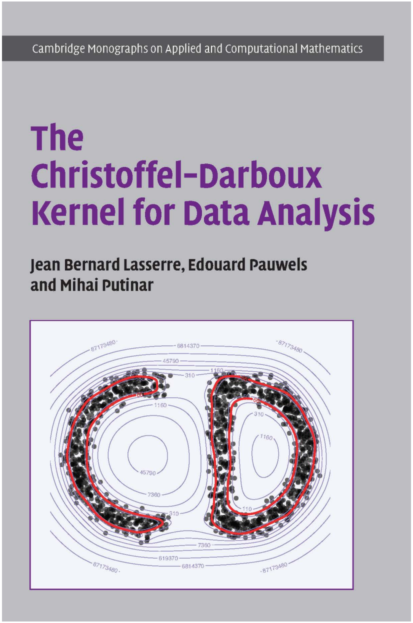
The Christoffel Function and some of its applications

Jean B. Lasserre*

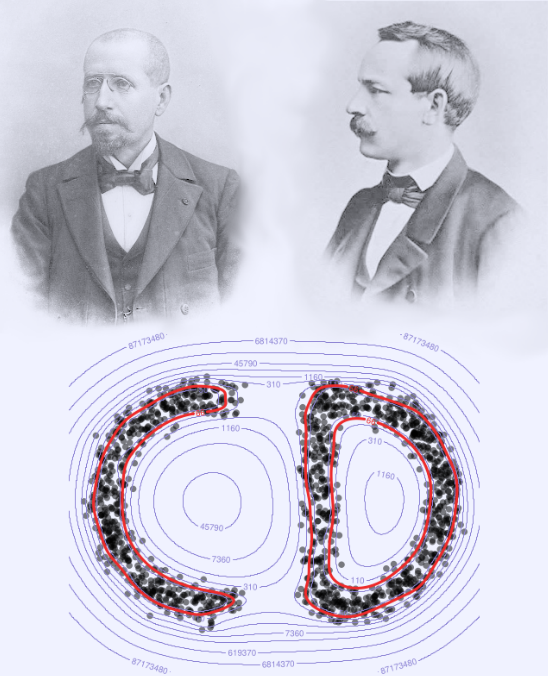
LAAS-CNRS, Institute of Mathematics & ANITI, Toulouse, France

NusCAP 2023, Lyon, May 2023

ANITI
ARTIFICIAL & NATURAL INTELLIGENCE
TOULOUSE INSTITUTE

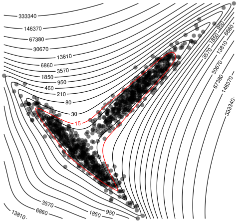

Cambridge Monographs on Applied and Computational Mathematics
The Christoffel-Darboux Kernel for Data Analysis
Jean Bernard Lasserre, Edouard Pauwels
and Mihai Putinar

- The Christoffel function
- Some applications in data analysis
- Connections

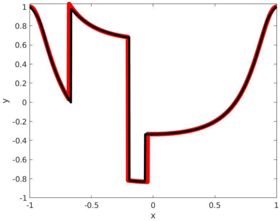

87173480
6814370
87173480
45790
310
1160
60
1160
310
1160
45790
7360
310
110
7360
619370
87173480
6814370
87173480

☞ 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.

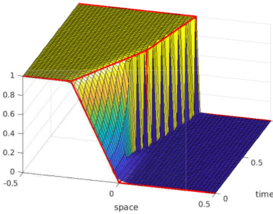
Outlier detection



Interpolation

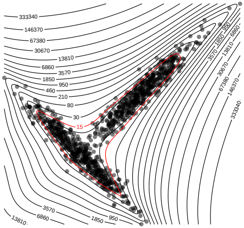


Recovery



Motivation

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



The red curve is the level set

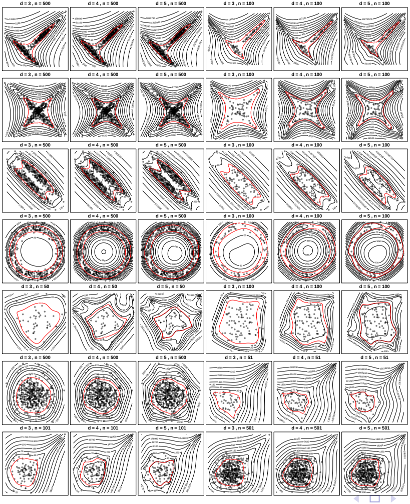
$$S_\gamma := \{ \mathbf{x} : Q_d(\mathbf{x}) \leq \gamma \}, \quad \gamma \in \mathbb{R}_+$$

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 $d = 1$:

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

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

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

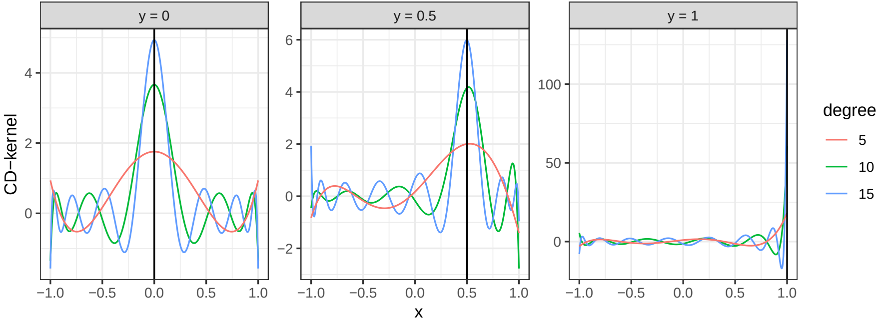
of f in $L^2(\mu)$. Indeed:

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

Just to visualize. On $[-1, 1]$, the polynomial

$$x \mapsto \mathbf{K}(y, x)\,, \quad x \in [-1, 1]\,;\; y = 0,\, 0.5\,, 1\,.$$

mimics the Dirac measure at **y** (same moments up to degree 5, 10, 15.



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

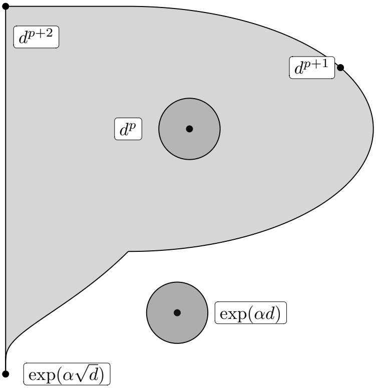
Let the support Ω of μ be compact with nonempty interior. Then:

- *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

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

  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

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

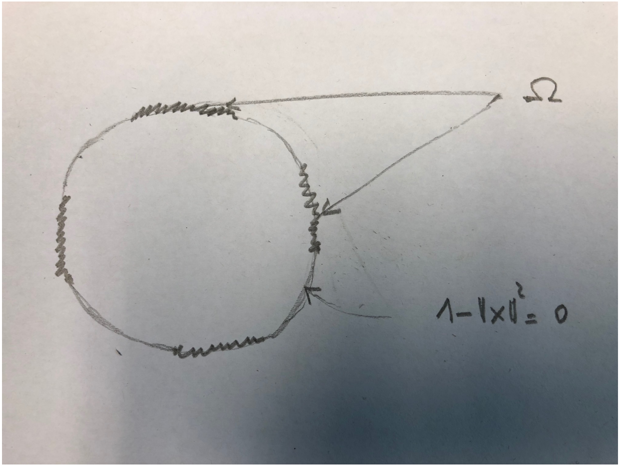
Manifold learning

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.

Intuitively, with

$$\Lambda_d^{\mu}(\mathbf{x})^{-1} \approx \mathbf{v}_d(\mathbf{x})^T\,(\mathbf{M}_d(\mu) + \varepsilon\mathbf{I})^{-1}\mathbf{v}_d(\mathbf{x})$$

$$\Lambda_d^{\mu}(\mathbf{x})^{-1} = \underbrace{\sum_{\lambda_\alpha=0} \frac{P_\alpha(\mathbf{x})^2}{\lambda_\alpha+\varepsilon}}_{\text{identifies the variety}} + \underbrace{\sum_{\lambda_\alpha>0} \frac{P_\alpha(\mathbf{x})^2}{\lambda_\alpha+\varepsilon}}_{\text{where on the variety}}.$$

☞ 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.

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

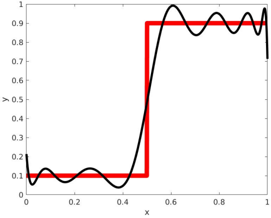
which can be solved quite efficiently.

A standard 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: step function

☞ 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 via a non-standard use of 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 via a non-standard use of 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 $[0,1] \in \mathbb{R}$,

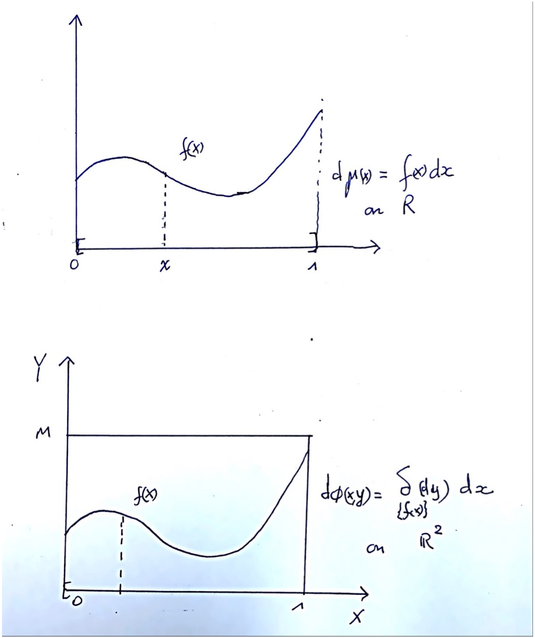
☞ Rather consider the graph $\Omega \subset \mathbb{R}^2$ of f, i.e., the set

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

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

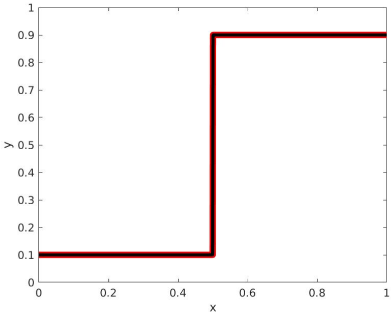
$$\varepsilon := 2^{3-\sqrt{d}}$$

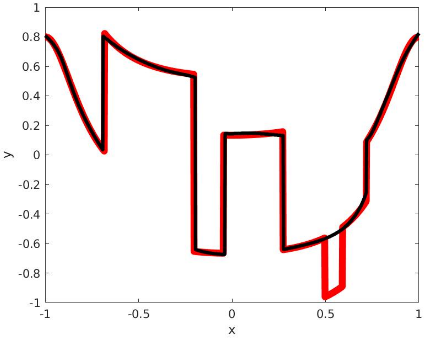
ensures convergence properties for bounded measurable functions, e.g. pointwise on open sets with no point of discontinuity.

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 .





Application in dynamical systems

☞ When solving Optimal Control problems (OCP) or some Nonlinear Partial Differential Equations (PDEs) via the Moment-SOS hierarchy, one ends up with moments up to some degree $2d$, of a measure μ supported on the trajectories

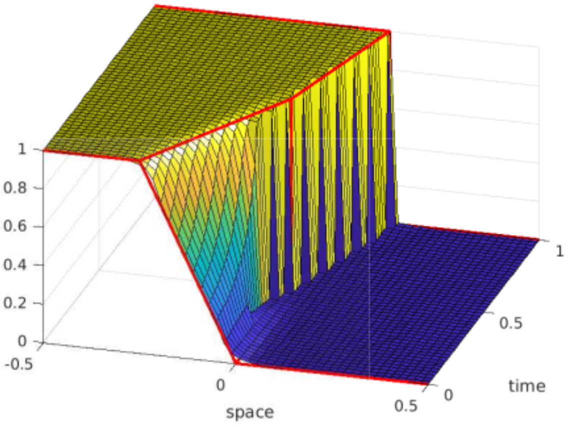
$$\begin{array}{rcl} t & \mapsto & x_i(t),\ u_j(t) \quad i=1,\ldots,n;\ j=1,\ldots m \quad (OCP) \\ (t,\mathbf{x}) & \mapsto & y(\mathbf{x},t) \quad (PDEs) \end{array}$$

So it remains to recover such functions from the sole knowledge of moments of μ, as $\mathbf{M}_d(\mu)$ is available!.

☞ CD kernel associated with μ !

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

Connections

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

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

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

☞ These two convex cones are at the core of the moment-SOS hierarchy in Polynomial Optimization to solve

$$f^* = \min_{\mathbf{x}} \{ f(\mathbf{x}) : \mathbf{x} \in \Omega \}.$$

☞ One instead solves the hierarchy of semidefinite programs

$$\rho_t = \sup_{\lambda, \sigma_j} \{ \lambda : f - \lambda \in Q_t(\Omega) \}, \quad t \in \mathbb{R},$$

and $\rho_t \uparrow f^*$ as t increases.

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

☞ These two convex cones are at the core of the moment-SOS hierarchy in Polynomial Optimization to solve

$$f^* \;=\; \min_{\mathbf{x}} \{\, f(\mathbf{x}) :\; \mathbf{x} \in \Omega \,\}\,.$$

☞ One instead solves the hierarchy of semidefinite programs

$$\rho_t \;=\; \sup_{\lambda,\sigma_j} \{\, \lambda :\; f-\lambda \in\; Q_t(\Omega)\,\}\,, \quad t \in \mathbb{R}\,,$$

and $\rho_t \uparrow f^*$ as t increases.

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

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

is well-defined.

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*

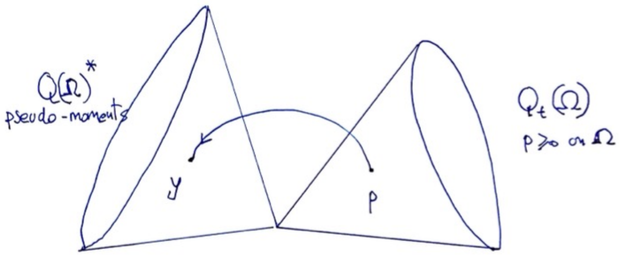
$$p(\mathbf{x}) = \sum_{j=0}^{m} \left(\mathbf{v}_{t-s_j}(\mathbf{x})^T \mathbf{M}_t(g_j\, y)^{-1}\, \mathbf{v}_{t-s_j}(\mathbf{x})\right)\, g_j(\mathbf{x})$$

$$= \sum_{j=0}^{m} \Lambda_{t-s_j}^{g_j \cdot y}(\mathbf{x})^{-1}\, g_j(\mathbf{x})$$

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


$Q(\Omega)^*$
pseudo-moments
y
p
$Q_t(\Omega)$
$p \geq 0$ 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. (2023)

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:

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

☞ 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 $\Lambda_{t-s_j}^{g_j \cdot \phi}(\mathbf{x})^{-1}$.

A prototype example

Let $\Omega = [-1, 1]$, $x \mapsto g(x) := 1 - x^2$, and let

- $(T_n)_{n\in\mathbb{N}}$, be the Chebyshev polynomials of first kind, orthogonal w.r.t. $\mu = dx/\sqrt{1-x^2}$
- $(U_n)_{n\in\mathbb{N}}$) be the Chebyshev polynomials of second kind, orthogonal w.r.t. $g \cdot \mu := \sqrt{1-x^2}\,dx$.

Pell's polynomial equation reads:

$$1 \;=\; T_n(x)^2 + (1-x^2)\,U_{n-1}(x)^2\,, \quad \forall n \in \mathbb{N}\,,\ \forall x \in \mathbb{R}\,.$$

☞ nothing less than Markov-Lukács decomposition of the constant polynomial "1" nonnegative on $[-1, 1]$!

A prototype example

Let $\Omega = [-1, 1]$, $x \mapsto g(x) := 1 - x^2$, and let

- $(T_n)_{n\in\mathbb{N}}$, be the Chebyshev polynomials of first kind, orthogonal w.r.t. $\mu = dx/\sqrt{1-x^2}$
- $(U_n)_{n\in\mathbb{N}})$ be the Chebyshev polynomials of second kind, orthogonal w.r.t. $g \cdot \mu := \sqrt{1-x^2}\,dx$.

Pell's polynomial equation reads:

$$1 \;=\; T_n(x)^2 + (1-x^2)\,U_{n-1}(x)^2\,,\quad \forall n\in\mathbb{N}\,,\ \forall x\in\mathbb{R}\,.$$

☞ nothing less than Markov-Lukács decomposition of the constant polynomial “1" nonnegative on $[-1, 1]$!

Then summing up yields

$$\begin{aligned} 2t+1 &= \underbrace{\sum_{n=0}^{t} \widehat{T}_n(x)^2}_{\Lambda_t^{\mu}(x)^{-1}} + (1-x^2) \underbrace{\sum_{n=0}^{t-1} \widehat{U}_{n-1}(x)^2}_{\Lambda_{t-1}^{g\cdot\mu}(x)^{-1}}, \quad \forall x\,,\ \forall n \\ &= \sigma_0(x) + (1-x^2)\,\sigma_1(x) \end{aligned}$$

☞ So for the interval $[-1,1]$ and $p=1$, one obtains that μ is the equilibrium measure $\frac{dx}{\sqrt{1-x^2}}$ of the interval $[-1,1]$!

☞ We have been able to extend this result to the unit box, the Euclidean unit ball, and the simplex of $\mathbb{R}^d$, but only for $t = 1, 2, 3$. We conjecture that it is also true for all $t \in \mathbb{N}$.

☞ Lass (2022) Pell's equation, sum-of-squares and equlibrium measure on a compact set, Comptes Rendus Math. (2023)

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 !