Moments, Positive Polynomials, and the Christoffel Function

Jean-Bernard Lasserre

LAAS – CNRS

12 décembre 2022
Moments, Positive Polynomials, and the Christoffel Function

Jean B. Lasserre*

LAAS-CNRS, Institute of Mathematics & ANITI, Toulouse, France

CERMICS COLLOQUIUM, December 2022
... in collaboration with
D. Henrion, M. Korda, V. Magron, S. Marx, E. Pauwels, M. Putinar, T. Weisser
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 context, mathematical finance, multivariate integration, etc., and examples are provided for each particular application.
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.
The Moment-SOS Hierarchy

Cambridge Monographs on Applied and Computational Mathematics

The Christoffel–Darboux Kernel for Data Analysis

Jean Bernard Lasserre, Edouard Pauwels, and Mihai Putinar

Jean B. Lasserre

Moments, Positive Polynomials, and the Christoffel Function
• Part I:
  • The Moment-SOS hierarchy
  • SOS-based CERTIFICATES of POSITIVITY
  • Illustration of the Moment-SOS hierarchy for POLYNOMIAL optimization

• Part II:
  • The Christoffel function
  • Applications and link with Optimization
• Part I:
  • The **Moment-SOS hierarchy**
  • **SOS-based CERTIFICATES of POSITIVITY**
    • Illustration of the Moment-SOS hierarchy for **POLYNOMIAL** optimization

• Part II:
  • The **Christoffel function**
  • Applications and link with Optimization
• Part I:
  - The Moment-SOS hierarchy
  - SOS-based CERTIFICATES of POSITIVITY
  - Illustration of the Moment-SOS hierarchy for POLYNOMIAL optimization

• Part II:
  - The Christoffel function
  - Applications and link with Optimization
• Part I:
  • The Moment-SOS hierarchy
  • SOS-based CERTIFICATES of POSITIVITY
  • Illustration of the Moment-SOS hierarchy for POLYNOMIAL optimization

• Part II:
  • The Christoffel function
  • Applications and link with Optimization
The Moment-SOS Hierarchy

A brief overview of the methodology on 3 examples

Let $P$ be the initial problem to solve, for instance

- **Optimization** $\min_x \{ f(x) : x \in K \}$
- **Optimal control**

$$\min_u \int_0^1 h(x(t), u(t)) \, dt$$

$$\dot{x}(t) = f(x(t), u(t)), \quad t \in (0, 1)$$

$$x(t) \in X; \quad u(t) \in U, \quad t \in [0, 1]$$

- **Compute** (or approximate)

$$\tau = \int_K x^\alpha \, d\lambda(x), \quad \alpha \in \Gamma,$$

and in particular, $\text{vol}(K)$ ($\Gamma = \{0\}$) where $K$ is compact basic semi-algebraic set.
basic strategy

(i) Search for a measure $\mu$ whose support is the solution

$$d\mu(x) = \delta_{x^*}; \quad d\mu(x, u, t) = \delta_{x(t), u(t)}(d(x, u)) \, dt; \quad d\mu(x) = \lambda_K(dx)$$

(ii) compute its moments, and

(iii) recover the solution from moments.

Implementation in Three steps:

I: LIFTING

I: Build up an infinite-dimensional LP with $\mu$ as unknown:

Constraints of the initial problem become

LINEAR constraints

on the unknown moments $(\mu_\alpha)_{\alpha \in \mathbb{N}^n}$ of $\mu$
II: Truncation

Consider only FINITELY MANY candidate moments $y_{\alpha}$:

- Semidefinite constraints on the scalars $y_{\alpha}$’s state necessary conditions to qualify them as moments of some measure $\mu$.
- Solve the resulting finite-dimensional convex (conic) optimization problem to obtain a guaranteed lower bound.

III: Iterate

- Increase the number of moments considered and iterate so as to obtain a monotone non increasing sequence of lower bounds which converges to the optimal value.
Consider the polynomial optimization problem:

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

for some polynomials \( f, g_j \in \mathbb{R}[x] \).
Consider the polynomial optimization problem:

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

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

After all ... \( \mathbf{P} \) is just a particular case of Non Linear Programming (\( \text{NLP} \))!

True!

... if one is interested with a \textbf{LOCAL} optimum only!!

- Many minimization algorithms do the job efficiently.
- The fact that \( f, g_j \) are \textbf{POLYNOMIALS} does not help much!
Why Polynomial Optimization?

After all ... \( P \) is just a particular case of Non Linear Programming (\textit{NLP})!

True!

... if one is interested with a \textbf{LOCAL} optimum only!!

Many minimization algorithms do the job efficiently.

The fact that \( f, g_j \) are \textbf{POLYNOMIALS} does not help much!
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!!

Many minimization algorithms do the job efficiently.

The fact that $f, g_j$ are POLYNOMIALS does not help much!
BUT for GLOBAL Optimization

... the picture is different!
For GLOBAL Optimization

... the picture is different!
Remember that for the \textbf{GLOBAL} minimum \( f^* \):

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

(Not true for a \textbf{LOCAL} minimum!)
Remember that for the **GLOBAL** minimum $f^*$:

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

(Not true for a **LOCAL** minimum!)
and so to compute (or approximate) $f^*$ ... one needs to handle **EFFICIENTLY** the difficult constraint

$$f(x) - \lambda \geq 0 \quad \forall x \in K, \quad (f - \lambda 1 \in \mathcal{P}(K)_+)$$

i.e. one needs

**TRACTABLE CERTIFICATES of POSITIVITY** on $K$ for the polynomial $x \mapsto f(x) - \lambda$!
and so to compute (or approximate) $f^*$ ... one needs to handle **EFFICIENTLY** the difficult constraint

$$f(x) - \lambda \geq 0 \quad \forall x \in K, \quad (f - \lambda 1 \in \mathcal{P}(K)_+)$$

i.e. one needs

**TRACTABLE CERTIFICATES of POSITIVITY** on $K$ for the polynomial $x \mapsto f(x) - \lambda$!
The Moment-SOS Hierarchy

REAL ALGEBRAIC GEOMETRY helps!!!!

Indeed, **POWERFUL CERTIFICATES OF POSITIVITY EXIST!**

Moreover .... and importantly,

Such certificates are amenable to **PRACTICAL COMPUTATION!**
REAL ALGEBRAIC GEOMETRY helps!!!!

Indeed, **POWERFUL CERTIFICATES OF POSITIVITY EXIST!**

Moreover .... and importantly,

Such certificates are amenable to **PRACTICAL COMPUTATION!**
A polynomial $p$ is a sum-of-squares (SOS) if and only if

$$p(x) = \sum_{k=1}^{s} q_k(x)^2, \quad \forall x \in \mathbb{R}^n,$$

for some polynomials $q_k$.

Detecting whether a given polynomial $p$ is SOS can be done efficiently by solving a SEMIDEFINITE PROGRAM.

A SEMIDEFINITE PROGRAM (SDP) is a CONIC, CONVEX OPTIMIZATION PROBLEM that can be solved EFFICIENTLY (up to arbitrary fixed precision).
SOS-based certificate

A polynomial $p$ is a sum-of-squares (SOS) if and only if

$$p(x) = \sum_{k=1}^{s} q_k(x)^2, \quad \forall x \in \mathbb{R}^n,$$

for some polynomials $q_k$.

Detecting whether a given polynomial $p$ is SOS can be done efficiently by solving a SEMIDEFINITE PROGRAM.

A SEMIDEFINITE PROGRAM (SDP) is a CONIC, CONVEX OPTIMIZATION PROBLEM that can be solved EFFICIENTLY (up to arbitrary fixed precision)
Let $v_d(t) = (1, t, t^2, \ldots, t^d)$ and let $p$ be of even degree $2d$. 

$$p(t) = \sum_{k=1}^{2d} p_k t^k \quad (= \langle p, v_{2d}(t) \rangle)$$

is SOS if and only if there exists $Q \succeq 0$ such that

$$p(t) = \begin{bmatrix} t^d \\ \vdots \\ t \\ 1 \end{bmatrix}^T \begin{bmatrix} a & b & c & \cdots \\ b & d & e & \cdots \\ c & e & f & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \begin{bmatrix} 1 \\ t \\ t^2 \\ \vdots \\ t^d \end{bmatrix}$$

\[Q \succeq 0\]
\[ Q \succeq 0 \Rightarrow Q = \sum_{j=1}^{s} q_k q_k^T \]

\[
\begin{bmatrix}
1 \\
t \\
t^2 \\
\vdots \\
t^d
\end{bmatrix}^T Q \begin{bmatrix}
1 \\
t \\
t^2 \\
\vdots \\
t^d
\end{bmatrix} = \sum_{k=1}^{s} \left( \begin{bmatrix}
1 \\
t \\
\vdots \\
t^d
\end{bmatrix}^T q_k \right) \left( q_k \begin{bmatrix}
1 \\
t \\
\vdots \\
t^d
\end{bmatrix} \right) = \sum_{k=1}^{s} q_k(t)^2
\]
Conversely if

\[ p(t) = \sum_{k=1}^{s} q_k(t)^2, \]

then write

\[
p(t) = \sum_{k=1}^{s} \left( \begin{bmatrix} 1 \\ t \\ \cdots \\ t^d \end{bmatrix} \right)^T \begin{bmatrix} q_k \\ q_k^T \begin{bmatrix} 1 \\ t \\ \cdots \\ t^d \end{bmatrix} \end{bmatrix}
\]

\[
= \sum_{k=1}^{s} \begin{bmatrix} 1 \\ t \\ \cdots \\ t^d \end{bmatrix}^T \left( \sum_{k=1}^{s} q_k q_k^T \right) \begin{bmatrix} 1 \\ t \\ \cdots \\ t^d \end{bmatrix}
\]
Example

Let $t \mapsto f(t) = 6 + 4t + 9t^2 - 4t^3 + 6t^4$. Is $f$ an SOS? Do we have

$$f(t) = \begin{bmatrix} 1 & t & t^2 \\ \end{bmatrix}^T \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix} \begin{bmatrix} 1 \\ t \\ t^2 \end{bmatrix}$$

We must have:

$$a = 6; \quad 2b = 4; \quad d + 2c = 9; \quad 2e = -4; \quad f = 6.$$
Example

Let \( t \mapsto f(t) = 6 + 4t + 9t^2 - 4t^3 + 6t^4 \). Is \( f \) an SOS? Do we have

\[
f(t) = \begin{bmatrix} 1 \\ t \\ t^2 \end{bmatrix}^T \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix} \begin{bmatrix} 1 \\ t \\ t^2 \end{bmatrix}
\]

\( Q \succeq 0 \)

We must have:

\[ a = 6; \quad 2b = 4; \quad d + 2c = 9; \quad 2e = -4; \quad f = 6. \]

And so we must find a scalar \( c \) such that

\[
Q = \begin{bmatrix} 6 & 2 & c \\ 2 & 9 - 2c & -2 \\ c & -2 & 6 \end{bmatrix} \succeq 0.
\]
With $c = -4$ we have

$$Q = \begin{bmatrix} 6 & 2 & -4 \\ 2 & 17 & -2 \\ -4 & -2 & 6 \end{bmatrix} \succeq 0.$$ 

et

$$Q = 2 \begin{bmatrix} \sqrt{2/2} \\ 0 \\ \sqrt{2/2} \end{bmatrix} \begin{bmatrix} \sqrt{2/2} \\ 0 \\ \sqrt{2/2} \end{bmatrix}^\prime + 9 \begin{bmatrix} 2/3 \\ -1/3 \\ -2/3 \end{bmatrix} \begin{bmatrix} 2/3 \\ -1/3 \\ -2/3 \end{bmatrix}^\prime$$

$$+ 18 \begin{bmatrix} 1/\sqrt{18} \\ 4/\sqrt{18} \\ -1/\sqrt{18} \end{bmatrix} \begin{bmatrix} 1/\sqrt{18} \\ 4/\sqrt{18} \\ -1/\sqrt{18} \end{bmatrix}^\prime$$
and so

\[ f(t) = (1 + t^2)^2 + (2 - t - 2t^2)^2 + (1 + 4t - t^2)^2 \]
Let $K := \{ x : g_j(x) \geq 0, \quad j = 1, \ldots, m \}$ be compact (with $g_1(x) = M - \|x\|^2$, so that $K \subset B(0, M)$).
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:

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

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

However ... In Putinar’s theorem
... nothing is said on the DEGREE of the SOS polynomials ($\sigma_j$)!

BUT ... GOOD news ..!!

Testing whether $\uparrow$ 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!
Immediate application

In ANY application where one need to impose that a polynomial $f$ (to be determined) must be positive on $K$, then:

### DECLARE

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

with the additional constraint $\deg(\sigma_j g_j) \leq 2t$ for all $j = 1, \ldots, m$.

where the degree-parameter $t$ is YOUR CHOICE!

Then identifying both sides of the identity yields:

- Linear constraints on the coefficients of $f$ and $\sigma_j$,
- Semidefinite constraints coming from SOS conditions on the $\sigma_j$'s
Immediate application

In ANY application where one need to impose that a polynomial $f$ (to be determined) must be positive on $K$, then:

DECLARE

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

with the additional constraint $\text{deg}(\sigma_j g_j) \leq 2t$ for all $j = 1, \ldots, m$.

where the degree-parameter $t$ is YOUR CHOICE!

Then identifying both sides of the identity yields:

- Linear constraints on the coefficients of $f$ and $\sigma_j$,
- Semidefinite constraints coming from SOS conditions on the $\sigma_j$'s
Immediate application

In ANY application where one need to impose that a polynomial \( f \) (to be determined) must be positive on \( K \), then:

**DECLARE**

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

with the additional constraint \( \deg(\sigma_j g_j) \leq 2t \) for all \( j = 1, \ldots, m \).

where the degree-parameter \( t \) is YOUR CHOICE!

Then identifying both sides of the identity yields:

- ⚫ **Linear constraints** on the coefficients of \( f \) and \( \sigma_j \),
- ⚫ **Semidefinite constraints** coming from SOS conditions on the \( \sigma_j \)'s
In ANY application where one needs to impose that a polynomial $f$ (to be determined) must be positive on $K$, then:

\[
\begin{align*}
\text{DECLARE} \\
\hat{f}(x) = \sigma_0(x) + \sum_{j=1}^{m} \sigma_j(x) g_j(x), \quad \forall x \in \mathbb{R}^n,
\end{align*}
\]

with the additional constraint $\deg(\sigma_j g_j) \leq 2t$ for all $j = 1, \ldots, m$.

where the degree-parameter $t$ is YOUR CHOICE!

Then identifying both sides of the identity yields:

- Linear constraints on the coefficients of $f$ and $\sigma_j$,
- Semidefinite constraints coming from SOS conditions on the $\sigma_j$'s
• In fact, polynomials **NONNEGATIVE ON A SET** $K \subset \mathbb{R}^n$ are ubiquitous. They also appear in many important applications (outside optimization),

among which

- Optimization, Probability, Optimal and Robust Control,
- non-linear PDEs, Game theory, Signal processing, multivariate integration, etc.
Given a real sequence $y = (y_{\alpha})$, $\alpha \in \mathbb{N}^n$, does there exist a Borel measure $\mu$ on $K$ such that

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

If yes then $y$ is said to have a representing measure supported on $K$. 
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 $g_j$. 
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_jy) \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 $M_d(g_jy)$ is called the **LOCALIZING MATRIX** associated with the sequence $y$ and the polynomial $g_j$. 

Jean B. Lasserre* 
Moments, Positive Polynomials, and the Christoffel Function
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 \( d = 0, 1, \ldots \)

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

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 \( K := \{ x : g_j(x) \geq 0, \ j = 1, \ldots, m \} \)
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 \( g_j \).
Remarkably,

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

Moment matrix \(M_1(y)\) in dimension 2 with \(d = 1\):

\[
M_1(y) = \begin{pmatrix}
1 & X_1 & X_2 \\
1 & y_{00} & y_{10} & y_{01} \\
X_1 & y_{10} & y_{20} & y_{11} \\
X_2 & y_{01} & y_{11} & y_{02}
\end{pmatrix}
\]
Remarkably,

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

\[ M_1(y) = \begin{pmatrix}
1 & X_1 & X_2 \\
1 & y_{00} & y_{10} & y_{01} \\
X_1 & y_{10} & y_{20} & y_{11} \\
X_2 & y_{01} & y_{11} & y_{02}
\end{pmatrix} \]

\[\text{Moment matrix } M_1(y) \text{ in dimension } 2 \text{ with } d = 1: \]
localizing matrix $M_1(g \mathbf{y})$ in dimension 2 with $d = 1$ and $g(x) = 1 - x_1^2 - x_2^2$:

$$
\begin{pmatrix}
1 & X_1 & X_2 \\
1 & y_{00} - y_{20} - y_{02} & y_{10} - y_{30} - y_{12} & y_{01} - y_{21} - y_{03} \\
X_1 & y_{10} - y_{30} - y_{12} & y_{20} - y_{40} - y_{22} & y_{11} - y_{31} - y_{13} \\
X_2 & y_{01} - y_{21} - y_{03} & y_{11} - y_{31} - y_{13} & y_{02} - y_{22} - y_{04}
\end{pmatrix}
$$
ALGEBRAIC SIDE

POSITIVITY ON $K$

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

$f > 0$ on $K$?

CHARACTERIZE THOSE $f$
The Moment-SOS Hierarchy

Jean B. Lasserre

Moments, Positive Polynomials, and the Christoffel Function

**ALGEBRAIC SIDE**

**Positivity on $K$**

$f(x) = \sum f_\alpha x^\alpha$

$f \geq 0$ on $K$?

**Characterize those $f$**

**Duality** \[ \langle f, y \rangle = \sum f_\alpha y_\alpha \]

**FUNCTIONAL ANALYSIS**

**The $K$-moment problem**

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

\[ y_\alpha \geq \int_K x^\alpha \, d\mu \quad \forall \alpha \]

for some $\mu$

**Characterize those $y$**
• In fact, polynomials \textbf{NONNEGATIVE ON A SET } \( K \subseteq \mathbb{R}^n \) are ubiquitous. They also appear in many important applications (outside optimization),

\[ \text{... modeled as} \]

\text{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: The primal view

The GMP is the infinite-dimensional LP:

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

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:

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

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 \(K_i, i = 1, \ldots, s\).
GMP: The dual view

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

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

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 $K_i, \ i = 1, \ldots, s$. 
The moment-SOS hierarchy

- is an **iterative numerical scheme** to (help) solve the **GMP**.

- It consists of using a certain type of **positivity certificate** (e.g., *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 SEMIDEFINITE PROGRAMS**

... of **increasing size**!.
The moment-SOS hierarchy

- is an **iterative numerical scheme** to (help) solve the **GMP**.
- It consists of using a certain type of **positivity certificate** (e.g., 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 SEMIDEFINITE PROGRAMS**

... of **increasing size**!
The Moment-SOS Hierarchy

Infinite-dimensional conic problem on moments

- Initial (difficult) problem with algebraic data:
  - nonconvex optimization
  - set approximation
  - nonlinear PDE,
  - etc.

Infinite-dimensional LP on measures

- Moment formulation

Infinite-dimensional conic problem on positive polynomials

- Positive Polynomials formulation

Infinite-dimensional LP on polynomials

- Finite-dimensional Primal SDPs of increasing size

- Truncation to finitely many semidefinite moment conditions

- Duality

Finite-dimensional Dual SDPs of increasing size

- Truncation to bounded degree SOS-based certificates of positivity

- Duality

Moment-SOS Hierarchy

Jean B. Lasserre
Moments, Positive Polynomials, and the Christoffel Function
The Moment-SOS Hierarchy

• Has already been proved 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
The Moment-SOS Hierarchy

Global optimization

Volume of semialgebraic set

Reachable set

Super resolution

Optimal control

Region of attraction

Maximum invariant sets

PDE analysis & control
Example: Global optimization

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

is the SIMPLEST example of the GMP

\[ f^* = \inf_{\mu \in \mathcal{M}(K)_+} \{ \int_K f d\mu : \int_K 1 d\mu = 1 \} \]

A GMP with only one unknown measure \( \mu \) and only one moment-constraint \( \int_K 1 d\mu = 1 \)
The Moment-SOS Hierarchy

Example: Global optimization

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

is the SIMPLEST example of the GMP

because ...

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

A GMP with only one unknown measure \( \mu \) and only one moment-constraint \( \int_K 1 \, d\mu = 1 \)
Example: Global optimization

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

is the SIMPLEST example of the GMP

because ...

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

\[ \blacklozenge \text{ A GMP with only one unknown measure } \mu \text{ and only one moment-constraint } \int_{K} 1 \, d\mu = 1 \]
Remember also that for the GLOBAL minimum $f^*$:

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

Then for each $t$ solve:

$$\rho_t = \sup_{\lambda, \sigma_j} \{ \lambda : f(x) - \lambda = \sigma_0(x) + \sum_{j=1}^{m} \sigma_j(x) g_j(x), \quad \forall x \in \mathbb{R}^n \} \quad \text{deg}(\sigma_j g_j) \leq t, \quad j = 0, \ldots, m \}$$

$$\rho_t \leq \rho_{t+1} \leq f^* \text{ for all } t \text{ and } \rho_t \uparrow f^* \text{ as } t \to \infty.$$
Alternatively, for each $t$ solve:

$$\rho_t^* = \inf_y \left\{ L_y(f) : \begin{array}{l} y_0 = 1 \\ M_t(y) \succeq 0 \\ M_{t-t_j}(g_j y) \succeq 0 \quad \forall j = 1, \ldots, m \end{array} \right\} \iff y_\alpha = \int_K x^\alpha d\mu$$

**Theorem (Lass 2000)**

$\rho_t \leq \rho_t^* \leq f^*$ for all $t$ and $\rho_t^* \uparrow f^*$ as $t \to \infty$.

Moreover, generically $\rho_t^* = f^*$ and one may extract global minimizers from the optimal (truncated moment) solution $y^*$. 
Alternatively, for each $t$ solve:

$\rho_t^* = \inf \{ L_y(f) : \text{think of } \int f \, d\mu \}$

\[
\begin{align*}
y_0 &= 1 \\
M_t(y) &\geq 0 \\
M_{t-t_j}(g_j y) &\geq 0 \quad \forall j = 1, \ldots, m
\end{align*}
\]

$\iff y_\alpha = \int_K x_\alpha \, d\mu$

**Theorem (Lass 2000)**

$\rho_t \leq \rho_t^* \leq f^* \text{ for all } t \text{ and } \rho_t^* \uparrow f^* \text{ as } t \to \infty.$

Moreover, generically $\rho_t^* = f^*$ and one may extract global minimizers from the optimal (truncated moment) solution $y^*$. 
Ex: Consider the optimization problem: \( \min \{ f(x) : x \in [0, 1] \} : \)

\[
x \mapsto f(x) := \sum_{j=1}^{4} a_j x^j ; \quad [0, 1] = \{ x : x(1-x) \geq 0 \},
\]

**SDP relaxation**

\[
\begin{align*}
\text{SDP} \quad f^* &= \min \{ \sum_{j=1}^{4} a_j y_j : & \begin{bmatrix} y_0 & y_1 & y_2 \\ y_1 & y_2 & y_3 \\ y_2 & y_3 & y_4 \end{bmatrix} \succeq 0 \\
\text{SDP} \quad & \begin{bmatrix} y_1 - y_2 \\ y_2 - y_3 \\ y_3 - y_4 \end{bmatrix} \succeq 0 ; \quad y_0 = 1. \}
\end{align*}
\]

\( y^* = (1, x^*, (x^*)^2, (x^*)^3, (x^*)^4) , \) and

\[
f(x) - f^* = \underbrace{\sigma_0(x)}_{\text{SOS of degree 4}} + \underbrace{\sigma_1(x)}_{\text{SOS of degree 2}} x(1-x).
\]
Ex: Consider the optimization problem: \( \min \{ f(x) : x \in [0, 1] \} \):

\[
x \mapsto f(x) := \sum_{j=1}^{4} a_j x^j ; \quad [0, 1] = \{ x : x(1 - x) \geq 0 \},
\]

**SDP relaxation**

\[
\begin{align*}
\text{SDP} \quad f^* &= \min_{y} \left\{ \sum_{j=1}^{4} a_j \ y_j : \begin{bmatrix} y_0 & y_1 & y_2 \\ y_1 & y_2 & y_3 \\ y_2 & y_3 & y_4 \end{bmatrix} \succeq 0 \right\} \\
\text{SDP} \quad &\quad \begin{bmatrix} y_1 - y_2 & y_2 - y_3 \\ y_2 - y_3 & y_3 - y_4 \end{bmatrix} \succeq 0 ; \quad y_0 = 1.
\end{align*}
\]

\( y^* = (1, x^*, (x^*)^2, (x^*)^3, (x^*)^4) \), and

\[
f(x) - f^* = \underbrace{\sigma_0(x)}_{\text{SOS of degree 4}} + \underbrace{\sigma_1(x)}_{\text{SOS of degree 2}} \ x(1 - x).
\]
Finite convergence is "generic"

Extraction of minimizers from an optimal solution of the dual (linear algebra)

If the problem is SOS-convex then convergence takes place at the first step of the hierarchy

The "same algorithm" for many combinatorial optimization problems (just use $x_i^2 = x_i$ to model boolean variables) which still provides better lower bounds than ad-hoc tailored algorithms.

Has become key tool to prove/disprove Khot’s Unique Games Conjecture in computational complexity.

The NPA-hierarchy is a non-commutative version of the Moment-SOS hierarchy to address some quantitative problems in Quantum Information.
In static optimization, an optimal solution is a point $x^* \in \mathbb{R}^n$.

**Generically**, some semidefinite relaxation at step $t$ of the Moment-SOS hierarchy is exact and:

- To recover $x^*$ from its optimal solution $y^* = (y^*_\alpha)_{\alpha \in \mathbb{N}_2^n}$ can be done via a linear algebra subroutine.
- If $x^*$ is unique then it is even trivial as $x^*$ is just the subvector of degree-1 moments of $y^*$. 
However,

in many other problems like optimal control, PDE’s, computational geometry, an optimal solution is a function $f : \Omega \rightarrow \mathbb{R}$ (e.g. a trajectory $\{x(t) : t \in [0, 1]\}$)

... and the Moment-SOS hierarchy provides a sequence of scalars $(\mu_{\alpha,j})_{\alpha,j}$ which approximates moments

$$\mu^{*}_{\alpha,j} = \int_{\Omega} x^\alpha y^j d\mu^{*}(x, y), \quad \alpha \in \mathbb{N}^n, \ j \in \mathbb{N},$$

of the measure $d\mu^{*}(x, y) = \delta_{\{f(x)\}}(dy) \phi(dx)$

whose support IS the graph $\{(x, f(x)) : x \in \Omega\}$ of the optimal solution $f$. 
For instance in optimal control (OC), one uses the weak formulation of OC

- infinite-dimensional LP on occupation (Young) measure $\mu$

- Controlled dynamics of OC
  - linear constraints on moments of $\mu$ via integration of polynomial test functions

- integral cost functional
  - linear criterion $\langle h, \mu \rangle$ on $\mu$

- state/control constraints = support constraints on $\mu$
  - semidefinite conditions on moments of $\mu$ (by Putinar theorem)
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.
Then it remains to extract $f$ from knowledge of the $(\mu_{\alpha,j})$ ... 

This can be done by several techniques (including $L^2$-polynomial approximation via a standard application of the Christoffel-Darboux kernel) not detailed here.
We claim that a non-standard application of the CD kernel provides a simple and easy to use tool (with no optimization involved) which can help solve problems not only in data analysis, but also in approximation and interpolation of (possibly discontinuous) functions. In particular one is able to recover a discontinuous function with no Gibbs phenomenon.
Part two:

The Christoffel function
The Moment-SOS Hierarchy

Jean B. Lasserre

Moments, Positive Polynomials, and the Christoffel Function

Cambridge Monographs on Applied and Computational Mathematics

The Christoffel–Darboux Kernel for Data Analysis

Jean Bernard Lasserre, Edouard Pauwels and Mihai Putinar

Jean B. Lasserre*
The Moment-SOS Hierarchy

Moments, Positive Polynomials, and the Christoffel Function
Consider the following cloud of 2D-points (data set) below.

The red curve is the level set

\[ S_\gamma := \{ x : Q_d(x) \leq \gamma \}, \quad \gamma \in \mathbb{R}_+ \]

of a certain polynomial \( Q_d \in \mathbb{R}[x_1, x_2] \) of degree 2\( d \).

Notice that \( S_\gamma \) captures quite well the shape of the cloud.
Not a coincidence!

\[\text{Surprisingly, low degree } d \text{ for } Q_d \text{ is often enough to get a pretty good idea of the shape of } \Omega \text{ (at least in dimension } p = 2, 3)\]
Perform the following simple operations on a preferred cloud of 2D-points: So let $d = 2$, $p = 2$ and $s(d) = \binom{p+d}{p}$.

- Let $v_d(x)^T = (1, x_1, x_2, x_1^2, x_1 x_2, \ldots, x_1 x_2^{d-1}, x_2^d)$. be the vector of all monomials $x_1^i x_2^j$ of total degree $i + j \leq d$
- Form the real symmetric matrix of size $s(d)$

$$M_d := \frac{1}{N} \sum_{i=1}^{N} v_d(x(i)) v_d(x(i))^T,$$

where the sum is over all points $(x(i))_{i=1,...,N} \subset \mathbb{R}^2$ of the data set.
Note that $M_d$ is the **MOMENT-matrix** $M_d(\mu^N)$ of the empirical measure

$$\mu^N := \frac{1}{N} \sum_{i=1}^{N} \delta_{x(i)}$$

associated with a sample of size $N$, drawn according to an unknown measure $\mu$. 

The (usual) notation $\delta_{x(i)}$ stands for the **DIRAC** measure supported at the point $x(i)$ of $\mathbb{R}^2$. 

Jean B. Lasserre*  
Moments, Positive Polynomials, and the Christoffel Function
Recall that the moment matrix $\mathbf{M}_d(\mu)$ is real symmetric with rows and columns indexed by $(x^\alpha)_{\alpha \in \mathbb{N}_d^p}$, and with entries

$$
\mathbf{M}_d(\mu)(\alpha, \beta) := \int_{\Omega} x^{\alpha+\beta} \, d\mu = \mu_{\alpha+\beta}, \quad \forall \alpha, \beta \in \mathbb{N}_d^p.
$$

Illustrative example in dimension 2 with $d = 1$:

$$
\mathbf{M}_1(\mu) := \begin{pmatrix}
1 & X_1 & X_2 \\
1 & \mu_{00} & \mu_{10} & \mu_{01} \\
X_1 & \mu_{10} & \mu_{20} & \mu_{11} \\
X_2 & \mu_{01} & \mu_{11} & \mu_{02}
\end{pmatrix}
$$

is the moment matrix of $\mu$ of "degree $d=1$".
Next, form the SOS polynomial:

\[
\mathbf{x} \mapsto Q_d(\mathbf{x}) := \mathbf{v}_d(\mathbf{x})^T \mathbf{M}_d^{-1}(\mu^N) \mathbf{v}_d(\mathbf{x}).
\]

\[
= (1, x_1, x_2, x_1^2, \ldots, x_2^d) \mathbf{M}_d^{-1}(\mu^N) \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_2^d \end{pmatrix}
\]

Plot some level sets

\[
S_\gamma := \{ \mathbf{x} \in \mathbb{R}^2 : Q_d(\mathbf{x}) = \gamma \}
\]

for some values of \( \gamma \), the thick one representing the particular value \( \gamma = \binom{2+d}{2} \).
The Christoffel function $\Lambda_d : \mathbb{R}^p \rightarrow \mathbb{R}_+$ is the reciprocal

$$x \mapsto Q_d(x)^{-1}, \quad \forall x \in \mathbb{R}^p$$

of the SOS polynomial $Q_d$.

It has a rich history in Approximation theory and Orthogonal Polynomials.

Among main contributors: Nevai, Totik, Króo, Lubinsky, Simon, ...
The Christoffel function $\Lambda_d : \mathbb{R}^p \rightarrow \mathbb{R}_+$ is the reciprocal

$$x \mapsto Q_d(x)^{-1}, \quad \forall x \in \mathbb{R}^p$$

of the SOS polynomial $Q_d$.

\[\Rightarrow\] It has a rich history in Approximation theory and Orthogonal Polynomials.

\[\Rightarrow\] Among main contributors: Nevai, Totik, Króo, Lubinsky, Simon, \ldots
Let $\Omega \subset \mathbb{R}^p$ be the compact support of $\mu$ with nonempty interior, and $(P_\alpha)_{\alpha \in \mathbb{N}^p}$ be a family of orthonormal polynomials w.r.t. $\mu$.

The vector space $\mathbb{R}[x]_d$ viewed as a subspace of $L^2(\mu)$ is a Reproducing Kernel Hilbert Space (RKHS). Its reproducing kernel

$$(x, y) \mapsto K_d^\mu(x, y) := \sum_{|\alpha| \leq d} P_\alpha(x) P_\alpha(y), \quad \forall x, y \in \mathbb{R}^p,$$

is called the Christoffel-Darboux kernel.
The reproducing property

\[ x \mapsto q(x) = \int_{\Omega} K^\mu_d(x, y) q(y) \, d\mu(y), \quad \forall q \in \mathbb{R}[x]_d. \]

useful to determinate the best degree-\(d\) polynomial approximation

\[ \inf_{q \in \mathbb{R}[x]_d} \| f - q \|_{L^2(\mu)} \]

of \(f\) in \(L^2(\mu)\). Indeed:

\[ x \mapsto \hat{f}_d(x) := \sum_{\alpha \in \mathbb{N}^p_d} \left( \int_{\Omega} f(y) P_\alpha(y) \, d\mu(y) \right) P_\alpha(x) \in \mathbb{R}[x]_d \]

\[ = \arg \min_{q \in \mathbb{R}[x]_d} \| f - q \|_{L^2(\mu)} \]
The Christoffel function $\Lambda_d^\mu : \mathbb{R}^p \rightarrow \mathbb{R}_+$ is defined by:

$$
\xi \mapsto \Lambda_d^\mu(\xi)^{-1} = \sum_{|\alpha| \leq d} P_\alpha(\xi)^2 = K_d^\mu(\xi, \xi), \quad \forall \xi \in \mathbb{R}^p,
$$

and it also satisfies the variational property:

$$
\Lambda_d^\mu(\xi) = \min_{P \in \mathbb{R}[x]_d} \left\{ \int_\Omega P^2 \, d\mu : P(\xi) = 1 \right\}, \quad \forall \xi \in \mathbb{R}^p.
$$

Alternatively

$$
\Lambda_d^\mu(\xi)^{-1} = v_d(\xi)^T M_d(\mu)^{-1} v_d(\xi), \quad \forall \xi \in \mathbb{R}^p.
$$
The Christoffel function $\Lambda_d^\mu : \mathbb{R}^p \to \mathbb{R}_+$ is defined by:

$$\xi \mapsto \Lambda_d^\mu(\xi)^{-1} = \sum_{|\alpha| \leq d} P_\alpha(\xi)^2 = K_d^\mu(\xi, \xi), \quad \forall \xi \in \mathbb{R}^p,$$

and it also satisfies the variational property:

$$\Lambda_d^\mu(\xi) = \min_{P \in \mathbb{R}[x]_d} \left\{ \int_\Omega P^2 \, d\mu : P(\xi) = 1 \right\}, \quad \forall \xi \in \mathbb{R}^p.$$

Alternatively

$$\Lambda_d^\mu(\xi)^{-1} = v_d(\xi)^T M_d(\mu)^{-1} v_d(\xi), \quad \forall \xi \in \mathbb{R}^p.$$
Importantly, and crucial for applications, the Christoffel function identifies the support $\Omega$ of the underlying measure $\mu$.

**Theorem**

Let the support $\Omega$ of $\mu$ be compact with nonempty interior. Then:

1. For all $x \in \text{int}(\Omega)$: $K^\mu_d(x, x) = O(d^p)$.
2. For all $x \in \text{int}(\mathbb{R}^p \setminus \Omega)$: $K^\mu_d(x, x) = \Omega(\exp(\alpha d))$ for some $\alpha > 0$.

In particular, as $d \to \infty$,

$$d^p \wedge^\mu_d(x) \to 0 \text{ very fast whenever } x \notin \Omega.$$
Growth rates for $K_d^\mu(x, x) = \Lambda_d^\mu(x)^{-1}$. 

\[ d^p \exp(\alpha d) \]

\[ d^p+1 \]

\[ d^p+2 \]

\[ \exp(\alpha \sqrt{d}) \]
Some other properties

- Under some (restrictive) assumption on $\Omega$ and $\mu$
  \[
  \lim_{d \to \infty} s(d) \Lambda_d^\mu(\xi) = f_\mu(\xi) \omega(\xi)^{-1}
  \]
  where $\omega$ is the density of an equilibrium measure intrinsically associated with $\Omega$.
  For instance with $\rho = 1$ and $\Omega = [-1, 1]$, $\omega(\xi) = \sqrt{1 - \xi^2}$.

- If $\mu$ and $\nu$ have same support $\Omega$ and respective densities $f_\mu$ and $f_\nu$ w.r.t. Lebesgue measure on $\Omega$, positive on $\Omega$, then:
  \[
  \lim_{d \to \infty} \frac{\Lambda_d^\mu(\xi)}{\Lambda_d^\nu(\xi)} = \frac{f_\mu(\xi)}{f_\nu(\xi)}, \quad \forall \xi \in \Omega.
  \]
  useful for density approximation
Some other properties

- Under some (restrictive) assumption on $\Omega$ and $\mu$

$$\lim_{d \to \infty} s(d) \wedge_d^\mu(\xi) = f_\mu(\xi) \omega(\xi)^{-1}$$

where $\omega$ is the density of an equilibrium measure intrinsically associated with $\Omega$.

For instance with $p = 1$ and $\Omega = [-1, 1]$, $\omega(\xi) = \sqrt{1 - \xi^2}$.

- If $\mu$ and $\nu$ have same support $\Omega$ and respective densities $f_\mu$ and $f_\nu$ w.r.t. Lebesgue measure on $\Omega$, positive on $\Omega$, then:

$$\lim_{d \to \infty} \frac{\wedge_d^\mu(\xi)}{\wedge_d^\nu(\xi)} = \frac{f_\mu(\xi)}{f_\nu(\xi)}, \quad \forall \xi \in \Omega.$$

useful for density approximation
For instance one may decide to classify as **outliers** all points \( \xi \) such that \( \wedge_d^{\mu^N}(\xi) < \binom{p+d}{p}^{-1} \).

Such a strategy (even with relatively low degree \( d \)) is as efficient as more elaborated techniques, with only one parameter (the degree \( d \)), and with no optimization involved.

For instance one may decide to classify as outliers all points \( \xi \) such that \( \Lambda_{d}^{\mu_{N}}(\xi) < \binom{p+d}{p}^{-1} \).

Such a strategy (even with relatively low degree \( d \)) is as efficient as more elaborated techniques, with only one parameter (the degree \( d \)), and with no optimization involved.

For instance one may decide to classify as outliers all points \( \xi \) such that \( \Lambda_d^{\mu_N}(\xi) < \binom{p+d}{p}^{-1} \).

Such a strategy (even with relatively low degree \( d \)) is as efficient as more elaborated techniques, with only one parameter (the degree \( d \)), and with no optimization involved.

A measure $\mu$ on compact set $\Omega$ is completely determined by its moments and therefore it should not be a surprise that its moment matrix $M_d(\mu)$ contains a lot of information.

We have already seen that its inverse $M_d(\mu)^{-1}$ defines the Christoffel function.

When $\mu$ is degenerate and its support $\Omega$ is contained in a zero-dimensional real algebraic variety $V$ then the kernel of $M_d(\mu)$ identifies the generators of a corresponding ideal of $\mathbb{R}[x]$ (the vanishing ideal of $V$).
For instance let $\Omega \subset \mathbb{S}^{p-1}$ (the Euclidean unit sphere of $\mathbb{R}^p$).
Then the kernel of $M_d(\mu)$ contains vectors of coefficients of polynomials in the ideal generated by the quadratic polynomial $x \mapsto g(x) := 1 - \|x\|^2$.

In fact and remarkably,

$$\text{rank } M_d(\mu) = p(d)$$

for some univariate polynomial $p$ (the Hilbert polynomial associated with the algebraic variety) which is of degree $t$ if $t$ is the dimension of the variety.

For instance $t = p - 1$ if the support is contained in the sphere $S^{p-1}$ of $\mathbb{R}^p$.

Again this illustrates how quite sophisticated concepts of algebraic geometry are hidden and encapsulated in the moment matrix $M_d(\mu)$.

They can be exploited to extract various useful information on the data set.

In addition, extraction of this information can be done via quite simple linear algebra techniques.
Again this illustrates how quite sophisticated concepts of algebraic geometry are hidden and encapsulated in the moment matrix $M_d(\mu)$.

They can be exploited to extract various useful information on the data set.

In addition, extraction of this information can be done via quite simple linear algebra techniques.
Again this illustrates how quite sophisticated concepts of algebraic geometry are hidden and encapsulated in the moment matrix $M_d(\mu)$.

They can be exploited to extract various useful information on the data set.

In addition, extraction of this information can be done via quite simple linear algebra techniques.
However

for non modest dimension of data, matrix inversion of $M_d^{-1}$ does not scale well ...

On the other hand

for evaluation $\Lambda_d^\mu(\xi)$ at a point $\xi \in \mathbb{R}^p$, the variational formulation

$$\Lambda_d^\mu(\xi) = \min_{P \in \mathbb{R}[x]_d} \left\{ \int_\Omega P^2 d\mu : P(\xi) = 1 \right\}, \quad \forall \xi \in \mathbb{R}^p.$$  

is the simple quadratic programming problem.

$$\min_{p \in \mathbb{R}^{s(d)}} \left\{ p^T M_d p : v_d(\xi)^T p = 1 \right\},$$

which can be solved quite efficiently.

Jean B. Lasserre*  
Moments, Positive Polynomials, and the Christoffel Function
However

for non modest dimension of data, matrix inversion of $M^{-1}_d$ does not scale well ...

On the other hand

for evaluation $\Lambda^\mu_d(\xi)$ at a point $\xi \in \mathbb{R}^p$, the variational formulation

$$
\Lambda^\mu_d(\xi) = \min_{P \in \mathbb{R}[x]_d} \left\{ \int_{\Omega} P^2 \, d\mu : P(\xi) = 1 \right\}, \quad \forall \xi \in \mathbb{R}^p.
$$

is the simple quadratic programming problem.

$$
\min_{p \in \mathbb{R}^{s(d)}} \left\{ p^T M_d p : \nu_d(\xi)^T p = 1 \right\},
$$

which can be solved quite efficiently.
The Christoffel function for approximation

A typical approach is to approximate $f : [0, 1] \rightarrow \mathbb{R}$ in some function space, e.g. its projection on $\mathbb{R}[x]_n \subset L^2([0, 1])$:

$$x \mapsto \hat{f}_n(x) := \sum_{j=0}^{n} \left( \int_0^1 f(y) L_j(y) \, dy \right) L_j(x),$$

with an orthonormal basis $\{L_j\}_{j \in \mathbb{N}}$ of $L^2([0, 1])$.

BUT ... Ex: Chebyshev interpolant

† Typical Gibbs phenomenon occurs.
Alternative **Positive Kernels** with better convergence properties have been proposed, still in the same framework:

- **Féjer, Jackson kernels, etc.**
  - Reproducing property of the CD kernel is **LOST**
  - Preserve positivity (e.g. when approximating a density)
  - Better convergence properties than the CD kernel, in particular uniform convergence (for continuous functions) on arbitrary compact subsets
An alternative approach, still via the CD-kernel

A counter-intuitive detour: Instead of considering \( f : [0, 1] \to \mathbb{R} \), and the associated measure

\[
d\mu(x) := f(x) \, dx
\]
on the real line, whose support is \([0, 1] \in \mathbb{R}\),

Rather consider the graph \( \Omega \subset \mathbb{R}^2 \) of \( f \), i.e., the set

\[
\Omega := \{(x, f(x)) : x \in [0, 1]\}.
\]
and the measure

\[
d\phi(x, y) := \delta_{f(x)}(dy) \, 1_{[0,1]}(x) \, dx
\]
on \( \mathbb{R}^2 \) with degenerate support \( \Omega \subset \mathbb{R}^2 \).
A counter-intuitive detour: Instead of considering $f : [0, 1] \to \mathbb{R}$, and the associated measure

$$d\mu(x) := f(x) \, dx$$

on the real line, whose support is $[0, 1] \subset \mathbb{R}$,

Rather consider the graph $\Omega \subset \mathbb{R}^2$ of $f$, i.e., the set

$$\Omega := \{ (x, f(x)) : x \in [0, 1] \}.$$

and the measure

$$d\phi(x, y) := \delta_{f(x)}(dy) \, 1_{[0,1]}(x) \, dx$$

on $\mathbb{R}^2$ with degenerate support $\Omega \subset \mathbb{R}^2$. 
Why should we do that as it implies going to $\mathbb{R}^2$ instead of staying in $\mathbb{R}$?

... because

- The support of $\phi$ is exactly the graph of $f$, and
- The CF $(x, y) \mapsto \Lambda_n^\phi(x, y)$ identifies the support of $\phi$!
So suppose that we know the moments

\[ \phi_{i,j} = \int x^i y^j d\phi(x, y) = \int_{[0,1]} x^i f(x)^j \, dx, \quad i + j \leq 2d, \]

and let \( \varepsilon > 0 \) and \( \lambda \) be the Lebesgue measure on \([0,1]\).

- Compute the degree-\( d \) moment matrix of \( \phi \):

\[ M_d(\phi) := \int v_d(x, y) v_d(x, y)^T \, d\phi(x, y), \]

- Compute the Christoffel function

\[ x \mapsto \Lambda_{d, \varepsilon}^{\phi, \varepsilon}(x, y)^{-1} := v_d(x, y)^T M_d(\phi + \varepsilon \lambda)^{-1} v_d(x, y). \]

- Approximate \( f(x) \) by \( \hat{f}_{d, \varepsilon}(x) := \arg \min_y \Lambda_{d, \varepsilon}^{\phi, \varepsilon}(x, y)^{-1} \).

\( \varepsilon \) minimize a univariate polynomial! (easy)
So suppose that we know the moments

\[ \phi_{i,j} = \int x^i y^j \, d\phi(x, y) = \int_{[0,1]} x^i f(x)^j \, dx, \quad i + j \leq 2d, \]

and let \( \varepsilon > 0 \) and \( \lambda \) be the Lebesgue measure on \([0, 1]\).

- **Compute the degree-\( d \) moment matrix of \( \phi \):**

\[ M_d(\phi) := \int \mathbf{v}_d(x, y) \mathbf{v}_d(x, y)^T \, d\phi(x, y), \]

- **Compute the Christoffel function**

\[ x \mapsto \Lambda_d^{\phi,\varepsilon}(x, y)^{-1} := \mathbf{v}_d(x, y)^T M_d(\phi + \varepsilon \lambda)^{-1} \mathbf{v}_d(x, y). \]

- Approximate \( f(x) \) by \( \hat{f}_{d,\varepsilon}(x) := \arg \min_y \Lambda_d^{\phi,\varepsilon}(x, y)^{-1}. \)

  minimize a univariate polynomial! (easy)
So suppose that we know the moments

\[ \phi_{i,j} = \int x^i y^j \, d\phi(x, y) = \int_{[0,1]} x^i f(x)^j \, dx, \quad i + j \leq 2d, \]

and let \( \varepsilon > 0 \) and \( \lambda \) be the Lebesgue measure on \([0, 1]\).

- Compute the degree-\( d \) moment matrix of \( \phi \):

\[
M_d(\phi) := \int v_d(x, y) v_d(x, y)^T \, d\phi(x, y),
\]

- Compute the Christoffel function

\[
x \mapsto \Lambda_{d,\varepsilon}^\phi(x, y)^{-1} := v_d(x, y)^T M_d(\phi + \varepsilon \lambda)^{-1} v_d(x, y). \]

- Approximate \( f(x) \) by \( \hat{f}_{d,\varepsilon}(x) := \arg \min_y \Lambda_{d,\varepsilon}^\phi(x, y) \).

\(\varepsilon\) minimize a univariate polynomial! (easy)
The Moment-SOS Hierarchy

Interpolation: same story

So suppose that you are given point evaluations \( \{ f(x_i) \}_{i \leq N} \) of an unknown function \( f \) on \([0, 1] \), and again let

\[
v_d(x, y) := (1, x, y, x^2, x y, y^2, \ldots, x y^{d-1}, y^d).
\]

Compute the degree-\( d \) empirical moment matrix:

\[
M_d(\phi) := \sum_{i=1}^{N} v_d((x_i, f(x_i))) v_d(x_i, f(x_i))^T,
\]

of the empirical measure \( d\phi(x, y) := \frac{1}{N} \sum_{i=1}^{N} \delta_{x(i), f(x(i))} \) on \( \mathbb{R}^2 \), by one pass over the data.

Compute the Christoffel function

\[
x \mapsto \Lambda_{d}^{\phi, \epsilon}(x, y)^{-1} := v_d(x, y)^T M_d(\phi + \epsilon \lambda)^{-1} v_d(x, y).
\]

Approximate \( f(x) \) by \( \hat{f}_{d, \epsilon}(x) := \arg \min_y \Lambda_{d}^{\phi, \epsilon}(x, y)^{-1}. \)

minimize a univariate polynomial! (easy)
Interpolation: same story

So suppose that you are given point evaluations \( \{ f(x_i) \}_{i \leq N} \) of an unknown function \( f \) on \([0, 1]\), and again let

\[
v_d(x, y) := (1, x, y, x^2, xy, y^2, \ldots, xy^{d-1}, y^d).
\]

Compute the degree-\( d \) empirical moment matrix:

\[
M_d(\phi) := \sum_{i=1}^{N} v_d((x_i, f(x_i))) v_d(x_i, f(x_i))^T,
\]

of the empirical measure \( d\phi(x, y) := \frac{1}{N} \sum_{i=1}^{N} \delta_{x(i), f(x(i))} \) on \( \mathbb{R}^2 \), by one pass over the data.

Compute the Christoffel function

\[
x \mapsto \Lambda_d^{\phi, \varepsilon}(x, y)^{-1} := v_d(x, y)^T M_d(\phi + \varepsilon \lambda)^{-1} v_d(x, y).
\]

Approximate \( f(x) \) by \( \hat{f}_{d, \varepsilon}(x) := \arg \min_y \Lambda_d^{\phi, \varepsilon}(x, y)^{-1} \).

minimize a univariate polynomial! (easy)
So suppose that you are given point evaluations \( \{ f(x_i) \}_{i \leq N} \) of an unknown function \( f \) on \([0, 1]\), and again let

\[
v_d(x, y) := (1, x, y, x^2, xy, y^2, \ldots, xy^{d-1}, y^d).
\]

Compute the degree-\( d \) empirical moment matrix:

\[
M_d(\phi) := \sum_{i=1}^{N} v_d((x_i, f(x_i))) v_d(x_i, f(x_i))^T,
\]

of the empirical measure \( d\phi(x, y) := \frac{1}{N} \sum_{i=1}^{N} \delta_{x(i), f(x(i))} \) on \( \mathbb{R}^2 \), by one pass over the data.

Compute the Christoffel function

\[
x \mapsto \Lambda_{d}^{\phi, \varepsilon}(x, y)^{-1} := v_d(x, y)^T M_d(\phi + \varepsilon \lambda)^{-1} v_d(x, y).
\]

Approximate \( f(x) \) by \( \hat{f}_{d, \varepsilon}(x) := \arg \min_y \Lambda_{d}^{\phi, \varepsilon}(x, y)^{-1} \).

minimize a univariate polynomial! (easy)
Interpolation: same story

So suppose that you are given point evaluations \( \{ f(x_i) \}_{i \leq N} \) of an unknown function \( f \) on \([0, 1]\), and again let

\[
v_d(x, y) := (1, x, y, x^2, xy, y^2, \ldots, xy^{d-1}, y^d).
\]

Compute the degree-\( d \) empirical moment matrix:

\[
M_d(\phi) := \sum_{i=1}^{N} v_d((x_i, f(x_i))) v_d(x_i, f(x_i))^T,
\]

of the empirical measure \( d\phi(x, y) := \frac{1}{N} \sum_{i=1}^{N} \delta_{x(i),f(x(i))} \) on \( \mathbb{R}^2 \), by one pass over the data

Compute the Christoffel function

\[
x \mapsto \Lambda_{d, \varepsilon}^{\phi, \varepsilon}(x, y)^{-1} := v_d(x, y)^T M_d(\phi + \varepsilon \lambda)^{-1} v_d(x, y).
\]

Approximate \( f(x) \) by \( \hat{f}_{d, \varepsilon}(x) := \arg \min_y \Lambda_{d, \varepsilon}^{\phi, \varepsilon}(x, y)^{-1} \).

minimize a univariate polynomial! (easy)
Choosing

\[ \varepsilon := 2^{3 - \sqrt{d}} \]

ensures convergence properties for bounded measurable functions, e.g. pointwise on open sets with no point of discontinuity.

Convergence properties as \( d \uparrow \)

- \( L^1 \)-convergence
- pointwise convergence on open sets with no point of discontinuity, and so almost uniform convergence.
- \( L^1 \)-convergence at a rate \( O(d^{-1/2}) \) for Lipschitz continuous \( f \).
In non trivial examples, the approximation is quite good with small values of $d$, and with no Gibbs phenomenon.
Ex: Recovery

Below: Recovery of a (discontinuous) solution of the Burgers Equation from knowledge of approximate moments of the occupation measure supported on the solution.
Again note the central role played by the Moment Matrix!

Let $\Omega \subset \mathbb{R}^n$ be the basic semi-algebraic set (with nonempty interior)

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

with $g_j \in \mathbb{R}[x]_{d_j}$ and let $s_j = \lceil \deg(g_j)/2 \rceil$. Let $g_0 = 1$ with $s_0 = 0$.

With $t$ fixed, its associated quadratic module

$$
Q_t(\Omega) := \left\{ \sum_{j=0}^{m} \sigma_j g_j : \quad \sigma_j \in \sum [x]_{t-s_j} \right\} \subset \mathbb{R}[x]
$$

is a convex cone with nonempty interior,
The Moment-SOS Hierarchy

and with dual convex cone of pseudo-moments

\[ Q_t(\Omega)^* := \{ y \in \mathbb{R}^{s(t)} : M_{t-s_j}(g_j y) \succeq 0, \quad j = 0, \ldots, m \}, \]

where \( s(t) = \binom{n+t}{n} \).

Notice that if \( M_t(y)^{-1} \succ 0 \) for all \( t \),
then one may define a family of polynomials \((P_\alpha)_{\alpha \in \mathbb{N}^n} \subset \mathbb{R}[x]\)
orthonormal w.r.t. \( y \), meaning that

\[ L_y(P_\alpha \cdot P_\beta) = \delta_{\alpha=\beta}, \quad \alpha, \beta \in \mathbb{N}^n, \]

and exactly as for measures, the Christoffel function \( \Lambda_t^y \)

\[ x \mapsto \Lambda_t^y(x)^{-1} := \sum_{|\alpha| \leq t} P_\alpha(x)^2. \]
The Moment-SOS Hierarchy

Theorem

For every \( p \in \text{int}(Q_t(\Omega)) \) there exists a sequence of pseudo-moments \( y \in \text{int}(Q_t(\Omega)^*) \) such that

\[
p(x) = \sum_{j=0}^{m} \left( v_{t-s_j}(x)^T M_t(g_j y)^{-1} v_{t-s_j}(x) \right) g_j(x)
\]

where \((g \cdot y)\) is the sequence of pseudo-moments

\[
(g \cdot y)_\alpha := \sum_{\gamma} g_\gamma y_{\alpha+\gamma}, \quad \alpha \in \mathbb{N}^n \quad \text{(if } g(x) = \sum_{\gamma} g_\gamma x^\gamma \text{)}. 
\]

In addition \( L_y(p) = \sum_{j=0}^{m} \binom{n+t-s_j}{n} \). 

Jean B. Lasserre

Moments, Positive Polynomials, and the Christoffel Function
The Moment-SOS Hierarchy

Moments, Positive Polynomials, and the Christoffel Function
The proof combines

- a result by Nesterov on a one-to-one correspondence between \( \text{int}(Q_t(\Omega)) \) and \( \text{int}(Q_t(\Omega)^*) \), and

- the fact that

\[
\mathbf{v}_{t-s_j}(x)^T \mathbf{M}_t(g_j \cdot y)^{-1} \mathbf{v}_{t-s_j}(x) = \Lambda_{t-s_j}(x)^{-1}.
\]

The proof combines

- a result by Nesterov on a one-to-one correspondence between $\text{int}(Q_t(\Omega))$ and $\text{int}(Q_t(\Omega)^*)$, and

- the fact that

$$v_{t-s_j}(x)^T M_t(g_jy)^{-1} v_{t-s_j}(x) = \Lambda_{t-s_j}^{g_j \cdot y}(x)^{-1}.$$  

In other words:

If \( p \in \text{int}(Q_t(\Omega)) \) then in Putinar’s certificate

\[
p = \sum_{j=0}^{m} \sigma_j g_j, \quad \sigma_j \in \mathbb{R}[x]_{t-s_j},
\]

of positivity of \( p \) on \( \Omega \),

one may always choose the SOS weights \( \sigma_j \) in the form

\[
\sigma_j(x) := \Lambda_{t-s_j}^{g_j \cdot y}(x)^{-1}, \quad j = 0, \ldots, m,
\]

for some sequence of pseudo-moments \( y \in \text{int}(Q_t(\Omega)^*) \).
In particular, every SOS polynomial $p$ of degree $2d$, in the interior of the SOS-cone, is the reciprocal of the CF of some linear functional $y \in \mathbb{R}[x]_{2d}^*$. That is:

$$p(x) = v_d(x)^T M_d(y)^{-1} v_x(x) = \Lambda_d^y(x)^{-1}, \quad \forall x \in \mathbb{R}^n.$$
What is the link between $p \in \text{int}(Q_t(\Omega))$ and the mysterious linear functional $y$?

**Theorem**

For some sets $\Omega$, $1 \in \text{int}(Q_t(\Omega))$ and

$$
1 = \frac{1}{\sum_{j=0}^{m} s(t - t_j)} \sum_{j=0}^{m} \Lambda_{t-s_j}^{g_j \cdot \phi}(x)^{-1} g_j(x) \quad (1)
$$

where $\phi$ is the equilibrium measure of $\Omega$.

(1) can be called a *generalized polynomial Pell’s equation* satisfied by the CFs $\Lambda_{t-s_j}^{g_j \cdot \phi}(x)^{-1}$. 
Recall that if $\mu$ is a measure on a Borel set $\Omega := X \times Y$, then it disintegrates as

$$d\mu(x, y) = \hat{\mu}(dy| x) \phi(dx)$$

with marginal $\phi$ on $X$ and conditional $\hat{\mu}(dy| x)$ on $Y$ given $x \in X$.

**Theorem (Lass (2022))**

The Christoffel function $\Lambda^\mu_d(x, y)$ disintegrates into

$$\Lambda^\mu_d(x, y) = \Lambda_d^\phi(x) \cdot \Lambda_{d, x, d}^{\nu}(y)$$

for some measure $\nu_{x, d}$ on $\mathbb{R}$. 
Recall that if $\mu$ is a measure on a Borel set $\Omega := X \times Y$, then it disintegrates as

$$d\mu(x, y) = \hat{\mu}(dy|x) \phi(dx)$$

with marginal $\phi$ on $X$ and conditional $\hat{\mu}(dy|x)$ on $Y$ given $x \in X$.

**Theorem (Lass (2022))**

The Christoffel function $\Lambda^\mu_d(x, y)$ disintegrates into

$$\Lambda^\mu_d(x, y) = \Lambda^\phi_d(x) \cdot \Lambda^{\nu_{x,d}}_d(y)$$

for some measure $\nu_{x,d}$ on $\mathbb{R}$.
Crucial in the proof is the use of the previous duality result of Nesterov.
THANK YOU!