



HAL
open science

A DISINTEGRATION OF THE CHRISTOFFEL FUNCTION

Jean-Bernard Lasserre

► **To cite this version:**

Jean-Bernard Lasserre. A DISINTEGRATION OF THE CHRISTOFFEL FUNCTION. Comptes Rendus. Mathématique, 2022, 360, pp.1071–1079. 10.5802/crmath.380 . hal-03624003v2

HAL Id: hal-03624003

<https://laas.hal.science/hal-03624003v2>

Submitted on 9 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



INSTITUT DE FRANCE
Académie des sciences

Comptes Rendus

Mathématique

Jean B. Lasserre

A disintegration of the Christoffel function

Volume 360 (2022), p. 1071-1079

<https://doi.org/10.5802/crmath.380>



This article is licensed under the
CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LICENSE.
<http://creativecommons.org/licenses/by/4.0/>



Les Comptes Rendus. Mathématique sont membres du
Centre Mersenne pour l'édition scientifique ouverte
www.centre-mersenne.org
e-ISSN : 1778-3569



Statistics / Statistiques

A disintegration of the Christoffel function

Jean B. Lasserre^a

^a LAAS-CNRS and Institute of Mathematics, BP 54200, 7 Avenue du Colonel Roche,
31031 Toulouse cédex 4, France

E-mail: lasserre@laas.fr

Abstract. We show that the Christoffel function (CF) factorizes (or can be disintegrated) as the product of two Christoffel functions, one associated with the marginal and the another related to the conditional distribution, in the spirit of “*the CF of the disintegration is the disintegration of the CFs*”. In the proof one uses an apparently overlooked property (but interesting in its own) which states that any sum-of-squares polynomial is the Christoffel function of some linear form (with a representing measure in the univariate case). The same is true for the convex cone of polynomials that are positive on a basic semi-algebraic set. This interpretation of the CF establishes another bridge between polynomials optimization and orthogonal polynomials.

Résumé. Nous montrons que la fonction de Christoffel (CF) se factorise en le produit de deux fonctions de Christoffel dont une est celle de la marginale et l'autre est liée à la probabilité conditionnelle. La démonstration utilise une propriété apparemment ignorée (mais intéressante en soi), qui stipule que tout polynôme qui est somme de carrés est aussi la fonction de Christoffel d'une forme linéaire (représentée par une mesure dans le cas univarié). Il en va de même pour le cône convexe des polynômes positifs sur un ensemble basique semi-algébrique. Cette interprétation de la fonction de Christoffel fournit un pont supplémentaire entre l'optimisation polynomiale et les polynômes orthogonaux.

2020 Mathematics Subject Classification. 42C05, 47B32.

Funding. J.B. Lasserre is 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.

Manuscript received 26 March 2022, accepted 19 May 2022.

1. Introduction

It is well-known that a probability measure μ on a Cartesian product $\mathbf{X} \times \mathbf{Y} \subset \mathbb{R}^n \times \mathbb{R}^p$ of Borel spaces, *disintegrates* into $\hat{\mu}(d\mathbf{y}|\mathbf{x})\phi(d\mathbf{x})$ with its marginal $\phi(d\mathbf{x})$ on \mathbf{X} and its conditional probability $\hat{\mu}(d\mathbf{y}|\mathbf{x})$ on \mathbf{Y} , given $\mathbf{x} \in \mathbf{X}$. That is:

$$\mu(\mathbf{A} \times \mathbf{B}) = \int_{\mathbf{X} \cap \mathbf{A}} \hat{\mu}(\mathbf{B}|\mathbf{x})\phi(d\mathbf{x}), \quad \forall \mathbf{A} \in \mathcal{B}(\mathbb{R}^n), \mathbf{B} \in \mathcal{B}(\mathbb{R}^p). \quad (1)$$

The goal of this note is to provide a similar disintegration (or factorization) for the family of its Christoffel functions $(\mathbf{x}, \mathbf{y}) \mapsto \Lambda_t^\mu(\mathbf{x}, \mathbf{y})$, $t \in \mathbb{N}$.

Contribution. Our contribution is twofold.

(i). Consider a probability measure μ on a compact subset $\Omega := \mathbf{X} \times \mathbf{Y} \subset \mathbb{R}^n \times \mathbb{R}$ and let $\Lambda_t^\mu : \mathbb{R}^{n+1} \rightarrow \mathbb{R}_+$ be its associated Christoffel function, i.e., with $\mathbb{N}_t^n = \{\boldsymbol{\beta} \in \mathbb{N}^n : |\boldsymbol{\beta}| \leq t\}$,

$$(\mathbf{x}, y) \mapsto \Lambda_t^\mu(\mathbf{x}, y) := \sum_{(\boldsymbol{\alpha}, j) \in \mathbb{N}_t^{n+1}} P_{\boldsymbol{\alpha}, j}(\mathbf{x}, y)^2, \quad \forall (\mathbf{x}, y) \in \mathbb{R}^n \times \mathbb{R},$$

where $(P_{\boldsymbol{\alpha}, j})_{(\boldsymbol{\alpha}, j) \in \mathbb{N}_t^{n+1}} \subset \mathbb{R}[\mathbf{x}, y]$ is a family of orthonormal polynomials with respect to (w.r.t.) μ .

Our main result states that Λ_t^μ disintegrates (or factorizes) into

$$\Lambda_t^\mu(\mathbf{x}, y) = \Lambda_t^\phi(\mathbf{x}) \cdot \Lambda_t^{\hat{\nu}_{\mathbf{x}, t}}(y), \quad \forall (\mathbf{x}, y) \in \mathbb{R}^{n+1}, \forall t \in \mathbb{N}, \tag{2}$$

where Λ_t^ϕ (resp. $\Lambda_t^{\hat{\nu}_{\mathbf{x}, t}}$) is the Christoffel function of the marginal ϕ of μ on \mathbf{X} (resp. of some probability measure $\nu_{\mathbf{x}, t}$ on \mathbb{R} , given $\mathbf{x} \in \mathbf{X}$). Moreover, for every fixed $\mathbf{x} \in \mathbf{X}$, one can compute explicitly the Hankel moment matrix of the measure $\nu_{\mathbf{x}, t}$ by solving a single convex optimization problem on positive definite matrices with $\log \det(\cdot)$ as objective function.

Notice how (2) mimics the disintegration (1). Indeed, as we should expect from the disintegration (2), it turns out that for each fixed $\mathbf{x} \in \mathbf{X}$, the family $(\Lambda_t^{\hat{\nu}_{\mathbf{x}, t}})_{t \in \mathbb{N}}$ shares asymptotic properties of the Christoffel function $\Lambda_t^{\hat{\mu}}(y)$ of the conditional probability $\hat{\mu}(dy | \mathbf{x})$ on \mathbf{Y} , given $\mathbf{x} \in \mathbf{X}$.

Actually, the same disintegration (2) holds if the conditioning is multivariate, i.e., on $\mathbf{y} \in \mathbb{R}^p$ given $\mathbf{x} \in \mathbb{R}^n$, with $p > 1$. The only difference is that now $\nu_{\mathbf{x}, t}$ is a linear functional on $\mathbb{R}[\mathbf{y}]_t$ not necessarily represented by a probability measure on \mathbb{R}^p .

(ii). Interestingly, the technique of proof relies on a certain one-to-one mapping between interiors of the convex cone of sum-of-squares polynomials and its dual cone of moment matrices. In particular, and as a by-product, it implies the following simple but apparently unnoticed result that every sum-of-squares polynomial is the reciprocal of a Christoffel function of some linear functional (guaranteed to have a representing measure in the univariate case).

2. Notation, definitions and preliminary results

2.1. Notation and definitions

Let $\mathbb{R}[\mathbf{x}]$ denote the ring of real polynomials in the variables $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbb{R}[\mathbf{x}]_t \subset \mathbb{R}[\mathbf{x}]$ be its subset of polynomials of total degree at most t . Let $\mathbb{N}_t^n := \{\boldsymbol{\alpha} \in \mathbb{N}^n : |\boldsymbol{\alpha}| \leq t\}$ (where $|\boldsymbol{\alpha}| = \sum_i \alpha_i$) with cardinal $s_n(t) = \binom{n+t}{n}$. Let $\mathbf{v}_t(\mathbf{x}) = (\mathbf{x}^\alpha)_{\boldsymbol{\alpha} \in \mathbb{N}_t^n}$ be the vector of monomials up to degree t .

Let $\Sigma[\mathbf{x}]_t \subset \mathbb{R}[\mathbf{x}]_{2t}$ be the convex cone of polynomials of total degree at most $2t$ which are sum-of-squares (in short SOS). For a real symmetric matrix $\mathbf{A} = \mathbf{A}^T$ the notation $\mathbf{A} \geq 0$ (resp. $\mathbf{A} > 0$) stands for \mathbf{A} is positive semidefinite (p.s.d.) (resp. positive definite (p.d.)). The support of a Borel measure μ on \mathbb{R}^n is the smallest closed set A such that $\mu(\mathbb{R}^n \setminus A) = 0$, and such a set A is unique.

Riesz functional. With a real sequence $\boldsymbol{\phi} = (\phi_\alpha)_{\alpha \in \mathbb{N}^n}$ is associated the Riesz linear functional $L_\boldsymbol{\phi} : \mathbb{R}[\mathbf{x}] \rightarrow \mathbb{R}$ defined by:

$$p \left(= \sum_{\alpha \in \mathbb{N}^n} p_\alpha \mathbf{x}^\alpha \right) \mapsto L_\boldsymbol{\phi}(p) := \sum_{\alpha \in \mathbb{N}^n} p_\alpha \phi_\alpha, \quad \forall p \in \mathbb{R}[\mathbf{x}].$$

A sequence $\boldsymbol{\phi} = (\phi_\alpha)_\alpha$ has a representing measure if and only if there exists a Borel measure ϕ on \mathbb{R}^n such that $\int \mathbf{x}^\alpha d\phi = \phi_\alpha$, for all $\alpha \in \mathbb{N}^n$.

Moment matrix. With a real sequence $\phi = (\phi_\alpha)_{\alpha \in \mathbb{N}^n}$ is associated its *moment matrix* $\mathbf{M}_t(\phi)$ of order (or degree) t . It is a real symmetric matrix with rows and columns indexed by \mathbb{N}_t^n , and with entries

$$\mathbf{M}_t(\phi)(\alpha, \beta) := L_\phi(\mathbf{x}^{\alpha+\beta}) = \phi_{\alpha+\beta}, \quad \alpha, \beta \in \mathbb{N}_t^n.$$

Importantly, $\mathbf{M}_t(\phi)$ depends only on moments ϕ_α with $|\alpha| \leq 2t$. If ϕ has a representing measure ϕ then we also write $\mathbf{M}_t(\phi)$ and necessarily $\mathbf{M}_t(\phi)$ is p.s.d. for all t , i.e., $\mathbf{M}_t(\phi) \geq 0$ for all t .

Christoffel function. Let $\phi = (\phi_\alpha)_{\alpha \in \mathbb{N}^n}$ be such that $\mathbf{M}_t(\phi) > 0$ for all t , and let $(P_\alpha)_{\alpha \in \mathbb{N}^n} \subset \mathbb{R}[\mathbf{x}]$ be a family of polynomials, orthonormal with respect to ϕ , i.e.,

$$L_\phi(P_\alpha P_\beta) = \delta_{\alpha=\beta}, \quad \forall \alpha, \beta \in \mathbb{N}^n.$$

Then the Christoffel function (CF) $\Lambda_t^\phi : \mathbb{R}^n \rightarrow \mathbb{R}_+$ associated with ϕ , is defined by

$$\mathbf{x} \mapsto \Lambda_t^\phi(\mathbf{x}) := \left[\sum_{\alpha \in \mathbb{N}_t^n} P_\alpha(\mathbf{x})^2 \right]^{-1}, \quad \forall \mathbf{x} \in \mathbb{R}^n, \tag{3}$$

and recalling that $\mathbf{M}_t(\phi)$ is nonsingular, it turns out that

$$\Lambda_t^\phi(\mathbf{x}) = [\mathbf{v}_t(\mathbf{x})^T \mathbf{M}_t(\phi)^{-1} \mathbf{v}_t(\mathbf{x})]^{-1}, \quad \forall \mathbf{x} \in \mathbb{R}^n. \tag{4}$$

An equivalent and variational definition is also

$$\Lambda_t^\phi(\mathbf{x}) = \inf_{p \in \mathbb{R}[\mathbf{x}]_t} \{L_\phi(p^2) : p(\mathbf{x}) = 1\}, \quad \forall \mathbf{x} \in \mathbb{R}^n. \tag{5}$$

In [2] the authors describe a way to obtain a family of orthonormal polynomials w.r.t. ϕ from the moment matrices $\mathbf{M}_t(\phi) > 0$ via simple determinant calculations. We will use this construction with a special ordering of the monomials that index the rows and columns of $\mathbf{M}_t(\phi)$.

If ϕ has a representing measure ϕ we also write its CF as Λ_t^ϕ . The CF is usually defined for measures ϕ on a compact set Ω rather than for linear functionals ϕ with $\mathbf{M}_t(\phi) > 0$ for all t . In this case one interesting and distinguishing feature of the CF is that as t increases, $\Lambda_t^\phi(\mathbf{x}) \downarrow 0$ exponentially fast for every \mathbf{x} outside the support of ϕ . In other words, Λ_t^ϕ identifies the support of ϕ when t is sufficiently large, a nice property that can be exploited for outlier detection in some data analysis applications; see for instance [6, 7]. In addition, at least in dimension $n = 2$ or $n = 3$, one may visualize this property even for small t , as the resulting superlevel sets $\Omega_\gamma := \{\mathbf{x} : \Lambda_t^\phi(\mathbf{x}) \geq \gamma\}$, $\gamma \in \mathbb{R}$, capture the shape of Ω quite well; see e.g. [5].

2.2. A specific family of orthonormal polynomials

Let μ be a Borel measure on a compact subset of $\mathbf{X} \times \mathbf{Y} \subset \mathbb{R}^n \times \mathbb{R}^p$, and let $\mathbf{M}_t(\mu)$ be the moment matrix of μ with rows and columns indexed by the monomials $(\mathbf{x}^\alpha \mathbf{y}^\beta)_{(\alpha, \beta) \in \mathbb{N}_t^{n+p}}$ listed according to some ordering noted “ \leq ” between monomials, defined as follows. First in the list, we find all monomials $(\mathbf{x}^\alpha)_{\alpha \in \mathbb{N}_t^n}$ (i.e. all monomials $\mathbf{x}^\alpha \mathbf{y}^\beta$ with $|\beta| = 0$) listed e.g. according to the lexicographic ordering. Then we find all monomials $\mathbf{x}^\alpha \mathbf{y}^\beta$ with $|\beta| = 1$, then followed by monomials $\mathbf{x}^\alpha \mathbf{y}^\beta$ with $|\beta| = 2$, etc. Below is displayed $\mathbf{M}_2(\mu)$ in the bivariate case $(n, p) = (1, 1)$.

$$\mathbf{M}_2(\mu) = \begin{bmatrix} \mu_{0,0} & \mu_{1,0} & \mu_{2,0} & \mu_{0,1} & \mu_{1,1} & \mu_{0,2} \\ \mu_{1,0} & \mu_{2,0} & \mu_{3,0} & \mu_{1,1} & \mu_{2,1} & \mu_{1,2} \\ \mu_{2,0} & \mu_{3,0} & \mu_{4,0} & \mu_{2,1} & \mu_{3,1} & \mu_{2,2} \\ \mu_{0,1} & \mu_{1,1} & \mu_{2,1} & \mu_{0,2} & \mu_{1,2} & \mu_{0,3} \\ \mu_{1,1} & \mu_{2,1} & \mu_{3,1} & \mu_{1,2} & \mu_{2,2} & \mu_{1,3} \\ \mu_{0,2} & \mu_{1,2} & \mu_{2,2} & \mu_{0,3} & \mu_{1,3} & \mu_{0,4} \end{bmatrix}. \tag{6}$$

Similarly, let $\mathbf{v}_t(\mathbf{x}, \mathbf{y})$ be the vector of monomials that form a basis of $\mathbb{R}[\mathbf{x}, \mathbf{y}]_t$ listed with the same above ordering “ \leq ”; for instance with $(n, p) = (1, 1)$ and $t = 2$, $\mathbf{v}_2(x, y) = (1, x, x^2, y, xy, y^2)$. Then by (4), the Christoffel function Λ_t^μ is given by

$$(\mathbf{x}, \mathbf{y}) \mapsto \Lambda_t^\mu(\mathbf{x}, \mathbf{y}) := \mathbf{v}_t(\mathbf{x}, \mathbf{y})^T \mathbf{M}_t(\mu)^{-1} \mathbf{v}_t(\mathbf{x}, \mathbf{y}), \quad \forall (\mathbf{x}, \mathbf{y}) \in \mathbb{R}^n \times \mathbb{R}^p. \tag{7}$$

We next see that with ordering “ \leq ” defined above, one may define a certain family of orthonormal polynomials $(P_{\alpha, \beta}) \subset \mathbb{R}[\mathbf{x}, \mathbf{y}]_t$ by following the recipe described in [2] and that we briefly summarize: To compute $P_{\alpha, \beta} \in \mathbb{R}[\mathbf{x}, \mathbf{y}]_t$ one proceeds in three steps:

- From $\mathbf{M}_t(\mu)$ extract its submatrix $S(\alpha, \beta)$ with rows and columns indexed by $(\gamma, \eta) \leq (\alpha, \beta)$.
- Delete the last row and replace it with the monomials $(\mathbf{x}^\gamma \mathbf{y}^\eta)$ with $(\gamma, \eta) \leq (\alpha, \beta)$.
- Define $\tilde{P}_{\alpha, \beta}(\mathbf{x}, \mathbf{y}) := \det(S(\alpha, \beta))$ and then normalize $P_{\alpha, \beta} = \tau \tilde{P}_{\alpha, \beta}$ with $\tau > 0$ such that $\tau^2 \int (\tilde{P}_{\alpha, \beta})^2 d\mu = 1$.

Lemma 1. *With the ordering “ \leq ” and the above construction, the orthonormal polynomials $(P_{\alpha, 0})_{\alpha \in \mathbb{N}_t^n}$ depend only on \mathbf{x} , and are orthonormal w.r.t. the marginal ϕ of μ .*

Proof. In the above construction the orthonormal polynomials $(P_{\alpha, 0})_{\alpha \in \mathbb{N}_t^n}$ are obtained from the submatrices $S(\alpha, 0)$ of $\mathbf{M}_t(\mu)$, $\alpha \in \mathbb{N}_t^n$, which are exactly the submatrices of $\mathbf{M}_t(\phi)$ since they are formed with only monomials \mathbf{x}^α (as $|\beta| = 0$). Hence the conclusion follows. \square

For illustration purpose, with $\mathbf{M}_2(\mu)$ as in (6),

$$\tilde{P}_{0,0}(x, y) = \mu_{0,0}; \quad \tilde{P}_{1,0}(x, y) = \det \left(\begin{bmatrix} \mu_{0,0} & \mu_{1,0} \\ 1 & x \end{bmatrix} \right); \quad \tilde{P}_{2,0}(x, y) = \det \begin{bmatrix} \mu_{0,0} & \mu_{1,0} & \mu_{2,0} \\ \mu_{1,0} & \mu_{2,0} & \mu_{3,0} \\ 1 & x & x^2 \end{bmatrix},$$

so that

$$\tilde{P}_{1,0}(x, y) = \mu_{0,0}x - \mu_{1,0}; \quad \tilde{P}_{2,0}(x, y) = (\mu_{0,0}\mu_{2,0} - \mu_{1,0}^2)x^2 - (\mu_{0,0}\mu_{3,0} - \mu_{1,0}\mu_{2,0})x + (\mu_{1,0}\mu_{3,0} - \mu_{2,0}^2),$$

Corollary 2. *Let μ be a Borel measure on $\mathbf{X} \times \mathbf{Y}$ with marginal ϕ on \mathbf{X} , and assume that $\mathbf{M}_t(\mu) > 0$ for all t . Let $(P_{\alpha, \beta})$, $(\alpha, \beta) \in \mathbb{N}_t^{n+p}$, be the family of orthonormal polynomials defined in Section 2.2, and let Λ_t^μ be as in (7). Then:*

$$\Lambda_t^\mu(\mathbf{x}, \mathbf{y})^{-1} = \Lambda_t^\phi(\mathbf{x})^{-1} + \sum_{(\alpha, \beta) \in \mathbb{N}_t^{n+p}, |\beta| \geq 1} P_{\alpha, \beta}(\mathbf{x}, \mathbf{y})^2. \tag{8}$$

Proof. By Lemma 1, the polynomials $(P_{\alpha, 0})_{\alpha \in \mathbb{N}_t^n}$ depend on \mathbf{x} only and are orthonormal w.r.t. ϕ . Therefore by (3):

$$\sum_{\alpha \in \mathbb{N}_t^n} P_{\alpha, 0}(\mathbf{x})^2 = \Lambda_t^\phi(\mathbf{x})^{-1}, \quad \forall \mathbf{x} \in \mathbb{R}^n, \quad t \in \mathbb{N}.$$

Then the result follows from

$$\begin{aligned} \Lambda_t^\mu(\mathbf{x}, \mathbf{y})^{-1} &= \sum_{(\alpha, \beta) \in \mathbb{N}_t^{n+p}} P_{\alpha, \beta}(\mathbf{x}, \mathbf{y})^2 = \sum_{\alpha \in \mathbb{N}_t^n, \beta=0} P_{\alpha, 0}(\mathbf{x}, \mathbf{y})^2 + \sum_{(\alpha, \beta) \in \mathbb{N}_t^{n+p}, |\beta| \geq 1} P_{\alpha, \beta}(\mathbf{x}, \mathbf{y})^2 \\ &= \Lambda_t^\phi(\mathbf{x})^{-1} + \sum_{(\alpha, \beta) \in \mathbb{N}_t^{n+p}, |\beta| \geq 1} P_{\alpha, \beta}(\mathbf{x}, \mathbf{y})^2. \quad \square \end{aligned}$$

2.3. Positive polynomials and Christoffel functions

Recall that $p \in \Sigma[\mathbf{x}]_t$ (i.e., p is an SOS of degree at most $2t$) if and only if there exists a real symmetric matrix $\mathbf{Q} \geq 0$ such that $p(\mathbf{x}) = \mathbf{v}_t(\mathbf{x})^T \mathbf{Q} \mathbf{v}_t(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$. Notice that except when $t = 1$, there are several possible choices for \mathbf{Q} which is called a Gram matrix of p . As we next see, one choice is particularly interesting. The dual cone Σ_t^* is the convex cone characterized by:

$$\Sigma_t^* = \{ \phi \in \mathbb{N}_{2t}^n : \mathbf{M}_t(\phi) \geq 0 \}.$$

Lemma 3.

- (i) Every SOS polynomial in the interior of Σ_t is the reciprocal of the Christoffel function Λ_t^ϕ of some linear functional L_ϕ , with $\phi \in \text{int}(\Sigma_t^*)$. That is, $p \in \text{int}(\Sigma_t)$ if and only if

$$p(\mathbf{x}) = \mathbf{v}_t(\mathbf{x})^T \mathbf{M}_t(\phi)^{-1} \mathbf{v}_t(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n,$$

for some moment sequence $\phi \in \text{int}(\Sigma_t^*)$. Therefore p^{-1} is the Christoffel function associated with some linear functional ϕ (with not necessarily a representing measure ϕ).

- (ii) In addition, if $p \in \text{int}(\Sigma_t)$ is univariate then ϕ has a representing measure ϕ , and so p^{-1} is the Christoffel function Λ_t^ϕ of some measure ϕ on the real line.

Proof. The first part of the statement is a direct consequence from Nesterov [8, Theorem 17.3, p. 412] which states that the respective interiors of Σ_t and its dual Σ_t^* are in one-to-one correspondence, and $-\log \det(\mathbf{A})$ is a τ -self-concordant barrier function associated with the convex cone Σ_t , with $\tau = \binom{n+t}{t}$. The second statement follows from the characterization (4) of Λ_t^ϕ . \square

Surprisingly, the fact that every (strictly positive) SOS polynomial of degree at most $2t$, is the Christoffel function Λ_t^ϕ of some linear functional L_ϕ on $\mathbb{R}[\mathbf{x}]_{2t}$ with $\mathbf{M}_t(\phi) > 0$, does not seem to have been noticed before, even though Nesterov’s result [8, Theorem 17.3] is quite classical in convex conic optimization. In addition, observe that Lemma 3 is the degree- t analogue of the well-known fact that the Gram matrix of every positive quadratic form is the covariance of a Gaussian measure (possibly after scaling). Finally, and said differently, the inverse of moment matrices $\mathbf{M}_t(\phi)$ associated with the Christoffel functions Λ_t^ϕ are crucial to characterize the *central path*¹ in optimization over the convex cone $(\Sigma_t)^*$ of moment matrices.

We even have a similar result in a more general context. With $g_j \in \mathbb{R}[\mathbf{x}]_{d_j}$, let $\mathbf{g}_j = (g_j \alpha)_{\alpha \in \mathbb{N}_{d_j}^n}$ denote its vector of coefficients, $j = 0, \dots, m$. Given a real sequence $\phi = (\phi_\alpha)_{\alpha \in \mathbb{N}^n}$, define the new sequences $g_j \cdot \phi := (\phi_{j,\alpha})_{\alpha \in \mathbb{N}^n}$, where for each $j = 0, \dots, m$,

$$\phi_{j,\alpha} := \sum_{\beta \in \mathbb{N}_{d_j}^n} g_j \beta \phi_{\alpha+\beta}, \quad \forall \alpha \in \mathbb{N}^n.$$

Lemma 4. With $g_j \in \mathbb{R}[\mathbf{x}]_{d_j}$ (and $g_0 = 1$), let $s_j := \lceil \deg(g_j)/2 \rceil$, $j = 0, \dots, m$, and for every $t \geq \max_j s_j$, let $\mathbf{K}_t \subset \mathbb{R}[\mathbf{x}]$ be the convex cone defined by:

$$\mathbf{K}_t := \left\{ \sum_{j=0}^m \sigma_j(\mathbf{x}) g_j(\mathbf{x}) : \sigma_j \in \Sigma_{t-s_j} \right\} \subset \mathbb{R}[\mathbf{x}]_{2t}. \tag{9}$$

If $p \in \text{int}(\mathbf{K}_t)$ then $p \geq 0$ on $S := \{\mathbf{x} : g_j(\mathbf{x}) \geq 0, j = 1, \dots, m\} \subset \mathbb{R}^n$ and

$$p(\mathbf{x}) = \sum_{j=0}^m \Lambda_{t-s_j}^{g_j \cdot \phi}(\mathbf{x})^{-1} g_j(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n \tag{10}$$

for some linear functional $\phi = (\phi_\alpha)_{\alpha \in \mathbb{N}_{2t}^n}$ that satisfies $\mathbf{M}_{t-s_j}(g_j \cdot \phi) > 0$ for all $j = 0, \dots, m$.

Proof. The proof is in the same spirit and again relies on the one-to-one mapping between the interior of the convex cone \mathbf{K}_t and that of its dual

$$\mathbf{K}_t^* = \{ \phi \in \mathbb{N}_{2t}^n : \mathbf{M}_{t-s_j}(g_j \cdot \phi) \geq 0, j = 0, \dots, m \},$$

described in [8, Theorem 17.6 (2), p. 416]. Translated in our notation, (17.9) in Nesterov [8, Theorem 17.7, p. 417] reads,

$$p(\mathbf{x}) = \sum_{j=0}^m \mathbf{v}_{t-s_j}(\mathbf{x})^T \mathbf{M}_{t-s_j}(g_j \cdot \phi)^{-1} \mathbf{v}_{t-s_j}(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n, \quad \text{for some } \phi \in \text{int}(\mathbf{K}_t^*). \tag{11} \quad \square$$

¹In convex optimization, the central path associated with a convex cone \mathbf{K} , plays a central role in the analysis of the computational complexity of interior points methods for optimizing over such a cone.

Again the inverse of moment matrices $\mathbf{M}_{t-s_j}(g_j \cdot \phi)$ associated with the Christoffel functions $\Lambda_{t-s_j}^{g_j \cdot \phi}$ are crucial to characterize the central path in optimization over the dual cone of \mathbf{K}_t in (9). For compact set S (with an additional Archimedean assumption), the cone \mathbf{K}_t is very important in the Moment-SOS hierarchy for polynomial optimization [4]. It is used to replace the intractable positivity constraint “ $p \geq 0$ ” on S , with the more restrictive constraint “ $p \in \mathbf{K}_t$ ” (and let t increase) because the latter being semidefinite representable, is tractable.

3. Main result

Let μ be a Borel measure on a compact set $\Omega \subset \mathbf{X} \times \mathbf{Y} \subset \mathbb{R}^n \times \mathbb{R}$ which disintegrates into its marginal ϕ on $\mathbf{X} \subset \mathbb{R}^n$ and its conditional probability $\hat{\mu}(dy | \mathbf{x})$ on $\mathbf{Y}_{\mathbf{x}} \subset \mathbf{Y}$ for every $\mathbf{x} \in \mathbf{X}$. Throughout the rest of the paper we assume that Ω has nonempty interior so that $\mathbf{M}_t(\mu) > 0$ for all $t \in \mathbb{N}$, where $\hat{\mu}_t(\mu)$ is constructed as in Section 2.2.

Theorem 5. *Let Λ_t^μ be as in (7) with $\mathbf{M}_t(\mu)$ constructed as indicated just above (7). Then for every $\mathbf{x} \in \mathbf{X}$ and $t \in \mathbb{N}$, there exists a probability measure $\nu_{\mathbf{x},t}$ on \mathbb{R} such that*

$$\Lambda_t^\mu(\mathbf{x}, y) = \Lambda_t^\phi(\mathbf{x}) \cdot \Lambda_t^{\nu_{\mathbf{x},t}}(y), \quad \forall \mathbf{x} \in \mathbf{X}, y \in \mathbb{R}. \tag{11}$$

Proof. Let $t \in \mathbb{N}$ and $\mathbf{x} \in \mathbb{R}^n$ be fixed. From (8) in Corollary 2 and as $p = 1$,

$$\begin{aligned} \frac{\Lambda_t^\mu(\mathbf{x}, y)^{-1}}{\Lambda_t^\phi(\mathbf{x})^{-1}} &= 1 + \Lambda_t^\phi(\mathbf{x}) \left[\sum_{(\alpha,j) \in \mathbb{N}_t^{n+1}, 1 \leq j \leq t} P_{\alpha,j}(\mathbf{x}, y)^2 \right] \\ &=: p_t(y; \mathbf{x}) \in \mathbb{R}[y]_{2t}. \end{aligned}$$

Hence for each fixed $\mathbf{x} \in \mathbb{R}^n$, $1 \leq p_t(y; \mathbf{x}) \in \mathbb{R}[y]$ is a strictly positive univariate SOS of degree $2t$ with a non-zero coefficient for y^{2t} ; hence $p_t \in \text{int}(\Sigma_t)$. Therefore by Lemma 3 (ii) there exists a Borel measure $\nu_{\mathbf{x},t}$ on \mathbb{R} such that $p_t(y; \mathbf{x})^{-1} = \Lambda_t^{\nu_{\mathbf{x},t}}(y)$, which yields (11). \square

When $\mathbf{x} \in \mathbf{X}$, notice how well (11) mimics the disintegration (1) of μ into its marginal ϕ on \mathbf{X} and its conditional $\hat{\mu}(dy | \mathbf{x})$ on $\mathbf{Y}_{\mathbf{x}}$, given $\mathbf{x} \in \mathbf{X}$. However when $\mathbf{x} \in \mathbf{X}$, it remains to relate the family of measures $(\nu_{\mathbf{x},t})_{t \in \mathbb{N}}$ on $\mathbf{Y}_{\mathbf{x}}$ with the conditional probability $\hat{\mu}(dy | \mathbf{x})$.

Computing the moment matrix of $\nu_{\mathbf{x},t}$. To obtain the moment matrix of $\nu_{\mathbf{x},t}$, for an arbitrary but fixed $\mathbf{x} \in \mathbb{R}^n$, is relatively easy. Let S_t be the space of $(t + 1) \times (t + 1)$ real symmetric matrices.

- Compute the polynomial $y \mapsto p_t(y; \mathbf{x}) := \Lambda_t^\mu(\mathbf{x}, y)^{-1} / \Lambda_t^\phi(\mathbf{x})^{-1}$ which is an SOS in “ y ” of degree $2t$. This is easy once moments of μ are available. Indeed one computes $\Lambda_t^\mu(\mathbf{x}, y)$ (resp. $\Lambda_t^\phi(\mathbf{x})$) via (4) with the moment matrix $\mathbf{M}_t(\mu)$ and $\mathbf{v}_t(\mathbf{x}, y)$ (resp. $\mathbf{M}_t(\phi)$ and $\mathbf{v}_t(\mathbf{x})$).
- Then following [8, p. 412], solve the convex optimization problem

$$\mathbf{M}_t(\nu_{\mathbf{x},t}) = \arg \min_{0 < \mathbf{Q} \in S_t} \{-\log \det(\mathbf{Q}) : p_t(y; \mathbf{x}) = \mathbf{v}_t(y)^T \mathbf{Q} \mathbf{v}_t(y), \forall y\}. \tag{12}$$

The optimization problem (12) is convex and can be solved by off-the-shelf solvers like e.g. CVX [1].

Multivariate conditional. If $p > 1$ and $\mathbf{Y} \subset \mathbb{R}^p$, then we still obtain the decomposition (11)

$$\Lambda_t^\mu(\mathbf{x}, \mathbf{y}) = \Lambda_t^\phi(\mathbf{x}) \cdot \Lambda_t^{\nu_{\mathbf{x},t}}(\mathbf{y}), \quad \forall \mathbf{x} \in \mathbf{X}, \mathbf{y} \in \mathbb{R}^p, \tag{13}$$

with exactly the same proof as that of Theorem 5. The difference with (11) is that $\nu_{\mathbf{x},t}$ in (13) is a linear functional on $\mathbb{R}[\mathbf{y}]_t$ which is not guaranteed to have a representing measure $\nu_{\mathbf{x},t}$ on \mathbb{R}^p .

3.1. Discussion

Define the scalar $s_n(t) := \binom{n+t}{t}$ for every integer t, n . Under some conditions on the sets Ω and $\mathbf{X}, \mathbf{Y}_{\mathbf{x}}$ and if μ has a density w.r.t. Lebesgue measure on Ω that also satisfies some conditions, then one may indeed relate the family $(v_{\mathbf{x},t})_{t \in \mathbb{N}}$ with the conditional probability $\widehat{\mu}(dy|\mathbf{x})$ on \mathbf{Y} , given $\mathbf{x} \in \mathbf{X}$. Under such conditions one may interpret the limit $s_{n+1}(t)\Lambda_t^\mu(\mathbf{x}, y)$ and $s_n(t)\Lambda_t^\phi(\mathbf{x})$, as t increases, in terms of the density of μ and an *equilibrium* measure intrinsically related to the respective supports Ω and \mathbf{X} . For such conditions the interested reader is referred to [3, 7] and the many references therein. For instance:

Corollary 6. *Let $\Omega = \mathbf{X} \times \mathbf{Y} \subset \mathbb{R}^{n+1}$ be compact with $\Omega = \overline{\text{int}(\Omega)}$, $\mathbf{X} = \overline{\text{int}(\mathbf{X})}$, and assume that μ has a density f w.r.t. Lebesgue on \mathbb{R}^{n+1} , bounded away from 0 on Ω .*

If $\mathbf{x} \in \text{int}(\mathbf{X})$ but $(\mathbf{x}, y) \notin \Omega$, then as t increases, $\Lambda_t^{v_{\mathbf{x},t}}(y) \downarrow 0$ exponentially fast (as would do the Christoffel function $\Lambda_t^\mu(y)$ of the conditional probability $\widehat{\mu}(dy|\mathbf{x})$).

Proof. By [6, 7], as $(\mathbf{x}, y) \notin \Omega$, $\Lambda_t^\mu(\mathbf{x}, y) \downarrow 0$ exponentially fast as t increases. On the other hand, as $\mathbf{x} \in \text{int}(\mathbf{X})$ and the density of ϕ w.r.t. Lebesgue on \mathbb{R}^n is bounded away from zero, $\Lambda_t^\phi(\mathbf{x})^{-1}$ increases with t not faster than $O(t^n)$. Therefore by (11), $\Lambda_t^{v_{\mathbf{x},t}}(y)^{-1}$ has to grow exponentially fast with t . The same conclusion holds for $\widehat{\mu}(dy|\mathbf{x})$; indeed let y be outside the support $\mathbf{Y}_{\mathbf{x}}$ of $\widehat{\mu}(dy|\mathbf{x})$. The density of $\widehat{\mu}(dy|\mathbf{x})$ which reads $y \mapsto f(\mathbf{x}, y) / \int_{\mathbf{Y}} f(\mathbf{x}, y) \widehat{\mu}(dy|\mathbf{x})$ on $\mathbf{Y}_{\mathbf{x}}$, is bounded away from zero. Therefore $\Lambda_t^\mu(y) \downarrow 0$ exponentially fast as t increases. \square

So Corollary 6 states that whenever $\mathbf{x} \in \mathbf{X}$ and $y \notin \text{supp}(\widehat{\mu}(dy|\mathbf{x}))$, then asymptotically the growth rate of $\Lambda_t^{v_{\mathbf{x},t}}(y)^{-1}$ is exponential as for the CF of the conditional probability $\widehat{\mu}(dy|\mathbf{x})$. To obtain precise asymptotic results when $(\mathbf{x}, y) \in \Omega$, additional conditions on μ are required. Below is such a typical result.

Lemma 7 (Kroó and Lubinsky [3]). *Let $S \subset \mathbb{R}^n$ be compact and assume that there exists a measure ψ_0 supported on S such that uniformly on compact subsets of $\text{int}(S)$, $\lim_{t \rightarrow \infty} s_n(t)\Lambda_t^{\psi_0}(\mathbf{x}) = W_0(\mathbf{x})$ where W_0 is continuous and positive on $\text{int}(S)$.*

If a measure ψ has continuous and positive density D w.r.t. ψ_0 on $\text{int}(S)$, then uniformly on compact subsets of $\text{int}(S)$, $\lim_{t \rightarrow \infty} s_n(t)\Lambda_t^\psi(\mathbf{x}) = D(\mathbf{x})W_0(\mathbf{x})$.

Given a compact set \mathcal{X} , let $\mathcal{C}(\mathcal{X})$ denote the space of continuous functions on \mathcal{X} . In our context of μ on a compact set $\Omega \subset \mathbf{X} \times \mathbb{R}$ with marginal ϕ on \mathbf{X} , we obtain the following result:

Theorem 8. *Assume that there exists a measure μ_0 on Ω with marginal ϕ_0 on \mathbf{X} and conditional $\widehat{\mu}_0(dy|\mathbf{x})$ on \mathbb{R} , such that uniformly on compact subsets of Ω (resp. \mathbf{X}):*

$$\lim_{t \rightarrow \infty} s_{n+1}(t)\Lambda_t^{\mu_0}(\mathbf{x}, y) = W_0(\mathbf{x}, y); \quad \lim_{t \rightarrow \infty} s_n(t)\Lambda_t^{\phi_0}(\mathbf{x}) = W_0'(\mathbf{x}).$$

In addition assume that the following Feller-type property holds:

$$\mathbf{x} \mapsto \int h(\mathbf{x}, y) \widehat{\mu}_0(dy|\mathbf{x}) \in \mathcal{C}(\mathbf{X}) \quad \text{whenever } h \in \mathcal{C}(\Omega).$$

Let μ be a measure on Ω with a continuous and positive density f w.r.t. μ_0 . Then, with $v_{\mathbf{x},t}$ being the measure on \mathbb{R} in Theorem 5:

$$\lim_{t \rightarrow \infty} t \Lambda_t^{v_{\mathbf{x},t}}(y) = (n+1) \frac{f(\mathbf{x}, y)}{g(\mathbf{x})} \frac{W_0(\mathbf{x}, y)}{W_0'(\mathbf{x})}, \quad \forall (\mathbf{x}, y) \in \text{int}(\Omega), \tag{14}$$

where $g(\mathbf{x}) := \int f(\mathbf{x}, y) \widehat{\mu}_0(dy|\mathbf{x})$.

Observe that for all $\mathbf{x} \in \text{int}(\mathbf{X})$, $f(\mathbf{x}, y) / g(\mathbf{x})$ is the density of $\widehat{\mu}(dy|\mathbf{x})$ w.r.t. $\widehat{\mu}_0(dy|\mathbf{x})$.

Proof. Disintegrating μ_0 yields $d\mu_0(\mathbf{x}, y) = \widehat{\mu}_0(dy|\mathbf{x}) \phi_0(d\mathbf{x})$. Therefore

$$d\mu(\mathbf{x}, y) = \frac{f(\mathbf{x}, y) \widehat{\mu}_0(dy|\mathbf{x})}{g(\mathbf{x})} g(\mathbf{x}) \phi_0(d\mathbf{x}); \quad g(\mathbf{x}) := \int f(\mathbf{x}, y) \widehat{\mu}_0(dy|\mathbf{x}),$$

and $\phi(d\mathbf{x}) = g(\mathbf{x}) \phi_0(d\mathbf{x})$. Moreover observe that for every $\mathbf{x} \in \mathbf{X}$,

$$\widehat{\mu}(dy|\mathbf{x}) = \frac{f(\mathbf{x}, y)}{g(\mathbf{x})} \widehat{\mu}_0(dy|\mathbf{x}).$$

That is, for every $\mathbf{x} \in \mathbf{X}$, $y \mapsto f(\mathbf{x}, y)/g(\mathbf{x})$ is the density of $\widehat{\mu}(dy|\mathbf{x})$ w.r.t. $\widehat{\mu}_0(dy|\mathbf{x})$, and by the Feller-like property, g is continuous and positive on \mathbf{X} . Next, by our hypotheses and from Theorem 5,

$$\begin{aligned} f(\mathbf{x}, y) W_0(\mathbf{x}, y) &= \lim_{t \rightarrow \infty} s_{n+1}(t) \Lambda_t^\mu(\mathbf{x}, y) = \lim_{t \rightarrow \infty} \left[s_n(t) (\Lambda_t^\phi(\mathbf{x}) \cdot \frac{s_{n+1}(t)}{s_n(t)} \Lambda_t^{v_{\mathbf{x},t}}(y)) \right] \\ &= g(\mathbf{x}) W_0'(\mathbf{x}) \lim_{t \rightarrow \infty} \frac{s_{n+1}(t)}{s_n(t)} \Lambda_t^{v_{\mathbf{x},t}}(y) = \frac{g(\mathbf{x}) W_0'(\mathbf{x})}{n+1} \lim_{t \rightarrow \infty} t \Lambda_t^{v_{\mathbf{x},t}}(y), \end{aligned}$$

for all $(\mathbf{x}, y) \in \text{int}(\Omega)$, which yields (14). □

So for instance with $n = 1 = p$, let $d\mu_0(x, y) = 1_{[-1,1]}(x) 1_{[a(x),b(x)]}(y) dx dy$ where $x \mapsto a(x)$ and $x \mapsto b(x)$ are strictly positive and continuous on $[-1, 1]$, and $b(x) - a(x) > 0$ on $\mathbf{X} = [-1, 1]$. Then (14) reads

$$\lim_{t \rightarrow \infty} t \Lambda_t^{v_{\mathbf{x},t}}(y) = \frac{f(x, y)}{\int_{a(x)}^{b(x)} f(x, y) dy} \frac{2 W_0(x, y)}{W_0'(x)}, \quad \forall (x, y) \in \text{int}(\Omega),$$

and $f(x, y)/\int_{a(x)}^{b(x)} f(x, y) dy$ is the density of $\widehat{\mu}(dy|x)$ w.r.t. Lebesgue on the interval $[a(x), b(x)]$, for every $\mathbf{x} \in [-1, 1]$.

As expected from the disintegration (11), convergence of $t \Lambda_t^{v_{\mathbf{x},t}}(y)$ as t increases, is towards the density of the conditional $\widehat{\mu}(dy|x)$ times a weight function intrinsic to the support Ω of μ , which is typical of convergence results for Christoffel functions (whenever convergence takes place).

4. Conclusion

We have shown that in quite general setup, the Christoffel function disintegrates (or factorizes) and mimics the disintegration of its associated measure on $\mathbf{X} \times \mathbf{Y}$ into its marginal on \mathbf{X} and its conditional on \mathbf{Y} , given $\mathbf{x} \in \mathbf{X}$. The result uses a straightforward (but novel) interpretation of a well-known intermediate result of convex optimization, which is of interest in its own. Namely that every SOS polynomial is the reciprocal of the Christoffel function associated with some linear functional (which always has a representing measure in the univariate case). A similar interpretation is valid for the cone of polynomials that are positive on a basic semi-algebraic set.

We think that a better understanding of the linear functional $v_{\mathbf{x},t}$ (which has a representing measure when $p = 1$) is needed. In particular, further investigation beyond the scope of the present note, could consider a more detailed (and non-asymptotic) comparison of $v_{\mathbf{x},t}$ with the conditional $\widehat{\mu}(dy|\mathbf{x})$ when $\mathbf{x} \in \mathbf{X}$, as well as understanding its meaning when $p > 1$, i.e., when it may not have a representing measure. For instance we conjecture (but have been unable to prove) that $v_{\mathbf{x},t}$ does *not* depend on t , and has a representing measure.

References

- [1] M. Grant, S. Boyd, "CVX: Matlab Software for Disciplined Convex Programming, version 2.1", 2014, <http://cvxr.com/cvx>.
- [2] J. W. Helton, J. B. Lasserre, M. Putinar, "Measures with zeros in the inverse of their moment matrix", *Ann. Probab.* **36** (2008), no. 4, p. 1453-1471.
- [3] A. Kroó, D. S. Lubinsky, "Christoffel functions and universality in the bulk for multivariate orthogonal polynomials", *Can. J. Math.* **65** (2013), no. 3, p. 600-620.
- [4] J. B. Lasserre, *Introduction to Polynomial and Semi-Algebraic Optimization*, Cambridge Texts in Applied Mathematics, Cambridge University Press, 2015.

- [5] J. B. Lasserre, E. Pauwels, "Sorting out typicality via the inverse moment matrix SOS polynomial", in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2016, p. 190-198.
- [6] ———, "The empirical Christoffel function with applications in data analysis", *Adv. Comput. Math.* **45** (2019), no. 3, p. 1439-1468.
- [7] J. B. Lasserre, E. Pauwels, M. Putinar, *The Christoffel-darboux Kernel for Data Analysis*, Cambridge Monographs on Applied and Computational Mathematics, vol. 38, Cambridge University Press, 2022.
- [8] Y. Nesterov, "Squared functional systems and optimization problems", in *High Performance Optimization*, Applied Optimization, vol. 33, Springer, 2000, p. 405-440.