SDP-relaxation) match those of the measure μ supported on the graph of $y(t,x)$ (with at least 4 digits of precision).

We also have an approximation procedure (not described here) to recover the solution $(t,x) \mapsto y(t,x)$ from finitely many moments of μ.

☞ S. Marx, E. Pauwels, T. Weisser, D. Henrion and J.B. Lass (2018). Tractable semi-algebraic approximation using Christoffel-Darboux kernel. arXiv:1807.02306

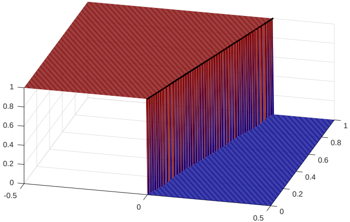
Ex: The Burgers equation

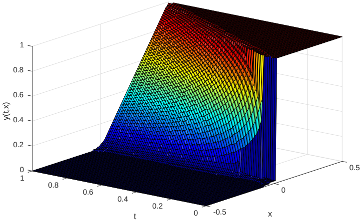
Two initial conditions: One yields a solution $y(t,x)$ with a discontinuity (shock) and the other yields a continuous solution $y(t,x)$ (rarefaction).

☞ At step 6 of the Moment-SOS hierarchy (i.e., with moments up to order 12), the moments (computed from the SDP-relaxation) match those of the measure μ supported on the graph of $y(t,x)$ (with at least 4 digits of precision).

We also have an approximation procedure (not described here) to recover the solution $(t,x) \mapsto y(t,x)$ from finitely many moments of μ.

☞ S. Marx, E. Pauwels, T. Weisser, D. Henrion and J.B. Lass (2018). Tractable semi-algebraic approximation using Christoffel-Darboux kernel. `arXiv:1807.02306`





Remarkably ... No Gibbs' phenomenon!

S. Marx, T. Weisser, D. Henrion, J.B. Lass (2018), A moment approach for entropy solutions to nonlinear hyperbolic PDEs. Math. Control and Related Fields (2019)

Extensions & Related works

- Compute polynomial Lyapunov Functions
- Approximate Regions Of Attraction (ROA) by sets of the form $\{\mathbf{x} : g(\mathbf{x}) \geq 0\}$ for some polynomial g.
- Convex Optimization of Non-Linear Feedback Controllers

By several authors ... Ahmadi, Henrion, Korda, Lass., Majumdar, Parrilo, Tedrake, Tobenkin, etc.

Extensions & Related works

☞ Compute polynomial Lyapunov Functions

☞ Approximate Regions Of Attraction (ROA) by sets of the form $\{\mathbf{x} : g(\mathbf{x}) \geq 0\}$ for some polynomial g.

☞ Convex Optimization of Non-Linear Feedback Controllers

By several authors ... Ahmadi, Henrion, Korda, Lass., Majumdar, Parrilo, Tedrake, Tobenkin, etc.

Extensions & Related works

☞ Compute polynomial Lyapunov Functions

☞ Approximate Regions Of Attraction (ROA) by sets of the form $\{\mathbf{x} : g(\mathbf{x}) \geq 0\}$ for some polynomial g.

☞ Convex Optimization of Non-Linear Feedback Controllers

By several authors ... Ahmadi, Henrion, Korda, Lass., Majumdar, Parrilo, Tedrake, Tobenkin, etc.

Extensions & Related works

☞ Compute polynomial Lyapunov Functions

☞ Approximate Regions Of Attraction (ROA) by sets of the form $\{\mathbf{x} : g(\mathbf{x}) \geq 0\}$ for some polynomial g.

☞ Convex Optimization of Non-Linear Feedback Controllers

By several authors ... Ahmadi, Henrion, Korda, Lass., Majumdar, Parrilo, Tedrake, Tobenkin, etc.

... and SDP-relaxations are also used:

☞ for Estimation problems (seen as Min-max optimization)

☞ for Robust Stability analysis and probabilistic $\mathcal{D}$-Stability Analysis

☞ for Detection of Anomalies and/or Causal Interactions in video sequences (Big data ...)

by several authors ... Benavoli, Lagoa, Lass., Piga, Regruto, Sznaier, ...

... and SDP-relaxations are also used:

☞ for Estimation problems (seen as Min-max optimization)

☞ for Robust Stability analysis and probabilistic $\mathcal{D}$-Stability Analysis

☞ for Detection of Anomalies and/or Causal Interactions in video sequences (Big data ...)

by several authors ... Benavoli, Lagoa, Lass., Piga, Regruto, Sznaier, ...

... and SDP-relaxations are also used:

☞ for Estimation problems (seen as Min-max optimization)

☞ for Robust Stability analysis and probabilistic $\mathcal{D}$-Stability Analysis

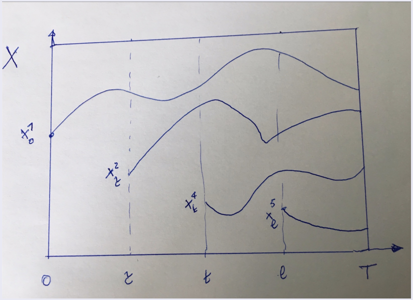
☞ for Detection of Anomalies and/or Causal Interactions in video sequences (Big data ...)

by several authors ... Benavoli, Lagoa, Lass., Piga, Regruto, Sznaier, ...

II. Inverse Optimal Control

Given:

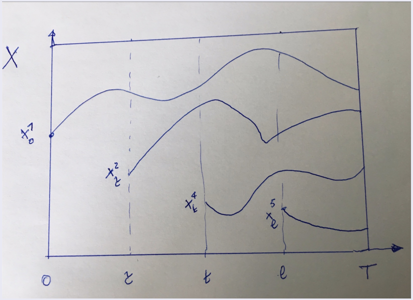
- ☞ a dynamical system $\dot{\mathbf{x}}(t) = f(\mathbf{x}(t), u(t)),\ t \in [0, T]$
- ☞ State and/or Control constraints $\mathbf{x}(t) \in X,\ u(t) \in U$,
- ☞ a database of recorded feasible trajectories $\{x(t; x_\tau), u(t; x_\tau)\}$ for several initial states $x_\tau \in X$,



II. Inverse Optimal Control

Given:

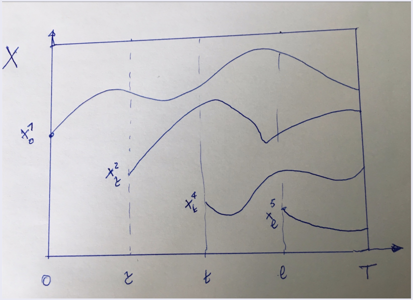
- ☞ a dynamical system $\dot{\mathbf{x}}(t) = f(\mathbf{x}(t), u(t)),\ t \in [0, T]$
- ☞ State and/or Control constraints $\mathbf{x}(t) \in X,\ u(t) \in U$,
- ☞ a database of recorded feasible trajectories $\{x(t; x_\tau), u(t; x_\tau)\}$ for several initial states $x_\tau \in X$,



II. Inverse Optimal Control

Given:

- ☞ a dynamical system $\dot{\mathbf{x}}(t) = f(\mathbf{x}(t), u(t)), t \in [0, T]$
- ☞ State and/or Control constraints $\mathbf{x}(t) \in X$, $u(t) \in U$,
- ☞ a database of recorded feasible trajectories $\{x(t; x_\tau), u(t; x_\tau)\}$ for several initial states $x_\tau \in X$,



compute a Lagrangian

$h : X \times U \to \mathbb{R}$ for which those trajectories are optimal.

☞ Key idea: I: Hamilton-Jacobi-Bellman (HJB) is the perfect tool to certify GLOBAL OPTIMALITY of the given trajectories in the database.

compute a Lagrangian

$h: X \times U \to \mathbb{R}$ for which those trajectories are optimal.

☞ Key idea: I: Hamilton-Jacobi-Bellman (HJB) is the perfect tool to certify GLOBAL OPTIMALITY of the given trajectories in the database.

Indeed suppose that two functions $\phi : [0,T] \times X \to \mathbb{R}$ and $h : X \times U \to \mathbb{R}$ satisfy:

$$(*) \quad \frac{\partial \phi}{\partial t} + \frac{\partial \phi}{\partial x} f(x,u) + h(x,u) \geq 0, \quad \forall (x,u,t) \in X \times U \times [0,T]$$

$$(**) \quad \phi(T,x) \leq 0 \quad \forall x \in X_T.$$

and †

$$\left(\frac{\partial \phi}{\partial t} + \frac{\partial \phi}{\partial x} f + h\right)(x(t;x_\tau), u(t;x_\tau), \tau) \leq 0; \quad \phi(T, x(T;x_\tau)) \geq 0,$$

for all $(x(t;x_\tau), u(t;x_\tau), \tau)$ in the database

Then

$$\phi(t,z) = \inf_{u} \int_t^T h(\mathbf{x}(s), u(s))\, ds$$

$$\text{s.t.} \quad \begin{aligned} \dot{\mathbf{x}}(s) &= f(\mathbf{x}(s), u(s)), \quad s \in [t,T] \\ \mathbf{x}(s) &\in X \subset \mathbb{R}^n;\ u(s) \in U \subset \mathbb{R}^m \\ \mathbf{x}(t) &= z \end{aligned}$$

☞ and all the trajectories $\{x(t; x_\tau), u(t; x_\tau)\}$ of the database are optimal solutions.

☞ That is: The Lagrangian h solves the INVERSE OCP and ϕ is the associated Optimal Value Function

Then

$$\phi(t, z) = \inf_{u} \int_t^T h(\mathbf{x}(s), u(s))\, ds$$

$$\text{s.t.} \quad \begin{aligned} \dot{\mathbf{x}}(s) &= f(\mathbf{x}(s), u(s)), \quad s \in [t, T] \\ \mathbf{x}(s) &\in X \subset \mathbb{R}^n;\ u(s) \in U \subset \mathbb{R}^m \\ \mathbf{x}(t) &= z \end{aligned}$$

☞ and all the trajectories $\{x(t; x_\tau), u(t; x_\tau)\}$ of the database are optimal solutions.

☞ That is: The Lagrangian h solves the INVERSE OCP and ϕ is the associated Optimal Value Function

Then

$$\phi(t,z) = \inf_{u} \int_t^T h(\mathbf{x}(s), u(s))\, ds$$

$$\text{s.t.} \quad \dot{\mathbf{x}}(s) = f(\mathbf{x}(s), u(s)), \quad s \in [t, T]$$
$$\mathbf{x}(s) \in X \subset \mathbb{R}^n; \; u(s) \in U \subset \mathbb{R}^m$$
$$\mathbf{x}(t) = z$$

☞ and all the trajectories $\{x(t; x_\tau), u(t; x_\tau)\}$ of the database are optimal solutions.

☞ That is: The Lagrangian h solves the INVERSE OCP and ϕ is the associated Optimal Value Function

☞ Key idea II: Look for POLYNOMIALS

$\phi \in \mathbb{R}[x,t]$ and $h \in \mathbb{R}[x,u]$

- that satisfy the relaxed HJB conditions (*) and (**)
- and also satisfy

$$(\dagger) \quad \left(\frac{\partial \phi}{\partial t} + \frac{\partial \phi}{\partial x} f + h\right)(x(t;x_\tau), u(t;x_\tau), \tau) \leq \epsilon$$

$$(\dagger\dagger) \quad \phi(T, x(T;x_\tau)) \geq -\epsilon,$$

for all $(x(t;x_\tau), u(t;x_\tau), \tau)$ in the database

☞ ... and SOLVE:

$$\rho_d = \min_{\phi,h} \{ \epsilon + \|h\|_1 : \text{s.t. } (*), (**), (\dagger), (\dagger\dagger);\ \deg(\phi), \deg(h) \leq 2d \}$$

where one replaces the nonnegativity conditions (*), (**), (†) and (††) by appropriate positivity certificates.

☞ a HIERARCHY of SEMIDEFINITE PROGRAMS (whose size increases with the degree d).

Pauwels E., Henrion D., Lasserre J.B. (2016) Linear Conic Optimization for Inverse Optimal Control, SIAM J. Control & Optim. 54, pp. 1798–1825.

☞ ... and SOLVE:

$$\rho_d = \min_{\phi,h} \{ \epsilon + \|h\|_1 : \text{s.t. } (*), (**), (\dagger), (\dagger\dagger);\ \deg(\phi), \deg(h) \leq 2d \}$$

where one replaces the nonnegativity conditions (*), (**), (†) and (††) by appropriate positivity certificates.

☞ a HIERARCHY of SEMIDEFINITE PROGRAMS (whose size increases with the degree d).

Pauwels E., Henrion D., Lasserre J.B. (2016) Linear Conic Optimization for Inverse Optimal Control, SIAM J. Control & Optim. 54, pp. 1798–1825.

III. 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:

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

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

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.

☞ Use Putinar Positivity Certificate to build up a hierarchy of semidefinite programs $(\mathbf{Q}_k)_{k\in N}$ of increasing size:

- An optimal solution of $\mathbf{Q}_k$ provides the coefficients of the polynomial J_k of degree $2k$.
- For every k:

$$\Theta_k \;:=\; \{\mathbf{x} \in \mathbf{B} \,:\, J_k(x) \,\leq\, 0\,\} \subset R_f \qquad \text{(inner approximations)}$$

- $\mathrm{vol}(R_f \setminus \Theta_k) \to 0$ as $k \to \infty$.

Lass. J.B. (2015) Tractable approximations of sets defined with quantifiers, Math. Program. 151, pp. 507–527.
Henrion D., Lass. J.B. (2006), Convergent relaxations of polynomial matrix inequalities and static output feedback, IEEE Trans. Auto. Control 51, pp. 192–202

☞ Use Putinar Positivity Certificate to build up a hierarchy of semidefinite programs $(\mathbf{Q}_k)_{k\in N}$ of increasing size:

- An optimal solution of $\mathbf{Q}_k$ provides the coefficients of the polynomial J_k of degree $2k$.
- For every k:

$$\Theta_k := \{\mathbf{x} \in \mathbf{B} : J_k(x) \leq 0\} \subset R_f \qquad \text{(inner approximations)}$$

- $\mathrm{vol}(R_f \setminus \Theta_k) \to 0$ as $k \to \infty$.

Lass. J.B. (2015) Tractable approximations of sets defined with quantifiers, Math. Program. 151, pp. 507–527.
Henrion D., Lass. J.B. (2006), Convergent relaxations of polynomial matrix inequalities and static output feedback, IEEE Trans. Auto. Control 51, pp. 192–202

☞ Use Putinar Positivity Certificate to build up a hierarchy of semidefinite programs $(\mathbf{Q}_k)_{k\in N}$ of increasing size:

- An optimal solution of $\mathbf{Q}_k$ provides the coefficients of the polynomial J_k of degree $2k$.
- For every k:

$$\Theta_k \;:=\; \{\mathbf{x} \in \mathbf{B} : \; J_k(x) \;\leq\; 0\} \subset R_f \qquad \text{(inner approximations)}$$

- $\mathrm{vol}(R_f \setminus \Theta_k) \to 0$ as $k \to \infty$.

Lass. J.B. (2015) Tractable approximations of sets defined with quantifiers, Math. Program. 151, pp. 507–527.
Henrion D., Lass. J.B. (2006), Convergent relaxations of polynomial matrix inequalities and static output feedback, IEEE Trans. Auto. Control 51, pp. 192–202

☞ Use Putinar Positivity Certificate to build up a hierarchy of semidefinite programs $(\mathbf{Q}_k)_{k\in N}$ of increasing size:

- An optimal solution of $\mathbf{Q}_k$ provides the coefficients of the polynomial J_k of degree $2k$.
- For every k:

$$\Theta_k := \{\mathbf{x} \in \mathbf{B} : J_k(x) \leq 0\} \subset R_f \quad \text{(inner approximations)}$$

- $\mathrm{vol}(R_f \setminus \Theta_k) \to 0$ as $k \to \infty$.

Lass. J.B. (2015) Tractable approximations of sets defined with quantifiers, Math. Program. 151, pp. 507–527.
Henrion D., Lass. J.B. (2006), Convergent relaxations of polynomial matrix inequalities and static output feedback, IEEE Trans. Auto. Control 51, pp. 192–202

IV. Convex Underestimators of Polynomials

☞ e.g., in the context of large scale MINLP the most efficient & popular strategy is to use BRANCH & BOUND combined with efficient LOWER BOUNDING techniques used at each node of the search tree.

• Typically, f is a sum $\sum_k f_k$ where each f_k “sees" only very few variables (say $3, 4$). The same observation is true for each g_j in the constraints:

Hence a very appealing idea is to pre-compute CONVEX UNDER-ESTIMATORS $\widehat{f_k} \leq f_k$ and $\widehat{g_j} \leq g_j$ for each non convex f_k and each non convex g_j, independently and separately!

$\rightarrow$ hence potentially many BUT LOW-DIMENSIONAL problems.

IV. Convex Underestimators of Polynomials

☞ e.g., in the context of large scale MINLP the most efficient & popular strategy is to use BRANCH & BOUND combined with efficient LOWER BOUNDING techniques used at each node of the search tree.

• Typically, f is a sum $\sum_k f_k$ where each f_k "sees" only very few variables (say $3, 4$). The same observation is true for each g_j in the constraints:

Hence a very appealing idea is to pre-compute CONVEX UNDER-ESTIMATORS $\widehat{f_k} \leq f_k$ and $\widehat{g_j} \leq g_j$ for each non convex f_k and each non convex g_j, independently and separately!

→ hence potentially many BUT LOW-DIMENSIONAL problems.

IV. Convex Underestimators of Polynomials

☞ e.g., in the context of large scale MINLP the most efficient & popular strategy is to use BRANCH & BOUND combined with efficient LOWER BOUNDING techniques used at each node of the search tree.

• Typically, f is a sum $\sum_k f_k$ where each f_k "sees" only very few variables (say $3, 4$). The same observation is true for each g_j in the constraints:

Hence a very appealing idea is to pre-compute CONVEX UNDER-ESTIMATORS $\widehat{f_k} \leq f_k$ and $\widehat{g_j} \leq g_j$ for each non convex f_k and each non convex g_j, independently and separately!

$\rightarrow$ hence potentially many BUT LOW-DIMENSIONAL problems.

Hence one has to solve the generic problem

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

Message:

"Good" CONVEX POLYNOMIAL UNDER-ESTIMATORS can be computed efficiently!

Hence one has to solve the generic problem

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

Message:

"Good" CONVEX POLYNOMIAL UNDER-ESTIMATORS can be computed efficiently!

I: Characterizing convex polynomial under-estimators

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 $\mathbf{x} \in \mathbf{B}$,

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

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

Hence we have the two "Positivity constraints"

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

I: Characterizing convex polynomial under-estimators

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 we have the two "Positivity constraints"

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

I: Characterizing convex polynomial under-estimators

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 $\mathbf{x} \in \mathbf{B}$,

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

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

Hence we have the two "Positivity constraints"

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

II: Characterizing "tightness"

One possibility is to evaluate the L_1-norm $\displaystyle\int_{\mathbf{B}} |f(\mathbf{x}) - p(\mathbf{x})|\, d\mathbf{x}$

$$\rightarrow \int_{\mathbf{B}} (f(\mathbf{x}) - p(\mathbf{x})\, d\mathbf{x} = \underbrace{\int_{\mathbf{B}} f(\mathbf{x}) d\mathbf{x}}_{constant} - \underbrace{\int_{\mathbf{B}} p(\mathbf{x})\, d\mathbf{x}}_{\text{linear in } p!}$$

Indeed, writing $p(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}^n} p_\alpha\, \mathbf{x}^\alpha$,

$$\int_{\mathbf{B}} p(\mathbf{x})\, d\mathbf{x} = \sum_{\alpha \in \mathbb{N}^n} p_\alpha \underbrace{\int_{\mathbf{B}} \mathbf{x}^\alpha\, d\mathbf{x}}_{\gamma_\alpha},$$

where γ_α is known (and easy to compute)!

II: Characterizing "tightness"

One possibility is to evaluate the L_1-norm $\int_{\mathbf{B}} |f(\mathbf{x}) - p(\mathbf{x})| \, d\mathbf{x}$

$$\rightarrow \int_{\mathbf{B}} (f(\mathbf{x}) - p(\mathbf{x}) \, d\mathbf{x} \,=\, \underbrace{\int_{\mathbf{B}} f(\mathbf{x}) d\mathbf{x}}_{\textit{constant}} - \underbrace{\int_{\mathbf{B}} p(\mathbf{x}) \, d\mathbf{x}}_{\text{linear in } p!}$$

Indeed, writing $p(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}^n} p_\alpha \, \mathbf{x}^\alpha$,

$$\int_{\mathbf{B}} p(\mathbf{x}) \, d\mathbf{x} \,=\, \sum_{\alpha \in \mathbb{N}^n} p_\alpha \underbrace{\int_{\mathbf{B}} \mathbf{x}^\alpha \, d\mathbf{x}}_{\gamma_\alpha},$$

where γ_α is known (and easy to compute)!

II: Characterizing "tightness"

One possibility is to evaluate the L_1-norm $\int_{\mathbf{B}} |f(\mathbf{x}) - p(\mathbf{x})| \, d\mathbf{x}$

$$\rightarrow \int_{\mathbf{B}} (f(\mathbf{x}) - p(\mathbf{x}) \, d\mathbf{x} = \underbrace{\int_{\mathbf{B}} f(\mathbf{x}) d\mathbf{x}}_{\textit{constant}} - \underbrace{\int_{\mathbf{B}} p(\mathbf{x}) \, d\mathbf{x}}_{\text{linear in } p!}$$

Indeed, writing $p(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}^n} p_\alpha \, \mathbf{x}^\alpha$,

$$\int_{\mathbf{B}} p(\mathbf{x}) \, d\mathbf{x} = \sum_{\alpha \in \mathbb{N}^n} p_\alpha \underbrace{\int_{\mathbf{B}} \mathbf{x}^\alpha \, d\mathbf{x}}_{\gamma_\alpha},$$

where γ_α is known (and easy to compute)!

Hence computing the best degree-d convex polynomial under-estimator of f reduces to solve the CONVEX optimization problem:

$$\mathbf{P}: \quad \rho = \inf_{p \in \mathbb{R}[\mathbf{x}]_d} \sum_{\alpha \in \mathbb{N}^n_d} p_\alpha \, \gamma_\alpha$$

$$\text{s.t.} \quad f(\mathbf{x}) - p(\mathbf{x}) \geq 0, \; \forall \, \mathbf{x} \in \mathbf{B}$$

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

☞ which has an optimal solution $p^* \in \mathbb{R}[\mathbf{x}]_d$

Replacing the positivity constraints with Putinar's positivity certificate

☞ yields a HIERARCHY of SEMIDEFINITE PROGRAMS, each with an optimal solution $p^*_\ell \in \mathbb{R}[\mathbf{x}]_d$, and:

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.

Replacing the positivity constraints with Putinar's positivity certificate

☞ yields a HIERARCHY of SEMIDEFINITE PROGRAMS, each with an optimal solution $p^*_\ell \in \mathbb{R}[\mathbf{x}]_d$, and:

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.

Suppose that an unknown SIGNED measure ϕ^* (signal) is supported on finitely many (few) 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 $I \subset \mathbb{R}$ (or equivalently on the torus $\mathbb{T} \subset \mathbb{C}$):

If the distance between any two atoms is sufficiently large and sufficiently many (few) moments are available then :

- ϕ^* is the unique solution of **P**, and
- exact recovery is obtained by solving a single SDP.

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

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 $I \subset \mathbb{R}$ (or equivalently on the torus $\mathbb{T} \subset \mathbb{C}$):

If the distance between any two atoms is sufficiently large and sufficiently many (few) moments are available then :
- ϕ^* is the unique solution of **P**, and
- 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

$$\inf_{\phi^+,\phi^-} \int_I d(\phi^+ + \phi^-) : \int_I \mathbf{x}^k\, d\phi^+(\mathbf{x}) - \int_I \mathbf{x}^k\, d\phi^+(\mathbf{x}) = q_\alpha, \quad \alpha \in \Gamma \}$$

... again an instance of the GMP!

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

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

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

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

P reads

$$\inf_{\phi^+,\phi^-} \int_I d(\phi^+ + \phi^-) : \int_I \mathbf{x}^k \, d\phi^+(\mathbf{x}) - \int_I \mathbf{x}^k \, d\phi^+(\mathbf{x}) = q_\alpha, \quad \alpha \in \Gamma \}$$

... again an instance of the GMP!

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

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

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

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

P reads

$$\inf_{\phi^+,\phi^-} \int_I d(\phi^+ + \phi^-) : \int_I \mathbf{x}^k \, d\phi^+(\mathbf{x}) - \int_I \mathbf{x}^k \, d\phi^+(\mathbf{x}) \,=\, q_\alpha, \quad \alpha \in \Gamma \,\}$$

... again an instance of the GMP!

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

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

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

VI: Optimal design in Statistics

In designing experiments one models the responses $z_1, \ldots, z_N$ of a random experiment whose inputs are represented by a vector $(t_i) \subset \mathbb{R}^n$, with respect to known regression functions $\mathbf{x} \mapsto \Phi(\mathbf{x}) = (\phi_1(\mathbf{x}), \ldots, \phi_d(\mathbf{x}))$, that is:

$$z_i = \sum_{j=1}^{d} \theta_j \, \phi_j(t_i) + \varepsilon_i, \quad i = 1, \ldots, N.$$

where (θ_j) are unknown parameters that the experimenter wants to estimate, ε_i is some noise and the (t_i) are chosen by the experimenter in a design space $\mathbf{X} \subset \mathbb{R}^n$.

A design

The goal is to find appropriate points $t_i \in \{\mathbf{x}_1, \ldots, \mathbf{x}_\ell\} \subset \mathbf{X}$ and associated frequency $\frac{n_i}{N}$ with which the point t_i is chosen for the experiment. Then:

$$\xi = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \ldots & \mathbf{x}_\ell \\ \frac{n_1}{N} & \frac{n_2}{N} & \ldots & \frac{n_\ell}{N} \end{pmatrix}$$

is called a design with associated information matrix

$$\mathbf{M}(\xi) := \sum_{i=1}^{\ell} \frac{n_i}{N} \, \Phi(\mathbf{x}_i) \, \Phi(\mathbf{x}_i)^T.$$

Optimal design

is concerned with finding a set of points in $\mathbf{X}$ that optimizes a certain statistical criterion $f(\mathbf{M}(\xi))$ where f must be real-valued, positively homogeneous, non constant, upper semi-continuous, and isotonic w.r.t. Loewner-ordering, and concave. An important choice is

$$f(\mathbf{M}(\xi)) \;:=\; \log\det(\mathbf{M}(\xi)).$$

Usually one ends up with a convex optimization problem AFTER some DISCRETIZATION of the design space $\mathbf{X}$.

In the Moment-SOS approach we DO NOT discretize **X** and rather search for an ATOMIC probability measure on **X**:

$$\mu := \sum_{k=1}^{m} \lambda_k \, \delta_{\mathbf{x}_k}, \quad \text{with unknown atoms } x_k \in \mathbf{X} \text{ and weights } \lambda_k > 0.$$

With base functions $\Phi(\mathbf{x}) = (\mathbf{x}^\alpha)_{\alpha \in \mathbb{N}^n_d}$ one solves the infinite-dimensional CONVEX optimization problem:

$$\sup_{\mu} \{ \log\det \mathbf{M}_d(\mu) : \quad \mu \in \mathscr{P}(\mathbf{X}) \}.$$

where $\mathscr{P}(\mathbf{X})$ is the space of probability measures on **X**, and $\mathbf{M}_d(\mu)$ is the (order-d) moment matrix of μ.

☞ Works remarkably well! De Castro, Gamboa, Henrion, Hess, and Lasserre, *Annals of Statistics*, to appear.

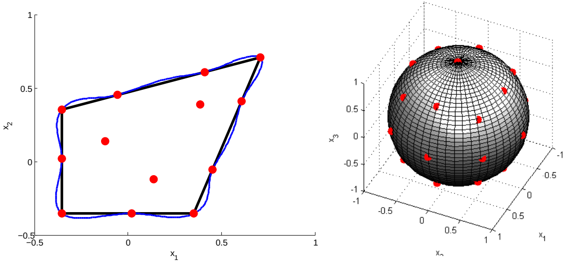


Figure: Polygon and Sphere for $d = 3$

VII. LP on spaces of measures: a rich framework

Consider the infinite dimensional LP:

$$\min_{\phi} \left\{ \int_{\mathbf{K}} f\, d\phi : \quad \phi \leq \mu; \quad \int_{\mathbf{K}} g\, d\phi = b, \forall g \in G \right\}$$

where :

- $\mathbf{K} \subset \mathbb{R}^n$ is a basic semi-algebraic set,
- The unknown ϕ is a Borel measure supported on $\mathbf{K}$
- The functions f, and $g \in G$ are polynomials
- All moments of the measure μ are available.

For instance this framework can be used :

- To compute Sharp Upper Bounds on $\mu(\mathbf{K})$ GIVEN some moments of μ.
- To approximate as closely as desired, from below and above, the Lebesgue volume of $\mathbf{K}$, or the Gaussian measure of $\mathbf{K}$ (for possibly non-compact $\mathbf{K}$)
- CHANCE-CONSTRAINTS: Given $\epsilon > 0$ and a prob. distribution μ, approximate AS CLOSELY AS DESIRED

$$\Omega_\epsilon := \{\,\mathbf{x} : \mathrm{Prob}_\omega(f(\mathbf{x},\omega) \leq 0) \geq 1-\epsilon\,\}$$

by sets of form : $\Omega_\epsilon^d := \{\,\mathbf{x} : h_d(\mathbf{x}) \leq 0\,\}$ for some polynomial h_d of degree d.

and more ! ☞ Henrion et al. (SIREV 2009), Lass. (Adv. Appl. Math. (2017)), Lass. (Adv. Comput. Math. (2016)), Lass. (2017) (IEEE Control Systems Letters), ...

For instance this framework can be used :

- To compute Sharp Upper Bounds on $\mu(\mathbf{K})$ GIVEN some moments of μ.
- To approximate as closely as desired, from below and above, the Lebesgue volume of $\mathbf{K}$, or the Gaussian measure of $\mathbf{K}$ (for possibly non-compact $\mathbf{K}$)
- CHANCE-CONSTRAINTS: Given $\epsilon > 0$ and a prob. distribution μ, approximate AS CLOSELY AS DESIRED

$$\Omega_\epsilon := \{\mathbf{x} : \mathrm{Prob}_\omega(f(\mathbf{x},\omega) \leq 0) \geq 1-\epsilon\}$$

by sets of form : $\Omega_\epsilon^d := \{\mathbf{x} : h_d(\mathbf{x}) \leq 0\}$ for some polynomial h_d of degree d.

and more ! ☞ Henrion et al. (SIREV 2009), Lass. (Adv. Appl. Math. (2017)), Lass. (Adv. Comput. Math. (2016)), Lass. (2017) (IEEE Control Systems Letters), ...

For instance this framework can be used :

- To compute Sharp Upper Bounds on $\mu(\mathbf{K})$ GIVEN some moments of μ.
- To approximate as closely as desired, from below and above, the Lebesgue volume of **K**, or the Gaussian measure of **K** (for possibly non-compact **K**)
- CHANCE-CONSTRAINTS: Given $\epsilon > 0$ and a prob. distribution μ, approximate AS CLOSELY AS DESIRED

$$\Omega_\epsilon \; := \; \{\, \mathbf{x} : \; \mathrm{Prob}_\omega(f(\mathbf{x},\omega) \leq 0) \; \geq \; 1 - \epsilon \,\}$$

by sets of form : $\Omega_\epsilon^d := \{\, \mathbf{x} : \; h_d(\mathbf{x}) \leq 0 \,\}$ for some polynomial h_d of degree d.

and more ! ☞ Henrion et al. (SIREV 2009), Lass. (Adv. Appl. Math. (2017)), Lass. (Adv. Comput. Math. (2016)), Lass. (2017) (IEEE Control Systems Letters), ...

For instance this framework can be used :

- To compute Sharp Upper Bounds on $\mu(\mathbf{K})$ GIVEN some moments of μ.
- To approximate as closely as desired, from below and above, the Lebesgue volume of **K**, or the Gaussian measure of **K** (for possibly non-compact **K**)
- CHANCE-CONSTRAINTS: Given $\epsilon > 0$ and a prob. distribution μ, approximate AS CLOSELY AS DESIRED

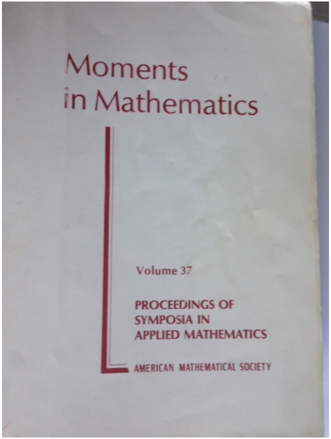
$$\Omega_\epsilon \;:=\; \{\, \mathbf{x} :\; \mathrm{Prob}_\omega(f(\mathbf{x},\omega) \leq 0) \;\geq\; 1-\epsilon \,\}$$

by sets of form : $\Omega_\epsilon^d := \{\, \mathbf{x} :\; h_d(\mathbf{x}) \leq 0 \,\}$ for some polynomial h_d of degree d.

and more ! ☞ Henrion et al. (SIREV 2009), Lass. (Adv. Appl. Math. (2017)), Lass. (Adv. Comput. Math. (2016)), Lass. (2017) (IEEE Control Systems Letters), ...

In fact .... the list of potential applications of the GMP is almost ENDLESS!

☞



THANK YOU!