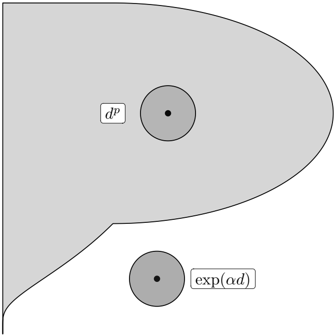
Introduction to Christoffel-Darboux kernels for polynomial optimization

Edouard Pauwels

POEMA Online Workshop

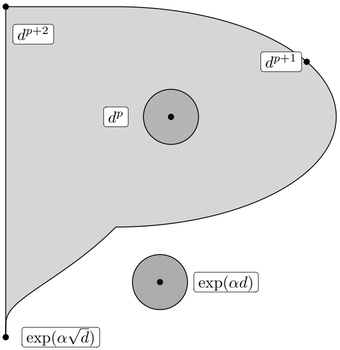
July 2020

μ: Lebesgue restricted to $S \subset \mathbb{R}^p$, compact, non-empty interior.



Exponential separation of the support

μ: Lebesgue restricted to $S \subset \mathbb{R}^p$, compact, non-empty interior.



Thresholding scheme: $C > 0$, $q > p$

$$\{\mathbf{x},\ \mathbf{v}_d(\mathbf{x})^T M_{\mu,d}^{-1} \mathbf{v}_d(\mathbf{x}) \leq C d^q\} \qquad \text{“} \underset{d\to\infty}{\to} \text{”} \qquad \mathrm{cl}(\mathrm{int}(S)).$$

Extends to positive densities on S. $=$

Outline

1. CD kernel, Christoffel function, orthogonal polynomials, moments
2. CD kernel captures measure theoretic properties: univariate case
3. Quantitative asymptotics
4. The singular case
5. Using approximate moments
6. An application to polynomial optimal control

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

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

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

Specificity of the singular case

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

Moment based computation

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

Then, for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$, $K_d^{\mu}(\mathbf{x}, \mathbf{y}) = \mathbf{v}_d(\mathbf{x})^T M_{\mu,d}^{-1} \mathbf{v}_d(\mathbf{y})$ ~~$\mathbf{v}_d(\mathbf{x})^T M_{\mu,d}^{-1} \mathbf{v}_d(\mathbf{y})$~~

Let $P(\mathbf{x}) = \sum_{i=1}^{s(d)} \mathbf{p}_i P_i(\mathbf{x}) P \in \mathbb{R}_d[X]$. We have

$$\int P^2 d\mu = \mathbf{p}^T M_{\mu,d} \mathbf{p}.$$

If P vanishes on S, if and only if $\mathbf{p} \in \ker(M_{\mu,d})$.
Singular moment matrix, morally, CD kernel should be $+\infty$.

Christoffel function to the rescue

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

Variational formulation: for all $\mathbf{z} \in \mathbb{R}^p$

$$\frac{1}{K_d^\mu(\mathbf{z},\mathbf{z})} = \Lambda_d^\mu(\mathbf{z}) = \min_{P \in \mathbb{R}_d[X]} \left\{ \int P^2 d\mu : \quad P(\mathbf{z}) = 1 \right\}.$$

$$\Lambda_d^\mu(\mathbf{z}) = \min_{P \in \mathbb{R}_d[X]} \left\{ \int P^2 d\mu : \quad P(\mathbf{z}) = 1 \right\}.$$

Given $\mathbf{z} \in \mathbb{R}^p$, such that there exists $P \in \mathbb{R}_d[X]$ such that

- $P(\mathbf{z}) \neq 0$
- P vanishes on S.

Then $\Lambda_d^\mu(\mathbf{z}) = 0$.

Getting the CD kernel back (and computation from moments)

μ: Borel probability measure in $\mathbb{R}^p$,compact support: S.
$\mathbb{R}_d[X]$: p-variate polynomials of degree at most d (of dimension $s(d) = \binom{p+d}{d}$).
V denotes the Zariski closure of S (smallest algebraic set containing S).

For d large enough, $V = \{\mathbf{z} \in \mathbb{R}^p, \Lambda_d^\mu(\mathbf{z}) > 0\}$.

Polynomials on V: $L^2_{\mu,d} = \mathbb{R}_d[X] \;/\; \{P \in \mathbb{R}_d[X],\, P \text{ vanishes on } V\}$.

RKHS: $(L^2_{\mu,d}, \langle\!\langle \cdot, \cdot \rangle\!\rangle_\mu)$ is a Hilbert space of functions on V. K_d^μ is its reproducing kernel (defined on V).
For any $\mathbf{x} \in V$ and $P \in L^2_{\mu,d}$, $P(\mathbf{x}) = \int P(\mathbf{y}) K_d^\mu(\mathbf{x}, \mathbf{y}) d\mu(\mathbf{y})$.

Relation with Christoffel function: $\Lambda_d^\mu(\mathbf{z}) K_d^\mu(\mathbf{z}, \mathbf{z}) = 1$, for $\mathbf{z} \in V$.

Pseudo inverse computation: let $\mathbf{v}_d$ be any basis of $\mathbb{R}_d[X]$, $M_{\mu,d}$ moment matrix:

$$\forall \mathbf{x}, \mathbf{y} \in V \qquad K_d^\mu(\mathbf{x}, \mathbf{y}) = \mathbf{v}_d(\mathbf{x}) M_{\mu,d}^\dagger \mathbf{v}_d(\mathbf{y}).$$

Average value and Hilbert function: $\int K_d^\mu(\mathbf{x}, \mathbf{x}) d\mu(\mathbf{x}) = \dim(L^2_{\mu,d}) \leq s(d)$.

Outline

1. CD kernel, Christoffel function, orthogonal polynomials, moments
2. CD kernel captures measure theoretic properties: univariate case
3. Quantitative asymptotics
4. The singular case
5. Using approximate moments
6. An application to polynomial optimal control

Motivation for approximate moments

"I am a Lasserre hierarchist, I work with pseudo-moments."

"I am a statistician, I work with empirical moments."

"I am a numerician, among others, I care about sensitivity to errors."

Choose a basis $\mathbf{v}_d$ of $\mathbb{R}_d[X]$.

Approximation of Christoffel function: Let $Q(\mathbf{x}, \mathbf{y}) = \mathbf{v}_d(\mathbf{x}) M^{-1} \mathbf{v}_d(\mathbf{y})$ where $M \in \mathbb{R}^{s(d) \times s(d)}$ is positive definite, then for all $\mathbf{x} \in \mathbb{R}^p$,

$$|Q(\mathbf{x}, \mathbf{x}) \Lambda_d^\mu(\mathbf{x}) - 1| \leq \|I - M_{\mu,d}^{\frac{1}{2}} M^{-1} M_{\mu,d}^{\frac{1}{2}}\|_{op}$$

If $M \simeq M_{\mu,d}$, then $\Lambda_d^\mu(\mathbf{x}) \simeq \frac{1}{Q(\mathbf{x},\mathbf{x})}$.

Regularization

"Using pseudo inverse is like saying $0 = +\infty$".

Regularization: Let μ_0 be a simple absolutely continuous measure (moments are easy to compute). Replace μ by $\mu + \beta\mu_0$, $\beta > 0$.

$$M_{\mu+\beta\mu_0,d} = M_{\mu,d} + \beta M_{\mu_0,d} \succ 0$$

$$\Lambda^d_{\mu+\beta\mu_0} \geq \Lambda^d_{\mu} + \beta\Lambda^d_{\mu_0}$$

$$\int (\Lambda^d_{\mu+\beta\mu_0})^{-1} d\mu \leq \int (\Lambda^d_{\mu+\beta\mu_0})^{-1} d(\mu + \beta\mu_0) = s(d) = O(d^p)$$

- The moment matrix is positive definite
- If $\Lambda^d_{\mu+\beta\mu_0}$ is small, then Λ^d_{μ} is also small.
- $\Lambda^d_{\mu+\beta\mu_0}$ stays reasonably big on the support of μ.
- $\Lambda^d_{\mu+\beta\mu_0}$ stays reasonably small outside the support of μ (if β is small).

Outline

The content of this section is taken from

Marx, S., Pauwels, E., Weisser, T., Henrion, D., & Lasserre, J. (2019). Tractable semi-algebraic approximation using Christoffel-Darboux kernel. arXiv preprint arXiv:1904.01833.

From the tutorial of Didier

Controled ODE,

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t)), \\ x(t) &\in X, \\ u(t) &\in U, \\ t &\in [0, 1], \\ x(0) &= 0 \end{aligned}$$

Occupation measure, given a classical trajectory

$$d\mu(x, u, t) = d\delta_{x(t)}(x) d\delta_{u(t)}(u) dt$$

Relaxation: Replace classical trajectories satisfying an ODE by measures satisfying a linear transport PDE.

A heuristic argument

Hierarchy: f polynomial, X, U basic semi-algebraic: level d provides pseudo-moments up to degree $2d$ in variables t, u, x.

$$PM_d$$

Heuristic: As d grows PM_d should get close to $M_{\mu,d}$ where μ is an occupation measure supported on optimal trajectories.

Use the Christoffel Darboux kernel:

$$\text{“}(x, u, t)^T PM_d^{-1}(x, u, t)\text{”}$$

- The measure is singular, we only have pseudo moments ...
- Morally, it is small on the support of μ and large outside the support.
- Morally, it is small on the optimal trajectory and large outside.

Hierarchy: f polynomial, X, U basic semi-algebraic: level d provides pseudo-moments up to degree $2d$ in variables t, u, x.

$$PM_d$$

Christoffel Darboux kernel:

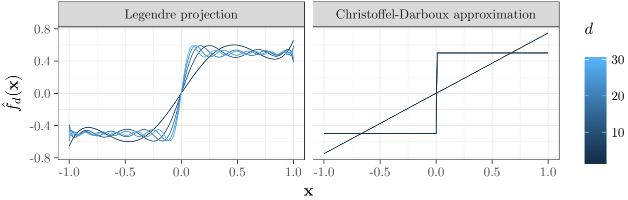
$$\text{“}(x,u,t)^T PM_d^{-1}(x,u,t)\text{”} = Q_d(x,u,t)$$

Morally, it is small on the optimal trajectory and large outside.

A semi-algebraic estimator: For all $t \in [0,1]$

$$(\hat{u}(t), \hat{x}(t)) \in \mathrm{argmin}_{(x,u)} Q(x,u,t).$$

An example with $x(t) = \mathrm{sign}(t)/2$ and exact moments



Convergence guaranties

A semi-algebraic estimator: $Q_d(x, u, t) =$ "$(x, u, t)^T P M_d^{-1} (x, u, t)$"

$$(\hat{u}, \hat{x})\colon t \mapsto (\hat{u}(t), \hat{x}(t)) \in \mathrm{argmin}_{(x,u)} Q(x, u, t).$$

Assumption: x, u in L^1, bounded, continuous almost everywhere, exact moments.
Strong convergence in L^1.

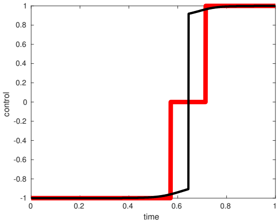
Assumption: x, u Lipschitz, exact moments.
Rate of order $O(1/\sqrt{d})$.

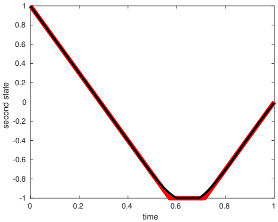
Assumption: x, u have bounded total variation, exact moments.
Conjecture: Rate of order $O(1/d^{\frac{1}{4}})$.

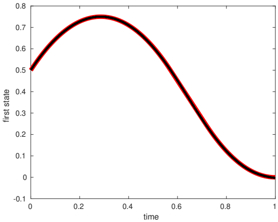
Illustration on the double integrator with constraints

Minimal time to reach the origin. $u \in [-1, 1]$, $x_1 \geq -1$.

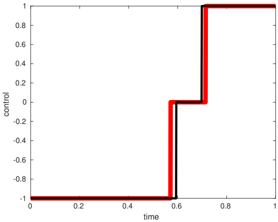
$$\dot{x}_2(t) = x_1(t)$$
$$\dot{x}_1(t) = u(t)$$



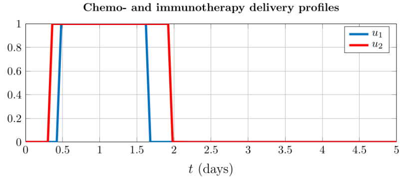


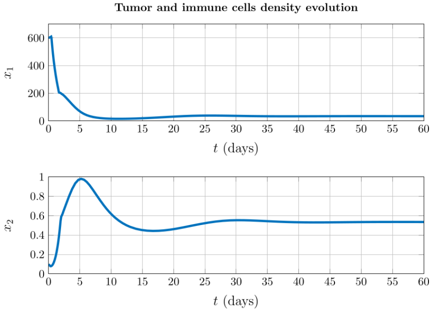


With True moments:

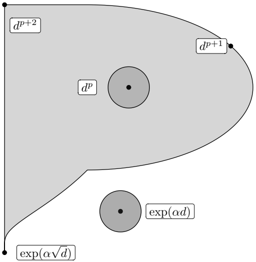


Moussa, K., Fiacchini, M., & Alamir, M. (2019). Robust Optimal Control-based Design of Combined Chemo-and Immunotherapy Delivery Profiles. IFAC-PapersOnLine, 52(26), 76-81.





Conclusion



- CD kernel is computed from moments of a measure μ.
- It captures the support of μ.
- Century old mathematical history and still active.
- Proper set up, proof guaranties, require some subtleties.
- Can be combined with Lassere's Hierarchy: example in polynomial optimal control.