Introduction to Christoffel-Darboux kernels for polynomial optimization

Edouard Pauwels

POEMA Online Workshop

June 2020

Acknowledgement

Currently working on a overview / application in data science project with Lasserre and Putinar.

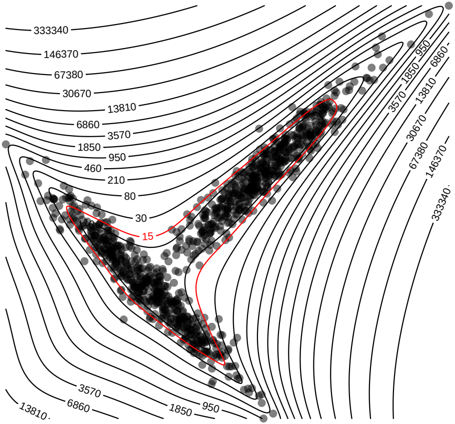
Motivation

Mathematical context:

- Christoffel Darboux (CD) kernels are as old as orthogonal polynomials ($\sim$ 19-th century).
- Fine properties of these objects have important consequences in approximation theory (convergence of generalized Fourier series).
- There is still a lot of activity going on.
- Recently used in polynomial optimization contexts.

Polynomial optimization?

- A CD kernel depdends on a (probability) measure μ on a Euclidean space $\mathbb{R}^p$
- It captures information on μ (support, density).
- It is easily computed from moments.
- Moments (or pseudomoments) of measures are typical outputs of Lassere's Hierarchy.



Here μ is an empirical average $\mu = \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}$. Applications in statistics also.

Plan for today:

Introduction of these objects and first properties.

Outline

1. CD kernel, Christoffel function, orthogonal polynomials, moments

2. CD kernel captures measure theoretic properties: univariate case

μ: Borel probability measure in $\mathbb{R}^p$ (compact support, absolutely continuous).
$\mathbb{R}_d[X]$: p-variate polynomials of degree at most d (of dimension $s(d) = \binom{p+d}{d}$).

$$(P, Q) \qquad \mapsto \qquad \langle\langle P, Q \rangle\rangle_\mu := \int PQd\mu,$$

defines a valid scalar product on $\mathbb{R}_d[X]$.

$(\mathbb{R}_d[X], \langle\langle \cdot, \cdot \rangle\rangle_\mu)$ is a *finite dimensional, Hilbert space* of functions from $\mathbb{R}^p$ to $\mathbb{R}$.

Remark: discussions and more general conditions in exercises.

Reproducing Kernel Hilbert Space (RKHS)

Hilbert space method (first half of 20-th century): Zarembda, Mercer, Moore, Szegö, Bergman, Bochner, Kolmogorov, Aronszajn ...

Reproducing property: For all $d \in \mathbb{N}$, there exists $K_d^\mu : \mathbb{R}^p \times \mathbb{R}^p \mapsto \mathbb{R}$, symmetric such that for all $\mathbf{z} \in \mathbb{R}^p$,

$$K_d^\mu(\mathbf{z}, \cdot) \in \mathbb{R}_d[X].$$

K_d^μ satifies the reproducing property, for all $P \in \mathbb{R}_d[X]$ and $\mathbf{z} \in \mathbb{R}^p$,

$$P(\mathbf{z}) \quad = \quad \langle\!\langle P(\cdot), K_d^\mu(\mathbf{z}, \cdot) \rangle\!\rangle_\mu \quad = \quad \int P(\mathbf{x}) K_d^\mu(\mathbf{z}, \mathbf{x}) d\mu(\mathbf{x})$$

$\mathcal{H} = (\mathbb{R}_d[X], \langle\!\langle \cdot, \cdot \rangle\!\rangle_\mu)$ is called a Reproducing Kernel Hilbert Space (RKHS).
Generalize to any Hilbert space of functions with continuous pointwise evaluation.

Christoffel-Darboux kernel: K_μ^d is *the* reproducing kernel of $\mathcal{H}$.

Computation from moments

μ: Borel probability measure in $\mathbb{R}^p$ (compact support, absolutely continuous).
$\mathbb{R}_d[X]$: p-variate polynomials of degree at most d (of dimension $s(d) = \binom{p+d}{d}$).

- Let $\{P_i\}_{i=1}^{s(d)}$ be any basis of $\mathbb{R}_d[X]$,
- $\mathbf{v}_d : \mathbf{x} \mapsto (P_1(\mathbf{x}), \ldots, P_{s(d)}(\mathbf{x}))^T$.
- $M_{\mu,d} = \int \mathbf{v}_d \mathbf{v}_d^T d\mu \in \mathbb{R}^{s(d)\times s(d)}$ (integral pointwise).

Then $M_{\mu,d}$ is invertible and for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$,

$$K_d^{\mu}(\mathbf{x}, \mathbf{y}) = \mathbf{v}_d(\mathbf{x})^T M_{\mu,d}^{-1} \mathbf{v}_d(\mathbf{y}).$$

Remark: If $\mathbf{v}_d$ is the monomial basis, then we recover the usual moment matrix (Tutorials by Mihai and Didier).
Remark: It does not depend on the choice of the basis.

μ: Borel probability measure in $\mathbb{R}^p$ (compact support, absolutely continuous).
$\mathbb{R}_d[X]$: p-variate polynomials of degree at most d (of dimension $s(d) = \binom{p+d}{d}$).

Relation with orthogonal polynomials

Let $\{P_i\}_{i=1}^{s(d)}$ be any orthonormal basis of $\mathbb{R}_d[X]$ (*w.r.t.* $\langle\langle \cdot, \cdot \rangle\rangle_\mu$), then for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$,

$$K_d^\mu(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{s(d)} P_i(\mathbf{x}) P_i(\mathbf{y}).$$

Remark: monomial basis, Gram-Schmitt provides a canonical way to construct such a basis. This is at the hear of the (rich) theory of orthogonal polynomials (see exercises).

Tip: Working in an orthonormal basis is often much more stable numerically. Inverting the moment matrix of the uniform measure on $[-1, 1]$ fails for $d = 23$.

Christoffel function

μ: Borel probability measure in $\mathbb{R}^p$ (compact support, absolutely continuous).
$\mathbb{R}_d[X]$: p-variate polynomials of degree at most d (of dimension $s(d) = \binom{p+d}{d}$).

Christoffel function

$$\begin{aligned} \Lambda_d^\mu : \mathbb{R}^p &\mapsto [0,1] \\ \mathbf{z} &\mapsto \min_{P \in \mathbb{R}_d[X]} \left\{ \int P^2 d\mu : \quad P(\mathbf{z}) = 1 \right\}. \end{aligned}$$

CD kernel and Christoffel function:

$$\Lambda_d^\mu(\mathbf{z}) = \frac{1}{K_d^\mu(\mathbf{z}, \mathbf{z})},$$

for all $\mathbf{z} \in \mathbb{R}^p$. The optimal solution in definition of Λ_d^μ is

$$P(\cdot) = \frac{K_d^\mu(\cdot, \mathbf{z})}{K_d^\mu(\mathbf{z}, \mathbf{z})}.$$

Affine invariance

Let $\mathcal{A}\colon \mathbb{R}^p \mapsto \mathbb{R}^p$ be an invertible affine map.

Push forward: $\mathcal{A}_*\mu$ such that $\mathcal{A}_*\mu = \mu(\mathcal{A}^{-1}(B))$, for all Borel sets B. Then for all measurable f

$$\int f(\mathbf{z}) d\mathcal{A}_*\mu(\mathbf{z}) = \int f(\mathcal{A}(\mathbf{x})) d\mu(\mathbf{x}).$$

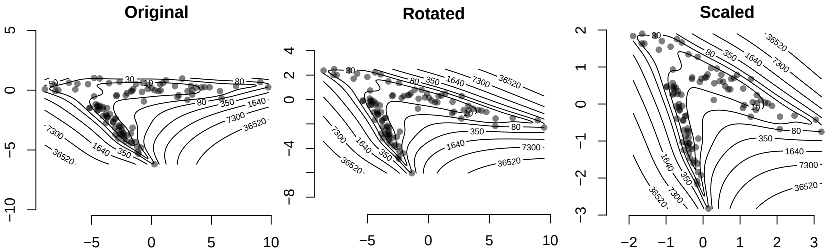
Invariance:

For all $\mathbf{x} \in \mathbb{R}^p$

$$\Lambda_d^{\mathcal{A}*\mu}(\mathcal{A}(\mathbf{x})) = \Lambda_d^{\mu}(\mathbf{x}).$$

Affine invariance

With an image



Here the push forward is simply the empirical average supported on images of the point cloud by the affine map.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}, \qquad \mathcal{A}_*\mu = \frac{1}{n}\sum_{i=1}^{n} \delta_{\mathcal{A}(x_i)}.$$

Pointwise evaluation and Bernstein-Markov property

Pointwise evaluation:
For all $P \in \mathbb{R}_d[X]$ and $\mathbf{z} \in \mathbb{R}^p$, $P(\cdot)/P(\mathbf{z})$ evaluates to 1 at $\mathbf{z}$.

$$P(\mathbf{z})^2 \leq K_\mu^d(\mathbf{z},\mathbf{z}) \int P^2 d\mu.$$

Bernstein-Markov property for μ with compact support S:

For all $P \in \mathbb{R}_d[X]$,

$$\sup_{\mathbf{z}\in S} |P(\mathbf{z})| \quad \leq \quad C(d)\|P\|_\mu$$

where $C(d)^{\frac{1}{d}} \to 1$ as $d \to \infty$.

$$\sup_{\mathbf{z}\in S} |P(\mathbf{z})| \quad \leq \sqrt{\sup_{\mathbf{z}\in S} K_d^\mu(\mathbf{z},\mathbf{z})}\|P\|_\mu$$

Historical remarks

Univariate case (complex and real) since beginning of 20-th century:

- quadrature, interpolation, approximation
- orthogonal polynomials
- potential theory
- random matrices/polynomials
- ...

A few contributors

- Szegö, Erdös, Turan, Freud, Totik, Máté, Nevai, ...

Still an object of very active research (asymptotics, multivariate case).

Outline

1. CD kernel, Christoffel function, orthogonal polynomials, moments

2. CD kernel captures measure theoretic properties: univariate case

Exercise: Let μ be a compactly supported probability measure on $\mathbb{R}^p$ and define Λ_d^μ, with its variational form. Show that

$$\lim_{d \to \infty} \Lambda_d^\mu(x_0) = \mu(\{x_0\}),$$

for all x_0 in $\mathbb{R}^p$.

Typical value

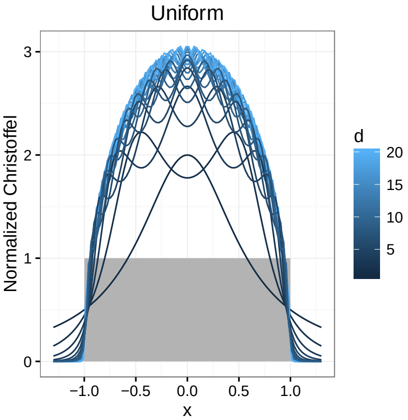
Exercise: Let μ be a compactly supported absolutely continuous probability measure on $\mathbb{R}^p$. Let Z be a random variable with distribution μ, show that

$$\mathbb{E}_{Z\sim\mu}\left[(\Lambda_d^{\mu}(Z))^{-1}\right] = \binom{d+p}{p} \sim d^p.$$

Asymptotics for the Christoffel function: sublinear on the support

Maté, Nevai and Totik, (1991): $p = 1$ and $d\mu = f$ on $[-1, 1]$ and 0 elsewhere, $f > 0$ continuous. For almost all x in $[-1, 1]$

$$\lim_{d \to \infty} \Lambda_{\mu,d}(x) d = \pi f(x) \sqrt{1 - x^2}$$



Asymptotics for the Christoffel function: linear outside the support

Stahl and Totik, (1992): $p = 1$ and $d\mu = f$ on $[-1, 1]$ and 0 elsewhere, $f > 0$, for all $x \notin [-1, 1]$,

$$\lim_{d \to \infty} \Lambda_{\mu,d}(x)^{\frac{1}{2d}} < 1$$