Colloquium du CERMICS

École des Ponts
ParisTech

Moments, Positive Polynomials, and the Christoffel Function

Jean-Bernard Lasserre

LAAS – CNRS

12 décembre 2022

Moments, Positive Polynomials, and the Christoffel Function

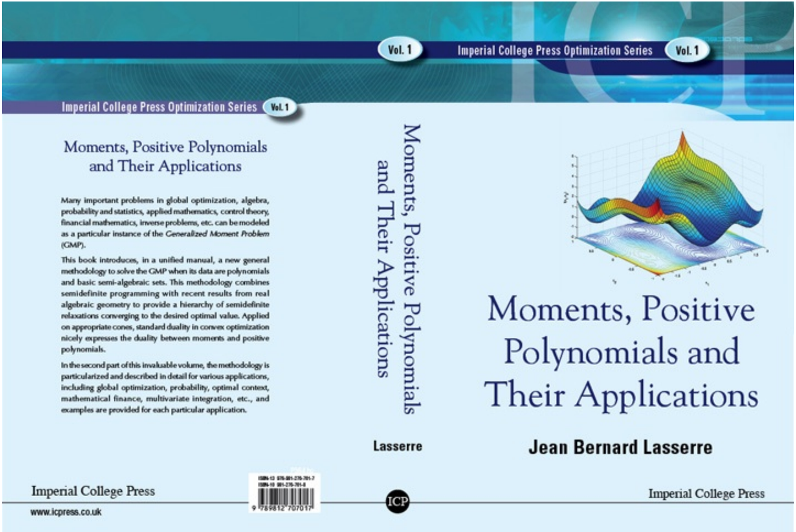
Jean B. Lasserre*

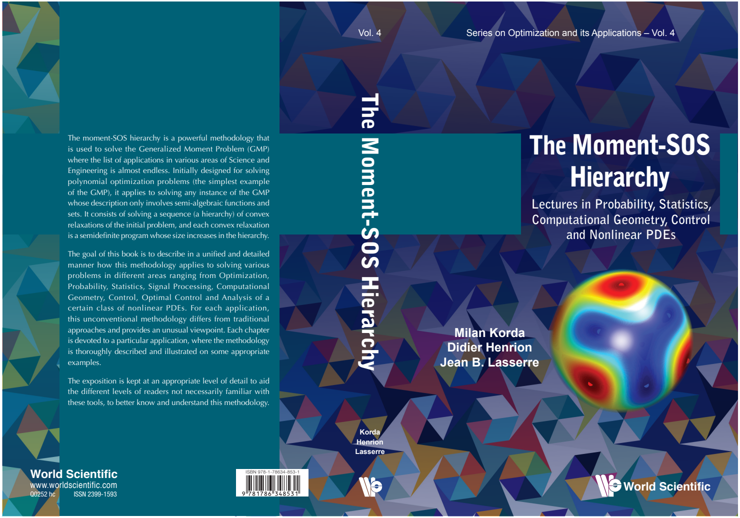
LAAS-CNRS, Institute of Mathematics & ANITI, Toulouse, France

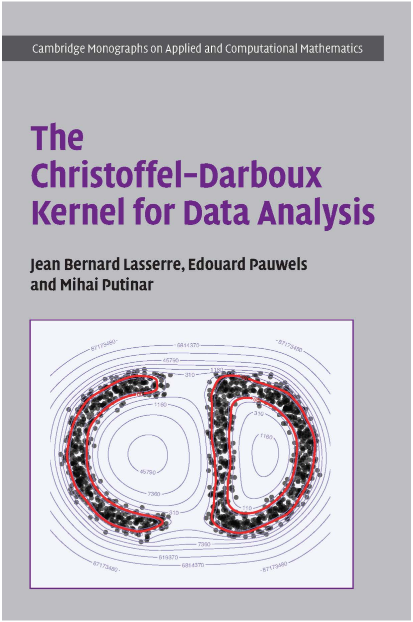
CERMICS COLLOQUIUM, December 2022

... in collaboration with

D. Henrion, M. Korda, V. Magron, S. Marx, E. Pauwels, M. Putinar, T. Weisser


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


The moment-SOS hierarchy is a powerful methodology that is used to solve the Generalized Moment Problem (GMP) where the list of applications in various areas of Science and Engineering is almost endless. Initially designed for solving polynomial optimization problems (the simplest example of the GMP), it applies to solving any instance of the GMP whose description only involves semi-algebraic functions and sets. It consists of solving a sequence (a hierarchy) of convex relaxations of the initial problem, and each convex relaxation is a semidefinite program whose size increases in the hierarchy.
The goal of this book is to describe in a unified and detailed manner how this methodology applies to solving various problems in different areas ranging from Optimization, Probability, Statistics, Signal Processing, Computational Geometry, Control, Optimal Control and Analysis of a certain class of nonlinear PDEs. For each application, this unconventional methodology differs from traditional approaches and provides an unusual viewpoint. Each chapter is devoted to a particular application, where the methodology is thoroughly described and illustrated on some appropriate examples.
The exposition is kept at an appropriate level of detail to aid the different levels of readers not necessarily familiar with these tools, to better know and understand this methodology.
World Scientific
www.worldscientific.com
00252 hc
ISSN 2399-1593
ISBN 978-1-78634-853-1
Vol. 4
The Moment-SOS Hierarchy
Korda
Henrion
Lasserre
Series on Optimization and its Applications – Vol. 4
The Moment-SOS Hierarchy
Lectures in Probability, Statistics, Computational Geometry, Control and Nonlinear PDEs
Milan Korda
Didier Henrion
Jean B. Lasserre
World Scientific


Cambridge Monographs on Applied and Computational Mathematics
The Christoffel–Darboux Kernel for Data Analysis
Jean Bernard Lasserre, Edouard Pauwels and Mihai Putinar

- 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

A brief overview of the methodology on 3 examples

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

- Optimization $\min_{\mathbf{x}} \{ f(\mathbf{x}) : \mathbf{x} \in \mathbf{K} \}$
- Optimal control

$$\begin{aligned} \min_{u} \quad & \int_0^1 h(\mathbf{x}(t), u(t))\, dt \\ & \dot{x}(t) = f(\mathbf{x}(t), u(t)), \quad t \in (0,1) \\ & \mathbf{x}(t) \in X; \quad u(t) \in U, \quad t \in [0,1] \end{aligned}$$

- Compute (or approximate)

$$\tau = \int_{\mathbf{K}} \mathbf{x}^{\alpha}\, d\lambda(\mathbf{x}), \quad \alpha \in \Gamma,$$

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

basic strategy

- ☞ (i) Search for a measure μ whose support is the solution

$$d\mu(\mathbf{x}) = \delta_{x^*}\,;\, d\mu(x,u,t) = \delta_{x(t),u(t)}(d(x,u))\,dt\,;\, d\mu(\mathbf{x}) = \lambda_{\mathbf{K}}(d\mathbf{x})$$

- ☞ (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 μ as unknown:

Constraints of the initial problem become
☞ LINEAR constraints
on the unknown moments $(\mu_\alpha)_{\alpha\in\mathbb{N}^n}$ of μ

II: Truncation

Consider only FINITELY MANY candidate moments y_α:

☞ Semidefinite constraints on the scalars y_α's state necessary conditions to qualify them as moments of some measure μ

☞ Solve the resulting finite-dimensional convex (conic) optimization problem to obtain a guaranteed lower bound.

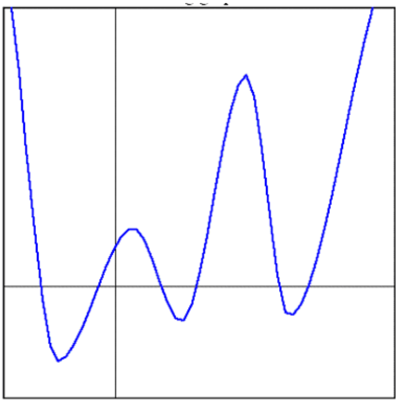
III: Iterate

. ☞ 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:

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

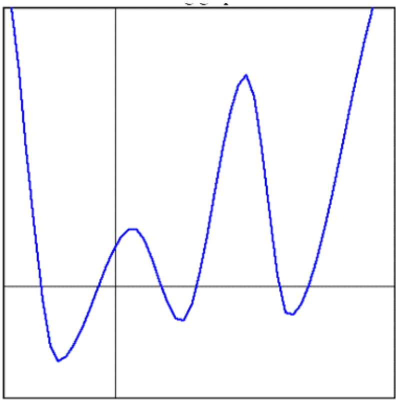
for some polynomials $f, g_j \in \mathbb{R}[\mathbf{x}]$.



Consider the polynomial optimization problem:

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

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



Why Polynomial Optimization?

After all ... **P** is just a particular case of Non Linear Programming (NLP)!

True!

... if one is interested with a LOCAL optimum only!!

☞ Many minimization algorithms do the job efficiently.

☞ The fact that f, g_j are 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!

Why Polynomial Optimization?

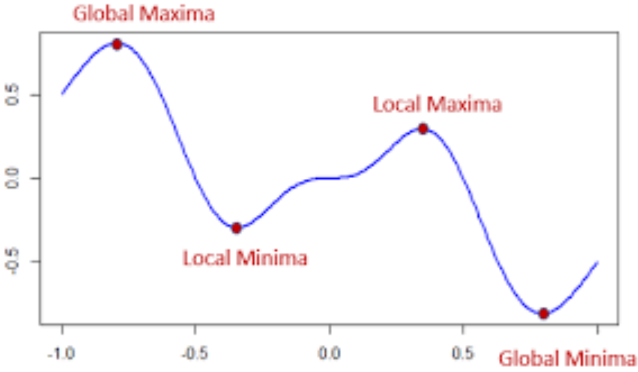
After all ... **P** is just a particular case of Non Linear Programming (NLP)!

True!

... if one is interested with a LOCAL optimum only!!

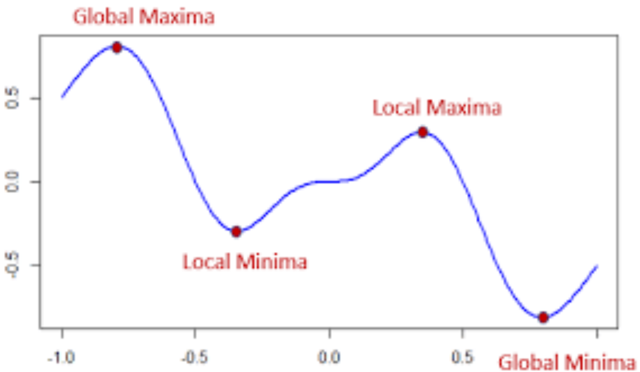
☞ Many minimization algorithms do the job efficiently.

☞ The fact that f, g_j are POLYNOMIALS does not help much!



BUT for GLOBAL Optimization

... the picture is different!



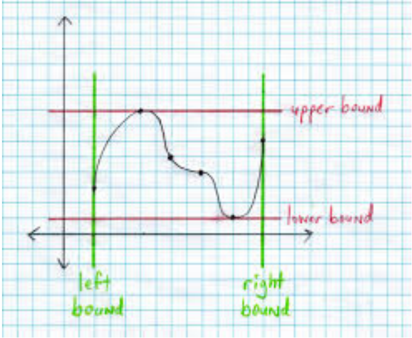
BUT for GLOBAL Optimization

... the picture is different!

Remember that for the GLOBAL minimum f^*:

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

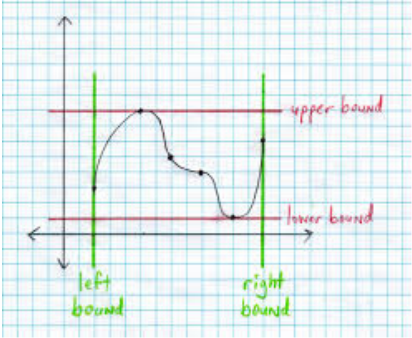
(Not true for a LOCAL minimum!)



Remember that for the GLOBAL minimum f^*:

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

(Not true for a LOCAL minimum!)



and so to compute (or approximate) f^* ...

☞ one needs to handle EFFICIENTLY the difficult constraint

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

i.e. one needs

☞ TRACTABLE CERTIFICATES of POSITIVITY on **K**
for the polynomial $\mathbf{x} \mapsto f(\mathbf{x}) - \lambda$!

and so to compute (or approximate) f^* ...

☞ one needs to handle EFFICIENTLY the difficult constraint

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

i.e. one needs

☞ TRACTABLE CERTIFICATES of POSITIVITY on $\mathbf{K}$ for the polynomial $\mathbf{x} \mapsto f(\mathbf{x}) - \lambda$!

REAL ALGEBRAIC GEOMETRY helps!!!!

Indeed, POWERFUL CERTIFICATES OF POSITIVITY EXIST!

Moreover .... and importantly,

Such certificates are amenable to PRACTICAL COMPUTATION!

REAL ALGEBRAIC GEOMETRY helps!!!!

Indeed, POWERFUL CERTIFICATES OF POSITIVITY EXIST!

Moreover .... and importantly,

Such certificates are amenable to PRACTICAL COMPUTATION!

SOS-based certificate

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

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

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

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

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)

Illustration for univariate polynomials

Let $v_d(t) = (1, t, t^2, \ldots, t^d)$ and let p be of even degree $2d$.

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

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

$$p(t) = \begin{bmatrix} 1 \\ t \\ \\ \ldots \\ t^d \end{bmatrix}^T \underbrace{\begin{bmatrix} a & b & c & \cdots \\ b & d & e & \cdots \\ c & e & f & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}}_{Q \succeq 0} \begin{bmatrix} 1 \\ t \\ t^2 \\ \ldots \\ t^d \end{bmatrix}$$

$$Q \succeq 0 \;\Rightarrow\; Q = \sum_{j=1}^{s} \mathbf{q_k}\, \mathbf{q_k}^T$$

$$\begin{aligned}
\begin{bmatrix} 1 \\ t \\ t^2 \end{bmatrix}^T Q \begin{bmatrix} 1 \\ t \\ \dots \\ t^d \end{bmatrix} &= \sum_{k=1}^{s} \left(\begin{bmatrix} 1 \\ t \\ \dots \\ t^d \end{bmatrix}^T \mathbf{q_k} \right) \left(\mathbf{q_k}^T \begin{bmatrix} 1 \\ t \\ \dots \\ t^d \end{bmatrix} \right) \\
&= \sum_{k=1}^{s} q_k(t)^2
\end{aligned}$$

Conversely if

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

then write

$$\begin{aligned} p(t) &= \sum_{k=1}^{s} \left(\begin{bmatrix} 1 \\ t \\ \dots \\ t^d \end{bmatrix}^T \mathbf{q_k} \right) \left(\mathbf{q_k}^T \begin{bmatrix} 1 \\ t \\ \dots t^d \end{bmatrix} \right) \\ &= \sum_{k=1}^{s} \begin{bmatrix} 1 \\ t \\ \dots \\ t^d \end{bmatrix}^T \left(\underbrace{\sum_{k=1}^{s} \mathbf{q_k}\,\mathbf{q_k}^T}_{Q \succeq 0} \right) \begin{bmatrix} 1 \\ t \\ \dots t^d \end{bmatrix} \end{aligned}$$

Example

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

$$f(t) = \begin{bmatrix} 1 \\ t \\ t^2 \end{bmatrix}^T \underbrace{\begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}}_{Q \succeq 0} \begin{bmatrix} 1 \\ t \\ t^2 \end{bmatrix}$$

We must have:

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

And so we must find a scalar c such that

$$Q = \begin{bmatrix} 6 & 2 & c \\ 2 & 9-2c & -2 \\ c & -2 & 6 \end{bmatrix} \succeq 0.$$

Example

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

$$f(t) = \left[\begin{array}{c} 1 \\ t \\ t^2 \end{array}\right]^T \underbrace{\left[\begin{array}{ccc} a & b & c \\ b & d & e \\ c & e & f \end{array}\right]}_{Q \succeq 0} \left[\begin{array}{c} 1 \\ t \\ t^2 \end{array}\right]$$

We must have:

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

And so we must find a scalar c such that

$$Q = \left[\begin{array}{ccc} 6 & 2 & c \\ 2 & 9 - 2c & -2 \\ c & -2 & 6 \end{array}\right] \succeq 0.$$

With $c = -4$ we have

$$Q = \left[\begin{array}{ccc} 6 & 2 & -4 \\ 2 & 17 & -2 \\ -4 & -2 & 6 \end{array}\right] \succeq 0.$$

et

$$Q = 2\left[\begin{array}{c} \sqrt(2/2) \\ 0 \\ \sqrt(2)/2 \end{array}\right]\left[\begin{array}{c} \sqrt(2/2) \\ 0 \\ \sqrt(2)/2 \end{array}\right]' + 9\left[\begin{array}{c} 2/3 \\ -1/3 \\ -2/3 \end{array}\right]\left[\begin{array}{c} 2/3 \\ 1/3 \\ -2/3 \end{array}\right]'$$

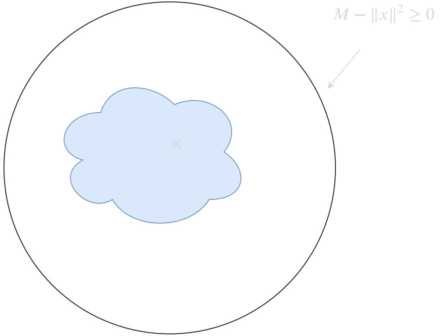
$$+18\left[\begin{array}{c} 1/\sqrt(18) \\ 4/\sqrt(18) \\ -1/\sqrt(18) \end{array}\right]\left[\begin{array}{c} 1/\sqrt(18) \\ 4/\sqrt(18) \\ -1/\sqrt(18) \end{array}\right]'$$

and so

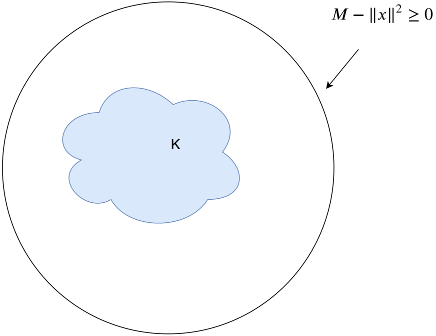
$$f(t) = (1+t^2)^2 + (2-t-2t^2)^2 + (1+4t-t^2)^2$$

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

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



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



Theorem (Putinar's Positivstellensatz)

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

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

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

However ... In Putinar's theorem

... nothing is said on the DEGREE of the SOS polynomials (σ_j)!

BUT ... GOOD news ..!!

☞ Testing whether † holds
for some SOS $(\sigma_j) \subset \mathbb{R}[\mathbf{x}]$ with a degree bound,

is SOLVING an SDP!

However ... In Putinar's theorem

... nothing is said on the DEGREE of the SOS polynomials (σ_j)!

BUT ... GOOD news ..!!

☞ Testing whether † holds
for some SOS $(\sigma_j) \subset \mathbb{R}[\mathbf{x}]$ with a degree bound,

is SOLVING an SDP!

Immediate application

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

DECLARE

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

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 σ_j,
- ☞ Semidefinite constraints coming from SOS conditions on the σ_j's

Immediate application

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

DECLARE

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

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 σ_j,
- ☞ Semidefinite constraints coming from SOS conditions on the σ_j's

Immediate application

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

DECLARE

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

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 σ_j,
- ☞ Semidefinite constraints coming from SOS conditions on the σ_j's

Immediate application

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

DECLARE

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

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 σ_j,
- ☞ Semidefinite constraints coming from SOS conditions on the σ_j's

• In fact, polynomials NONNEGATIVE ON A SET $\mathbf{K} \subset \mathbb{R}^n$ 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.

Dual side: The K-moment problem

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

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

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

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

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

Theorem (Dual side of Putinar's Theorem)

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

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

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

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

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

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

Theorem (Dual side of Putinar's Theorem)

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

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

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

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

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

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

Theorem (Dual side of Putinar's Theorem)

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

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

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

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

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

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

Theorem (Dual side of Putinar's Theorem)

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

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

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

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

Remarkably,

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

☞ Moment matrix $\mathbf{M}_1(\mathbf{y})$ in dimension 2 with $d = 1$:

$$\mathbf{M}_1(\mathbf{y}) = \begin{array}{c|ccc} & 1 & X_1 & X_2 \\ \hline 1 & y_{00} & y_{10} & y_{01} \\ X_1 & y_{10} & y_{20} & y_{11} \\ X_2 & y_{01} & y_{11} & y_{02} \end{array}$$

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 $\mathbf{M}_1(\mathbf{y})$ in dimension 2 with $d = 1$:

$$\mathbf{M}_1(\mathbf{y}) \;=\; \begin{array}{c|ccc} & 1 & X_1 & X_2 \\ \hline 1 & y_{00} & y_{10} & y_{01} \\ X_1 & y_{10} & y_{20} & y_{11} \\ X_2 & y_{01} & y_{11} & y_{02} \end{array}$$

☞ localizing matrix $\mathbf{M}_1(g\,\mathbf{y})$ in dimension 2 with $d = 1$ and $g(\mathbf{x}) = 1 - x_1^2 - x_2^2$:

$$\begin{array}{cc} & \begin{array}{ccc} \quad\quad 1 \quad\quad\quad & \quad\quad X_1 \quad\quad & \quad\quad X_2 \quad\quad \end{array} \\ \begin{array}{c} 1 \\ X_1 \\ X_2 \end{array} & \left(\begin{array}{ccc} y_{00} - y_{20} - y_{02} & y_{10} - y_{30} - y_{12} & y_{01} - y_{21} - y_{03} \\ y_{10} - y_{30} - y_{12} & y_{20} - y_{40} - y_{22} & y_{11} - y_{31} - y_{13} \\ y_{01} - y_{21} - y_{03} & y_{11} - y_{31} - y_{13} & y_{02} - y_{22} - y_{04} \end{array}\right) \end{array}$$

ALGEBRAIC SIDE

POSITIVITY ON K

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

$f > 0$ on K ?

CHARACTERIZE THOSE f

ALGEBRAIC SIDE

POSITIVITY ON K

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

$f > 0$ on K ?

CHARACTERIZE THOSE f

FUNCTIONAL ANALYSIS

THE K-MOMENT PROBLEM

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

$y_{\alpha} \overset{?}{=} \int_{K} x^{\alpha}\, d\mu \quad \forall \alpha$

for some μ

CHARACTERIZE THOSE y

DUALITY $\langle f, y \rangle = \sum_{\alpha} f_{\alpha}\, y_{\alpha}$

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

. . . modeled as

particular instances of the so called
Generalized Moment Problem, among which:
Probability, Optimal and Robust Control, Game theory, Signal processing, multivariate integration, etc.

GMP: The primal view

The GMP is the infinite-dimensional LP:

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

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

GMP: The dual view

The DUAL GMP* is the infinite-dimensional LP:

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

And one can see that ...

the constraints of *GMP** state that the functions

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

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

GMP: The dual view

The DUAL GMP* is the infinite-dimensional LP:

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

And one can see that ...

the constraints of *GMP** state that the functions

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

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

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!.

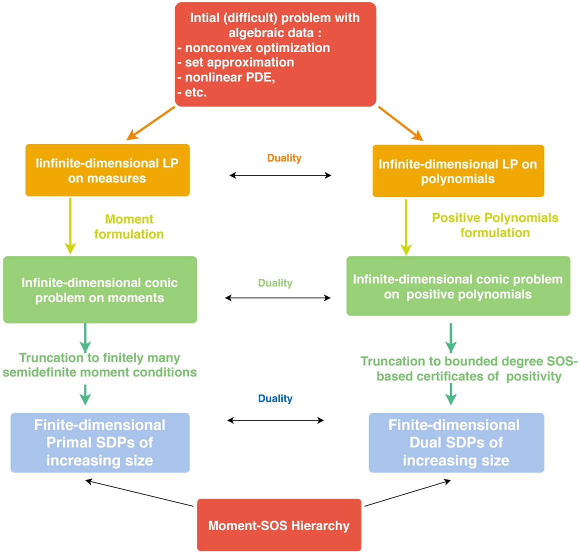
The moment-SOS hierarchy

- 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

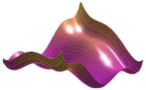
... of increasing size!.


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

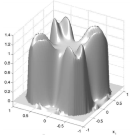
• Has already been proved successful in applications with modest problem size, notably in optimization, control, robust control, optimal control, estimation, computer vision, etc.

☞ If sparsity then problems of larger size can be addressed

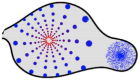
Global optimization



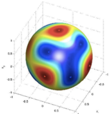
Volume of semialgebraic set



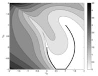
Reachable set



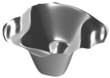
Super resolution



Optimal control

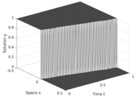


Region of attraction



Maximum invariant sets

PDE analysis & control



Example: Global optimization

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

is the SIMPLEST example of the GMP

because ...

$$f^* = \inf_{\mu \in \mathscr{M}(\mathbf{K})_+} \{ \int_{\mathbf{K}} f \, d\mu : \int_{\mathbf{K}} 1 \, d\mu = 1 \}$$

☞ A GMP with only one unknown measure μ and only one moment-constraint $\int_{\mathbf{K}} 1 \, d\mu = 1$

Example: Global optimization

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

is the SIMPLEST example of the GMP

because ...

$$f^* = \inf_{\mu \in \mathscr{M}(\mathbf{K})_+} \{ \int_{\mathbf{K}} f \, d\mu : \int_{\mathbf{K}} 1 \, d\mu = 1 \}$$

☞ A GMP with only one unknown measure μ and only one moment-constraint $\int_{\mathbf{K}} 1 \, d\mu = 1$

Example: Global optimization

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

is the SIMPLEST example of the GMP

because ...

$$f^* = \inf_{\mu \in \mathscr{M}(\mathbf{K})_+} \{ \int_{\mathbf{K}} f \, d\mu : \int_{\mathbf{K}} 1 \, d\mu = 1 \}$$

☞ A GMP with only one unknown measure μ and only one moment-constraint $\int_{\mathbf{K}} 1 \, d\mu = 1$

Remember also that for the GLOBAL minimum f^*:

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

Then for each t solve:

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

☞ $\rho_t \leq \rho_{t+1} \leq f^*$ for all t and $\rho_t \uparrow f^*$ as $t \to \infty$.

Alternatively, for each t solve:

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

$$\left.\begin{array}{rl} y_0 & = 1 \\ \mathbf{M}_t(y) & \succeq 0 \\ \mathbf{M}_{t-t_j}(g_j\, y) & \succeq 0 \quad \forall j = 1, \ldots, m \end{array}\right\} \Leftarrow y_\alpha = \int_{\mathbf{K}} x^\alpha d\mu \}$$

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:

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

$$\left.\begin{array}{r} y_0 = 1 \\ \mathbf{M}_t(y) \succeq 0 \\ \mathbf{M}_{t-t_j}(g_j\, y) \succeq 0 \quad \forall j = 1, \ldots, m \end{array}\right\} \Leftarrow y_\alpha = \int_{\mathbf{K}} x^\alpha d\mu\}$$

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^*.

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

$$\text{SDP} \quad f^* = \min_{y} \quad \{\sum_{j=1}^{4} a_j\, y_j \;:\; \begin{bmatrix} y_0 & y_1 & y_2 \\ y_1 & y_2 & y_3 \\ y_2 & y_3 & y_4 \end{bmatrix} \succeq 0$$

$$\text{SDP} \quad \begin{bmatrix} y_1 - y_2 & y_2 - y_3 \\ y_2 - y_3 & y_3 - y_4 \end{bmatrix} \succeq 0\,;\; y_0 = 1\,.\}$$

☞ $y^* = (1, x^*, (x^*)^2, (x^*)^3, (x^*)^4)$, 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

$$\text{SDP} \quad f^* = \min_{y} \quad \{ \sum_{j=1}^{4} a_j\, y_j \; : \; \begin{bmatrix} y_0 & y_1 & y_2 \\ y_1 & y_2 & y_3 \\ y_2 & y_3 & y_4 \end{bmatrix} \succeq 0$$

$$\text{SDP} \quad \begin{bmatrix} y_1 - y_2 & y_2 - y_3 \\ y_2 - y_3 & y_3 - y_4 \end{bmatrix} \succeq 0\,;\; y_0 = 1\,.\}$$

☞ $y^* = (1, x^*, (x^*)^2, (x^*)^3, (x^*)^4)$, 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.

A recovery issue

In static optimization, an optimal solution is a point $\mathbf{x}^* \in \mathbb{R}^n$.

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

☞ To recover $\mathbf{x}^*$ from its optimal solution $\mathbf{y}^* = (y^*_\alpha)_{\alpha\in\mathbb{N}^n_{2t}}$ can be done via a linear algebra subroutine.

☞ If $\mathbf{x}^*$ is unique then it is even trivial as $\mathbf{x}^*$ is just the subvector of degree-1 moments of $\mathbf{y}^*$

However,

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

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

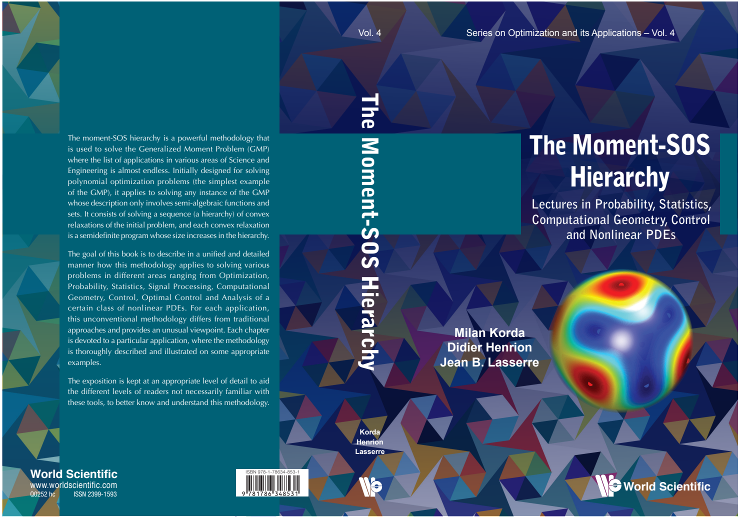
$$\mu^*_{\alpha,j} = \int_\Omega \mathbf{x}^\alpha\, y^j\, d\mu^*(\mathbf{x},y)\,, \quad \alpha \in \mathbb{N}^n\,,\, j \in \mathbb{N}\,,$$

of the measure $d\mu^*(\mathbf{x},y) = \delta_{\{f(\mathbf{x})\}}(dy)\,\phi(d\mathbf{x})$
whose support IS the graph
$\{(\mathbf{x}, f(\mathbf{x})) : \mathbf{x} \in \Omega\}$ 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 μ

- Controlled dynamics of OC
  ☞ linear constraints on moments of μ via integration of polynomial test functions
- integral cost functional
  ☞ linear criterion $\langle h, \mu \rangle$ on μ.
- state/control constraints = support constraints on μ
  ☞ semidefinite conditions on moments of μ (by Putinar theorem)

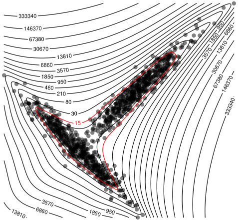

The moment-SOS hierarchy is a powerful methodology that is used to solve the Generalized Moment Problem (GMP) where the list of applications in various areas of Science and Engineering is almost endless. Initially designed for solving polynomial optimization problems (the simplest example of the GMP), 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, Statistics, Probability, Statistics, Signal Processing, Computational Geometry, Control, Optimal Control and Analysis of a certain class of nonlinear PDEs. For each application, this unconventional methodology differs from traditional approaches and provides an unusual viewpoint. Each chapter is devoted to a particular application, where the methodology is thoroughly described and illustrated on some appropriate examples.
The exposition is kept at an appropriate level of detail to aid the different levels of readers not necessarily familiar with these tools, to better know and understand this methodology.
World Scientific
www.worldscientific.com
00252 hc
ISSN 2399-1593
ISBN 978-1-78634-853-1
Vol. 4
The Moment-SOS Hierarchy
Korda
Henrion
Lasserre
Series on Optimization and its Applications – Vol. 4
The Moment-SOS Hierarchy
Lectures in Probability, Statistics, Computational Geometry, Control and Nonlinear PDEs
Milan Korda
Didier Henrion
Jean B. Lasserre
World Scientific

Then it remains to extract f from knowledge of the $(\mu_{\alpha,j})$ ...

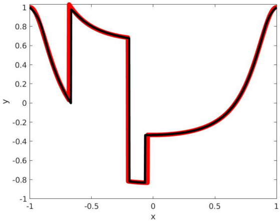
☞ This can be done by several techniques (including L^2-polynomial approximation via a standard application of the Christoffel-Darboux kernel) not detailed here.

☞ 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.

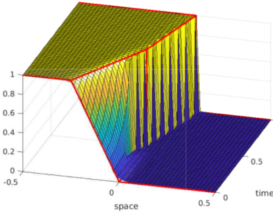
Outlier detection



Interpolation

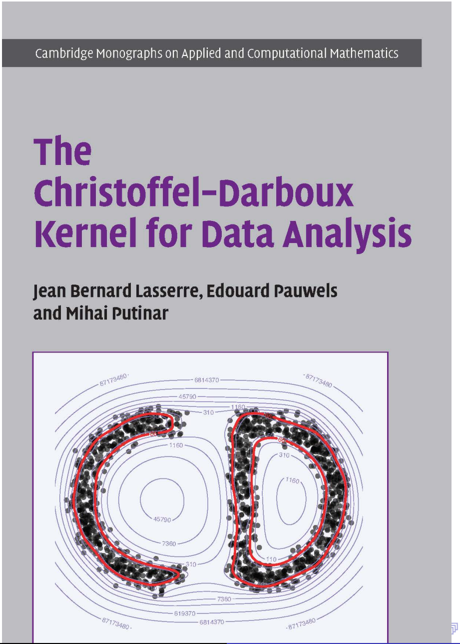


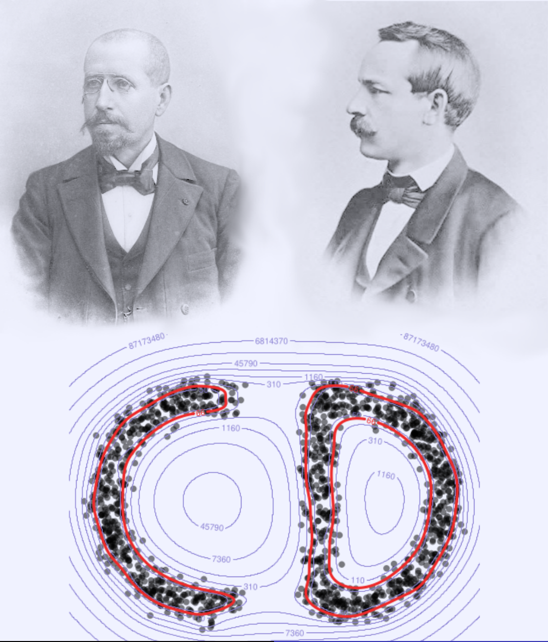
Recovery



Part two:

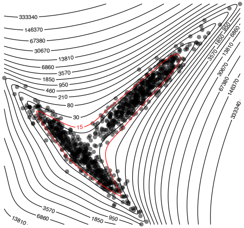
The Christoffel function


Cambridge Monographs on Applied and Computational Mathematics
The Christoffel–Darboux Kernel for Data Analysis
Jean Bernard Lasserre, Edouard Pauwels and Mihai Putinar


87173480
6814370
87173480
45790
1160
310
66
66
310
1160
1160
45790
7360
310
110
7360

Motivation

Consider the following cloud of $2D$-points (data set) below



The red curve is the level set

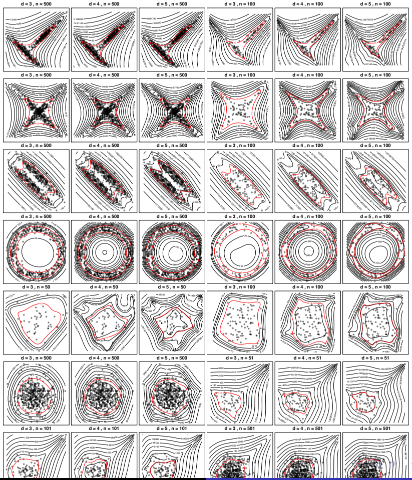
$$S_\gamma := \{ \mathbf{x} : Q_d(\mathbf{x}) \leq \gamma \}, \quad \gamma \in \mathbb{R}_+$$

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

☞ Notice that S_γ captures quite well the shape of the cloud.

Not a coincidence!

☞ Surprisingly, low degree d for Q_d is often enough to get a pretty good idea of the shape of Ω (at least in dimension $p = 2, 3$)



Cook up your own convincing example

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 $\mathbf{v}_d(\mathbf{x})^T = (1, x_1, x_2, x_1^2, x_1 x_2, \ldots, x_1 x_2^{d-1}, x_2^d)$. 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 $s(d)$

$$\mathbf{M}_d := \frac{1}{N} \sum_{i=1}^{N} \mathbf{v}_d(\mathbf{x}(i))\, \mathbf{v}_d(\mathbf{x}(i))^T,$$

where the sum is over all points $(\mathbf{x}(i))_{i=1\ldots,N} \subset \mathbb{R}^2$ of the data set.

☞ Note that $\mathbf{M}_d$ is the MOMENT-matrix $\mathbf{M}_d(\mu^N)$ of the empirical measure

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

associated with a sample of size N, drawn according to an unknown measure μ.

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

Recall that the moment matrix $\mathbf{M}_d(\mu)$ is real symmetric with rows and columns indexed by $(\mathbf{x}^\alpha)_{\alpha\in\mathbb{N}^p_d}$, and with entries

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

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

$$\mathbf{M}_1(\mu) \,:=\, \begin{array}{c} \\ 1 \\ X_1 \\ X_2 \end{array} \!\!\begin{array}{c} \begin{array}{ccc} 1 & X_1 & X_2 \end{array} \\ \left(\begin{array}{ccc} \mu_{00} & \mu_{10} & \mu_{01} \\ \mu_{10} & \mu_{20} & \mu_{11} \\ \mu_{01} & \mu_{11} & \mu_{02} \end{array}\right) \end{array}$$

is the *moment matrix of* μ *of "degree d=1"*.

- Next, form the SOS polynomial:

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

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

- Plot some level sets

$$S_\gamma := \{\,\mathbf{x} \in \mathbb{R}^2 : \; Q_d(\mathbf{x}) = \gamma\,\}$$

for some values of γ, the thick one representing the particular value $\gamma = \binom{2+d}{2}$.

The Christoffel function $\Lambda_d : \mathbb{R}^p \to \mathbb{R}_+$ is the reciprocal

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

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 $\Lambda_d : \mathbb{R}^p \to \mathbb{R}_+$ is the reciprocal

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

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, . . .

Let $\Omega \subset \mathbb{R}^p$ be the compact support of μ with nonempty interior, and $(P_\alpha)_{\alpha\in\mathbb{N}^p}$ be a family of orthonormal polynomials w.r.t. μ.

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

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

is called the *Christoffel-Darboux kernel*.

The reproducing property

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

☞ useful to determine the best degree-d polynomial approximation

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

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

$$\begin{aligned}
\mathbf{x} \mapsto \widehat{f_d}(\mathbf{x}) \;&:=\; \sum_{\alpha\in\mathbb{N}^p_d} (\overbrace{\int_{\Omega} f(y)\, P_\alpha(y)\, d\mu}^{\widehat{f_{d,\alpha}}})\, P_\alpha(\mathbf{x}) \,\in\, \mathbb{R}[\mathbf{x}]_d \\
&=\; \arg\min_{q\in\mathbb{R}[\mathbf{x}]_d} \|f - q\|_{L^2(\mu)}
\end{aligned}$$

Theorem

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:

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

☞ Alternatively

$$\Lambda_d^\mu(\xi)^{-1} \;=\; \mathbf{v}_d(\xi)^T \mathbf{M}_d(\mu)^{-1}\, \mathbf{v}_d(\xi)\,,\quad \forall \xi \in \mathbb{R}^p\,.$$

Theorem

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:

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

☞ Alternatively

$$\Lambda_d^\mu(\xi)^{-1} \;=\; \mathbf{v}_d(\xi)^T \mathbf{M}_d(\mu)^{-1}\, \mathbf{v}_d(\xi)\,, \quad \forall \xi \in \mathbb{R}^p\,.$$

☞ Importantly, and crucial for applications, the Christoffel function identifies the support Ω of the underlying measure μ.

Theorem

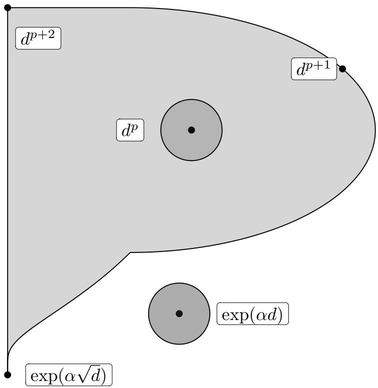
Let the support Ω of μ be compact with nonempty interior. Then:

- *For all $\mathbf{x} \in \mathrm{int}(\Omega)$: $K_d^{\mu}(\mathbf{x}, \mathbf{x}) = O(d^p)$.*
- *For all $\mathbf{x} \in \mathrm{int}(\mathbb{R}^p \setminus \Omega)$: $K_d^{\mu}(\mathbf{x}, \mathbf{x}) = \Omega(\exp(\alpha d))$ for some $\alpha > 0$.*

☞ In particular, as $d \to \infty$,

$$d^p \Lambda_d^{\mu}(\mathbf{x}) \to 0 \text{ very fast whenever } \mathbf{x} \notin \Omega.$$

Growth rates for $K_d^\mu(\mathbf{x},\mathbf{x}) = \Lambda_d^\mu(\mathbf{x})^{-1}$.



Some other properties

- Under some (restrictive) assumption on Ω and μ

$$\lim_{d\to\infty} s(d)\,\Lambda^{\mu}_{d}(\xi) \;=\; f_{\mu}(\xi)\,\omega(\xi)^{-1}$$

where ω is the density of an equilibrium measure intrinsically associated with Ω.
For instance with $p = 1$ and $\Omega = [-1, 1]$, $\omega(\xi) = \sqrt{1-\xi^2}$.

- If μ and ν have same support Ω and respective densities f_μ and f_ν w.r.t. Lebesgue measure on Ω, positive on Ω, then:

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

☞ useful for density approximation

Some other properties

- Under some (restrictive) assumption on Ω and μ

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

  where ω is the density of an equilibrium measure intrinsically associated with Ω.
  For instance with $p = 1$ and $\Omega = [-1, 1]$, $\omega(\xi) = \sqrt{1 - \xi^2}$.

- If μ and ν have same support Ω and respective densities f_μ and f_ν w.r.t. Lebesgue measure on Ω, positive on Ω, 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

☞ For instance one may decide to classify as outliers all points ξ 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.

☞ Lass. & Pauwels (2016) Sorting out typicality via the inverse moment matrix SOS polynomial, NIPS 2016.
Lass. & Pauwels (2019) The empirical Christoffel function with applications in data analysis, Adv. Comp. Math. 45, pp. 1439–1468
Lass. (2022) On the Christoffel function and classification in data analysis. Comptes Rendus Mathematique 360, pp 919–928

☞ For instance one may decide to classify as outliers all points ξ 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.

☞ Lass. & Pauwels (2016) Sorting out typicality via the inverse moment matrix SOS polynomial, NIPS 2016.
Lass. & Pauwels (2019) The empirical Christoffel function with applications in data analysis, Adv. Comp. Math. 45, pp. 1439–1468
Lass. (2022) On the Christoffel function and classification in data analysis. Comptes Rendus Mathematique 360, pp 919–928

☞ For instance one may decide to classify as outliers all points ξ 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.

☞ Lass. & Pauwels (2016) Sorting out typicality via the inverse moment matrix SOS polynomial, NIPS 2016.
Lass. & Pauwels (2019) The empirical Christoffel function with applications in data analysis, Adv. Comp. Math. 45, pp. 1439–1468
Lass. (2022) On the Christoffel function and classification in data analysis. Comptes Rendus Mathematique 360, pp 919–928

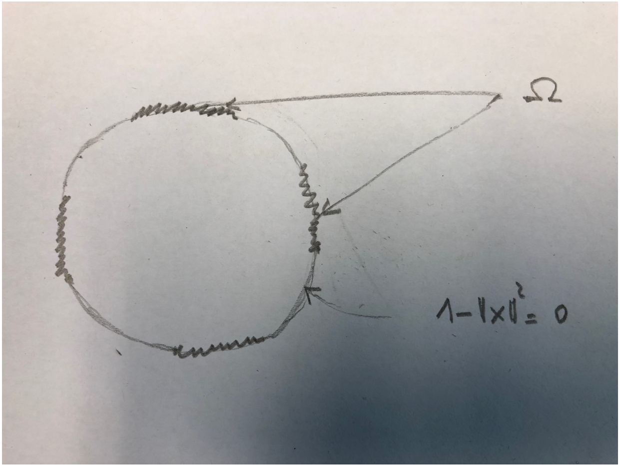
Manifold learning

A measure μ on compact set Ω is completely determined by its moments and therefore it should not be a surprise that its moment matrix $\mathbf{M}_d(\mu)$ contains a lot of information.

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

☞ When μ is degenerate and its support Ω is contained in a zero-dimensional real algebraic variety V then the kernel of $\mathbf{M}_d(\mu)$ identifies the generators of a corresponding ideal of $\mathbb{R}[\mathbf{x}]$ (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 $\mathbf{M}_d(\mu)$ contains vectors of coefficients of polynomials in the ideal generated by the quadratic polynomial $\mathbf{x} \mapsto g(\mathbf{x}) := 1 - \|\mathbf{x}\|^2$.

In fact and remarkably,

$$\mathrm{rank}\,\mathbf{M}_d(\mu) \;=\; p(d)$$

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 $\mathbb{S}^{p-1}$ of $\mathbb{R}^p$.

☞ **Pauwels E., Putinar M., Lass. J.B. (2021).** Data analysis from empirical moments and the Christoffel function, Found. Comput. Math. 21, pp. 243–273.

☞ Again this illustrates how quite sophisticated concepts of algebraic geometry are hidden and encapsulated in the moment matrix $\mathbf{M}_d(\mu)$.

☞ 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 $\mathbf{M}_d(\mu)$.
- ☞ 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 $\mathbf{M}_d(\mu)$.

☞ 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 $\mathbf{M}_d^{-1}$ 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

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

is the simple quadratic programming problem.

$$\min_{p \in \mathbb{R}^{s(d)}} \{ p^T \mathbf{M}_d p : \quad \mathbf{v}_d(\xi)^T p = 1 \},$$

which can be solved quite efficiently.

☞ However

for non modest dimension of data, matrix inversion of $\mathbf{M}_d^{-1}$ 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

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

is the simple quadratic programming problem.

$$\min_{p \in \mathbb{R}^{s(d)}} \{ p^T \, \mathbf{M}_d p \; : \quad \mathbf{v}_d(\xi)^T p = 1 \},$$

which can be solved quite efficiently.

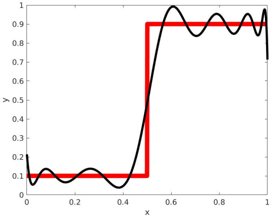
The Christoffel function for approximation

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

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

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

$$d\mu(x) \;:=\; f(\mathbf{x})\,dx$$

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

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

and the measure

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

on $\mathbb{R}^2$ with degenerate support $\Omega \subset \mathbb{R}^2$.

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

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

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

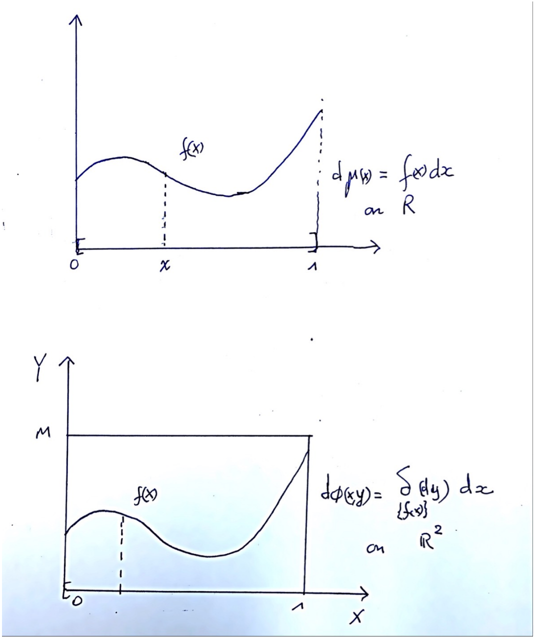
☞ Rather consider the graph $\Omega \subset \mathbb{R}^2$ of f, i.e., the set

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

and the measure

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

on $\mathbb{R}^2$ with degenerate support $\Omega \subset \mathbb{R}^2$.


f(x)
dμ(x) = f(x)dx
on R
0
x
1
Y
M
f(x)
dφ(x,y) = δ_{f(x)}(dy) dx
on ℝ²
0
1
X

Why should we do that as it implies going to $\mathbb{R}^2$ instead of staying in $\mathbb{R}$?

☞ ... because

- The support of ϕ is exactly the graph of f, and
- The CF $(x, y) \mapsto \Lambda_n^{\phi}(x, y)$ identifies the support of ϕ!

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 λ be the Lebesgue measure on $[0,1]$.

- ☞ Compute the degree-d moment matrix of ϕ:

$$\mathbf{M}_d(\phi) := \int \mathbf{v}_d(x,y) \, \mathbf{v}_d(x,y)^T \, d\phi(x,y),$$

- ☞ Compute the Christoffel function

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

- 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 λ be the Lebesgue measure on $[0,1]$.

- ☞ Compute the degree-d moment matrix of ϕ:

$$\mathbf{M}_d(\phi) := \int \mathbf{v}_d(x,y) \, \mathbf{v}_d(x,y)^T \, d\phi(x,y),$$

- ☞ Compute the Christoffel function

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

- 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 λ be the Lebesgue measure on $[0,1]$.

- ☞ Compute the degree-d moment matrix of ϕ:

$$\mathbf{M}_d(\phi) := \int \mathbf{v}_d(x,y) \, \mathbf{v}_d(x,y)^T \, d\phi(x,y),$$

- ☞ Compute the Christoffel function

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

- Approximate $f(x)$ by $\hat{f}_{d,\varepsilon}(x) := \arg\min_y \Lambda_d^{\phi,\varepsilon}(x,y)^{-1}$.
  ☞ minimize a univariate polynomial! (easy)

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

$$\mathbf{v}_d(x,y) := (1, x, y, x^2, x\,y, y^2, \ldots, x\,y^{d-1}, y^d)\,.$$

- ☞ Compute the degree-d empirical moment matrix:

$$\mathbf{M}_d(\phi) \;:=\; \sum_{i=1}^N \mathbf{v}_d((x_i, f(x_i))\,\mathbf{v}_d(x_i, f(x_i))^T\,,$$

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

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

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

  ☞ minimize a univariate polynomial! (easy)

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

$$\mathbf{v}_d(x,y) := (1, x, y, x^2, x\,y, y^2, \ldots, x\,y^{d-1}, y^d)\,.$$

- ☞ Compute the degree-d empirical moment matrix:

$$\mathbf{M}_d(\phi) \;:=\; \sum_{i=1}^N \mathbf{v}_d((x_i, f(x_i))\,\mathbf{v}_d(x_i, f(x_i))^T\,,$$

  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

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

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

  ☞ minimize a univariate polynomial! (easy)

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

$$\mathbf{v}_d(x,y) := (1, x, y, x^2, x\,y, y^2, \ldots, x\,y^{d-1}, y^d)\,.$$

- ☞ Compute the degree-d empirical moment matrix:

$$\mathbf{M}_d(\phi) \;:=\; \sum_{i=1}^{N} \mathbf{v}_d((x_i, f(x_i))\,\mathbf{v}_d(x_i, f(x_i))^T\,,$$

  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

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

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

☞ minimize a univariate polynomial! (easy)

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

$$\mathbf{v}_d(x,y) := (1,x,y,x^2,x\,y,y^2,\ldots,x\,y^{d-1},y^d)\,.$$

- ☞ Compute the degree-d empirical moment matrix:

$$\mathbf{M}_d(\phi) \;:=\; \sum_{i=1}^N \mathbf{v}_d((x_i,f(x_i))\,\mathbf{v}_d(x_i,f(x_i))^T\,,$$

  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

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

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

  ☞ minimize a univariate polynomial! (easy)

Choosing

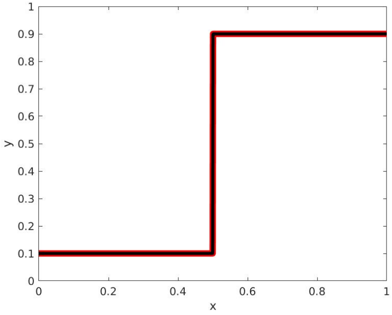
$$\varepsilon := 2^{3-\sqrt{d}}$$

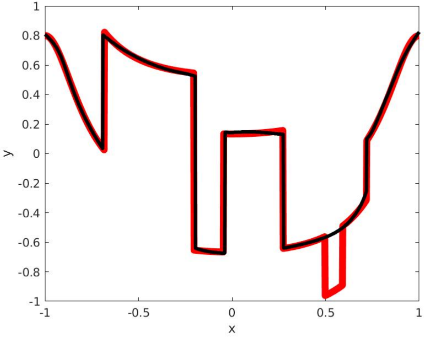
ensures convergence properties for bounded measurable functions, e.g. pointwise on open sets with no point of discontinuity.

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.

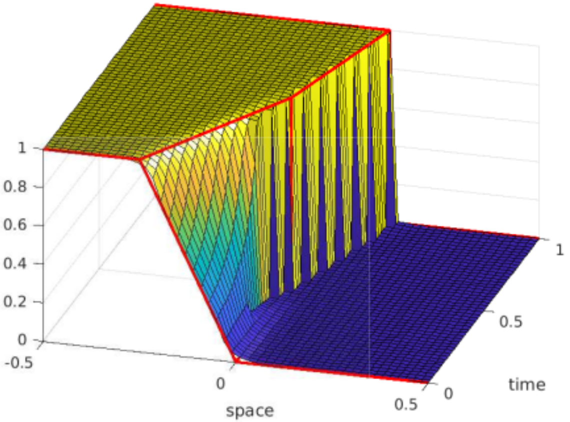
In non trivial exemples, the approximation is quite good with small values of d, and with no Gibbs phenomenon .





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!

S. Marx, E. Pauwels, T. Weisser, D. Henrion, J.B. Lass.
Semi-algebraic approximation using Christoffel-Darboux kernel,
Constructive Approximation, 2021

Christoffel function and Positive polynomials

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

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

with $g_j \in \mathbb{R}[\mathbf{x}]_{d_j}$ 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

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

is a convex cone with nonempty interior,

and with dual convex cone of pseudo-moments

$$Q_t(\Omega)^* := \{ y \in \mathbb{R}^{s(t)} \,:\, \mathbf{M}_{t-s_j}(g_j\, y) \succeq 0\,, \quad j=0,\ldots,m \},$$

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

Notice that if $\mathbf{M}_t(y)^{-1} \succ 0$ for all t,

then one may define a family of polynomials $(P_\alpha)_{\alpha\in\mathbb{N}^n} \subset \mathbb{R}[\mathbf{x}]$ 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 Λ_t^y

$$\mathbf{x} \mapsto \Lambda_t^y(\mathbf{x})^{-1} \,:=\, \sum_{|\alpha|\leq t} P_\alpha(\mathbf{x})^2\,.$$

Theorem

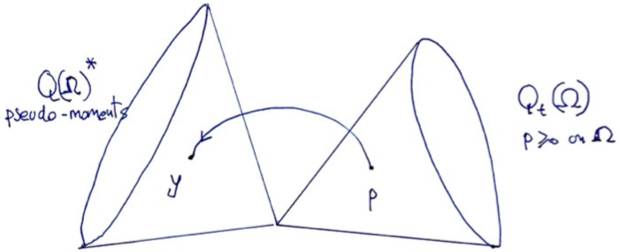
For every $p \in \mathrm{int}(Q_t(\Omega))$ *there exists a sequence of pseudo-moments* $y \in \mathrm{int}(Q_t(\Omega)^*)$ *such that*

$$
\begin{aligned}
p(\mathbf{x}) &= \sum_{j=0}^{m} \left(\mathbf{v}_{t-s_j}(\mathbf{x})^T \mathbf{M}_t(g_j\, y)^{-1}\, \mathbf{v}_{t-s_j}(\mathbf{x})\right)\, g_j(\mathbf{x}) \\
&= \sum_{j=0}^{m} \Lambda_{t-s_j}^{g_j \cdot y}(\mathbf{x})^{-1}\, g_j(\mathbf{x})
\end{aligned}
$$

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

$$
(g \cdot y)_\alpha := \sum_\gamma g_\gamma\, y_{\alpha+\gamma}\,, \quad \alpha \in \mathbb{N}^n \quad \textit{(if } g(\mathbf{x}) = \textstyle\sum_\gamma g_\gamma\, \mathbf{x}^\gamma \textit{)}.
$$

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


Q(Ω)*
pseudo-moments
y
p
Q_t(Ω)
p ≥ 0 on Ω

The proof combines

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

- ☞ the fact that

$$\mathbf{v}_{t-s_j}(\mathbf{x})^T\mathbf{M}_t(g_j\,y)^{-1}\,\mathbf{v}_{t-s_j}(\mathbf{x})\;=\;\Lambda^{g_j\cdot y}_{t-s_j}(\mathbf{x})^{-1}\,.$$

.

☞ Lass (2022) A Disintegration of the Christoffel function, Comptes Rendus Math. (2022)

The proof combines

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

- ☞ the fact that

$$\mathbf{v}_{t-s_j}(\mathbf{x})^T \mathbf{M}_t(g_j\, y)^{-1}\, \mathbf{v}_{t-s_j}(\mathbf{x}) \;=\; \Lambda^{g_j \cdot y}_{t-s_j}(\mathbf{x})^{-1}\,.$$

.

☞ Lass (2022) A Disintegration of the Christoffel function, Comptes Rendus Math. (2022)

In other words:

If $p \in \mathrm{int}(Q_t(\Omega))$ then in Putinar's certificate

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

of positivity of p on Ω,

☞ one may always choose the SOS weights σ_j in the form

$$\sigma_j(\mathbf{x}) \;:=\; \Lambda^{g_j\cdot y}_{t-s_j}(\mathbf{x})^{-1}\,, \quad j = 0, \ldots, m\,,$$

for some sequence of pseudo-moments $y \in \mathrm{int}(Q_t(\Omega)^*)$.

☞ 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 $y \in \mathbb{R}[\mathbf{x}]_{2d}^*$. That is:

$$p(\mathbf{x}) \,=\, \mathbf{v}_d(\mathbf{x})^T \mathbf{M}_d(\mathbf{y})^{-1} \mathbf{v}_x(\mathbf{x}) \,=\, \Lambda_d^y(\mathbf{x})^{-1}\,, \quad \forall \mathbf{x} \in \mathbb{R}^n\,.$$

CF – Pell's equation – equilibrium measure

☞ What is the link between $p \in \mathrm{int}(Q_t(\Omega))$ and the mysterious linear functional y?

Theorem

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

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

where ϕ is the equilibrium measure of Ω.

(1) can be called a *generalized polynomial Pell's equation* satisfied by the CFs $\Lambda_{t-s_j}^{g_j \cdot \phi}(\mathbf{x})^{-1}$.

Disintegration

Recall that if μ is a measure on a Borel set $\Omega := X \times Y$, then it disintegrates as

$$d\mu(x,y) \,=\, \underbrace{\hat{\mu}(dy\,|\,x)}_{\textit{conditional}}\;\underbrace{\phi(dx)}_{\textit{marginal}}$$

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

Theorem (Lass (2022))

The Christoffel function $\Lambda_d^{\mu}(x,y)$ disintegrates into

$$\Lambda_d^{\mu}(x,y) \,=\, \Lambda_d{}^{\phi}(x) \cdot \Lambda_d^{\nu_{x,d}}(y)$$

for some measure $\nu_{x,d}$ on $\mathbb{R}$.

Disintegration

Recall that if μ is a measure on a Borel set $\Omega := X \times Y$, then it disintegrates as

$$d\mu(x,y) \;=\; \underbrace{\hat{\mu}(dy\,|\,x)}_{\textit{conditional}}\;\underbrace{\phi(dx)}_{\textit{marginal}}$$

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

Theorem (Lass (2022))

The Christoffel function $\Lambda_d^{\mu}(x,y)$ disintegrates into

$$\Lambda_d^{\mu}(x,y) \;=\; \Lambda_d{}^{\phi}(x) \cdot \Lambda_d^{\nu_{x,d}}(y)$$

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 !