Moments & Positive Polynomials in and outside Optimization

Jean B. Lasserre*

LAAS-CNRS and Institute of Mathematics, Toulouse, France

Master-class, LAAS, Toulouse, June 2018

★ Research funded by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement 666981 TAMING)
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.
How to handle sparsity
How to handle sparsity

Jean B. Lasserre

semidefinite characterization
Why POLYNOMIAL optimization?
- LP- and SDP- CERTIFICATES of POSITIVITY
- The moment-LP and moment-SOS approaches
- Some applications
Why POLYNOMIAL optimization?

LP- and SDP- CERTIFICATES of POSITIVITY

The moment-LP and moment-SOS approaches

Some applications
Why POLYNOMIAL optimization?
LP- and SDP- CERTIFICATES of POSITIVITY
The moment-LP and moment-SOS approaches
Some applications
Why POLYNOMIAL optimization?

LP- and SDP- CERTIFICATES of POSITIVITY

The moment-LP and moment-SOS approaches

Some applications
Consider the polynomial optimization problem:

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

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

Why Polynomial Optimization?

After all ... \( P \) is just a particular case of Non Linear Programming (NLP)!
Consider the polynomial optimization problem:

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

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

**Why Polynomial Optimization?**

After all ... \( P \) is just a particular case of Non Linear Programming (NLP)!
How to handle sparsity

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 $x \in K$, i.e., LOCALLY AROUND $x \in K$, and in general, no GLOBAL property of $x \in K$ can be inferred.

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

Jean B. Lasserre*

semidefinite characterization
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 $x \in K$, i.e., **LOCALLY AROUND** $x \in K$, and in general, no **GLOBAL** property of $x \in K$ 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!
How to handle sparsity

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 $x \in K$, i.e., **LOCALLY AROUND** $x \in K$, and in general, no **GLOBAL** property of $x \in K$ can be inferred.

The fact that $f$ and $g_j$ are **POLYNOMIALS** does not help much!
How to handle sparsity

BUT for GLOBAL Optimization

... the picture is different!

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

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

(Not true for a LOCAL minimum!)

and so to compute $f^*$ ...

one needs to handle EFFICIENTLY the difficult constraint

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

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

... the picture is different!

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

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

(Not true for a LOCAL minimum!)

and so to compute $f^*$ ...

one needs to handle EFFICIENTLY the difficult constraint

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

i.e. one needs TRACTABLE CERTIFICATES of POSITIVITY on $K$

for the polynomial $x \mapsto f(x) - \lambda$!
How to handle sparsity

BUT for GLOBAL Optimization

... the picture is different!

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

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

(Not true for a LOCAL minimum!)

and so to compute $f^*$ ...

one needs to handle EFFICIENTLY the difficult constraint

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

i.e. one needs TRACTABLE CERTIFICATES of POSITIVITY on $K$

for the polynomial $x \mapsto f(x) - \lambda$!
How to handle sparsity

REAL ALGEBRAIC GEOMETRY helps!!!!

Indeed, POWERFUL CERTIFICATES OF POSITIVITY EXIST!

Moreover .... and importantly,

Such certificates are amenable to PRACTICAL COMPUTATION!

(⋆ 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!**

(★ 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!

(⋆ Stronger Positivstellensatzë exist for analytic functions but (so far) are useless from a computational viewpoint.)
Let $\mathbf{K} := \{ \mathbf{x} : g_j(\mathbf{x}) \geq 0, \ j = 1, \ldots, m \}$ 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:*

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

*for some* SOS polynomials ($\sigma_j \subset \mathbb{R}[\mathbf{x}]$).
Let $K := \{ x : g_j(x) \geq 0, \quad j = 1, \ldots, m \}$ be compact (with $g_1(x) = M - \|x\|^2$, so that $K \subset B(0, M)$).

**Theorem (Putinar’s Positivstellensatz)**

If $f \in \mathbb{R}[x]$ is strictly positive ($f > 0$) on $K$ then:

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

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

Jean B. Lasserre*  
semidefinite characterization
However ... In Putinar’s theorem
... nothing is said on the DEGREE of the SOS polynomials $\sigma_j$!

BUT ... GOOD news ..!!

Testing whether $\dagger$ holds for some SOS $\sigma_j \subset \mathbb{R}[x]$ with a degree bound,
is SOLVING an SDP!
However ... In Putinar’s theorem
... nothing is said on the **DEGREE** of the SOS polynomials \((\sigma_j)!\)

**BUT** ... **GOOD** news ...!!

💡 Testing whether \(\dagger\) holds for some **SOS** \((\sigma_j) \subset \mathbb{R}[x]\) with a **degree** bound,

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

$$
\int_K x_1^{\alpha_1} \cdots x_n^{\alpha_n} \, d\mu, \quad \forall \alpha \in \mathbb{N}^n
$$

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

Jean B. Lasserre

*semidefinite characterization*
Let \( K := \{ x : g_j(x) \geq 0, \quad j = 1, \ldots, m \} \) be compact (with \( g_1(x) = M - \|x\|^2 \), so that \( K \subset B(0, M) \)).

**Theorem (Dual side of Putinar’s Theorem)**

A sequence \( y = (y_\alpha), \alpha \in \mathbb{N}^n \), has a representing measure supported on \( K \) **IF AND ONLY IF** for every \( d = 0, 1, \ldots \)

\[
(*) \quad M_d(y) \succeq 0 \quad \text{and} \quad M_d(g_j y) \succeq 0, \quad j = 1, \ldots, m.
\]

The real symmetric matrix \( M_2(y) \) is called the **MOMENT MATRIX** associated with the sequence \( y \).

The real symmetric matrix \( M_d(g_j y) \) is called the **LOCALIZING MATRIX** associated with the sequence \( y \) and the polynomial \( g_j \).
Let \( K := \{ \mathbf{x} : g_j(\mathbf{x}) \geq 0, \quad j = 1, \ldots, m \} \) be compact (with \( g_1(\mathbf{x}) = M - \|\mathbf{x}\|^2 \), so that \( K \subset B(0, M) \)).

**Theorem (Dual side of Putinar’s Theorem)**

A sequence \( \mathbf{y} = (y_\alpha), \quad \alpha \in \mathbb{N}^n \), has a representing measure supported on \( K \) **IF AND ONLY IF** for every \( d = 0, 1, \ldots \)

\[
(\star) \quad M_d(\mathbf{y}) \succeq 0 \quad \text{and} \quad M_d(g_j \mathbf{y}) \succeq 0, \quad j = 1, \ldots, m.
\]

\( \text{The real symmetric matrix } M_2(\mathbf{y}) \text{ is called the MOMENT MATRIX associated with the sequence } \mathbf{y} \)

\( \text{The real symmetric matrix } M_d(g_j \mathbf{y}) \text{ is called the LOCALIZING MATRIX associated with the sequence } \mathbf{y} \text{ and the polynomial } g_j. \)
Let \( K := \{ x : g_j(x) \geq 0, \ j = 1, \ldots, m \} \) be compact (with \( g_1(x) = M - \|x\|^2 \), so that \( K \subset B(0, M) \)).

Theorem (Dual side of Putinar’s Theorem)

A sequence \( y = (y_\alpha), \ \alpha \in \mathbb{N}^n \), has a representing measure supported on \( K \) IF AND ONLY IF for every \( d = 0, 1, \ldots \)

\[
(\star) \quad M_d(y) \succeq 0 \quad \text{and} \quad M_d(g_j y) \succeq 0, \ j = 1, \ldots, m.
\]

The real symmetric matrix \( M_2(y) \) is called the **MOMENT MATRIX** associated with the sequence \( y \)

The real symmetric matrix \( M_d(g_j y) \) is called the **LOCALIZING MATRIX** associated with the sequence \( y \) and the polynomial \( g_j \).
How to handle sparsity

Let $K := \{ x : g_j(x) \geq 0, \quad j = 1, \ldots, m \}$
be compact (with $g_1(x) = M - \|x\|^2$, so that $K \subset B(0, M)$).

Theorem (Dual side of Putinar’s Theorem)

A sequence $y = (y_\alpha), \alpha \in \mathbb{N}^n$, has a representing measure
supported on $K$ IF AND ONLY IF for every $d = 0, 1, \ldots$

\[ (*) \quad M_d(y) \succeq 0 \quad \text{and} \quad M_d(g_j y) \succeq 0, \quad j = 1, \ldots, m. \]

The real symmetric matrix $M_2(y)$ is called the \textbf{MOMENT MATRIX} associated with the sequence $y$.

The real symmetric matrix $M_d(g_j y)$ is called the \textbf{LOCALIZING MATRIX} associated with the sequence $y$ and the polynomial $g_j$.
Remarkably, the **Necessary & Sufficient conditions** (⋆) 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 \( M_2(y) \) reads

\[
M_2(y) = \begin{bmatrix}
1 & x_1 & x_2 & x_1^2 & x_1 x_2 & x_2^2 \\
\gamma_{00} & \gamma_{10} & \gamma_{01} & \gamma_{20} & \gamma_{11} & \gamma_{02} \\
\gamma_{10} & \gamma_{20} & \gamma_{11} & \gamma_{30} & \gamma_{21} & \gamma_{12} \\
\gamma_{01} & \gamma_{11} & \gamma_{02} & \gamma_{21} & \gamma_{12} & \gamma_{03} \\
\gamma_{20} & \gamma_{30} & \gamma_{21} & \gamma_{40} & \gamma_{31} & \gamma_{22} \\
\gamma_{11} & \gamma_{12} & \gamma_{21} & \gamma_{31} & \gamma_{22} & \gamma_{13} \\
\gamma_{02} & \gamma_{12} & \gamma_{03} & \gamma_{22} & \gamma_{13} & \gamma_{04}
\end{bmatrix}
\]
Remarkably, the **Necessary & Sufficient conditions** (⋆) 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 \( M_2(y) \) reads

\[
M_2(y) = \begin{bmatrix}
1 & X_1 & X_2 & X_1^2 & X_1X_2 & X_2^2 \\
\gamma_{00} & \gamma_{10} & \gamma_{01} & \gamma_{20} & \gamma_{11} & \gamma_{02} \\
\gamma_{10} & \gamma_{20} & \gamma_{11} & \gamma_{30} & \gamma_{21} & \gamma_{12} \\
\gamma_{01} & \gamma_{11} & \gamma_{02} & \gamma_{21} & \gamma_{12} & \gamma_{03} \\
\gamma_{10} & \gamma_{20} & \gamma_{11} & \gamma_{30} & \gamma_{21} & \gamma_{12} \\
\gamma_{02} & \gamma_{12} & \gamma_{03} & \gamma_{22} & \gamma_{13} & \gamma_{04}
\end{bmatrix}
\]
How to handle sparsity

Jean B. Lasserre

ALGEBRAIC SIDE

Positivity on K

\[ f(x) = \sum \alpha f_\alpha x^\alpha \]

\[ f > 0 \text{ on } K? \]

Characterize those \( f \)
How to handle sparsity

Jean B. Lasserre

ALGEBRAIC SIDE

Positivity on K

\[ f(x) = \sum_{\alpha} f_{\alpha} x^{\alpha} \]

\( f > 0 \) on K?

CHARACTERIZE THOSE \( f \)

FUNCTIONAL ANALYSIS

THE K-FACTORIAL MOMENT PROBLEM

\[ y = (y_{\alpha}), \quad x \in \mathbb{N}^n \]

\[ y_{\alpha} = \int_{K} x^{\alpha} d\mu \quad \forall \alpha \]

for some \( \mu \)

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

\textbf{…modeled as}

particular instances of the so called \textbf{Generalized Moment Problem}, among which: Probability, Optimal and Robust Control, Game theory, Signal processing, multivariate integration, etc.
The GMP is the infinite-dimensional LP:

\[
\text{GMP : } \inf_{\mu_i \in M(K_i)} \left\{ \sum_{i=1}^{s} \int_{K_i} f_i \, d\mu_i : \sum_{i=1}^{s} \int_{K_i} h_{ij} \, d\mu_i \geq b_j, \quad j \in J \right\}
\]

with \( M(K_i) \) space of Borel measures on \( K_i \subset \mathbb{R}^{n_i}, \ i = 1, \ldots, s \).
GMP: The dual view

The DUAL $GMP^*$ is the infinite-dimensional LP:

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

And one can see that ...

the constraints of $GMP^*$ state that the functions

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

must be NONNEGATIVE on certain sets $K_i, \ i = 1, \ldots, s$. 

Jean B. Lasserre * semidefinite characterization
GMP: The dual view

The **DUAL GMP** is the infinite-dimensional LP:

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

And one can see that ...

the constraints of **GMP** state that the functions

\[
x \mapsto f_i(x) - \sum_{j \in J} \lambda_j h_{ij}(x)
\]

must be **NONNEGATIVE** on certain sets \(K_i, \ i = 1, \ldots, s\).
How to handle sparsity

Several examples will follow .... and

Global OPTIM \rightarrow f^* = \inf_x \{ f(x) : x \in K \}

is the SIMPLEST example of the GMP

because ...

\[ f^* = \inf_{\mu \in M(K)} \left\{ \int_K f \, d\mu : \int_K 1 \, d\mu = 1 \right\} \]

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

• On the other hand, for every \( x \in K \) the probability measure \( \mu := \delta_x \) is such that \( \int f \, d\mu = f(x) \).
Several examples will follow .... and

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

is the SIMPLEST example of the GMP

because ...

$$f^* = \inf_{\mu \in M(K)} \left\{ \int_K f \, d\mu : \int_K 1 \, d\mu = 1 \right\}$$

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

• On the other hand, for every $x \in K$ the probability measure $\mu := \delta_x$ is such that $\int f \, d\mu = f(x)$. 
Several examples will follow .... and

Global OPTIM \[ \rightarrow f^* = \inf_{x} \{ f(x) : x \in K \} \]

is the SIMPLEST example of the GMP

because ...

\[ f^* = \inf_{\mu \in M(K)} \left\{ \int_{K} f \, d\mu : \int_{K} 1 \, d\mu = 1 \right\} \]

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

• On the other hand, for every \( x \in K \) the probability measure \( \mu := \delta_x \) is such that \( \int f \, d\mu = f(x) \).
How to handle sparsity

Several examples will follow ... and

Global OPTIM \rightarrow f^* = \inf_x \{ f(x) : x \in K \}

is the SIMPLEST example of the GMP

because ...

\[ f^* = \inf_{\mu \in M(K)} \left\{ \int_K f \, d\mu : \int_K 1 \, d\mu = 1 \right\} \]

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

• On the other hand, for every \( x \in K \) the probability measure \( \mu := \delta_x \) is such that \( \int f \, d\mu = f(x) \).
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!.
How to handle sparsity

LP- and SDP-hierarchies for optimization

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

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

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

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

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

Jean B. Lasserre*  semidefinite characterization
Replace $f^* = \sup_{\lambda} \{ \lambda : f(x) - \lambda \geq 0 \quad \forall x \in K \}$ with:

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

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

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

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

**Theorem**

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

\[
\begin{align*}
    f^* &= \lim_{d \to \infty} f_d^* \\
    \theta_d &= \lim_{d \to \infty} \theta_d.
\end{align*}
\]

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 \(f^* = f_d^*\).
How to handle sparsity

Theorem

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

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

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!
• 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 (→ 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 (→ 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**

\[
\begin{align*}
\mathbf{x} \in \mathbf{K}; \ |f(\mathbf{x})| & \iff \mathbf{x} \in \mathbf{K}; \ f(\mathbf{x})^2 - z^2 = 0; \ z \geq 0. \\
f(\mathbf{x}) \geq 0 \text{ on } \mathbf{K}; \ \sqrt{f(\mathbf{x})} & \iff \mathbf{x} \in \mathbf{K}; \ f(\mathbf{x}) - z^2 = 0; \ z \geq 0.
\end{align*}
\]

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

\[
(f_1(\mathbf{x}) - f_2(\mathbf{x}))^2 - z^2 = 0; \ z \geq 0 \iff g(\mathbf{x}) = \frac{z}{2} + \frac{f_1(\mathbf{x}) + f_2(\mathbf{x})}{2}
\]

\[
\underbrace{z = |f_1(\mathbf{x}) - f_2(\mathbf{x})|}_{z = |f_1(\mathbf{x}) - f_2(\mathbf{x})|}
\]

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

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

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

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

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

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

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 → (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 → (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
  -> (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 → (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
- → (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 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 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 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 ?
How to handle sparsity

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 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 ?
• 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 $\kappa$, 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 $\kappa$, 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
Consider the **OPTIMAL CONTROL** (OCP) problem:

\[
\rho = \inf_{u} \int_{0}^{T} h(x(t), u(t)) \, dt
\]

s.t. \( \dot{x}(t) = f(x(t), u(t)), \quad t \in [0, T] \),
\( x(0) = x_0 \),
\( x(t) \in X \subset \mathbb{R}^n; \ u(t) \in U \subset \mathbb{R}^m, \)

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

In general **OCP** problems are hard to solve, and particularly when **STATE CONSTRAINTS** \( x(t) \in X \) 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 $\rho$ 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 $\rho$ 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 $\rho$ 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 $\rho$ of the OCP.

How to handle sparsity

Extensions & Related works

Compute polynomial Lyapunov Functions

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

Convex Optimization of Non-Linear Feedback Controllers

By several authors ... Ahmadi, Henrion, Korda, Lass., Majumdar, Parrilo, Tedrake, Tobenkin, etc.
How to handle sparsity

Extensions & Related works

Compute polynomial Lyapunov Functions

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

Convex Optimization of Non-Linear Feedback Controllers

By several authors ... Ahmadi, Henrion, Korda, Lass., Majumdar, Parrilo, Tedrake, Tobenkin, etc.
How to handle sparsity

Extensions & Related works

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

By several authors ... Ahmadi, Henrion, Korda, Lass., Majumdar, Parrilo, Tedrake, Tobenkin, etc.
How to handle sparsity

Extensions & Related works

- Compute polynomial Lyapunov Functions
- Approximate Regions Of Attraction (ROA) by sets of the form \( \{ x : g(x) \geq 0 \} \) 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 $\dot{x}(t) = f(x(t), u(t)), t \in [0, T]$
- State and/or Control constraints $x(t) \in X, u(t) \in U$,
- 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 \( \dot{x}(t) = f(x(t), u(t)), \ t \in [0, T] \)
- State and/or Control constraints \( x(t) \in X, \ u(t) \in U \),
- A database of recorded feasible trajectories \( \{x(t; x_\tau), u(t; x_\tau)\} \) for several initial states \( x_\tau \in X \),

Jean B. Lasserre

Semidefinite characterization
II. Inverse Optimal Control

Given:

- a dynamical system \( \dot{x}(t) = f(x(t), u(t)), t \in [0, T] \)
- State and/or Control constraints \( x(t) \in X, u(t) \in U \),
- 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

\[ h : X \times U \rightarrow \mathbb{R} \] 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.
How to handle sparsity

compute a Lagrangian

\[ h : X \times U \rightarrow \mathbb{R} \] 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 \[\]

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

for all $(x(t; x_T), u(t; x_T), \tau)$ in the database.
Then

\[
\phi(t, z) = \inf_u \int_t^T h(x(s), u(s)) \, ds
\]

s.t.

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

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 \( \phi \) is the associated Optimal Value Function.
Then

$$\phi(t, z) = \inf_u \int_t^T h(x(s), u(s)) \, ds$$

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

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 \( \phi \) is the associated Optimal Value Function.
Then

\[ \phi(t, z) = \inf_u \int_t^T h(x(s), u(s)) \, ds \]

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

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 \( \phi \) 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:

\[
\rho_d = \min_{\phi, h} \left\{ \epsilon + \| h \|_1 : \text{s.t.} \ (\ast), (\ast\ast), (\dagger), (\dagger\dagger) ; \ \text{deg}(\phi), \text{deg}(h) \leq 2d \right\}
\]

where one replaces the nonnegativity conditions (\ast), (\ast\ast), (\dagger) and (\dagger\dagger) by appropriate positivity certificates.

\[\text{a HIERARCHY of SEMIDEFINITE PROGRAMS (whose size increases with the degree } d).\]

... and SOLVE:

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

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

A HIERARCHY of SEMIDEFINITE PROGRAMS (whose size increases with the degree $d$).

III. Approximation of sets with quantifiers

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

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

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

Approximate the set:

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

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

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

for some polynomials $J_k$. 
Use Putinar Positivity Certificate to build up a hierarchy of semidefinite programs \((Q_k)_{k \in \mathbb{N}}\) of increasing size:

- An optimal solution of \(Q_k\) provides the coefficients of the polynomial \(J_k\) of degree \(2k\).
- For every \(k\):
  \[
  \Theta_k := \{ x \in B : J_k(x) \leq 0 \} \subset R_f \quad \text{(inner approximations)}
  \]
- \(\text{vol}(R_f \setminus \Theta_k) \to 0\) as \(k \to \infty\).


How to handle sparsity

Use **Putinar Positivity Certificate** to build up a hierarchy of semidefinite programs \((Q_k)_{k \in \mathbb{N}}\) of increasing size:

- An optimal solution of \(Q_k\) provides the coefficients of the polynomial \(J_k\) of degree \(2k\).
- For every \(k\):
  \[
  \Theta_k := \{ x \in B : J_k(x) \leq 0 \} \subset R_f \quad \text{(inner approximations)}
  \]
  \[
  \text{vol}(R_f \setminus \Theta_k) \to 0 \text{ as } k \to \infty.
  \]

Use Putinar Positivity Certificate to build up a hierarchy of semidefinite programs \((Q_k)_{k \in \mathbb{N}}\) of increasing size:

- An optimal solution of \(Q_k\) provides the coefficients of the polynomial \(J_k\) of degree \(2k\).
- For every \(k:\)
  \[\Theta_k := \{x \in B : J_k(x) \leq 0\} \subset R_f\] (inner approximations)
  \[\text{vol}(R_f \setminus \Theta_k) \to 0 \text{ as } k \to \infty.\]

Use Putinar Positivity Certificate to build up a hierarchy of semidefinite programs \((Q_k)_{k \in \mathbb{N}}\) of increasing size:

- An optimal solution of \(Q_k\) provides the coefficients of the polynomial \(J_k\) of degree \(2k\).
- For every \(k\):
  \[
  \Theta_k := \{ \mathbf{x} \in \mathcal{B} : J_k(\mathbf{x}) \leq 0 \} \subset R_f \quad \text{(inner approximations)}
  \]
  \[
  \text{vol}(R_f \setminus \Theta_k) \to 0 \text{ as } k \to \infty.
  \]

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 $\hat{f}_k \leq f_k$ and $\hat{g}_j \leq g_j$ for each non convex $f_k$ and each non convex $g_j$, independently and separately!

→ hence potentially many BUT LOW-DIMENSIONAL problems.
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 $\hat{f}_k \leq f_k$ and $\hat{g}_j \leq g_j$ 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 $\hat{f}_k \leq f_k$ and $\hat{g}_j \leq g_j$ for each non convex $f_k$ and each non convex $g_j$, independently and separately!

→ hence potentially many BUT LOW-DIMENSIONAL problems.
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 $B \subset \mathbb{R}^n$.

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 $B \subset \mathbb{R}^n$.

Message:

“Good" CONVEX POLYNOMIAL UNDER-ESTIMATORS can be computed efficiently!
I: Characterizing convex polynomial under-estimators

1. \( p(x) \leq f(x) \) for every \( x \in B \).
2. \( p \) convex on \( B \rightarrow \nabla^2 p(x) \succeq 0 \) for all \( x \in B \),
   \[
   \Leftrightarrow \quad u^T \nabla^2 p(x) u \geq 0, \quad \forall (x, u) \in B \times U,
   \]
   where \( U := \{ u : \|u\|^2 \leq 1 \} \).

Hence we have the two "Positivity constraints"

\[
\begin{align*}
    f(x) - p(x) & \geq 0, \quad \forall x \in B \\
    u^T \nabla^2 p(x) u & \geq 0, \quad \forall (x, u) \in B \times U.
\end{align*}
\]
I: Characterizing convex polynomial under-estimators

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

\[ \iff u^T \nabla^2 p(x) u \geq 0, \forall (x, u) \in B \times U, \]

where \( U := \{ u : \|u\|^2 \leq 1 \} \).

Hence we have the two "Positivity constraints"

\[ f(x) - p(x) \geq 0, \forall x \in B \]
\[ u^T \nabla^2 p(x) u \geq 0, \forall (x, u) \in B \times U. \]
I: Characterizing convex polynomial under-estimators

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

\[ \iff \quad u^T \nabla^2 p(x) u \geq 0, \quad \forall (x, u) \in B \times U, \]

where \( U := \{ u : \|u\|^2 \leq 1 \} \).

Hence we have the two "Positivity constraints"

\[ f(x) - p(x) \geq 0, \quad \forall x \in B \]
\[ u^T \nabla^2 p(x) u \geq 0, \quad \forall (x, u) \in B \times U. \]
II: Characterizing "tightness"

One possibility is to evaluate the $L_1$-norm $\int_B |f(x) - p(x)| \, dx$

$\to \int_B (f(x) - p(x)) \, dx = \int_B f(x) \, dx - \int_B p(x) \, dx$

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

$$\int_B p(x) \, dx = \sum_{\alpha \in \mathbb{N}^n} p_\alpha \int_B x^\alpha \, dx,$$

where $\gamma_\alpha$ is known (and easy to compute)!
II: Characterizing "tightness"

One possibility is to evaluate the $L_1$-norm $\int_B |f(x) - p(x)| \, dx$

$$\rightarrow \int_B (f(x) - p(x)) \, dx = \int_B f(x) \, dx - \int_B p(x) \, dx$$

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

$$\int_B p(x) \, dx = \sum_{\alpha \in \mathbb{N}^n} p_\alpha \int_B x^\alpha \, dx,$$

where $\gamma_\alpha$ is known (and easy to compute)!
II: Characterizing "tightness"

One possibility is to evaluate the $L_1$-norm
\[ \int_B |f(x) - p(x)| \, dx \]

\[ \rightarrow \int_B (f(x) - p(x)) \, dx = \int_B f(x) \, dx - \int_B p(x) \, dx \]

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

\[ \int_B p(x) \, dx = \sum_{\alpha \in \mathbb{N}^n} p_{\alpha} \int_B x^\alpha \, dx, \]

where $\gamma_{\alpha}$ is known (and easy to compute)!
How to handle sparsity

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

$$
\textbf{P}:
\rho = \inf_{p \in \mathbb{R}[x]_d} \sum_{\alpha \in \mathbb{N}^n_d} p_{\alpha} \gamma_{\alpha}
\text{s.t. } f(x) - p(x) \geq 0, \ \forall x \in B
\text{ and } u^T \nabla^2 p(x) u \geq 0, \ \forall (x, u) \in B \times U.
$$

which has an optimal solution $p^* \in \mathbb{R}[x]_d$
Replacing the positivity constraints with Putinar’s positivity certificate yields a HIERARCHY of SEMIDEFINITE PROGRAMS, each with an optimal solution $p^*_ℓ \in \mathbb{R}[x]_d$, and:

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

$p^*_ℓ \to p^* \in \mathbb{R}[x]_d$, as $ℓ \to \infty$

→ Provides the best results in the comparison:

Replacing the positivity constraints with Putinar’s positivity certificate yields a HIERARCHY of SEMIDEFINITE PROGRAMS, each with an optimal solution $p^*_\ell \in \mathbb{R}[x]_d$, and:

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

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

→ Provides the best results in the comparison:

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

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

The goal is to find the SUPPORT \( (x(k))_{k=1}^{p} \subset K \) and WEIGHTS \( (\gamma_k)_{k=1}^{p} \) from only FINITELY MANY MEASUREMENTS (moments)

\[
q_\alpha = \int_{K} x^\alpha \, d\phi^*(x), \quad \alpha \in \Gamma.
\]
How to handle sparsity

Solve the infinite-dimensional LP

\[ P : \inf_{\phi} \{ \|\phi\|_{TV} : \int_K x^\alpha \, d\phi(x) = q_\alpha, \alpha \in \Gamma \}. \]

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:

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

How to handle sparsity

Solve the infinite-dimensional LP

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

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:

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

How to handle sparsity

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

The dual $P^*$ reads:

$$\sup_{p \in \mathbb{R}[x]} \{ \langle p, q \rangle : \sup_{x \in I} |p(x)| \leq 1 \}.$$ 

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

How to handle sparsity

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

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

... again an instance of the GMP!

The dual $P^*$ reads:

$$\sup_{p \in \mathbb{R}[x]} \{ \langle p, q \rangle : \sup_{x \in I} |p(x)| \leq 1 \}.$$  

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

How to handle sparsity

Writing the signed measure \( \phi \) on \( I \) as \( \phi^+ - \phi^- \),

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

... again an instance of the GMP!

The dual \( P^* \) reads:

\[
\sup_{p \in \mathbb{R}[x]} \{ \langle p, q \rangle : \sup_{x \in I} |p(x)| \leq 1 \}.
\]

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

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 $x \mapsto \Phi(x) = (\phi_1(x), \ldots, \phi_d(x))$, that is:

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

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

The goal is to find appropriate points $t_i \in \{x_1, \ldots, x_\ell\} \subset X$ and associated frequency $\frac{n_i}{N}$ with which the point $t_i$ is chosen for the experiment. Then:

$$\xi = \left( \begin{array}{cccc} x_1 & x_2 & \cdots & x_\ell \\ \frac{n_1}{N} & \frac{n_2}{N} & \cdots & \frac{n_\ell}{N} \end{array} \right)$$

is called a design with associated information matrix

$$M(\xi) := \sum_{i=1}^{\ell} \frac{n_i}{N} \Phi(x_i) \Phi(x_i)^T.$$
Optimal design

is concerned with finding a set of points in $\mathbf{X}$ that optimizes a certain statistical criterion $f(M(\xi))$ 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

$$f(M(\xi)) := \log \det (M(\xi)).$$

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

Jean B. Lasserre*  
semidefinite characterization
In the Moment-SOS approach we DO NOT discretize $X$ and rather search for an ATOMIC probability measure on $X$:

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

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

$$\sup_{\mu} \left\{ \log \det M_d(\mu) : \mu \in \mathcal{P}(X) \right\}.$$ 

where $\mathcal{P}(X)$ is the space of probability measures on $X$, and $M_d(\mu)$ is the (order-$d$) moment matrix of $\mu$.

Figure: Polygon and Sphere for $d = 3$
Consider the infinite dimensional LP:

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

where:

- $K \subset \mathbb{R}^n$ is a basic semi-algebraic set,
- The unknown $\phi$ is a Borel measure supported on $K$,
- The functions $f$, and $g \in G$ are polynomials,
- All moments of the measure $\mu$ are available.
For instance this framework can be used:

- To compute **Sharp Upper Bounds on** \( \mu(K) \) **GIVEN** some moments of \( \mu \).
- 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 \( \mu \), approximate **AS CLOSELY AS DESIRED**

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

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

For instance this framework can be used:

- To compute **Sharp Upper Bounds** on $\mu(K)$ **GIVEN** some moments of $\mu$.
- 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 $\mu$, approximate **AS CLOSELY AS DESIRED**

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

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

For instance this framework can be used:

- To compute **Sharp Upper Bounds** on $\mu(K)$ **GIVEN** some moments of $\mu$.
- 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 $\mu$, approximate **AS CLOSELY AS DESIRED**

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

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

For instance this framework can be used:

- To compute **Sharp Upper Bounds on** $\mu(K)$ **GIVEN** some moments of $\mu$.
- 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 $\mu$, approximate **AS CLOSELY AS DESIRED**

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

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

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