Numerical Algebra, Control and Optimization
Vol. 16, 2026, pp. 1-18
doi:10.3934/naco.2024011

THE CHRISTOFFEL FUNCTION: APPLICATIONS, CONNECTIONS AND EXTENSIONS

Jean B. Lasserre✉
LAAS-CNRS and Toulouse School of Economics (TSE), France

Abstract. We provide an introduction to the Christoffel function (CF), a well-known (and old) tool from the theory of approximation and orthogonal polynomials. We then describe how it provides a simple and easy-to-use tool to address some problems in data analysis and mining, and in approximation and interpolation of discontinuous functions. Finally we also reveal some surprising links of the CF with seemingly different disciplines, including polynomial optimization, positivity certificates, and equilibrium measures of compact sets.

1. **Introduction.** This paper provides a brief introduction to the Christoffel function (CF), a well-known tool from the theory of approximation and orthogonal polynomials. One reason to reconsider this "old" tool is that surprisingly, while well-known for a long time, only recently some of its properties have been shown to be particularly interesting and practical to address some problems in data analysis and mining, notably for support inference, outlier detection, and density approximation, as described in e.g. [16, 21, 22, 23, 31]. Indeed, quoting [23]:

> *Among the many positive definite kernels appearing in classical analysis, approximation theory, probability, mathematical physics, control theory and more recently in machine learning, Christoffel-Darboux kernel (CD kernel in short) stands aside by its numerical accessibility from raw data and its versatility in encoding/decoding fine properties of the generating measure. ...*
>
> *On the practical side, the CD kernel can be regarded as a bridge between the underlying Reproducing Kernel Hilbert Space (RKHS) carried by a finite dimensional function space and the unknown measure. This basic feature allows to import in the data analysis context an existing, rich theory describing the relation between the Christoffel-Darboux kernel and the underlying measure.*

In addition in [10, 25] a non-standard application of the Christoffel-Darboux kernel has also been proved quite efficient in some approximation and interpolation of functions (notably discontinuous).

2020 *Mathematics Subject Classification.* Primary: 42C05 65D05 65D15; Secondary: 90C23 90C22.

Key words and phrases. Theory of Approximation, orthogonal polynomials, Reproducing Kernel Hilbert Space, polynomial optimization, certificates positivity.

The author is supported by the ANR-NuSCAP 20-CE-48-0014 and ANR-19-PI3A-0004 program of the (french) national agency ANRS.

This paper is handled by Jiawang Nie as guest editor.

Another reason is that again recently, we have revealed some (in the author's opinion surprising) connections of the CF with seemingly different fields, including positivity certificates in real algebraic geometry and polynomial optimization on one side, and (a generalized) Pell's polynomial equation and equilibrium measure in pluripotential theory on the other side [18, 24]. At last, we also report on a simple modification (or regularization) of the CF due to the author [19], with a new and simple asymptotic property in which a factor containing the unknown equilibrium measure has disappeared. It thus provides a numerical means to approximate the unknown underlying density of the measure associated with the CF.

By its very nature this article is a condensed review of some previous results of the literature (most including the author as a contributor) and therefore sometimes results are simply re-stated from the article where they are published (with a reference), and their proof is omitted.

The paper is organized as follows: After introducing notation and definitions in Section 2, we briefly introduce the CF and some of its important asymptotic properties in Section 3. Then in Section 4 we briefly describe some applications of the CF in data analysis and approximation and interpolation of discontinuous functions. In Section 5 we introduce some connections of the CF with positive polynomials and with what we call a generalized polynomial Pell's equation. Finally, in Section 6 we describe an extension of the CF via a certain regularization.

2. Notation, definitions and some preliminaries.

2.1. **Notation, definitions.** Let $\mathbb{R}[\mathbf{x}]$ denote the ring of polynomials in the variables $\mathbf{x}=(x_1,\ldots,x_d)$ and let $\mathbb{R}[\mathbf{x}]_n$ be the vector space of polynomials of degree at most n (whose dimension is $s(n):=\binom{d+n}{d}$). For every $n\in\mathbb{N}$, let $\mathbb{N}^d_n:=\{\boldsymbol{\alpha}\in\mathbb{N}^d: |\boldsymbol{\alpha}|\,(=\sum_i\alpha_i)\le n\}$, and let $\mathbf{v}_n(\mathbf{x})=(\mathbf{x}^{\boldsymbol{\alpha}})$, $\boldsymbol{\alpha}\in\mathbb{N}^d$, be the vector of monomials of the canonical basis $(\mathbf{x}^{\boldsymbol{\alpha}})$ of $\mathbb{R}[\mathbf{x}]_n$. Given a closed set $\mathcal{X}\subseteq\mathbb{R}^n$, let $\mathcal{P}(\mathcal{X})\subset\mathbb{R}[\mathbf{x}]$ (resp. $\mathcal{P}_n(\mathcal{X})\subset\mathbb{R}[\mathbf{x}]_n$) be the convex cone of polynomials (resp. polynomials of degree at most n) that are nonnegative on $\mathcal{X}$. A polynomial $f\in\mathbb{R}[\mathbf{x}]_n$ is written

$$\mathbf{x}\mapsto f(\mathbf{x}) \;=\; \sum_{\alpha\in\mathbb{N}^d} f_{\boldsymbol{\alpha}}\,\mathbf{x}^{\boldsymbol{\alpha}} \;=\; \langle\mathbf{f},\mathbf{v}_n(\mathbf{x})\rangle\,,$$

with vector of coefficients $\mathbf{f}=(f_{\boldsymbol{\alpha}})\in\mathbb{R}^{s(n)}$ in the canonical basis of monomials $(\mathbf{x}^{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^d}$. For real symmetric matrices, let $\langle\mathbf{B},\mathbf{C}\rangle:=\text{trace}\,(\mathbf{B}\,\mathbf{C})$ while the notation $\mathbf{B}\succeq 0$ stands for $\mathbf{B}$ is positive semidefinite (psd) whereas $\mathbf{B}\succ 0$ stands for $\mathbf{B}$ is positive definite (pd). Denote by $\mathcal{S}^n$ the space of real $n\times n$ symmetric matrices and $\mathcal{S}^n_+$ its subset of psd matrices.

With a closed set $S\subset\mathbb{R}^d$, denote by $\mathscr{M}(S)$ the space of finite signed Borel measures on S, and $\mathscr{M}(S)_+\subset\mathscr{M}(S)$ (resp. $\mathscr{P}(S)$) the convex cone of finite positive Borel measures (resp. probability measures) on S. The support $\text{supp}\,(\mu)$ of a Borel measure μ on $\mathbb{R}^d$ is the smallest closed set $\Omega\subset\mathbb{R}^d$ such that $\mu(\mathbb{R}^d\setminus\Omega)=0$.

Riesz linear functional. Given a sequence $\boldsymbol{\phi}=(\phi_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^d}$ (in bold), its associated Riesz linear functional is the linear mapping $\phi:\mathbb{R}[\mathbf{x}]\to\mathbb{R}$ (not in bold) defined by:

$$f\,(=\sum_{\boldsymbol{\alpha}} f_{\boldsymbol{\alpha}}\,\mathbf{x}^{\alpha}) \quad\mapsto \phi(f) \;=\; \sum_{\boldsymbol{\alpha}\in\mathbb{N}^d} f_{\boldsymbol{\alpha}}\,\phi_{\boldsymbol{\alpha}} \;=\; \langle\mathbf{f},\boldsymbol{\phi}\rangle\,, \tag{1}$$

A sequence $\boldsymbol{\phi}$ has a *representing* measure if its associated Riesz linear functional ϕ is a (positive) Borel measure on $\mathbb{R}^d$, in which case,

$$\phi_{\boldsymbol{\alpha}} \,=\, \phi(\mathbf{x}^{\boldsymbol{\alpha}}) \,=\, \int_{\mathbb{R}^d} \mathbf{x}^{\boldsymbol{\alpha}}\, d\phi\,, \quad \forall \boldsymbol{\alpha} \in \mathbb{N}^d\,.$$

Given a sequence $\boldsymbol{\phi} = (\phi_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^d}$ and a polynomial $g \in \mathbb{R}[\mathbf{x}]$, $\mathbf{x} \mapsto g(\mathbf{x}) = \sum_{\boldsymbol{\gamma}} g_{\boldsymbol{\gamma}}\, \mathbf{x}^{\boldsymbol{\gamma}}$, define the new sequence $g \cdot \boldsymbol{\phi}$ such that

$$(g \cdot \boldsymbol{\phi})_{\boldsymbol{\alpha}} \,:=\, \phi(\mathbf{x}^{\boldsymbol{\alpha}}\, g) \,=\, \sum_{\boldsymbol{\gamma}\in\mathbb{N}^d} g_{\boldsymbol{\gamma}}\, \phi_{\boldsymbol{\alpha}+\boldsymbol{\gamma}}\,, \quad \forall \boldsymbol{\alpha} \in \mathbb{N}^d\,,$$

and therefore its associated Riesz linear functional, denoted by $g \cdot \phi$, satisfies

$$g \cdot \phi(f) \,=\, \phi(g\, f)\,, \quad \forall f \in \mathbb{R}[\mathbf{x}]\,.$$

In particular, if $\boldsymbol{\phi}$ has a representing measure ϕ and g is nonnegative, then the Riesz linear functional $g \cdot \phi$ is a representing measure, i.e.,

$$g \cdot \phi(f) \,=\, \phi(g\, f) \,=\, \int_{\mathbb{R}^d} f\, g\, d\phi\,, \quad \forall f \in \mathbb{R}[\mathbf{x}]\,.$$

Moment matrix. The (degree-n) *moment* matrix associated with a sequence $\boldsymbol{\phi} = (\phi_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^d}$ (or, equivalently, with the Riesz linear functional ϕ), is the real symmetric matrix denoted $\mathbf{M}_n(\boldsymbol{\phi})$ (or $\mathbf{M}_n(\phi)$) with rows and columns indexed by $\mathbb{N}^d_n$, and whose entry $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is just $\phi_{\boldsymbol{\alpha}+\boldsymbol{\beta}}$, for every $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}^d_n$. So $\mathbf{M}_n(\boldsymbol{\phi})$ depends only on moments $\phi_{\boldsymbol{\alpha}}$ of degree at most $2n$. Alternatively, if one introduces the real symmetric matrices $(\mathbf{B}^1_{\boldsymbol{\alpha}}) \subset \mathcal{S}^{s(n)}$ defined by

$$\mathbf{v}_n(\mathbf{x})\, \mathbf{v}_n(\mathbf{x})^T \,=\, \sum_{\boldsymbol{\alpha}\in\mathbb{N}^d_{2n}} \mathbf{B}^1_{\boldsymbol{\alpha}}\, \mathbf{x}^{\boldsymbol{\alpha}}, \qquad \forall \mathbf{x} \in \mathbb{R}^d\,, \tag{2}$$

then $\mathbf{M}_n(\boldsymbol{\phi}) = \sum_{\boldsymbol{\alpha}\in\mathbb{N}^d_{2n}} \phi_{\boldsymbol{\alpha}}\, \mathbf{B}^1_{\boldsymbol{\alpha}}$. Moreover, if $\boldsymbol{\phi}$ has a representing measure ϕ then $\mathbf{M}_n(\boldsymbol{\phi}) \succeq 0$ because $\langle \mathbf{f}, \mathbf{M}_n(\boldsymbol{\phi})\, \mathbf{f}\rangle = \int f^2 d\phi \geq 0$, for all $f \in \mathbb{R}[\mathbf{x}]_n$.

A measure whose all moments are finite, is *moment determinate* if there is no other measure with same moments.

Localizing matrix. With $\boldsymbol{\phi}$ as above and $g \in \mathbb{R}[\mathbf{x}]$ (with $g(\mathbf{x}) = \sum_\gamma g_\gamma \mathbf{x}^\gamma$), the localizing matrix associated with $\boldsymbol{\phi}$ and g is the moment matrix $\mathbf{M}_n(g\cdot\boldsymbol{\phi})$ associated with the sequence $g \cdot \boldsymbol{\phi}$. That is, $\mathbf{M}_n(g \cdot \boldsymbol{\phi})$ is the real symmetric matrix with rows and columns indexed by $\mathbb{N}^d_n$, and whose entry $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is just $(g \cdot \phi)_{\boldsymbol{\alpha}+\boldsymbol{\beta}}$, that is, $\mathbf{M}_n(g \cdot \boldsymbol{\phi})(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_\gamma g_{\boldsymbol{\gamma}} \phi_{\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\gamma}}$, for every $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}^d_n$.

Alternatively, letting $d_g := \lceil \deg(g)/2 \rceil$, and introducing the real symmetric matrices $\mathbf{B}^g_{\boldsymbol{\alpha}} \in \mathcal{S}^{s(n)}$, $\boldsymbol{\alpha} \in \mathbb{N}^d$, defined by

$$g(\mathbf{x})\, \mathbf{v}_n(\mathbf{x})\, \mathbf{v}_n(\mathbf{x})^T \,=\, \sum_{\boldsymbol{\alpha}\in\mathbb{N}^d_{2(n+d_g)}} \mathbf{B}^g_{\boldsymbol{\alpha}}\, \mathbf{x}^{\boldsymbol{\alpha}}, \qquad \forall \mathbf{x} \in \mathbb{R}^d\,, \tag{3}$$

one obtains $\mathbf{M}_n(g \cdot \boldsymbol{\phi}) = \sum_{\boldsymbol{\alpha}\in\mathbb{N}^d_{2(n+d_g)}} \phi_{\boldsymbol{\alpha}}\, \mathbf{B}^g_{\boldsymbol{\alpha}}$.

If $\boldsymbol{\phi}$ has a representing measure ϕ whose support is contained in the set $\{\mathbf{x} : g(\mathbf{x}) \geq 0\}$, then $\mathbf{M}_n(g \cdot \boldsymbol{\phi}) \succeq 0$ for all n, because

$$\langle \mathbf{f}, \mathbf{M}_n(g \cdot \boldsymbol{\phi})\, \mathbf{f}\rangle \,=\, g \cdot \phi(f^2) \,=\, \phi(f^2 g) \,=\, \int f^2\, g\, d\phi \,\geq\, 0\,, \quad \forall f \,\in \mathbb{R}[\mathbf{x}]_n\,.$$

2.2. **SOS polynomials and quadratic modules.** A polynomial $f \in \mathbb{R}[\mathbf{x}]$ is a Sum-of-Squares (SOS) if there exist $s \in \mathbb{N}$, and $f_1, \ldots, f_s \in \mathbb{R}[\mathbf{x}]$, such that $f(\mathbf{x}) = \sum_{k=1}^{s} f_k(\mathbf{x})^2$, for all $\mathbf{x} \in \mathbb{R}^d$. Denote by $\Sigma[\mathbf{x}]$ (resp. $\Sigma[\mathbf{x}]_n$) the set of SOS polynomials (resp. SOS polynomials of degree at most $2n$). Of course every SOS polynomial is nonnegative. However the converse is not true.

Membership in $\Sigma[\mathbf{x}]_n$. Checking whether a given polynomial f is nonnegative on $\mathbb{R}^d$ is difficult whereas, and this is crucial for the Moment-SOS hierarchy [9], checking whether f is SOS is much easier and can be done efficiently. Indeed let $f \in \mathbb{R}[\mathbf{x}]_{2n}$ (for f to be SOS its degree must be even), $\mathbf{x} \mapsto f(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}^d_{2n}} f_{\boldsymbol{\alpha}}\, \mathbf{x}^{\alpha}$. Then $f \in \mathbb{R}[\mathbf{x}]_{2n}$ is SOS if and only if there exists a real symmetric matrix $\mathbf{X}^T = \mathbf{X}$ of size $s(n) = \binom{d+n}{d}$, such that:

$$\mathbf{X} \succeq 0; \quad f_{\boldsymbol{\alpha}} \;=\; \langle \mathbf{X}, \mathbf{B}^1_{\boldsymbol{\alpha}} \rangle, \qquad \forall \boldsymbol{\alpha} \in \mathbb{N}^d_{2n}, \tag{4}$$

where the matrices $\mathbf{B}^1_{\boldsymbol{\alpha}}$ have been introduced in (2). It turns out that (4) defines the feasible set of what is called a *semidefinite program*[1] (in short, SDP).

Quadratic module. Introduce the constant polynomial $\mathbf{x} \mapsto g_0(\mathbf{x}) := 1$ for all $\mathbf{x} \in \mathbb{R}^d$ (also denoted $g_0 = \mathbf{1}$). With a family $(g_1, \ldots, g_m) \subset \mathbb{R}[\mathbf{x}]$ is associated the *quadratic module* $Q(g)\,(= Q(g_1, \ldots, g_m)) \subset \mathbb{R}[\mathbf{x}]$ defined by:

$$Q(g) \;:=\; \left\{ \sum_{j=0}^{m} \sigma_j\, g_j \,:\, \sigma_j \in \Sigma[\mathbf{x}],\; j = 0, \ldots, m \right\}, \tag{5}$$

and its degree-$2n$ *truncated* version

$$Q_n(g) \;:=\; \left\{ \sum_{j=0}^{m} \sigma_j\, g_j \,:\, \sigma_j \in \Sigma[\mathbf{x}]_{n-d_j},\; j = 0, \ldots, m \right\}, \tag{6}$$

where $d_j := \lceil \deg(g_j)/2 \rceil$, $j = 0, \ldots, m$. Observe that $Q_n(g) \subset \mathbb{R}[\mathbf{x}]_{2n}$ because indeed in (6), $\deg(\sigma_j\, g_j) \leq 2n$, for all $j = 0, \ldots, m$. Obviously both $Q(g)$ and its truncated version $Q_n(g)$ are convex cones of $\mathbb{R}[\mathbf{x}]$.

Definition 2.1. The quadratic module $Q(g)$ is said to be *Archimedean* if there exists $M > 0$ such that the quadratic polynomial $\mathbf{x} \mapsto M - \|\mathbf{x}\|^2$ belongs to $Q(g)$ (i.e., belongs to $Q_n(g)$ for some n).

If $Q(g)$ is Archimedean then necessarily, the set

$$S \;:=\; \{\, \mathbf{x} \in \mathbb{R}^d \,:\, g_j(\mathbf{x}) \geq 0\,,\; j = 1, \ldots, m \,\} \tag{7}$$

is compact but the reverse is not true. The Archimedean condition depends on the *representation* of S and can be seen as an *algebraic certificate* that S is compact.
Dual cone. The dual cone $Q^*_n(g)$ of $Q_n(g)$ is the convex cone of $\mathbb{R}^{s(2n)}$ defined by:

$$Q^*_n(g) \;=\; \{\, \boldsymbol{\phi} \in \mathbb{R}^{s(2n)} \,:\, \mathbf{M}_{n-d_j}(g_j \cdot \boldsymbol{\phi}) \;\succeq\; 0\,,\; j = 0, \ldots, m \,\}, \tag{8}$$

where $\mathbf{M}_n(g_j \cdot \boldsymbol{\phi})$ is the localizing matrix associated with the polynomial g_j and the sequence $\boldsymbol{\phi}$, defined in Section 2.1.

For more details on the above notions of moment and localizing matrix, quadratic module, as well as their use in potential applications, the interested reader is referred

[1] A semidefinite program (SDP) is a convex and conic optimization problem which can be solved (up to fixed arbitrary precision) in time polynomial in its input size; see e.g. [1] and also [30].

to [27] and [12, 13]. In particular, both convex cones $Q_n(g)$ and $Q_n^*(g)$ play a crucial role in the *Moment-SOS hierarchy* of lower bounds in polynomial optimization [12].

2.3. **Certificate of positivity (Positivstellensatz).** We next describe a particular certificate of positivity in real algebraic geometry. It is important because it provides a theoretical justification (or rationale) behind convergence of the sequence of semidefinite relaxations at the core of the so-called Moment-SOS hierarchy for global polynomial optimization [12, 13, 14].

Theorem 2.2 ([32]). *Let $S \subset \mathbb{R}^d$ be as in (7) and assume that $Q(g)$ is Archimedean.*
(i) If a polynomial $f \in \mathbb{R}[\mathbf{x}]$ is (strictly) positive on S then $f \in Q(g)$, that is,

$$f \,=\, \sum_{j=0}^{m} \sigma_j \, g_j \,, \tag{9}$$

for some SOS polynomials $\sigma_j \in \Sigma[\mathbf{x}]$, $j = 0, \ldots, m$ (and so $f \in Q_n(g)$ for some $2n \geq \deg(f)$).
(ii) A sequence $\boldsymbol{\phi} = (\phi_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{N}^d} \subset \mathbb{R}$ has a representing Borel measure on S if and only if $\phi(f^2\, g_j) \geq 0$ for all $f \in \mathbb{R}[\mathbf{x}]$, and all $j = 0, \ldots, m$. Equivalently, if and only if $\mathbf{M}_n(g_j \cdot \boldsymbol{\phi}) \succeq 0$ for all $j = 0, \ldots, m$, and all $n \in \mathbb{N}$.

In fact Theorem 2.2 (Putinar's Positivstellensatz) is a refinement of an earlier theorem by Schmüdgen [33] two years before (where the Archimedean condition is not needed). Notice that Theorem 2.2 has two facets (i) and (ii): The former is the algebraic facet (certificate of positivity) while the latter with a real analysis flavor is related to the S-moment problem [33]. Both facets are a nice illustration of the duality between moments and positive polynomials.

3. **The Christoffel function.** The Christoffel Function (CF in short) is usually defined for a measure μ with moments $\boldsymbol{\mu} = (\mu_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{N}^d}$ whose support $S \subset \mathbb{R}^d$ is compact and such that its moment matrix $\mathbf{M}_n(\boldsymbol{\mu})$ (or equivalently, $\mathbf{M}_n(\mu)$) is positive definite for every degree $n \in \mathbb{N}$. However it can also be defined for a Riesz linear functional $\phi \in \mathbb{R}[\mathbf{x}]^*$ (with $\boldsymbol{\phi} = (\phi_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{N}^d}$) such that $\mathbf{M}_n(\boldsymbol{\phi}) \succ 0$ for every $n \in \mathbb{N}$, not necessarily coming from a measure μ. For a more detailed treatment and historical developments, the interested reader is referred to [6, 11, 23, 22, 29, 34, 35, 36] and the many references therein.

3.1. **Christoffel-Darboux kernel and Christoffel Function.** So let $\phi \in \mathbb{R}[\mathbf{x}]^*$ (with $\boldsymbol{\phi} = (\phi_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{N}^d}$) be such that $\mathbf{M}_n(\boldsymbol{\phi}) \succ 0$ for every $n \in \mathbb{N}$. After fixing some ordering of monomials in $\mathbb{N}^d$, and since $\mathbf{M}_n(\boldsymbol{\phi}) \succ 0$ for every n, let $(P_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{N}^d}$ be a family of polynomials that are *orthonormal* w.r.t. ϕ, that is, such that

$$\phi(P_{\boldsymbol{\alpha}}\, P_{\boldsymbol{\beta}}) \,=\, \delta_{\boldsymbol{\alpha}=\boldsymbol{\beta}} \,, \quad \forall \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}^d \,, \tag{10}$$

where $\delta_{\cdot=\cdot}$ is the usual Kronecker symbol (with value 1, if $\boldsymbol{\alpha} = \boldsymbol{\beta}$ and 0 otherwise).

For every $n \in \mathbb{N}$, the Christoffel-Darboux (CD) kernel $K_n^{\phi} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, associated with ϕ, is then defined by:

$$(\mathbf{x}, \mathbf{y}) \mapsto K_n^{\phi}(\mathbf{x}, \mathbf{y}) \,:=\, \sum_{\boldsymbol{\alpha} \in \mathbb{N}^d_n} P_{\boldsymbol{\alpha}}(\mathbf{x})\, P_{\boldsymbol{\alpha}}(\mathbf{y}) \,, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d \,, \quad n \in \mathbb{N} \,, \tag{11}$$

and the Christoffel function (CF) $\Lambda_n^{\phi} : \mathbb{R}^d \to \mathbb{R}_+$ is defined by

$$\mathbf{x} \mapsto \Lambda_n^{\phi}(\mathbf{x})^{-1} \,:=\, K_n^{\phi}(\mathbf{x}, \mathbf{x}) \,=\, \sum_{\boldsymbol{\alpha} \in \mathbb{N}^d_n} P_{\boldsymbol{\alpha}}(\mathbf{x})^2 \,, \quad \forall \mathbf{x} \in \mathbb{R}^d \,, \quad n \in \mathbb{N} \,, \tag{12}$$

i.e., the CF is the reciprocal of the "diagonal" $K_n^\phi(\mathbf{x},\mathbf{x})$ of the CD kernel. Hence by construction $1/\Lambda_n^\phi$ is an SOS polynomial of degree $2n$.

3.2. **Alternative formulations of the CF.** Alternatively, the CF can also be defined by:

$$\Lambda_n^\phi(\boldsymbol{\xi})^{-1} \;=\; \mathbf{v}_n(\boldsymbol{\xi})^T \mathbf{M}_n(\boldsymbol{\phi})^{-1}\, \mathbf{v}_n(\boldsymbol{\xi})\,, \quad \forall \boldsymbol{\xi} \in \mathbb{R}^d\,, \tag{13}$$

which is the ABC theorem in [34], and it also has the variational formulation:

$$\Lambda_n^\phi(\boldsymbol{\xi}) \;=\; \min_{p \in \mathbb{R}[\mathbf{x}]_n} \{\, \phi(p^2) : p(\boldsymbol{\xi}) \;=\; 1 \,\}\,, \quad \forall \boldsymbol{\xi} \in \mathbb{R}^d\,. \tag{14}$$

In particular, observe that (14) can be rewritten

$$\Lambda_n^\phi(\boldsymbol{\xi}) \;=\; \min_{\mathbf{p} \in \mathbb{R}^{s(n)}} \{\, \mathbf{p}^T \mathbf{M}_n(\boldsymbol{\phi})\mathbf{p} : \langle \mathbf{p}, \mathbf{v}_n(\boldsymbol{\xi})\rangle \;=\; 1 \,\}\,, \quad \forall \boldsymbol{\xi} \in \mathbb{R}^d\,,$$

a convex quadratic optimization problem which can be solved efficiently even for large dimension d. After some algebra, the unique optimal solution $p^* \in \mathbb{R}[\mathbf{x}]_n$ of (14) reads

$$\mathbf{x} \mapsto p^*(\mathbf{x}) \;=\; \frac{K_n^\phi(\boldsymbol{\xi},\mathbf{x})}{K_n^\phi(\boldsymbol{\xi},\boldsymbol{\xi})}\,, \quad \forall \mathbf{x} \in \mathbb{R}^d\,.$$

The reproducing property. Let $p \in \mathbb{R}[\mathbf{x}]_n$ and as $(P_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{N}^d_n}$ in (10) form a basis of $\mathbb{R}[\mathbf{x}]_n$, write

$$\mathbf{x} \mapsto p(\mathbf{x}) \;=\; \sum_{\boldsymbol{\alpha} \in \mathbb{N}^d_n} p_{\boldsymbol{\alpha}}\, P_{\boldsymbol{\alpha}}(\mathbf{x}) \quad \forall \mathbf{x} \in \mathbb{R}^d\,,$$

for some vector of coefficients $\mathbf{p} = (p_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{N}^d_n}$ in $\mathbb{R}^{s(n)}$ (with $s(n) = \binom{n+d}{d}$).

With $\mathbf{x} \in \mathbb{R}^d$ fixed, $\mathbf{y} \mapsto K_n^\phi(\mathbf{x},\mathbf{y}) \in \mathbb{R}[\mathbf{y}]_n$, and we have

$$\begin{aligned} \phi(K_n^\phi(\mathbf{x},\cdot)\, p) &= \phi\left((\sum_{\boldsymbol{\alpha} \in \mathbb{N}^d_n} P_{\boldsymbol{\alpha}}(\mathbf{x})\, P_{\boldsymbol{\alpha}}(\mathbf{y})) \cdot (\sum_{\boldsymbol{\beta} \in \mathbb{N}^d_n} p_{\boldsymbol{\beta}}\, P_{\boldsymbol{\beta}}(\mathbf{y}))\right) \\ &= \sum_{\boldsymbol{\alpha} \in \mathbb{N}^d_n} p_{\boldsymbol{\alpha}}\, P_{\boldsymbol{\alpha}}(\mathbf{x}) \;=\; p(\mathbf{x})\,, \quad \forall p \in \mathbb{R}[\mathbf{x}]_n\,, \end{aligned} \tag{15}$$

where we have used that

$$\phi(p_{\boldsymbol{\beta}} P_{\boldsymbol{\beta}}(\mathbf{y})\, P_{\boldsymbol{\alpha}}(\mathbf{x}) P_{\boldsymbol{\alpha}}(\mathbf{y})) \;=\; p_{\boldsymbol{\beta}}\, P_{\boldsymbol{\alpha}}(\mathbf{x})\, \phi(P_{\boldsymbol{\beta}}\, P_{\boldsymbol{\alpha}}) \;=\; p_{\boldsymbol{\beta}}\, P_{\boldsymbol{\alpha}}(\mathbf{x})\, \delta_{\boldsymbol{\beta}=\boldsymbol{\alpha}}\,.$$

For this reason, if μ is a measure on $S \subset \mathbb{R}^d$ with $\mathbf{M}_n(\boldsymbol{\mu}) \succ 0$ for all n, and $L^2(\mu)$ is the Hilbert space of square integrable functions w.r.t. μ, with scalar product

$$\langle f, g\rangle \;=\; \int_S f\, g\, d\mu\,, \quad \forall f, g \;\in\; L^2(\mu)\,,$$

then $(\mathbb{R}[\mathbf{x}]_n, \langle\cdot,\cdot\rangle) \subset L^2(\mu)$ is called a Reproducing Kernel Hilbert Space (RKHS) with kernel K_n^μ, because

$$\int_S K_n^\mu(\mathbf{x},\mathbf{y})\, p(\mathbf{y})\, d\mu(\mathbf{y}) \;=\; p(\mathbf{x})\,, \quad \forall p \in \mathbb{R}[\mathbf{x}]_n\,.$$

3.3. **Equilibrium measure.** A Borel measure μ supported on a compact set $S \subset \mathbb{R}^d$ satisfies the Bernstein-Markov property if there exists a sequence of positive numbers $(M_n)_{n\in\mathbb{N}}$ such that for all n and $p \in \mathbb{R}[\mathbf{x}]_n$,

$$\sup_{\mathbf{x}\in S} |p(\mathbf{x})| \;\leq\; M_n \cdot \left(\int_S p^2\, d\mu\right)^{1/2}, \quad \text{and} \quad \lim_{n\to\infty} \log(M_n)/n \;=\; 0 \tag{16}$$

(see e.g. [23, Section 4.3.3]). The Bernstein-Markov property allows qualitative description for asymptotics of the Christoffel function Λ_n^μ as n grows.

The notion of *equilibrium measure* associated to a given set, originates from logarithmic potential theory (working in $\mathbb{C}$ in the univariate case) to minimize some energy functional. For instance, the (Chebsyshev) measure $d\mu := dx/\pi\sqrt{1-x^2}$ is the equilibrium measure of the interval $[-1, 1]$. Some generalizations have been obtained in the multivariate case via pluripotential theory in $\mathbb{C}^d$. In particular if $S \subset \mathbb{R}^d \subset \mathbb{C}^d$ is compact then its equilibrium measure (let us denote it by λ_S) is equivalent to Lebesgue measure on compact subsets of $\mathrm{int}(S)$. It has an even explicit expression if S is convex and symmetric about the origin; see e.g. [2] [3, Theorem 1.1 and Theorem 1.2].

For a brief account on equilibrium mesures see [2, 3], the discussion in [23, Section 4-5, pp. 56–60] while for more detailed expositions see some of the references therein.

3.4. **Some asymptotic properties of the CF.** As we next see, the CF has some very nice asymptotic properties when n increases, and in the next section we will see how to use some of these properties in various applications.

Support identification. A crucial property of the CFs $(\Lambda_n^\mu)_{n\in\mathbb{N}}$ associated with a measure μ on a compact set $S \subset \mathbb{R}^d$, is to identify the support of μ. Indeed its decay with the degree n exhibits the following interesting dichotomy:

- $\forall \boldsymbol{\xi} \in \mathrm{supp}(\mu)$, $\Lambda_n^\mu(\boldsymbol{\xi})^{-1}$ grows at most as a polynomial in n.
- $\forall \boldsymbol{\xi} \notin \mathrm{supp}(\mu)$, $\Lambda_n^\mu(\boldsymbol{\xi})^{-1}$ grows at least as an exponential in n.

This property has been exploited in data analysis to provide a simple and easy-to-use tool (with no tuning of parameters), e.g. to detect outliers, with similar (and sometimes better) performance as state-of-the-art techniques; see [23, 21].

Relating the CF with the equilibrium measure. If μ is a Borel probability measure on S and (S, μ) has the Bernstein-Markov property (16) then the sequence of probability measures $(\nu_n)_{n\in\mathbb{N}}$ with $\nu_n = \frac{\mu}{s(n)\Lambda_n^\mu(\mathbf{x})}$ for all $n \in \mathbb{N}$, converges to λ_S for the weak-$\star$ topology and therefore in particular:

$$\lim_{n\to\infty} \int_S \mathbf{x}^{\boldsymbol{\alpha}}\, d\nu_n \;=\; \lim_{n\to\infty} \int_S \frac{\mathbf{x}^{\boldsymbol{\alpha}}\, d\mu(\mathbf{x})}{s(n)\Lambda_n^\mu(\mathbf{x})} \;=\; \int_S \mathbf{x}^{\boldsymbol{\alpha}}\, d\lambda_S\,, \quad \forall \boldsymbol{\alpha} \in \mathbb{N}^d\,. \tag{17}$$

(See e.g. [23, Theorem 4.4.4].) In addition, if the compact $S \subset \mathbb{R}^d$ is regular then (S, λ_S) has the Bernstein-Markov property.

Next, if μ has a density f w.r.t. Lebesgue measure, and under some additional regularity properties of μ and its support S, one may link f and the density of the equilibrium measure of S. Namely:

$$\lim_{n\to\infty} s(n)\, \Lambda_n^\mu(\boldsymbol{\xi}) \;=\; f(\boldsymbol{\xi})/\mu_E(\boldsymbol{\xi})\,, \tag{18}$$

uniformly on compact subsets of $\mathrm{int}(S)$, where μ_E is the density of the *equilibrium measure* of S, with respect to Lebesgue measure.

Observe that (18) is quite interesting as for large degree n, the CF allows to approximate the unknown density f *but only if* one already knows the density μ_E of the equilibrium measure of S. Unfortunately, except for sets $S \subset \mathbb{R}^d$ with special geometry (e.g. simplex, unit box, Euclidean unit ball), the density μ_E is not known. However we will see in Section 6 that with an appropriate regularization of the CF one may get rid of μ_E in its asymptotic limit.

3.5. **A disintegration property of the CF.** It is well-know that a probability measure μ on a cartesian product $X \times Y$ disintegrates into

$$\mu(A \times B) \;=\; \int_{X\cap A} \hat{\mu}(B \,|\, \mathbf{x})\, \nu(d\mathbf{x})\,, \quad \forall A \in \mathcal{B}(X)\,, B \in \mathcal{B}(Y)\,, \tag{19}$$

where

- ν is the marginal (probability) measure of μ on X, and
- for every $\mathbf{x} \in X$, $\hat{\mu}(d\mathbf{y} \,|\, \mathbf{x})$ is the conditional probability on Y, given $\mathbf{x} \in X$.

Below we provide a similar disintegration of the CF. So with $X \subset \mathbb{R}^d$, and $Y \subset \mathbb{R}$, let $S \subset X \times Y$ be a compact set and μ be a probability measure on S with marginal ν on X, such that $\mathbf{M}_n(\mu) \succ 0$ for all n.

Theorem 3.1. *Let Λ_n^μ be the Christoffel function associated with μ. Then for every $\mathbf{x} \in X$ and $n \in \mathbb{N}$, there exists a probability measure $\nu_{\mathbf{x},n}$ on $\mathbb{R}$ such that*

$$\Lambda_n^\mu(\mathbf{x}, y) \;=\; \Lambda_n^\nu(\mathbf{x}) \cdot \Lambda_n^{\nu_{\mathbf{x},n}}(y)\,, \quad \forall \mathbf{x} \in \mathbb{R}^d\,,\; y \in \mathbb{R}\,. \tag{20}$$

Observe that (20) has the flavor of the disintegration (19) of μ, as Λ_n^μ disintegrates into the product of the Christoffel Λ_n^ν of the marginal ν with the Christoffel of a measure $\nu_{\mathbf{x},n}$ on the real line. However notice that $\nu_{\mathbf{x},n}$ depends not only on $\mathbf{x} \in \mathbb{R}^d$ but also on the degree n (whereas ideally one would have liked $\nu_{\mathbf{x},n}$ to depend only on $\mathbf{x}$).

Asymptotics. To see what the asymptotic behavior of $\Lambda_n^{\nu_{\mathbf{x},n}}$ as n increases looks like, consider the case where $\mu = f(\mathbf{x}, y)\, d\mathbf{x}\, dy$ so that $\nu = h(\mathbf{x}) d\mathbf{x}$ with $\mathbf{x} \mapsto h(\mathbf{x}) := \int_Y f(\mathbf{x}, y)\, dy$, for some densities $f(\mathbf{x}, y) > 0$ on S and $h(\mathbf{x}) > 0$ on X. Let $s_k(n) = \binom{k+n}{n}$ for all $k \in \mathbb{N}$. Then under some regularity assumptions on S, μ and ν, recall that by (18),

$$\begin{aligned} \lim_{n\to\infty} s_{d+1}(n)\, \Lambda_n^\mu(\mathbf{x}, y) &= f(\mathbf{x}, y)/\mu_E(\mathbf{x}, y)\,,\; \forall (\mathbf{x}, y) \in \mathrm{int}(S)\,; \\ \lim_{n\to\infty} s_d(n)\, \Lambda_n^\nu(\mathbf{x}) &= h(\mathbf{x})/\mu_E(\mathbf{x})\,,\; \forall (\mathbf{x}) \in \mathrm{int}(X)\,. \end{aligned}$$

Hence for every $(\mathbf{x}, y) \in \mathrm{int}(S)$, using $s_{d+1}(n) = s_d(n)(n+d+1)/(d+1)$,

$$\lim_{n\to\infty} s_1(n)\, \Lambda_n^{\nu_{\mathbf{x},n}}(y) \;=\; \frac{f(\mathbf{x}, y)}{h(\mathbf{x})} \cdot \frac{(d+1)\,\mu_E(\mathbf{x})}{\mu_E(\mathbf{x}, y)}\,. \tag{21}$$

Observe that for every fixed $\mathbf{x} \in \mathrm{int}(X)$, the function $y \mapsto f(\mathbf{x}, y)/h(\mathbf{x})$, is the density of the conditional probability $\hat{\mu}(dy|\mathbf{x})$. So, and as expected, (21) has the flavor of (18) but not totally because the term $(d+1)\,\mu_E(\mathbf{x})/\mu_E(\mathbf{x}, y)$ is not necessarily the equilibrium measure of the support $Y_\mathbf{x} \subset Y$ of the conditional probability $\hat{\mu}(dy|\mathbf{x})$.

In the simple case where $S = X \times [a, b]$, and if $\mu_E(\mathbf{x}, y) = \mu_E(\mathbf{x}) \cdot \mu_E(y)$ then (21) becomes

$$\lim_{n\to\infty} s_1(n)\, \Lambda_n^{\nu_{\mathbf{x},n}}(y) \;=\; \frac{f(\mathbf{x}, y)}{h(\mathbf{x})} \cdot \frac{(d+1)\,\mu_E(\mathbf{x})}{\mu_E(\mathbf{x}, y)} \;=\; \frac{f(\mathbf{x}, y)}{h(\mathbf{x})} \cdot \frac{d+1}{\mu_E(y)}\,,$$

which also has the flavor of (18) but with a slightly different corrective term.

4. **Some applications of the Christoffel function.** In addition to being an interesting mathematical object in its own right, the above asymptotic properties are also very useful and can be exploited in various applications, e.g. in data analysis and mining for outlier detection and support inference, interpolation and approximation of discontinuous functions .

4.1. **Outlier detection.** In data analysis and mining one is given a set of N data points $(\mathbf{x}(i))_{i=1,\ldots,N}$ (usually a sample drawn from some unknown distribution μ on $S \subset \mathbb{R}^d$), with possibly some *outlier* points. Therefore an important issue is to provide relatively simple tests to detect whether or not a given point of the cloud is an outlier.

A natural object related to μ and associated with the cloud of points is the empirical probability measure:

$$\mu^N := \frac{1}{N}\sum_{i=1}^{N} \delta_{\mathbf{x}(i)}\,, \tag{22}$$

where $\delta_{\mathbf{x}(i)}$ is the Dirac measure at the point $\mathbf{x}(i)$.

Let $\boldsymbol{\mu}^N := (\mu^N_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^d_{2n}}$. If S has nonempty interior and N is sufficiently large, then $\mathbf{M}_n(\mu^N) \succ 0$ and one may thus define orthonormal polynomials w.r.t. $\boldsymbol{\mu}^N$ and the degree-n CF $\Lambda_n^{\mu_N}$ associated with $\boldsymbol{\mu}^N$. As shown in [22, 23], with n fixed, $\mathbf{M}_n(\boldsymbol{\mu}^N)$ inherits the properties of $\mathbf{M}_n(\mu)$ almost surely w.r.t. samples of μ, as N increases.

So for instance, a naive test to detect outliers is to declare a point $\mathbf{x}(i)$ outlier if $\binom{n+d}{n}\Lambda_n^{\mu^N}(\mathbf{x}(i)) > 1$, that is, one uses the CF $\Lambda_n^{\mu^N}$ as a score function. With this simple naive approach, one obtains performances as good as those obtained with more sophisticated tests (e.g. based on isolation forest) that require a fine tuning of some parameters for practical efficiency. Notice that there is no tuning of parameters in this approach even if other choices than the score $\binom{n+d}{n}$ may be tested if needed.

In particular, in [5] the CF is also used for outlier detection but with a more sophisticated test based on the growth of $\Lambda_n^{\mu^N}$ with its degree n. Indeed the idea exploited in [5] is to choose and fix in advance finitely many degrees $n, n+1, \ldots, n+k$ (for some n and k), and to declare that a point $\mathbf{x}(i)$ as an outlier if the growth in its $k+1$ scores $\Lambda_n^{\mu^N}(\mathbf{x}(i))$, $\Lambda_{n+1}^{\mu^N}(\mathbf{x}(i))$, …, $\Lambda_{n+k}^{\mu^N}(\mathbf{x}(i))$ "looks like" exponential (as it should be if $\mathbf{x}(i)$ is not in the support of μ and N is sufficiently large). In doing so the approach gets rid of the threshold level; see [5] for more details.

4.2. **Interpolation and approximation.** With $X \subset \mathbb{R}$, let μ be a measure on $S \subset [0,1]\times X \subset \mathbb{R}^2$, defined by

$$d\mu(t,x) \;=\; \delta_{\{f(t)\}}(dx)\,1_{[0,1]}(t)\,dt\,, \tag{23}$$

for some unknown measurable function $f : [0,1] \to X$. That is, the support of μ is the graph $\{(t, f(t)) : t \in [0,1]\}$ of f.

Problem. *Recover or approximate f from the sole knowledge of moments $\boldsymbol{\mu} = (\mu_{ij})_{(i,j)\in\mathbb{N}^2}$ where :*

$$\mu_{ij} \;=\; \int t^i x^j\, d\mu(t,x) \;=\; \int_0^1 t^i f(t)^j\, dt\,, \quad (i,j) \in \mathbb{N}^2\,. \tag{24}$$

This generic problem has a lot of applications. For instance, when one applies the Moment-SOS hierarchy to optimal control problems and some nonlinear PDEs as well [9], at each semidefinite relaxation of the hierarchy, one obtains an approximation of finitely moments of a measure μ supported on state and control trajectories $x_i(t), u_j(t)$, $t \in [0,T]$, $i = 1,\ldots,d$, $j = 1,\ldots,m$. Then it remains to recover each trajectory $t \mapsto x_i(t)$ (or $t \mapsto u_i(t)$) from finitely moments of the (bivariate) marginal of μ w.r.t. x_i and t (or w.r.t. u_j and t); see e.g. [20, Section 7.3].

In a now classical approach, one may approximate f in (24) by a degree-n polynomial $p_n^* \in \mathbb{R}[t]_n$ that minimizes the mean squared error, i.e.,

$$p_n^* \;:=\; \arg\min_{p\in\mathbb{R}[t]_n} \int_0^1 (p-f)^2\,dt \;=\; \arg\min_{p\in\mathbb{R}[t]_n} \|p-f\|^2_{L^2([0,1])}\,. \tag{25}$$

This best L^2-norm approximation of f is obtained via a *standard* application of the CD-kernel K_n^ν associated with the univariate measure $\nu = dt$ on $[0,1]$. Indeed the minimizer p_n^* in (25) reads:

$$t \mapsto p_n^*(t) \;=\; \int_0^1 K_n^\nu(t,x)\,f(x)\,dx\,. \tag{26}$$

However it is well-known that if f is discontinuous then the polynomial approximant p_n^* has some annoying drawbacks. In particular is suffers from a classical Gibbs phenomenon at points of discontinuity; see .e.g. Figure 1 (left).

A non-standard application of the CD Kernel. We claim that a non-standard application of the CF can provide f with a better approximant than the polynomial p_n^* in (25). Its rationale is based on the simple previous observation that the CF is a good tool to identify the *support* of its associated measure. Since in the present framework the support of μ in (23) is precisely the *graph* of the unknown function f, then Λ_n^μ should provide us with an adequate tool to recover or approximate f from the sole knowledge of the moments $\boldsymbol{\mu}$ of μ in (24). (Notice that in (26) the support $[0,1]$ of ν has nothing to do with f.)

Here observe that μ in (23) is a degenerate measure on $[0,1]\times X$, i.e. its support has Lebesgue measure zero on $[0,1]\times X$. Therefore its moment matrix $\mathbf{M}_n(\mu)$ can be ill-conditioned and even singular if f is a polynomial (because then the vector of coefficients of $f \in \mathbb{R}[t]$ is in the kernel of $\mathbf{M}_n(\mu)$ when n is sufficiently large). So one first "perturbates" (or regularizes) $\mathbf{M}_n(\mu)$ to $\mathbf{M}_n(\mu)+\varepsilon\mathbf{I}$ with $\mathbf{I}$ the identity matrix and $\varepsilon > 0$ a small regularization parameter. Then define a new perturbated Christoffel function $\hat\Lambda_n^\mu$ by:

$$(t,x) \mapsto \hat\Lambda_n^\mu(t,x)^{-1} \;:=\; \mathbf{v}_n(t,x)^T(\mathbf{M}_n(\mu)+\varepsilon\mathbf{I})^{-1}\mathbf{v}_n(t,x)\,,\quad \forall (t,x)\in\mathbb{R}^2\,, \tag{27}$$

and introduce the following n-approximant $f_n : [0,1] \to X$ of f by:

$$t \mapsto f_n(t) \;:=\; \arg\min_{x\in\mathbb{R}} \hat\Lambda_n^\mu(t,x)^{-1}\,,\quad t\in[0,1]\,. \tag{28}$$

(In case of several minimizers in (28) then as a tie-breaker rule just take the smallest one.) Note that for every fixed $t \in [0,1]$, $f_n(t)$ can be obtained efficiently because $x \mapsto \hat\Lambda_n^\mu(t,x)^{-1}$ is a *univariate* SOS polynomial in x.

As n increases, pointwise convergence (except at points of discontinuity) and L^1-norm convergence to f are proved in [25]. It is crucial that the f_n approximant (28) is semi-algebraic and *not* a polynomial. Indeed by its very semi-algebraic nature, it is able approximate quite well some discontinuous functions with no Gibbs phenomenon.

Example 4.1. Let $f : [0,1] \to [0,1]$ be the discontinuous step function $t \mapsto f(t) = 0$ if $t \in [0, 1/2]$ and $f(t) = 1$ if $t \in (1/2, 1]$. In Figure 1 (left) f (in red) is approximated by $p_n^* \in \mathbb{R}[t]_n$ in (25) (in black) via a standard application of the CD-kernel K_n^ν associated with the univariate measure $\nu = dt$ on $[0,1]$. Observe that even with degree $n = 12$, the resulting polynomial approximation p_{12}^* is not satisfactory with significant oscillations associated with the Gibbs phenomenon. On the other hand, with $\varepsilon > 0$ very small and f_n as in (28), the step function is recovered almost exactly (in black) with no Gibbs phenomenon and with small degree $n = 4$. This is what one may call a *non-standard application* of the CD kernel as one considers the degenerate bivariate measure μ on $[0,1] \times X$ instead of the univariate measure $\nu = f(t)dt$ on $[0,1]$.

Similarly in Figure 2, two discontinuous Eckhoff functions from [7] in red are also recovered (in black) with very good precision via f_n in (28) with $n = 10$, and again with no Gibbs phenomenon; for more details the reader is referred to [25].

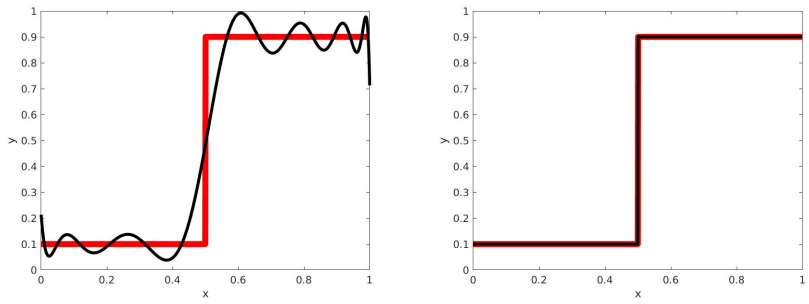


FIGURE 1. Left: Degree-12 polynomial approximation of step function by p_n^* in (25) with Gibbs phenomenon and right: step function approximated by f_4 in (28). © Reprinted from [25]

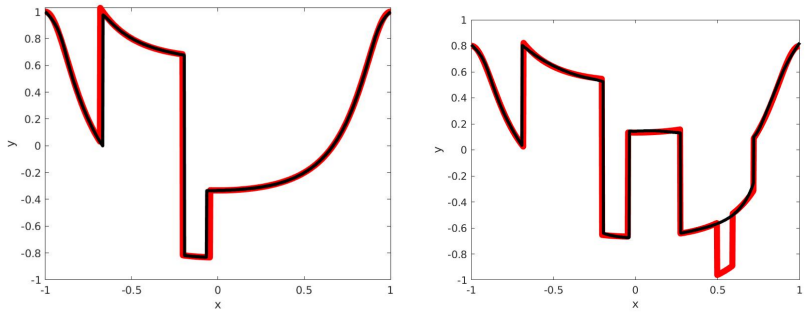


FIGURE 2. Two Eckhoff functions [7] approximated by f_{10} in (28). © Reprinted from [25]

Remark 4.2. This recovery of the unknown f is possible if one has access to the moment information (24). This is the case in interpolation with the moments $(\mu_{ij}^N)_{i,j}$

of the empirical measure μ^N supported on the points $(t(k), f(t(k)))$, $k = 1, \ldots, N$, with $(t(k))_{k\le N} \subset [0,1]$, that is,

$$\mu^N_{ij} := \frac{1}{N}\sum_{k=1}^{N} t(k)^i\, f(t(k))^j\,, \quad i,j \in N\,.$$

In some applications one only has access (e.g. via numerical measurements) to partial moments

$$\mu_{i1} = \int t^i x\, d\mu(t,x) = \int_0^1 t^i\, f(t)\, dt\,, \quad i \in \mathbb{N}\,.$$

So to apply the above recovery procedure one has to first reconstruct the *missing moments*

$$\mu_{ij} = \int t^i\, x^j\, d\mu(t,x) = \int_0^1 t^i\, f(t)^j\, dt\,, \quad i \in \mathbb{N}\,, j \ge 2\,.$$

Fortunately, good approximations of missing moments (in principle as closely as desired) can be obtained via solving a hierarchy of semidefinite relaxations described in [10] (as a special application of the Moment-SOS hierarchy).

4.3. **Supervised learning.** In supervised classification with noiseless deterministic labels, the objects of interest $\mathbf{x} \in X$ belong to m classes with supports $X_j \subset X \subset \mathbb{R}^n$, $j \in [m]$ (with $[m] = \{1, \ldots, m\} =: Y$). The supports satisfy $X_i \cap X_j = \emptyset$ for all i, j with $i \ne j$. The data set consists of clouds of finitely many points $(\mathbf{x}(i)) \subset X_j$ sampled from an underlying distribution ϕ_j on X_j, $j \in [m]$. In this situation, an *exact* classifier $f : X \to Y$, selects $j =: f(\mathbf{x})$ whenever $\mathbf{x} \in X_j$. When constructing a classifier from a sample of data points, as e.g. in machine learning, a sensitive issue is its generalization properties when applied on a test set different from the training set. For the reader interested in recent developments on various techniques and issues in supervised and unsupervised classification, we refer to e.g. the book [4] and the many references therein.

In [16] the author has introduced a simple and natural ideal classifier $f_n : X \to Y$, with nice asymptotic properties as n increases. It is based on the Christoffel function Λ^μ_n associated with the joint distribution $d\mu(\mathbf{x}, y)$ on $X \times Y$. As μ is supported on the graph $\{(\mathbf{x}, f(\mathbf{x})) : \mathbf{x} \in X\}$ of the exact classifier f, recent results of [25] also used in Section 4.2 and now transported in our context, suggest that the classifier

$$f_n(\mathbf{x}) := \arg\max_{y\in Y} \Lambda^\mu_n(\mathbf{x}, y)\,, \quad \mathbf{x} \in X\,, \tag{29}$$

should approximate f nicely. This is the case and indeed, by a slight modification $\hat\Lambda^\mu_n$ of the initial definition of Λ^μ_n, we show that f_n is simply expressed in terms of the Christoffel functions $\Lambda^{\phi_j}_n$ of the ϕ_j; namely,

$$f_n(\mathbf{x}) := \arg\max_{y\in Y} \hat\Lambda^\mu_n(\mathbf{x}, y) = \arg\max_j \Lambda^{\phi_j}_n(\mathbf{x})\,, \quad \mathbf{x} \in X\,. \tag{30}$$

So the classifier f_n simply selects the class $f(\mathbf{x}) := \{j\}$ whose value (score) of its Christoffel function $\Lambda^{\phi_j}_n(\mathbf{x})$ at $\mathbf{x} \in X$, is maximum over all classes. Notice that the simple form (30) of f_n mathematically justifies for (noiseless) supervised classification, the intuitive argument that $\Lambda^{\phi_j}_n(\mathbf{x}) > \Lambda^{\phi_k}_n(\mathbf{x})$, $\forall k \ne j$, whenever n is sufficiently large and $\mathbf{x} \in X_j$. Indeed from Section 3.4, as $\mathbf{x} \in X_j$ is outside the support X_k of ϕ_k, for every $k \ne j$, the "score" $\Lambda^{\phi_k}_n(\mathbf{x})$ is close to zero for sufficiently large n, as it decreases exponentially fast to zero (while the decrease of $\Lambda^{\phi_j}_n(\mathbf{x})$ is

at most polynomial in n). For more details and extension to the case where data are corrupted by noise, the interested reader is referred to [16].

5. Positive polynomials and Christoffel function.

5.1. A distinguished representation of positive polynomials.

From its definition (12) the CF is the reciprocal of a sum-of-squares (SOS). But in fact we next show that the CF also appears in a certain distinguished representation of polynomials that are positive on a semi-algebraic set $S \subset \mathbb{R}^d$ as in (7), extensively used in the Moment-SOS hierarchy [12, 13]. In particular:

Every SOS polynomial p in the interior of the convex cone $\Sigma[\mathbf{x}]_n$ of degree-$2n$ SOS polynomials, is the reciprocal of the CF of some linear functional ϕ in $\mathbb{R}[\mathbf{x}]_{2n}^*$.

This result is a consequence of a duality result of Nesterov [28] which can be translated as : Among all Gram matrices of a given polynomial $p \in \text{int}(\Sigma[\mathbf{x}]_n)$ there is one that maximizes the "log-barrier" function of the cone $\Sigma[\mathbf{x}]_n$, and it is the inverse of the moment matrix $\mathbf{M}_n(\boldsymbol{\phi}) \succ 0$ of some $\phi \in \text{int}(\Sigma[\mathbf{x}]_n^*)$. If $n = 2$ then ϕ has a clear interpretation in terms of a Gaussian measure but in the general case, the link between p and ϕ is only partially understood and remains to be interpreted.

The above result extends to a far more general setting.

Theorem 5.1. *Let $Q_n(g)$ be as in (6) with dual $Q_n^*(g)$ as in (8). If $p \in \text{int}(Q_n(g))$ then there exists $\boldsymbol{\phi} \in \text{int}(Q_n^*(g))$ such that*

$$p(\mathbf{x}) \;=\; \sum_{j=0}^{m} \Lambda_{n-d_j}^{g_j\cdot\boldsymbol{\phi}}(\mathbf{x})^{-1}\, g_j(\mathbf{x})\,, \quad \forall \mathbf{x} \in \mathbb{R}^d\,, \tag{31}$$

or, equivalently:

$$\text{int}(Q_n(g)) \;=\; \{\, \sum_{j=0}^{m} (\Lambda_{n-d_j}^{g_j\cdot\boldsymbol{\phi}})^{-1} g_j \;:\; \boldsymbol{\phi} \in \text{int}(Q_n^*(g))\,\}\,. \tag{32}$$

Theorem 5.1 is an interpretation in [17] of a duality result of Nesterov [28]. Remarkably, it states that every p in the interior of $Q_n(g)$ has a distinguished certificate of its positivity on S, with very specific SOS weights $\sigma_j = (\Lambda_{n-d_j}^{g_j\cdot\boldsymbol{\phi}})^{-1}$ in its Putinar's representation (9). Those specific weights are all coming from a *unique* element $\boldsymbol{\phi} \in \text{int}(Q_n^*(g))$ and its Christoffel functions associated with the Riesz linear functionals $g_j \cdot \boldsymbol{\phi}$, $j = 0, \ldots, m$. It also turns out that those weights have an extremal property: Consider the optimization problem:

$$\begin{aligned}\rho_n \;=\; \inf_{\boldsymbol{\phi}\in\mathbb{R}^{s(2n)}} \Big\{ -\sum_{j=0}^{m} \log\det(\mathbf{M}_{n-d_j}(g_j\cdot\boldsymbol{\phi})) : \\ \phi(p) \;=\; 1,\, \mathbf{M}_{n-d_j}(g_j\cdot\boldsymbol{\phi}) \;\succeq\; 0\,,\; \forall j = 0,\ldots,m \Big\}.\end{aligned} \tag{33}$$

It is a convex optimization problem which has an explicit dual, namely

$$\begin{aligned}\rho_n^* \;=\; \sup_{\mathbf{Q}_j} \Big\{ \sum_{j=0}^{m} \log\det(\mathbf{Q}_j) \;:\; \mathbf{Q}_j \succeq 0\,, \forall j = 0,\ldots,m \\ p(\mathbf{x}) \sum_{j=0}^{m} s(n-d_j) = \sum_{j=0}^{m} g_j(\mathbf{x}) \cdot \mathbf{v}_{n-d_j}(\mathbf{x})^T \mathbf{Q}_j \mathbf{v}_{n-d_j}(\mathbf{x}), \forall \mathbf{x} \in \mathbb{R}^d \Big\},\end{aligned} \tag{34}$$

where the supremum is taken over real symmetric matrices $\mathbf{Q}_j$ of respective size $s(n-d_j)$, $j=0,\ldots,m$ (and recall that $d_j = \lceil \deg(g_j)/2 \rceil$). The criterion to maximize in (33) is minus the log-barrier of the convex cone $Q_n^*(g)$. Both problems (33) and (34) are convex and can be solved by some standard software packages like e.g. [8].

Theorem 5.2. *With $n \in \mathbb{N}$ fixed, Problems (33) and (34) have the same finite optimal value $\rho_n = \rho_n^*$ if and only if $p \in \mathrm{int}(Q_n(g))$. Moreover, both have a unique optimal solution $\boldsymbol{\phi}_{2n}^* \in \mathbb{R}^{s(2n)}$ and $(\mathbf{Q}_j^*)_{j=0,\ldots,m}$ respectively, which satisfy $\mathbf{Q}_j^* = \mathbf{M}_{n-d_j}(g_j \cdot \boldsymbol{\phi}_{2n}^*)^{-1}$ for all $j=0,\ldots,m$. And, as a consequence,*

$$p(\mathbf{x}) \;=\; \frac{1}{\sum_{j=0}^{m} s(n-d_j)} \sum_{j=0}^{m} g_j(\mathbf{x})\, \Lambda_{n-d_j}^{g_j\cdot\boldsymbol{\phi}_{2n}^*}(\mathbf{x})^{-1}\,, \quad \forall \mathbf{x} \in \mathbb{R}^d\,. \tag{35}$$

Notice that $\boldsymbol{\phi}$ in (31) is just $(\sum_{j=0}^m s(n-d_j))\,\boldsymbol{\phi}_{2n}^*$ with $\boldsymbol{\phi}_{2n}^*$ as in (35).

Of course, Theorem 5.1 immediately raises the following question: *given $p \in \mathrm{int}(Q_n(g))$ what is this linear functional $\phi \in \mathbb{R}[\mathbf{x}]_{2n}^*$ with associated moment sequence $\boldsymbol{\phi} \in \mathbb{R}^{s(2n)}$ in Theorem 5.1?* It turns out that there is a simple and remarkable answer for special sets S and the constant polynomial $p=\mathbf{1}$. Before turning to this point we make a little detour and introduce the polynomial *Pell's equation*.

5.2. **Polynomial Pell's equation.** Let μ be the Chebyshev measure $dx/\pi\sqrt{1-x^2}$ on the interval $[-1,1]$, and let $(T_n)_{n\in\mathbb{N}}$ (resp. $(U_n)_{n\in\mathbb{N}}$) be the Chebyshev polynomials of first kind (resp. of second kind). Then polynomial Pell's equation states that

$$T_j(x)^2 + (1-x^2)\,U_{j-1}(x)^2 \;=\; 1\,, \quad \forall x \in \mathbb{R}\,, \quad j \geq 1\,. \tag{36}$$

Eq. (36) is an identity, but one also says that the triplet $(T_j,\,(1-x^2),\,U_{j-1})$ solves polynomial Pell's equation[2] (Pell's equation originates in algebraic number theory).

When looking at (36) with glasses from real algebraic geometry, one also immediately recognizes in (36) the Markov-Lukács representation of the constant polynomial $\mathbf{1} \in \mathbb{R}[x]$, which is positive on $[-1,1]$. Even more, letting $x \mapsto g(x) := (1-x^2)$ and scaling $\widehat{T}_j := T_j/\sqrt{2}$ (resp. $\widehat{U}_j := U_j/\sqrt{2}$) for all j (except $\widehat{T}_0 = T_0$) to make $(\widehat{T}_n)_n$ orthonormal w.r.t. μ (resp. $(\widehat{U}_n)_n$ orthonormal w.r.t. $\nu := g\cdot\mu$), one obtains:

$$\begin{aligned}(2n+1) \;&=\; \sum_{j=0}^{n} \widehat{T}_j(x)^2 + (1-x^2)\sum_{j=1}^{n} \widehat{U}_{j-1}(x)^2\,, \quad \forall x \in \mathbb{R}\,.\\ &=\; \Lambda_n^{\mu}(x)^{-1} + g(x)\,\Lambda_{n-1}^{g\cdot\mu}(x)^{-1}\,, \quad \forall x \in \mathbb{R}\,.\end{aligned} \tag{37}$$

Hence (37) is just (31) in Theorem 5.1 for the constant polynomial $p(x) = 2n+1$ for all x on $S = [-1,1] = \{\,x : g(x) \geq 0\}$. So here the mysterious linear functional $\boldsymbol{\phi} \in \mathbb{R}[x]_{2n}^* \in \mathrm{int}(Q_n^*(g))$ in Theorem 5.1 is the vector of moments $\boldsymbol{\mu} = (\mu_j)_{j\leq 2n}$ up to degree $2n$, of the Chebyshev measure $\mu = dx/\pi\sqrt{1-x^2}$, which is the equilibrium measure of S.

[2] A triplet (A,B,C) solves polynomial Pell's equation if $A,B,C \in \mathbb{Z}[x]$, and $A^2 - B\,C^2 = 1$. For more details on the polynomial Pell's equation the interested reader is referred to [3, 26, 37].

A multivariate extension. In [18] and later in [24], we have been able to extend this result to three cases of sets $S \subset \mathbb{R}^d$ and their equilibrium measure μ, namely:

Unit ball: $S = \{\mathbf{x} \in \mathbb{R}^d : 1 - \|\mathbf{x}\|^2 \geq 0\}$ where μ is proportional to

$$\frac{d\mathbf{x}}{\sqrt{1-\|\mathbf{x}\|^2}}\,. \tag{38}$$

Unit box: $S = \mathbf{B}_d := [-1,1]^d$ where μ is proportional to

$$\frac{d\mathbf{x}}{\sqrt{1-x_1^2}\cdots\sqrt{1-x_n^2}}\,. \tag{39}$$

Simplex: $S = \{\mathbf{x} \in \mathbb{R}^d_+ : 1 - \sum_{j=1}^n x_j \geq 0\}$ where μ is proportional to

$$\frac{d\mathbf{x}}{\sqrt{x_1}\cdots\sqrt{x_n}\sqrt{1-\sum_{j=1}^n x_j}}\,. \tag{40}$$

For instance consider the simplex S, and let

$$\mathbf{x} \mapsto g_i(\mathbf{x}) := x_i\,,\ i = 1,\ldots,d\,; \quad \mathbf{x} \mapsto g_{d+1}(\mathbf{x}) := 1 - \sum_{i=1}^d x_i\,,$$

so that $S = \{\mathbf{x} : g_j(\mathbf{x}) \geq 0\,,\ j = 1,\ldots,d+1\}$. Next, with $\varepsilon \in \{0,1\}^{d+1}$, define

$$\mathbf{x} \mapsto g_\varepsilon(\mathbf{x}) := x_1^{\varepsilon_1}\cdots x_d^{\varepsilon_d}\cdot(1-\sum_{i=1}^d x_i)^{\varepsilon_{d+1}}\,, \quad \forall \mathbf{x} \in \mathbb{R}^d\,.$$

Theorem 5.3 ([24]). *Let μ be the equilibrium measure of S in (40), re-scaled to be a probability measure. Then for every $j \in 2\mathbb{N}$:*

$$\sum_{\varepsilon\in\{0,1\}^{d+1};\,\varepsilon\in 2\mathbb{N};\,|\varepsilon|\leq 2j} g_\varepsilon(\mathbf{x})\,\tilde{K}^{g_\varepsilon\cdot\mu}_{j-\varepsilon/2}(\mathbf{x},\mathbf{x}) \;=\; \frac{(d+2)_{2j}}{2j!}\,, \quad \forall \mathbf{x} \in \mathbb{R}^d\,, \tag{41}$$

and as a consequence, for every $n \in \mathbb{N}$:

$$\sum_{\varepsilon\in\{0,1\}^{d+1};\,\varepsilon\in 2\mathbb{N};\,|\varepsilon|\leq n} g_\varepsilon(\mathbf{x})\,\Lambda^{g_\varepsilon\cdot\mu}_{n-|\varepsilon|/2}(\mathbf{x})^{-1} \;=\; \sum_{j=0}^n \frac{(d+2)_{2j}}{2j!}\,, \quad \forall \mathbf{x} \in \mathbb{R}^d \tag{42}$$

(where $(a)_n = a(a+1)(a+2)\cdots(a+n-1)$ is the Pochhammer symbol).

As we can see in (42), again the constant polynomial $\mathbf{1} \in \mathrm{int}(Q_n(g_1,\ldots,g_{d+1}))$ is related to the equilibrium mesure μ of S via the CFs of μ and $g_\varepsilon \cdot \mu$ in its distinguished Putinar's certificate of positivity (31). For an analogue of Theorem 5.3 for the unit box and the Euclidean unit box, the reader is referred to [24].

6. **A regularization of the Christoffel function.** Let $S \subset \mathbb{R}^d$ be compact with nonempty interior and let μ be a Borel measure with $\mathrm{supp}(\mu) = S$. Recall that by (18), if $\mu = f\,d\mathbf{x}$ with $f > 0$ on S, then under some regularity conditions on the couple (S,μ), one may obtain the asymptotic result

$$\lim_{n\to\infty} s(n)\,\Lambda^\mu_n(\mathbf{x}) \;=\; f(\mathbf{x})/\mu_E(\mathbf{x})\,, \quad \forall \mathbf{x} \in \mathrm{int}(S)\,, \tag{43}$$

where μ_E is the density of the equilibrium measure of S and the above convergence is uniform on compact subsets of $\mathrm{int}(S)$.

This result is quite interesting as its relates the Christoffel function with the density f of μ and the density μ_E of the equilibrium measure of S. However, for

practical purpose it does not allow to identify $f(\mathbf{x})$ via $\Lambda^\mu_n(\mathbf{x})$ for $\mathbf{x} \in \mathrm{int}(S)$ because in general the density μ_E of the equilibrium measure of S is *not* known. In the sequel we describe how to provide a remedy.

Let $\varepsilon > 0$ be fixed and define $\mathbf{B}_\infty(\mathbf{x},\varepsilon) := \{\mathbf{y} : \|\mathbf{y}-\mathbf{x}\|_\infty < \varepsilon/2\}$ so that $\mathrm{vol}(\mathbf{B}_\infty(\mathbf{x};\varepsilon)) = \varepsilon^d$. Then introduce the following function

$$(\mathbf{x},\varepsilon) \mapsto \hat{\Lambda}^\mu_n(\mathbf{x},\varepsilon) := \inf_{p\in\mathbb{R}[\mathbf{x}]_n} \{ \int p^2\, d\mu : \int_{\mathbf{B}_\infty(\mathbf{x},\varepsilon)} p(\mathbf{y})\, \frac{d\mathbf{y}}{\varepsilon^d} = 1 \}, \tag{44}$$

for all $(\mathbf{x},\varepsilon) \in \mathbb{R}^d \times \mathbb{R}_+$ and $n \in \mathbb{N}$. Observe that (44) is a *regularization* of the variational definition (14) of the CF Λ^μ_n. Even more, letting

$$\hat{\mathbf{v}}_n(\mathbf{x},\varepsilon) := \int_{\mathbf{B}_\infty(\mathbf{x},\varepsilon)} \mathbf{v}_n(\mathbf{y})\, \frac{d\mathbf{y}}{\varepsilon^d}, \quad n \in \mathbb{N},$$

we also obtain

$$\hat{\Lambda}^\mu_n(\mathbf{x},\varepsilon)^{-1} = \hat{\mathbf{v}}_n(\mathbf{x},\varepsilon)^T \mathbf{M}_n(\mu)^{-1}\, \hat{\mathbf{v}}_n(\mathbf{x},\varepsilon), \quad \forall \mathbf{x} \in \mathbb{R}^d, \tag{45}$$

which is also an analogue of (13). Interestingly, $\hat{\mathbf{v}}_n \in \mathbb{R}[\mathbf{x},\varepsilon]_n$, i.e., $\hat{\mathbf{v}}_n$ is a *polynomial* in the variables $(\mathbf{x},\varepsilon)$, which can be obtained explicitly in closed form. As a result, $1/\hat{\Lambda}^\mu_n \in \Sigma[\mathbf{x},\varepsilon]_n$, i.e., $1/\hat{\Lambda}^\mu_n$ is an SOS polynomial in the variables $(\mathbf{x},\varepsilon)$ which can be also obtained explicitly in closed-form.

As we next see, when $1/\hat{\Lambda}^\mu_n$ is seen as a a polynomial of $\mathbf{x}$, parametrized by $\varepsilon > 0$ *fixed*, we can obtain a nice characterization of its asymptotic behavior as n grows, in terms of the sole density f of μ without a term involving the unknown density μ_E of the equilibrium measure of S (as is the case in (43)).

Theorem 6.1. *Let $S \subset \mathbb{R}^d$ be compact with nonempty interior and $\varepsilon > 0$ be fixed. Let μ be a measure on S with density f w.r.t. Lebesgue measure and with $f > 0$ on S. Then for every $\boldsymbol{\xi} \in S$ such that $\mathbf{B}_\infty(\boldsymbol{\xi},\varepsilon) \subset S$,*

$$\lim_{n\to\infty} \varepsilon^d\, \hat{\Lambda}^\mu_n(\boldsymbol{\xi},\varepsilon)^{-1} = \int_{\mathbf{B}_\infty(\boldsymbol{\xi},\varepsilon)} \frac{1}{f}\, \frac{d\mathbf{x}}{\varepsilon^d} \tag{46}$$

and if f is continuous on S then

$$\lim_{n\to\infty} \varepsilon^{-d}\, \hat{\Lambda}^\mu_n(\boldsymbol{\xi},\varepsilon) = \left(\int_{\mathbf{B}_\infty(\boldsymbol{\xi},\varepsilon)} \frac{1}{f}\, \frac{d\mathbf{x}}{\varepsilon^d} \right)^{-1} = f(\boldsymbol{\zeta}), \tag{47}$$

for some $\boldsymbol{\zeta} \in \mathbf{B}_\infty(\boldsymbol{\xi},\varepsilon)$ (and $\lim_{n\to\infty} \varepsilon^{-d}\, \hat{\Lambda}^\mu_n(\boldsymbol{\xi},\varepsilon) \approx f(\boldsymbol{\xi})$ for ε sufficiently small).

As the reader may notice, in contrast with (18) the scaling ε^{-d} of $\hat{\Lambda}^\mu_n$ in (47) is fixed, and not growing with n. Moreover if f is continuous and $\varepsilon > 0$ is sufficiently small, the limit in (47) provides a simple approximation of the density $f(\boldsymbol{\xi})$ provided that $\boldsymbol{\xi} \in \mathrm{int}(S)$ is at least at ε-distance of the boundary ∂S.

7. **Conclusion.** We have provided a brief introduction to the Christoffel function (CF) and some of its recent applications in data analysis and mining, interpolation and approximation. We have also proposed a regularized version of the CF parametrized by a fixed (arbitrary) scalar $\varepsilon > 0$, with same computational cost and nice asymptotic property when the degree increases (and ε is small). We hope that we have convinced the reader that in view of all its properties and its (surprising) links with seemingly unrelated fields, the CF should suscitate further interest and investigations by the optimization research community.

Acknowledgments. This work is also supported by the AI Interdisciplinary Institute ANITI funding through the french program "Investing for the Future PI3A" under the grant agreement number ANR-19-PI3A-0004. This research is also part of the programme DesCartes and is supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme.

REFERENCES

[1] M. Anjos and J. B. Lasserre, *Handbook of Semidefinite, Conic and Polynomial Optimization*, Springer, New York, 2011.
[2] M. Baran, Complex equilibrium measure and Bernstein type theorems for compact sets in $\mathbb{R}^n$, *Proc. Amer. Math. Soc.*, **123** (1995), 485-494.
[3] E. Bedford and B. A. Taylor, The complex equilibrium measure of a symmetric convex set in $\mathbb{R}^n$, *Trans. Amer. Math. Soc.*, **294** (1986), 705-717.
[4] S. L. Brunton and J. N. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*, Cambridge University Press, 2019.
[5] K. Ducharlet, L. Travé-Massuyès, J. B. Lasserre, M.-V. Le Lann and Y. Miloudi, *Leveraging the Christoffel Function for Outlier Detection in Data Streams*, hal-03562614, 2022.
[6] C. F. Dunkl and Yuan Xu, *Orthogonal Polynomials of Several Variables*, 2^{nd} edition, Cambridge University Press, Cambridge, UK, 2014.
[7] K. S. Eckhoff, Accurate and efficient reconstruction of discontinuous functions from truncated series expansions, *Math. Comput.*, **61** (1993), 745-763.
[8] M. Grant and S. Boyd, *CVX: Matlab Software for Disciplined Convex Programming*, version 2.1., `http://cvxr.com/cvx`, 2014.
[9] D. Henrion, M. Korda and J. B. Lasserre, *The Moment-SOS Hierarchy: Lectures in Probability, Statistics, Computational Geometry, Control and Nonlinear PDEs*, World Scientific, Singapore, 2022.
[10] D. Henrion and J. B. Lasserre, Graph recovery from incomplete moment information, *Constr. Approx.*, **56** (2022), 165-187.
[11] A. Kroó and D. S. Lubinsky, Christoffel functions and universality in the bulk for multivariate orthogonal polynomials, *Canad. J. Math.*, **65** (2012), 600-620.
[12] J. B. Lasserre, Optimisation globale et théorie des moments, *C. R. Acad. Sci. Paris, Série I*, **331** (2000), 929-934.
[13] J. B. Lasserre, *Moments, Positive Polynomials and Their Applications*, Imperial College Press, London, 2009.
[14] J. B. Lasserre, *Introduction to Polynomial and Semi-Algebraic Optimization*, Cambridge University Press, Cambridge, UK, 2015.
[15] J. B. Lasserre, The Moment-SOS hierarchy and the Christoffel-Darboux kernel, *Optim. Letters*, **15** (2021), 1835-1845.
[16] J. B. Lasserre, On the Christoffel function and classification in data analysis, *Comptes Rendus Mathématique*, **360** (2022), 919-928.
[17] J. B. Lasserre, A disintegration of the Christoffel function, *Comptes Rendus Mathématique*, **360** (2022), 1071-1079.
[18] J. B. Lasserre, Pell's equation, sum-of-squares and equilibrium measures of compact sets, *Comptes Rendus Mathématique*, **361** (2023), 935-952.
[19] J. B. Lasserre, A modified Christoffel function and its asymptotic properties, *J. Approx. Theory*, **295** (2023), Article ID: 105955.
[20] J. B. Lasserre, The Moment-SOS hierarchy: Applications and related topics, *Acta Numerica*, (2023), to appear.
[21] J. B. Lasserre and E. Pauwels, Sorting out typicality via the inverse moment matrix SOS polynomial, in *Advances in Neural Information Processing Systems* (eds. D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon and R. Garnett), Curran Associates, Inc., (2016), 190-198.
[22] J. B. Lasserre and E. Pauwels, The empirical Christoffel function with applications in data analysis, *Adv. Comput. Math.*, **45** (2019), 1439-1468.

[23] J. B. Lasserre, E. Pauwels and M. Putinar, *The Christoffel-Darboux Kernel for Data Analysis*, Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, UK, 2022.
[24] J. B. Lasserre and Yuan Xu, A generalized Pell's equation for a class of orthogonal polynomials, arXiv:2307.10668.
[25] S. Marx, E. Pauwels, T. Weisser, D. Henrion and J. B. Lasserre, Semi-Algebraic approximation using Christoffel-Darboux kernel, *Constr. Approx.*, **54** (2021), 391-429.
[26] J. Mc Laughlin, Multivariable-polynomial solutions to Pell's equation and fundamental units in real quadratic fields, *Pacific J. Math.*, **210** (2002), 335-348.
[27] M. Laurent, Sums of squares, moment matrices and optimization over polynomials, in *Emerging Applications of Algebraic Geometry* (eds. M. Putinar and S. Sullivant), Springer, New York, (2008), 157-270.
[28] Y. Nesterov, Squared functional systems and optimization problems, in *High Performance Optimization* (eds.H. Frenk, K. Roos, T. Terlaky and Shuzong Zhang), Applied Optimization Series vol 33, Springer, Boston MA, (2000), 405-440.
[29] P. Nevai and G. Freud, Orthogonal polynomials and Christoffel functions, *J. Approx. Theory*, **48** (1986), 3-167.
[30] R. O'Donnell, SoS is not fully automatisable, even approximately, in *8th Innovations in Theoretical Computer Science Conference (ITCS 2017), Leibniz Intern. Proc. Informatics (LIPIcs)* (ed. H. Papadimitriou), Vol. **67** (2017), Dagstuhl, Germany, 59:1-59:10.
[31] E. Pauwels, M. Putinar and J. B. Lasserre, Data analysis from empirical moments and the Christoffel function, *Found. Comput. Math.*, **21** (2021), 243-273.
[32] M. Putinar, Positive polynomials on compact semi-algebraic sets, *Indiana Univ Math. J.*, **42** (1993), 969-984.
[33] K. Schmüdgen, *The Moment Problem*, Springer Cham, 2017.
[34] B. Simon, The Christoffel-Darboux kernel, in *Perspective in Partial Ddifferential Erations, Harmonic Analysis and Applications*, Proc. Symp. Pure Math. **79** (2008), 295-335, American Mathematical Society, Providence, RI, 2008.
[35] H. Stahl and V. Totik, *General Orthogonal Polynomials*, Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, 1992.
[36] V. Totik, Universality and fine zero spacing on general sets, *Ark. Mat.*, **47** (2009), 361-391.
[37] W. Webb and Hisashi Yokota, Polynomial Pell's equation, *Proc. Amer. Math. Soc.*, **131** (2002), 993-1006.

Received November 2023; 1^{st} revision February 2024; final revision March 2024; early access March 2024.