



HAL
open science

Spectral analysis of moment data

Edouard Pauwels, Mihai Putinar, Jean-Bernard Lasserre

► **To cite this version:**

Edouard Pauwels, Mihai Putinar, Jean-Bernard Lasserre. Spectral analysis of moment data. 2018.
hal-01845137v1

HAL Id: hal-01845137

<https://hal.science/hal-01845137v1>

Preprint submitted on 20 Jul 2018 (v1), last revised 2 Jan 2020 (v3)

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.

Spectral analysis of moment data

Edouard Pauwels *Mihai Putinar †Jean-Bernard Lasserre ‡

July 20, 2018

Abstract

Spectral features of the empirical moment matrix constitute a resourceful tool for unveiling properties of a cloud of points, among which, density, support and latent structures. It is already well known that the empirical moment matrix encodes a great deal of subtle attributes of the underlying measure. Starting from this object as base of observations we combine ideas from statistics, real algebraic geometry, orthogonal polynomials and approximation theory for opening new insights relevant for Machine Learning (ML) problems with data supported on singular sets.

Refined concepts and results from real algebraic geometry and approximation theory are empowering a *simple* tool (the empirical moment matrix) for the task of solving non-trivial questions in data analysis. We provide (1) theoretical validation, (2) numerical experiments and, (3) connections to real world data as a validation of the stamina of the empirical moment matrix approach.

Inference of low dimensional structures

Intrinsic dimension has a long history in signal processing expressing the idea that most empirical high dimensional signals are actually structured and can be approximated by a small number of entities [6, 24, 45, 52]. Dimension reduction in data analysis has witnessed a considerable renewal of interest in the early 2000's with the advent of non linear low dimensional structure estimation algorithms [46, 50, 4] and follow up works [16, 34, 20]. Graph Laplacian based methods were proposed in [5, 29] and intrinsic dimension estimation was revisited in [35, 28]. In a manifold learning context the questions of finite sample efficiency were treated in [26, 25, 31, 1] and a statistical test was proposed [23]. A spectral support estimator was described in [54].

Getting access to topological properties of a data distribution through computational topology is of rising interest [43, 19, 11, 27]. This was cast in a statistical framework in [12, 13, 10] and in a machine learning framework by [43, 44].

Classical real algebraic geometry [7] and its related computational tools [14] constitute a well established mathematical branch dealing with varieties described by polynomials. Our work streams from the potential of exploiting these results with the specific aim at inferring latent structures. A fundamental view in modern algebraic geometry is to study a finite

dimensional set S via duality, by investigating the algebra of polynomial functions acting on S . We put at work the idea to capture geometric characteristics of S from polynomials on a cloud of points spread across S , essentially a spectral property of the corresponding empirical moment matrix.

Christoffel-Darboux kernel

Given a positive, absolutely continuous, rapidly decaying Borel measure μ on euclidean space, integration with respect to μ defines a scalar product on the space of polynomials. The reproducing kernel associated to the Hilbert space of polynomials up to a given degree is called the Christoffel-Darboux kernel [48] and is usually computed from a suitable orthonormal family [49, 18]. This kernel only depends on the moments of μ and has been intensively studied for more than a century. It captures refined properties of the original measure [42, 53] and allows to decode information concerning the support and the density of μ . The univariate case was treated in [40, 41, 51], an application in an ergodic context is given in [32].

In the multivariate setting, explicit computations is known only for simple sets [8, 55, 9, 56] or under abstract assumptions [33]. As claimed in [37, 38], these tools constitute a promising research direction for data analysis. For latent structure inference, it is required to understand the properties of these objects in the *singular* setting, which is hardly tackled in the literature.

Results

Under natural assumptions, we show that the space of polynomials on a real algebraic set and the space of polynomials on a large enough generic sample on the same set are essentially the same spaces. As a result, global geometric properties of the support can be inferred *with probability one* from only finite samples. This relies on the rigidity of the polynomial setting, and constitutes a significant departure from more classical forms of statistical inference which most often hold non deterministically. We describe how the growth of the rank *empirical* moment matrix is related to the dimension of the support of the underlying measure.

Our second main result extends the weighted asymptotic convergence of the Christoffel function to the density of the underlying measure, cf. [33]. The work of [33] relies on the variational formulation of the Christoffel function, still valid in the singular case. Starting from simple assumptions, we cover the cases of the sphere, the ellipsoids and canonical operations of them, such as fibre products. The proof provides

*IRIT, AOC, Université Toulouse 3 Paul Sabatier, 118 route de Narbonne, 31062 Toulouse, France.

†University of California, Santa Barbara, Newcastle University, Newcastle upon Tyne, UK.

‡LAAS-CNRS and Institute of Mathematics, University of Toulouse, LAAS, 7 avenue du Colonel Roche 31077 Toulouse, France.

explicit convergence rates in supremum norm. All proof arguments are postponed to the appendix.

Numerical experiments

We illustrate on simulations that the growth of the rank of the empirical moment matrix is related to the dimensionality of the support of the underlying measure. Connection with the density is illustrated on three real world datasets featuring symmetry and periodicity. We map them to higher dimensional algebraic sets which captures naturally the corresponding symmetries and deploy the Christoffel function machinery illustrating how this simple tool can be used to estimate densities on the circle, the sphere and the bi-torus. Experiment details are found in the appendix.

1 The absolutely continuous case

Henceforth we fix the dimension of the ambient euclidean space to be p . For example, we will consider vectors in \mathbb{R}^p as well as p -variate polynomials with real coefficients.

1.1 Notations and definition

We denote by $\mathbb{R}[X]$ the algebra of p -variate polynomials with real coefficients. For $d \in \mathbb{N}$, $\mathbb{R}_d[X]$ stands for the set of p -variate polynomials of degree less than or equal to d . We set $s(d) = \binom{p+d}{d} = \dim \mathbb{R}_d[X]$.

Let $\mathbf{v}_d(X)$ denotes the vector of monomials of degree less than or equal to d . A polynomial $P \in \mathbb{R}_d[X]$ can be written as $P(X) = \langle \mathbf{p}, \mathbf{v}_d(X) \rangle$ for a given $\mathbf{p} \in \mathbb{R}^{s(d)}$. For a positive Borel measure μ on \mathbb{R}^p denote by $\text{supp}(\mu)$ its support, i.e., the smallest closed subset whose complement has measure zero. We will only consider rapidly decaying measures, that is measures whose all moments are finite.

Moment matrix For any $d \in \mathbb{N}$, the moment matrix of μ , with moment up to $2d$, is given by

$$M_{\mu,d} = \int_{\mathbb{R}^p} \mathbf{v}_d(\mathbf{x}) \mathbf{v}_d(\mathbf{x})^T d\mu(\mathbf{x})$$

where the integral is understood element-wise. Actually it is useful to interpret the moment matrix as representing the bilinear form on $\mathbb{R}[X]$, $\langle \cdot, \cdot \rangle_\mu : (P, Q) \mapsto \int PQ d\mu$, restricted to polynomials of degree up to d . Indeed, if $\mathbf{p}, \mathbf{q} \in \mathbb{R}^{s(d)}$ are the vectors of coefficients of any two polynomials P and Q of degree up to d , one has $\mathbf{p}^T M_{\mu,d} \mathbf{q} = \langle P, Q \rangle_\mu$. This entails that $M_{\mu,d}$ is positive semidefinite for all $d \in \mathbb{N}$.

1.2 Christoffel-Darboux kernel and Christoffel function

In this section we assume that the probability measure μ is absolutely continuous with respect to Lebesgue measure on \mathbb{R}^p . Fix d a positive integer. In this case, the bilinear form $\langle \cdot, \cdot \rangle_\mu$ is positive definite on $\mathbb{R}_d[X]$ which is a finite dimensional Hilbert space; we denote by $\| \cdot \|_\mu$ the corresponding norm. For any $\mathbf{x} \in \mathbb{R}^p$, the evaluation functional on $\mathbb{R}_d[X]$, $P \mapsto P(\mathbf{x})$, is continuous with respect to $\| \cdot \|_\mu$. Hence $\mathbb{R}_d[X]$

is a Reproducing Kernel Hilbert Space (RKHS), it admits a unique reproducing kernel, $\kappa_{\mu,d}$ [2], called the *Christoffel-Darboux kernel* [48]. The reproducing property and definition of $M_{\mu,d}$ ensure that for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$,

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

The *Christoffel function* is defined as follows:

$$\Lambda_{\mu,d} : \mathbf{x} \mapsto \inf_{P \in \mathbb{R}_d[X]} \left\{ \int P(\mathbf{z})^2 d\mu(\mathbf{z}), \quad \text{s.t.} \quad P(\mathbf{x}) = 1 \right\} \quad (2)$$

A well known crucial link between the Christoffel-Darboux kernel and the Christoffel function is given for any $\mathbf{x} \in \mathbb{R}^p$,

$$\kappa_{\mu,d}(\mathbf{x}, \mathbf{x}) \Lambda_{\mu,d}(\mathbf{x}) = 1. \quad (3)$$

2 The singular case

Throughout this section μ denotes a Borel probability measure possessing all moments, not necessarily absolutely continuous with respect to Lebesgue measure.

2.1 Departure from the regular case

Relations (6) and (8) do not need to hold in the singular case as the bilinear form $\langle \cdot, \cdot \rangle_\mu$ may fail to be positive definite. In this situation the definition of polynomial bases requires additional care and the RKHS interpretation is not valid over the whole space \mathbb{R}^p . There are essentially two ways to circumvent this difficulty:

- Construct, using a quotient map, the space of polynomials restricted on $\text{supp}(\mu)$, construct the kernel based on this family and extend it to the whole euclidean space.
- Consider only the variational formulation as in the left hand side of (7).

These are equivalent on the support of μ but do not lead the same definition outside of the support. Identity (8) ceases to hold in the singular case and both constructions provide a valid extension outside of the support. We will focus on the second choice and start by stating the following basic Lemma.

Lemma 1 *Let $p \in \mathbb{N}^*$, $M \in \mathbb{R}^{p \times p}$ be symmetric semidefinite and $\mathbf{u} \in \mathbb{R}^p$. Let M^\dagger denotes the Moore-Penrose pseudo inverse, we have*

$$\inf_{\mathbf{x} \in \mathbb{R}^p} \{ \mathbf{x}^T M \mathbf{x}; \quad \mathbf{x}^T \mathbf{u} = 1 \} = \begin{cases} \frac{1}{\mathbf{u}^T M^\dagger \mathbf{u}}, & \text{if } \text{proj}_{\ker(M)}(\mathbf{u}) = 0, \\ 0, & \text{otherwise} \end{cases}$$

2.2 Support of the measure

If $M_{\mu,d}$ does not have full rank, then this means that its support is contained in an algebraic set. We let V be the Zariski closure of $\text{supp}(\mu)$, that is the smallest algebraic set which contains $\text{supp}(\mu)$. It is well known that an algebraic set is always equal to the common zero set of finitely many polynomials. Note that in the present article we exclusively work over the real field of coefficients, a setting which is more intricate than that offered by an algebraically closed field (for instance Hilbert's Nullstellensatz is not valid in its original form over a real field).

Lemma 2 Denote by \mathcal{I} the set of polynomials P which satisfy $\int P^2 d\mu = 0$. The set \mathcal{I} coincides with the ideal of polynomials vanishing on V .

We conclude this section by illustrating important relation between the moment matrix and geometric properties of V . The proof combines [39, Proposition 2] and relation between Hilbert’s polynomial and dimension [14, Chapter 9]. An illustration of the relation between the growth of the rank of the moment matrix and the dimension of the underlying set is given in Section D with numerical simulations.

Proposition 3 For all $d \in \mathbb{N}^*$, $\text{rank}(M_{\mu,d}) = \text{HF}(d)$, where HF denotes the Hilbert function of V and gives the dimension of the space of polynomials of degree up to d on V . For d large enough, $\text{rank}(M_{\mu,d})$ is a polynomial in d whose degree is the dimension of V and $\ker(M_{\mu,d})$ provides a basis generating the ideal \mathcal{I} .

2.3 Construction of the Christoffel-Darboux kernel

We denote by $L_d^2(\mu)$ the space of polynomials on V of degree up to d with inner product and norm induced by μ denoted by $\langle \cdot, \cdot \rangle_\mu$ and $\| \cdot \|_\mu$ respectively. This can be constructed by considering that two elements are equivalent if and only if they agree on V . $L_d^2(\mu)$ is the quotient space $\mathbb{R}_d[X]/(\mathcal{I} \cap \mathbb{R}_d[X])$, and it can be easily checked that $\langle \cdot, \cdot \rangle_\mu$ is positive definite on $L_d^2(\mu)$ and hence induces a genuine scalar product.

As in the absolutely continuous setting, pointwise evaluation is continuous with respect to $\| \cdot \|_\mu$ and $L_d^2(\mu)$ is a RKHS [2]. The symmetric reproducing kernel is defined for all $\mathbf{x} \in V$, by $\kappa_{\mu,d}(\cdot, \mathbf{x}) \in L_d^2(\mu)$ such that for all $P \in L_d^2(\mu)$,

$$\int P(\mathbf{y}) \kappa_{\mu,d}(\mathbf{y}, \mathbf{x}) d\mu(\mathbf{y}) = P(\mathbf{x}).$$

We consider the variational formulation of the Christoffel function in (7). Considering the restriction of $\Lambda_{\mu,d}$ to V amounts to replace $\mathbb{R}_d[X]$ by $L_d^2(\mu)$, so that for all $\mathbf{z} \in V$

$$\Lambda_{\mu,d}(\mathbf{z}) = \inf_{P \in L_d^2(\mu)} \left\{ \int (P(\mathbf{x}))^2 d\mu(\mathbf{x}) : P(\mathbf{z}) = 1 \right\}. \quad (4)$$

(8) still holds for all $\mathbf{z} \in V$, $\Lambda_{\mu,d}(\mathbf{z}) \kappa_{\mu,d}(\mathbf{z}, \mathbf{z}) = 1$.

By definition of the Zariski closure, a polynomial vanishing on $\text{supp}(\mu)$ also vanishes on V and we deduce that:

$$\{\mathbf{z} \in \mathbb{R}^p, \Lambda_{\mu,d}(\mathbf{z}) = 0\} \cap V = \emptyset. \quad (5)$$

Therefore if $\mathbf{z} \in V$ then $\Lambda_{\mu,d}(\mathbf{z}) > 0$ and Lemma 1 ensures that $\Lambda_{\mu,d}(\mathbf{z}) = (\mathbf{v}_d(\mathbf{z})^T M_{\mu,d}^\dagger \mathbf{v}_d(\mathbf{z}))^{-1}$. Since V is algebraic, it is the solution set of a finite number of polynomial equations. Let d_V be the maximum degree of such equations. Then for any $d \geq d_V$: $V = \{\mathbf{x} \in \mathbb{R}^d : \Lambda_{\mu,d}(\mathbf{x}) > 0\}$.

3 Finite sample approximation

Given a positive Borel measure μ one may consider a finite sample $\{\mathbf{x}_i\}_{i=1}^n$ of independant random vectors drawn from μ and replace μ by its empirical counterpart $\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{\mathbf{x}_i}$

where $\delta_{\mathbf{a}}$ denotes the Dirac measure at \mathbf{a} . In this situation, for $i = 1, \dots, n$, \mathbf{x}_i is a random variable, hence μ_n is a random measure, $M_{\mu_n,d}$ is a random matrix, and the law of large numbers states that for every fixed d , $\|M_{\mu_n,d} - M_{\mu,d}\| \rightarrow 0$ almost surely. This entails that $\text{rank}(M_{\mu_n,d}) \rightarrow \text{rank}(M_{\mu,d})$ almost surely. Classical approaches to strengthen this result and provide quantitative estimates rely on concentration of measure for matrices or operators. In the present section, we investigate another direction and provide a different view of this convergence phenomenon. *Under suitable assumptions, the rigidity of polynomials ensure that convergence of the rank occurs almost surely for a finite sample size.*

3.1 Assumptions and main result

Recall that μ is a Borel probability measure on \mathbb{R}^p with all its moments finite and V denote the Zariski closure of $\text{supp}(\mu)$. An algebraic set is called irreducible if it is not the union of two algebraic proper subsets. Given an algebraic set W , we define a canonical area measure σ_W . Intuitively, an algebraic set consists of a finite union of a smooth submanifolds of dimension p_W "glued" along a lower dimensional singular locus. The area measure is constructed with geometric integration theory techniques on the smooth part.

Construction of the area measure: Let $W \subset \mathbb{R}^p$ be an irreducible real algebraic set of dimension p_W . According to Proposition 3.3.14, Proposition 3.3.10 and Definition 3.3.4 in [7], there exists a lower dimensional algebraic subset Y of dimension strictly smaller than p_W , such that the set $Z = W \setminus Y$ can be seen as a p_W dimensional smooth submanifold of \mathbb{R}^p (possibly not connected). There is a natural Euclidean p_W dimensional density on Z induced from the euclidean embedding, see for example Section 7.3 in [17]. This defines integration of continuous functions on Z ; then Riesz representation theorem yields a regular positive Borel measure representing this very integration linear functional. The resulting measure is called the *area measure* of W and it is denoted in short by σ_W .

Assumption 4 μ is a Borel measure on \mathbb{R}^p with finite moments and V is the Zariski closure of $\text{supp}(\mu)$ endowed with the area measure σ_V . They satisfy the following constraints:

1. V is an irreducible algebraic set,
2. μ is absolutely continuous with respect to σ_V .

Under this assumption one infers the following finite sample stabilization result. The statement formalizes the intuition that the set of polynomials on $\text{supp}(\mu)$ and the set of polynomials on a finite sample coincide almost surely as long as the sample size reaches the dimension of the space $L_d^2(\mu)$.

Theorem 5 For all $d \geq 1$ and all $n \geq \text{rank}(M_{\mu,d})$, it holds almost surely that,

- $\text{rank}(M_{\mu_n,d}) = \text{rank}(M_{\mu,d})$,

It is the combination of Theorem 18 and Proposition 16 which shows that, in principle, it is possible to fully characterize algebraic geometric properties of the Zariski closure of $\text{supp}(\mu)$ from a finite sample, with probability one.

It is obvious that $\text{rank}(M_{\mu_n, d}) \leq \text{rank}(M_{\mu, d})$. The main idea of the proof is to show that rank deficiency implies that $\{\mathbf{x}_i\}_{i=1}^n$ are in the zero locus of a polynomial P on V^n not vanishing everywhere on V^n . We then use the fact that a polynomial P vanishing in an euclidean neighborhood in V^n actually vanishes on the whole V^n . This fact is deduced from Rückert's complex analytic Nullstellensatz [30, Proposition 1.1.29], which provides a local characterization of P . Completion of the local ring of rational functions allows to deduce global properties of P [47]. The irreducibility hypothesis and the definition of the area measure allow precisely to switch from real to complex variables.

We remark that the assumptions on V and μ in 17 are mandatory for the validity of the above theorem. Two simple examples illustrate our claim.

Example Consider the following generative process in \mathbb{R}^2 : x , y and z are drawn independently from the uniform measure on $[-1, 1]$. If $z \leq 0$, return $(0, y)$, otherwise return $(x, 0)$.

The underlying measure μ is the "uniform" measure on the set $\{(x, y) \in [-1, 1]^2; xy = 0\}$ and hence we have $V = \{(x, y) \in \mathbb{R}^2; xy = 0\}$. Draw an independent sample from μ , $\{(x_i, y_i)\}_{i=1}^n$, the event $y_i = 0$ for $i = 1, \dots, n$ occurs with probability $1/2^n$. Hence there is a nonzero probability that our sample actually belong to the set $\tilde{V} = \{(x, y) \in \mathbb{R}^2; y = 0\}$. This will result in $\text{rank}(M_{\mu_n, d}) < \text{rank}(M_{\mu, d})$. Since this event holds with non zero probability, this shows that Theorem 18 may not hold for any value of n if V is not irreducible.

Example Similarly, absolute continuity is necessary. Consider for example a mixture between an absolutely continuous measure and a dirac. There is a non zero probability that the sample only contains a singleton which will induce a rank deficiency in the moment matrix.

4 Reference measure with uniform asymptotic behavior

This section describes an asymptotic relation between the Christoffel function and the density associated with the underlying measure. The proof is mostly adapted from [33, Theorem 1.1], we provide explicit details and use a simplified set of assumptions which allows us to deal with singular measures.

4.1 Main assumptions and examples

Our main hypothesis is related to a special property of a reference measure whose existence is assumed.

Assumption 6 Let Z be a compact subset of \mathbb{R}^p and assume that there exists a reference Borel probability measure λ whose support is Z and a polynomial function $N: \mathbb{R}_+ \mapsto \mathbb{R}_+^*$ such that

$$\limsup_{d \rightarrow \infty} \sup_{z \in Z} |N(d)\Lambda_{\lambda, d}(z) - 1| = 0.$$

Remark The constant 1 is arbitrary and could be replaced by a continuous and strictly positive function of z .

The principal example that fits Assumption 21 is the $p - 1$ dimensional sphere in \mathbb{R}^p , denoted \mathbb{S}^{p-1} . The dimension $N(d)$ of the vector space of polynomials over \mathbb{S}^{p-1} is given by:

$$\binom{p+d-1}{p-1} + \binom{p+d-2}{p-1} = \left(1 + \frac{2d}{p-1}\right) \binom{d+p-2}{p-2}$$

which grows with d like $\frac{2d^{p-1}}{p-1}$ for a fixed value of p . As a function of d it is exactly the Hilbert polynomial associated with the real algebraic set \mathbb{S}^{p-1} . Here we may choose for λ the rotation invariant probability measure on \mathbb{S}^{p-1} (as normalized area measure). In this case $\Lambda_{\lambda, d}(\mathbf{z}) = \frac{1}{N(d)}$ for all $\mathbf{z} \in \mathbb{S}^{p-1}$ and all d by rotational invariance of both the sphere and the Christoffel function.

The case of the sphere is important because it helps construct many more situations which satisfy Assumption 21:

- Product of spheres with products of area measures: the bi-torus in \mathbb{R}^4 with the corresponding area measure.
- Affine transformations of such sets with the push forward of the reference measure with respect to the affine map; for instance the ellipsoid,

or rational embeddings of the sphere in higher dimensional space, to mention only a few natural choice of admissible operations.

4.2 Main result

Given a reference measure λ as in Assumption 21, and another measure $\mu \sim \lambda$ (i.e. $\lambda \ll \mu$ and $\mu \ll \lambda$), one can describe a precise relation with the underlying density.

Theorem 7 Let Z and λ satisfy Assumption 21. Let μ be a Borel probability measure on \mathbb{R}^p , absolutely continuous with respect to λ , with density $f: Z \rightarrow \mathbb{R}_+^*$ which is continuous and positive. Then:

$$\limsup_{d \rightarrow \infty} \sup_{\mathbf{x} \in Z} |N(d)\Lambda_{\mu, d}(\mathbf{x}) - f(\mathbf{x})| = 0.$$

In the case of the sphere, Assumption 21 does not hold only in the limit and from the proof of Theorem 22 we are able to obtain explicit quantitative bounds and a convergence rate, which to the best of our knowledge is the first estimate of this type for the Christoffel function.

Corollary 8 Let $f: \mathbb{R}^p \rightarrow \mathbb{R}_+$ be Lipschitz on the unit ball with $0 < c \leq f \leq C < +\infty$ on \mathbb{S}^{p-1} and assume that μ has density f with respect to the uniform measure on the sphere. Then we have $\sup_{\mathbf{x} \in \mathbb{S}^{p-1}} |N(d)\Lambda_{\mu, d}(\mathbf{x}) - f(\mathbf{x})| = O(d^{1/4})$.

5 Numerical experiments

5.1 Rank of the moment matrix

We illustrate Proposition 16 by the following numerical experiment:

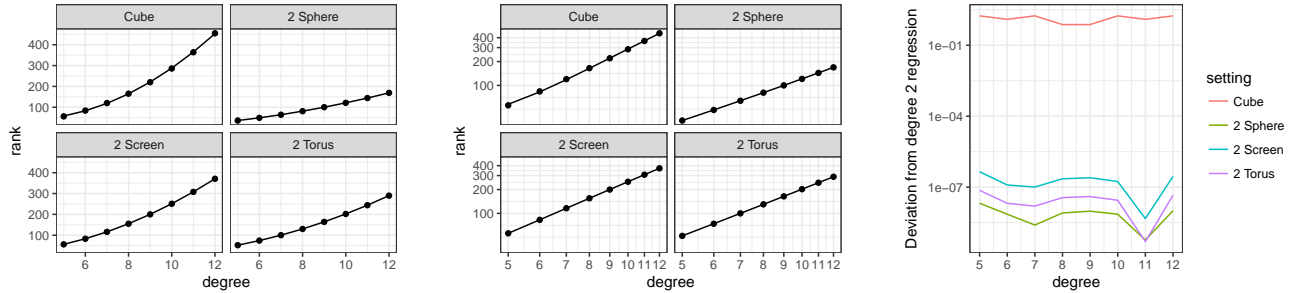


Figure 1: Relation between the rank of the moment matrix and the corresponding degree bound. For different sets, the dots represent the measured rank and the curve is the degree-two interpolation. On the left, the raw relation, we see that the cube has the highest rank. The same plot is in log log scale in the middle. The difference between measured rank and estimated degree-two interpolation is hardly visible. On the right, we represent the residuals between degree-two interpolation and measured ranks. The degree-rank relation is well interpolated for two dimensional sets while this is not the case for the cube.

- Sample 20000 points on a chosen set $\Omega \subset \mathbb{R}^3$, from a density with respect to the area measure on Ω .
- For $d = 5, \dots, 12$, compute the rank of the empirical moment matrix.
- Fit a degree-2 regression polynomial interpolating the relation between the degree and the rank.

We choose four different subsets of \mathbb{R}^3 : unit cube, unit sphere, TV screen, torus. The first one is 3-dimensional while all the others are 2-dimensional. From Proposition 16, in the first case it is expected that the computed rank grows like a third degree polynomial while for the remaining cases, it should grow like a quadratic. Hence the interpolation of the rank-degree relation should be of good quality for the last three cases and not for the first case. This is what we observed, see Figure 5.

5.2 Density estimation on an algebraic set

We present multivariate datasets whose topological characteristics suggest to map them to algebraic sets capturing symmetries. Practitioners have developed density estimation tools which mostly rely on the ability to compute a distance like divergence between two points which respects the topology of the data. As we next show, a *unifying and generic approach* allows to treat all these cases using the same computational tool: the empirical Christoffel function.

The first step consists in mapping the data of interest on an algebraic set whose topology reflects the intrinsic topology of the data, namely: the circle for periodic data, the sphere for celestial data, and the torus for bi-periodic data. Then, we evaluate the empirical Christoffel function on the chosen set and use it as a proxy to density. We use the pseudo-inverse of the empirical moment matrix.

The Christoffel function highly depends on the geometry of the boundary of the support. The algebraic sets considered here *do not* have boundaries (as manifolds) and isotropy properties ensure that the Christoffel function associated to the *uniform* measure on these sets is *constant*.

5.3 Dragonfly orientation: the circle

The dataset was described in [3] and consists of measurements of the orientation of 214 dragonflies with respect to the

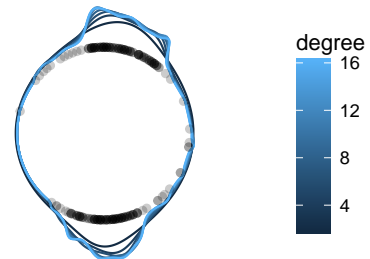


Figure 2: Dragon fly orientation with respect to the sun, on the torus. The curves represent the empirical Christoffel function and the dots are observations.

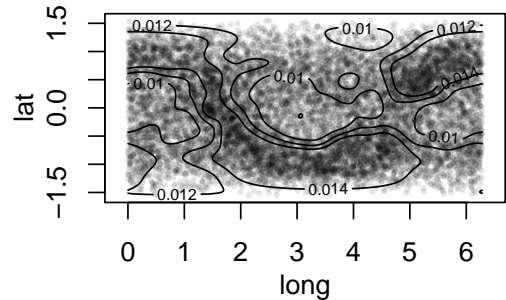


Figure 3: Each point represents the observation of a double star on the celestial sphere, associated to longitude and latitude. The level sets represent the empirical Christoffel functions on the sphere in \mathbb{R}^3 (degree 8). The highlighted band corresponds to the Milky Way.

azimuth of the sun. The orientation is an angle which has a periodicity and as such is naturally mapped to the circle. The dataset and the corresponding Christoffel function are displayed in Figure 2. As the degree increases, the Christoffel function captures regions densely populated by observations and regions without any observation. As was already observed in [3], dragonflies tend to sit in a direction perpendicular to the sun.

5.4 Double stars: the sphere

We reproduce the experiment performed in [15]. The dataset is provided by the European Space agency and was acquired

by Hipparcos satellite [21]. The data consists of the position of 12176 double stars on the celestial sphere described by spherical coordinates. Double stars are of interest in astronomy because of their connection with the formation and evolution of single stars. A natural question is that of the uniformity of the distribution of these double stars on the celestial map. The dataset and corresponding Christoffel function are displayed in Figure 3 using equirectangular projection. Firstly we note that the displayed level lines nicely capture the geometry of the sphere without distortion at the poles. Secondly the Christoffel function allows to detect a higher density region which corresponds to the Milky Way.

5.5 Amino-acid dihedral angles: the bi-torus

We reproduce the manipulations performed in [36]. Proteins are amino acid chains whose 3D structure can be described by ϕ and ψ backbone dihedral angles of amino acid residues. The 3D structure of a protein is extremely relevant as it relates to the molecular and biological function of a protein. Ramachandran plots consist of a scatter plot of these angles for different amino acids and allow to visualize energetically allowed configurations for each amino acid.

It is worth emphasizing that being able to describe typical regions in Ramachandran plots is of great relevance as a tool for *protein structure validation* [36]. Since the data consist of angles, it has a bi-periodic structure and therefore naturally maps to the bi-Torus in \mathbb{R}^4 . A Ramachandran plot for 7705 Glycine amino acids as well as the corresponding Christoffel function estimate is displayed in Figure 4. The Christoffel function is able to identify highly populated areas and its level set nicely fits the specific geometry of the torus. We refer the reader to [36] for more details about this dataset.

6 Discussion: perturbation of the moment matrix

We import real algebraic geometry and approximation theory results as tools to infer qualitative, topological properties from datasets. This is intrinsically connected to the notion of algebraic sets and allows to leverage more rigidity than provided by usual smoothness assumptions. This is suited for data restricted to algebraic sets.

However, in a singular situation, we are faced with a numerical instability issue. Aside from exact real arithmetic computation, the notion of rank in finite precision arithmetic is actually ill-defined which makes the singularity assumption questionable. We believe that the objects considered in this paper have sufficient stability properties to treat “near-singular” cases. We describe an heuristic argument, while more quantitative results are subject of future work.

Lemma 9 For a given $m \in \mathbb{N}$, let $A, M \in \mathbb{R}^{m \times m}$ be symmetric semidefinite. Then, for any $\mathbf{v} \in \mathbb{R}^m$ (uniformly on any compact)

$$\begin{aligned} & \liminf_{l \rightarrow 0_+} \inf_{\mathbf{x} \in \mathbb{R}^m} \{ \mathbf{x}^T (M + lA) \mathbf{x} \quad \text{s.t. } \mathbf{x} \in \mathbb{R}^m, \mathbf{x}^T \mathbf{v} = 1 \} \\ = & \inf \{ \mathbf{x}^T M \mathbf{x} \quad \text{s.t. } \mathbf{x} \in \mathbb{R}^m, \mathbf{x}^T \mathbf{v} = 1 \}. \end{aligned}$$

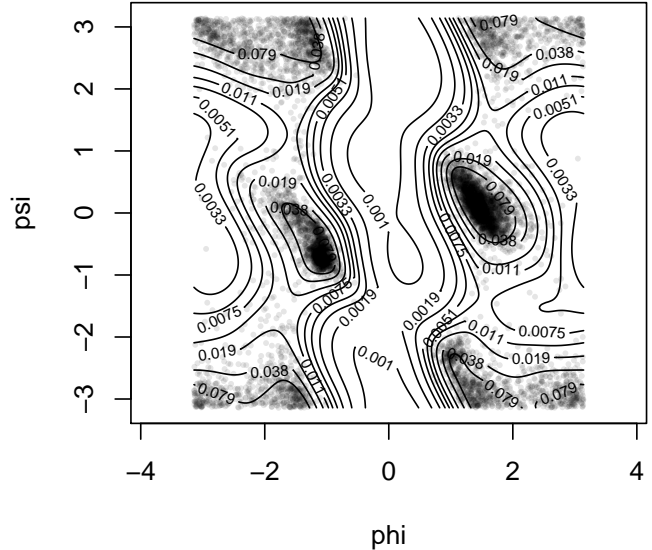


Figure 4: Each point represents two dihedral angles for a Glycine amino acid. These angles are used to describe the global three dimensional shape of a protein. They live on the bitorus. The level sets are those of the empirical Christoffel functions evaluated on the sphere in \mathbb{R}^4 . The degree is 4.

By Lemma 23, the Christoffel function associated with a slight perturbation of the moment matrix is very close to the actual Christoffel function. This justifies the variational formulation (4) as it can be seen as the limit of perturbations of μ making it non singular. By continuity of eigenvalues an appropriate thresholding scheme should lead to a correct evaluation of the rank of the moment matrix.

An important application of Lemma 23 is the addition of noise. Consider the following random process $\mathbf{y} = \mathbf{x} + \epsilon$, where \mathbf{x} is distributed according to μ and ϵ is independent small noise. Measuring the impact of ϵ on the moments of \mathbf{y} compared to moments of \mathbf{x} will help using our tools in the “close to singular” case. Understanding of the singular situation is a key to investigate robust variants suited to more general and practical *manifold learning* situations.

Acknowledgements

The research of J.B. Lasserre was funded by the European Research Council (ERC) under the European Union Horizon 2020 research and innovation program (grant agreement 666981 TAMING). This work was partially supported by CIMI (Centre International de Mathématiques et d’Informatique). Mihai Putinar is grateful to LAAS and Université Paul Sabatier for support. Jean-Bernard Lasserre and Edouard Pauwels would like to thank Yousouf Emin for early discussions on the topic.

References

- [1] E. Aamari and C. Levrard (2018). Non-Asymptotic Rates for Manifold, Tangent Space, and Curvature Estimation. *Annals of statistics*, in Press.

- [2] Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American mathematical society*, 68(3), 337–404.
- [3] E. Batschelet (1981). *Circular statistics in biology*. Academic press, New York.
- [4] M. Belkin and P. Niyogi (2002). Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in neural information processing systems*.
- [5] M. Belkin and P. Niyogi (2005). Towards a theoretical foundation for Laplacian-based manifold methods. In *International Conference on Computational Learning Theory*.
- [6] R. Bennett, R. (1969). The intrinsic dimensionality of signal collections. *IEEE Transactions on Information Theory*, 15(5):517–525.
- [7] J. Bochnak and M. Coste and M.F. Roy (1998). *Real algebraic geometry* (36). Springer Science & Business Media.
- [8] L. Bos (1994). Asymptotics for the Christoffel function for Jacobi like weights on a ball in \mathbb{R}^m . *New Zealand Journal of Mathematics*, 23(99):109–116.
- [9] L. Bos, B. D. Vecchia, and G. Mastroianni (1998). On the asymptotics of Christoffel functions for centrally symmetric weight functions on the ball in \mathbb{R}^d . *Rend. Circ. Mat. Palermo*, 2(52):277–290.
- [10] P. Bubenik (2015). Statistical topological data analysis using persistence landscapes. *The Journal of Machine Learning Research*, 16(1):77–102.
- [11] G. Carlsson (2009). Topology and data. *Bulletin of the American Mathematical Society*, 46(2), 255–308.
- [12] F. Chazal and D. Cohen-Steiner and Q. Mérigot (2011). Geometric inference for probability measures. *Foundations of Computational Mathematics*, 11(6):733–751.
- [13] F. Chazal and M. Glisse and C. Labruère and B. Michel (2015). Convergence rates for persistence diagram estimation in topological data analysis. *The Journal of Machine Learning Research*, 16(1):3603–3635.
- [14] D. Cox and J. Little and D. O’shea (2007). *Ideals, varieties, and algorithms* (3). Springer New York.
- [15] A. Cuevas and W. González-Manteiga and A. Rodríguez-Casal (2006). Plug-in estimation of general level sets. *Australian & New Zealand Journal of Statistics*, 48(1), 7–19.
- [16] D.L. Donoho and C. Grimes (2003). Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proceedings of the National Academy of Sciences*, 100(10):5591–5596.
- [17] J.J. Duistermaat and J.A. Kolk (2004). *Multidimensional real analysis II: integration*. Cambridge University Press.
- [18] C. Dunkl and Y. Xu (2001). *Orthogonal polynomials of several variables*. Cambridge University Press.
- [19] H. Edelsbrunner and J. Harer (2008). *Persistent homology—a survey*. *Contemporary mathematics*, 453, 257–282.
- [20] E. Elhamifar and R. Vidal (2011). Sparse manifold clustering and embedding. In *Advances in neural information processing systems*.
- [21] European Space Agency (1997). The Hipparcos and Tycho catalogues. <http://cdsweb.u-strasbg.fr/hipparcos.html>.
- [22] E. Facco and M. d’Errico and A. Rodriguez and A. Lai (2017). Estimating the intrinsic dimension of datasets by a minimal neighborhood information. *Scientific reports*, 7(1), 12140.
- [23] C. Fefferman and S. Mitter and H. Narayanan (2016). Testing the manifold hypothesis. *Journal of the American Mathematical Society*, 29(4):983–1049.
- [24] K. Fukunaga and D. R. Olsen (1971). An algorithm for finding intrinsic dimensionality of data. *IEEE Transactions on Computers*, 100(2):176–183.
- [25] C.R. Genovese and M. Perone-Pacifico and Verdinelli and L. Wasserman (2012). Manifold estimation and singular deconvolution under Hausdorff loss. *The Annals of Statistics*, 40(2):941–963.
- [26] C. Genovese and M. Perone-Pacifico and I. Verdinelli and L. Wasserman (2012). Minimax manifold estimation. *Journal of machine learning research*, 13(May), 1263–1291.
- [27] R. Ghrist (2008). Barcodes: the persistent topology of data. *Bulletin of the American Mathematical Society*, 45(1):61–75.
- [28] M. Hein and J. Y. Audibert (2005). Intrinsic dimensionality estimation of submanifolds in \mathbb{R}^d . In *Proceedings of the international conference on Machine learning*.
- [29] M. Hein and J.Y. Audibert and U. Von Luxburg (2005). From graphs to manifolds-weak and strong pointwise consistency of graph Laplacians. In *International Conference on Computational Learning Theory*.
- [30] D. Huybrechts (2006). *Complex geometry: an introduction*. Springer Science & Business Media.
- [31] A.K. Kim and H. H. Zhou (2015). Tight minimax rates for manifold estimation under Hausdorff loss. *Electronic Journal of Statistics*, 9(1):1562–1582.
- [32] M. Korda and M. Putinar and I. Mezic (2017). Data-driven spectral analysis of the Koopman operator. *arXiv preprint arXiv:1710.06532*.
- [33] A. Kròò and D. S. Lubinsky (2012). Christoffel functions and universality in the bulk for multivariate orthogonal polynomials. *Canadian Journal of Mathematics* 65(3):600620.

- [34] M.H. Law A.K. Jain (2006). Incremental nonlinear dimensionality reduction by manifold learning. *IEEE transactions on pattern analysis and machine intelligence*, 28(3):377–391.
- [35] E. Levina and P. J. Bickel (2005). Maximum likelihood estimation of intrinsic dimension. In *Advances in neural information processing systems*.
- [36] S.C. Lovell and I.W. Davis and W.B. Arendall and P.I. De Bakker and J.M. Word and M.G. Prisant and D.C. Richardson (2003). Structure validation by $C\alpha$ geometry: ϕ, ψ and $C\beta$ deviation. *Proteins: Structure, Function, and Genetics*, 50(3):437–450.
- [37] J. B. Lasserre and E. Pauwels (2016). Sorting out typicality via the inverse moment matrix SOS polynomial. In *Advances in Neural Information Processing Systems 29 (NIPS 2016, Barcelona)*, D.D. Lee and M. Sugiyama and U.V. Luxburg and I. Guyon and R. Garnett, Eds., Curran Associates, Inc. 2016, pp. 190–198.
- [38] J. B. Lasserre and E. Pauwels (2017). The empirical Christoffel function with applications in Machine Learning. arXiv:1701.02886. Submitted
- [39] M. Laurent and P. Rostalski (2012). The approach of moments for polynomial equations. In *Handbook on Semidefinite, Conic and Polynomial Optimization*, 25–60. Springer, Boston, MA.
- [40] A. Máté and P. Nevai (1980). Bernstein’s Inequality in L^p for $0 < p < 1$ and $(C, 1)$ Bounds for Orthogonal Polynomials. *Annals of Mathematics*, 111(1):145–154.
- [41] A. Máté, P. Nevai, and V. Totik (1991). Szego’s extremum problem on the unit circle. *Annals of Mathematics*, 134(2):433–53.
- [42] P. Nevai (1986). Géza Freud, orthogonal polynomials and Christoffel functions. A case study. *Journal of Approximation Theory*, 48(1):3–167.
- [43] P. Niyogi and S. Smale and S. Weinberger (2008). Finding the homology of submanifolds with high confidence from random samples. *Discrete & Computational Geometry*, 39(1-3):419–441.
- [44] P. Niyogi and S. Smale and S. Weinberger (2011). A topological view of unsupervised learning from noisy data. *SIAM Journal on Computing*, 40(3):646–663.
- [45] K. W. Pettis and T. A. Bailey and A. K. Jain and R. C. Dubes (1979). An intrinsic dimensionality estimator from near-neighbor information. *IEEE Transactions on pattern analysis and machine intelligence*, 1:125–37.
- [46] S. T. Roweis and L. K. Saul (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326.
- [47] J. P. Serre. Géométrie algébrique et géométrie analytique *Annales de l’institut Fourier*, 6:1–42
- [48] B. Simon (2008). The Christoffel-Darboux kernel. *Perspectives in PDE, Harmonic Analysis and Applications*, a volume in honor of VG Maz’ya’s 70th birthday, *Proceedings of Symposia in Pure Mathematics (79)*, 295–335.
- [49] G. Szegő (1974). *Orthogonal polynomials*. In *Colloquium publications*, AMS, (23), fourth edition.
- [50] J. B. Tenenbaum and V. De Silva and J. C. Langford (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323.
- [51] V. Totik (2000). Asymptotics for Christoffel functions for general measures on the real line. *Journal d’Analyse Mathématique*, 81(1):283–303.
- [52] G. V. Trunk (1968). Statistical estimation of the intrinsic dimensionality of data collections. *Information and Control*, 12(5):508–525.
- [53] W. V. Assche (1987). *Asymptotics for orthogonal polynomials*. Springer-Verlag Berlin Heidelberg.
- [54] E. D. Vito, L. Rosasco, and A. Toigo (2014). Learning sets with separating kernels. *Applied and Computational Harmonic Analysis*, 37(2):185–217.
- [55] Y. Xu (1996). Asymptotics for orthogonal polynomials and Christoffel functions on a ball. *Methods and Applications of Analysis*, 3(2):257–272.
- [56] Y. Xu (1999). Asymptotics of the Christoffel Functions on a Simplex in \mathbb{R}^d . *Journal of approximation theory*, 99(1):122–133.

A Introduction

This appendix contains supplementary information for the paper *Spectral analysis of moment data*. Section B provides additional details regarding the notations and constructions employed in the main text. In Section C, we describe proof arguments for all the results presented in the main text. More details about the numerical simulations are given in Section D. More details about the final discussion are given in Section E. Throughout this note, we repeat important elements from the main text.

B Polynomials and the moment matrix

In this section μ denotes a positive Borel measure on \mathbb{R}^p with finite moments and for any $d \in \mathbb{N}$, $M_{\mu,d}$ denotes its moment matrix with moments of degree up to $2d$. As a matter of fact we will soon restrict our attention to probability measures, which is a minor constraint to impose on the constructs below.

Henceforth we fix the dimension of the ambient euclidean space to be p . For example, vectors in \mathbb{R}^p as well as p -variate polynomials with real coefficients. We denote by X the tuple of p variables X_1, \dots, X_p which appear in mathematical expressions involving polynomials. Monomials from the canonical basis of p -variate polynomials are identified with their exponents in \mathbb{N}^p : specifically $\alpha = (\alpha_i)_{i=1\dots p} \in \mathbb{N}^p$ is associated to the monomial $X^\alpha := X_1^{\alpha_1} X_2^{\alpha_2} \dots X_p^{\alpha_p}$ of degree $\deg(\alpha) := \sum_{i=1}^p \alpha_i = |\alpha|$. The notations \leq_{gl} and \leq_{gl} stand for the graded lexicographic order, a well ordering over p -variate monomials. This amounts to, first, use the canonical order on the degree and, second, break ties in monomials with the same degree using the lexicographic order with $X_1 = a, X_2 = b \dots$. For example, the monomials in two variables X_1, X_2 , of degree less than or equal to 3 listed in this order are given by: 1, X_1 , X_2 , X_1^2 , X_1X_2 , X_2^2 , X_1^3 , $X_1^2X_2$, $X_1X_2^2$, X_2^3 . We focus here on the graded lexicographic order to provide a concrete example, but any ordering compatible with the degree would work similarly.

By definition \mathbb{N}_d^p is the set $\{\alpha \in \mathbb{N}^p; \deg(\alpha) \leq d\}$, while $\mathbb{R}[X]$ is the algebra of p -variate polynomials with real coefficients. The degree of a polynomial is the highest of the degrees of its monomials with nonzero coefficients¹. The notation $\deg(\cdot)$ applies a polynomial as well as to an element of \mathbb{N}^p . For $d \in \mathbb{N}$, $\mathbb{R}_d[X]$ stands for the set of p -variate polynomials of degree less than or equal to d . We set $s(d) = \binom{p+d}{d} = \dim \mathbb{R}_d[X]$; this is of course the number of monomials of degree less than or equal to d .

From now on $\mathbf{v}_d(X)$ denotes the vector of monomials of degree less or equal to d (sorted using \leq_{gl}), i.e., $\mathbf{v}_d(X) := (X^\alpha)_{\alpha \in \mathbb{N}_d^p} \in \mathbb{R}_d[X]^{s(d)}$. With this notation, one can write a polynomial $P \in \mathbb{R}_d[X]$ as $P(X) = \langle \mathbf{p}, \mathbf{v}_d(X) \rangle$ for some real vector of coefficients $\mathbf{p} = (p_\alpha)_{\alpha \in \mathbb{N}_d^p} \in \mathbb{R}^{s(d)}$ ordered using \leq_{gl} . Given $\mathbf{x} = (x_i)_{i=1\dots p} \in \mathbb{R}^p$, $P(\mathbf{x})$ denotes the evaluation of P with respect to the assignments $X_1 = x_1, X_2 = x_2, \dots, X_p = x_p$. Given a Borel probability measure μ and $\alpha \in \mathbb{N}^p$, $y_\alpha(\mu)$ denotes the moment α of μ , i.e., $y_\alpha(\mu) = \int_{\mathbb{R}^p} \mathbf{x}^\alpha d\mu(\mathbf{x})$. Throughout the paper we will only consider rapidly decaying measures, that is measure whose all moments are finite. For a positive Borel measure μ on \mathbb{R}^p denote by $\text{supp}(\mu)$ its support, i.e., the smallest closed set $\Omega \subset \mathbb{R}^p$ such that $\mu(\mathbb{R}^p \setminus \Omega) = 0$.

Moment matrix The moment matrix of μ , $M_d(\mu)$, is a matrix indexed by monomials of degree at most d ordered with respect to \leq_{gl} . For $\alpha, \beta \in \mathbb{N}_d^p$, the corresponding entry in $M_d(\mu)$ is defined by $M_d(\mu)_{\alpha,\beta} := y_{\alpha+\beta}(\mu)$, the moment $\int \mathbf{x}^{\alpha+\beta} d\mu$ of μ . For example, in the case $p = 2$, letting $y_\alpha = y_\alpha(\mu)$ for $\alpha \in \mathbb{N}_4^2$, one finds:

$$M_2(\mu) : \begin{array}{cccccc} & 1 & X_1 & X_2 & X_1^2 & X_1X_2 & X_2^2 \\ \begin{array}{c} 1 \\ X_1 \\ X_2 \\ X_1^2 \\ X_1X_2 \\ X_2^2 \end{array} & 1 & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} \\ & y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} \\ & y_{01} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} \\ & y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22} \\ & y_{11} & y_{21} & y_{12} & y_{31} & y_{22} & y_{13} \\ & y_{02} & y_{12} & y_{03} & y_{22} & y_{13} & y_{04} \end{array} .$$

¹For the null polynomial, we use the convention that its degree is 0.

The matrix $M_d(\mu)$ is positive semidefinite for all $d \in \mathbb{N}$. Indeed, for any $p \in \mathbb{R}^{s(d)}$, let $P \in \mathbb{R}_d[X]$ be the polynomial with vector of coefficients \mathbf{p} ; then $\mathbf{p}^T M_d(\mu) \mathbf{p} = \int_{\mathbb{R}^p} P(\mathbf{x})^2 d\mu(\mathbf{x}) \geq 0$. We also have the identity $M_d(\mu) = \int_{\mathbb{R}^p} \mathbf{v}_d(\mathbf{x}) \mathbf{v}_d(\mathbf{x})^T d\mu(\mathbf{x})$ where the integral is understood element-wise. Actually it is useful to interpret the moment matrix as representing the bilinear form

$$\begin{aligned} \langle \cdot, \cdot \rangle_\mu : \mathbb{R}[X] \times \mathbb{R}[X] &\mapsto \mathbb{R} \\ (P, Q) &\mapsto \int_{\mathbb{R}^p} P(\mathbf{x}) Q(\mathbf{x}) d\mu(\mathbf{x}), \end{aligned}$$

restricted to polynomials of degree up to d . Indeed, if $\mathbf{p}, \mathbf{q} \in \mathbb{R}^{s(d)}$ are the vectors of coefficients of any two polynomials P and Q of degree up to d , one has $\mathbf{p}^T M_d(\mu) \mathbf{q} = \langle P, Q \rangle_\mu$

C Proofs

C.1 Christoffel-Darboux kernel in the absolutely continuous case

Recall the definition of the Christoffel-Darboux kernel and the Christoffel function

$$\kappa_{\mu,d}(\mathbf{x}, \mathbf{y}) = \mathbf{v}_d(\mathbf{x})^T M_{\mu,d}^{-1} \mathbf{v}_d(\mathbf{y}). \quad (6)$$

$$\Lambda_{\mu,d}: \mathbf{x} \mapsto \inf_{P \in \mathbb{R}_d[X]} \left\{ \int_{\mathbb{R}^p} (P(\mathbf{z}))^2 d\mu(\mathbf{z}), \quad \text{s.t.} \quad P(\mathbf{x}) = 1 \right\} \quad (7)$$

Lemma 10 *For any $\mathbf{x} \in \mathbb{R}^p$, we have*

$$\kappa_{\mu,d}(\mathbf{x}, \mathbf{x}) \Lambda_{\mu,d}(\mathbf{x}) = 1, \quad (8)$$

and the solution of (7) is given by $P^*: \mathbf{x} \mapsto \frac{\kappa_{\mu,d}(\mathbf{z}, \mathbf{x})}{\kappa_{\mu,d}(\mathbf{z}, \mathbf{z})}$.

Proof We give a short proof for completeness. For any P feasible for (7), we have $\int P(\mathbf{x}) \kappa_{\mu,d}(\mathbf{z}, \mathbf{x}) d\mu(\mathbf{x}) = P(\mathbf{z}) = 1$. Cauchy-Schwartz inequality leads

$$1 \leq \int (P(\mathbf{x}))^2 d\mu(\mathbf{x}) \int (\kappa_{\mu,d}(\mathbf{z}, \mathbf{y}))^2 d\mu(\mathbf{y}) = \kappa_{\mu,d}(\mathbf{z}, \mathbf{z}) \int (P(\mathbf{x}))^2 d\mu(\mathbf{x}).$$

Choosing $P = P^*$, we obtain equality in the above inequality which proves the desired result. \square

C.2 Support of the measure

The following alternate definition of the closed support of a measure is useful in the sequel.

Definition 11 *A point x belongs to $\text{supp}(\mu)$ if and only if, for any open neighborhood, N of x , $\mu(N) > 0$.*

We denote by \mathcal{I} the set of polynomials P which satisfy $\int P^2 d\mu = 0$. We let V be the Zariski closure of $\text{supp}(\mu)$, that is the smallest algebraic set which contains $\text{supp}(\mu)$.

Lemma 12 *The set \mathcal{I} coincides with the ideal of polynomials vanishing on V .*

Proof First, $P \in \mathcal{I}$ entails that P vanishes on $\text{supp}(\mu)$ by continuity of P . If P did not vanish on V , then one would construct a strictly smaller algebraic subset of V containing $\text{supp}(\mu)$ which is contradictory. Conversely, let $P \in \mathbb{R}_d[X]$ vanish on V , then P vanishes on $\text{supp}(\mu)$ and $P(\mathbf{x})^2 = 0$ for all $\mathbf{x} \in \text{supp}(\mu)$ and therefore $\int P^2 d\mu = 0$, which in turn implies $P \in \mathcal{I}$. \square

C.3 Rank of the moment matrix

We introduce the following notion of genericity, borrowed from [39]. For any $d \in \mathbb{N}$, let $\mathcal{I}_d := \mathcal{I} \cap \mathbb{R}_d[x]$ be the intersection of the ideal \mathcal{I} with the vector space of polynomials of degree at most d .

Definition 13 For any fixed $d \in \mathbb{N}$, denote by \mathcal{K}_d the set

$$\mathcal{K}_d = \left\{ M \in \mathbb{R}^{s(d) \times s(d)}, M_{11} = 1, M \succeq 0, \mathbf{p}^T M \mathbf{p} = 0, \forall \mathbf{p} \in \mathcal{I}_d \right\}.$$

A matrix $T \in \mathbb{R}^{s(t) \times s(t)}$ is called generic if $T \in \mathcal{K}_d$ and $\text{rank}(T) \geq \text{rank}(M)$ for all $M \in \mathcal{K}_d$.

Lemma 14 For any $d \in \mathbb{N}^*$, $M_{\mu,d}$ is generic.

Proof First, for any polynomial P of degree at most d , with coefficient vector \mathbf{p} , we have $\mathbf{p}^T M_{\mu,d} \mathbf{p} = \int P^2 d\mu$. This quantity vanishes for $P \in \mathcal{I}_d$ and hence $M_{\mu,d} \in \mathcal{K}_d$. Choose $M \in \mathcal{K}_d$ and $\mathbf{p} \in \mathbb{R}^{s(d)}$ such that $M_{\mu,d} \mathbf{p} = 0$. We have that $\mathbf{p}^T M_{\mu,d} \mathbf{p} = \int P^2 d\mu = 0$ so that $P \in \mathcal{I}_d$ and hence $M \mathbf{p} = 0$. We have shown that $\ker(M_{\mu,d}) \subset \ker(M)$ so that $\text{rank}(M_{\mu,d}) \geq \text{rank}(M)$. \square

We introduce the Hilbert function and the Hilbert polynomial. We refer the reader to [14, Chapter 9] for a general presentation.

Definition 15 The Hilbert Function of \mathcal{I} is defined by

$$\begin{aligned} \text{HF}: \mathbb{N} &\mapsto \mathbb{N} \\ t &\mapsto \dim \mathbb{R}_t[x] / \mathcal{I}_t. \end{aligned}$$

For large values of t , HF is a polynomial.

Intuitively, the Hilbert function associates to d the dimension of the space of polynomials of degree up to d on $\text{supp}(\mu)$. This space is constructed using a natural quotient map. More details are given in the main text.

The following proposition combines [39, Proposition 2] and relation between Hilbert's polynomial and dimension [14, Chapter 9].

Proposition 16 For all $d \in \mathbb{N}^*$, $\text{rank}(M_{\mu,d}) = \text{HF}(d)$. For d large enough, $\text{rank}(M_{\mu,d})$ is a polynomial in d whose degree is the dimension of V and $\ker(M_{\mu,d})$ provides a basis for \mathcal{I} .

C.4 Finite sample approximation

Recall that for any $d \in \mathbb{N}^*$, $L_d^2(\mu)$ is identified with $\mathbb{R}_d[x] / \mathcal{I}_d$, two polynomials are considered equivalent if they agree on V . $(L_d^2(\mu), \langle \cdot, \cdot \rangle_\mu)$ is a finite dimensional RKHS which reproducing kernel is symmetric positive definite and defined for all $\mathbf{x} \in V$, by $\kappa_{\mu,d}(\cdot, \mathbf{x}) \in L_d^2(\mu)$ such that for all $P \in L_d^2(\mu)$,

$$\int P(\mathbf{y}) \kappa_{\mu,d}(\mathbf{y}, \mathbf{x}) d\mu(\mathbf{y}) = P(\mathbf{x}).$$

For any $\mathbf{x} \in V$, $\kappa_{\mu,d}(\cdot, \mathbf{x})$ is a polynomial and, by symmetry, $\kappa_{\mu,d}$ is itself a polynomial on V^2 which satisfies $\kappa_{\mu,d}(\mathbf{x}, \mathbf{x}) > 0$ for all $\mathbf{x} \in V$. Furthermore, we have $L_d^2(\mu) = \text{span} \{ \kappa_{\mu,d}(\cdot, \mathbf{x}), \mathbf{x} \in \text{supp}(\mu) \}$. This space being finite dimensional, there exists $N(d) \leq s(d)$ such that one can find a basis $\mathbf{e}_1, \dots, \mathbf{e}_{N(d)} \in \text{supp}(\mu)$ with

$$L_d^2(\mu) = \text{span} \{ \kappa_{\mu,d}(\cdot, \mathbf{e}_i), i = 1, \dots, N(d) \}. \quad (9)$$

Furthermore, let $\{P_i\}_{i=1}^{N(d)}$ be an orthonormal basis of $L_d^2(\mu)$. Then the identity $\int \kappa_{\mu,d}(x, y) P_i(y) d\mu(y) = P_i(x)$, for all $\mathbf{x} \in V$ and all i , yields:

$$\kappa_{\mu,d}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{N(d)} P_i(\mathbf{x}) P_i(\mathbf{y}), \quad \forall \mathbf{x}, \mathbf{y} \in V. \quad (10)$$

Assumption 17 μ is a Borel measure on \mathbb{R}^p with finite moments and V is the Zariski closure of $\text{supp}(\mu)$ endowed with the area measure σ_V . They satisfy the following constraints:

1. V is an irreducible algebraic set,
2. μ is absolutely continuous with respect to the area measure σ_V .

The following statement formalizes the intuition that the set of polynomials on $\text{supp}(\mu)$ and the set of polynomials on a finite sample coincide almost surely as long as the sample size reaches the dimension of the set of polynomials.

Theorem 18 For all $d \geq 1$ and all $n \geq \text{rank}(M_{\mu,d})$, it holds almost surely that,

- $\text{rank}(M_{\mu_n,d}) = \text{rank}(M_{\mu,d})$,
- $\{P \in \mathbb{R}_d[X]; \sum_{i=1}^n (P(\mathbf{x}_i))^2 = 0\} = \mathcal{I}_d$.

We begin by a technical Lemma which is a consequence of the reproducing kernel construction and its relation to orthogonal polynomials.

Lemma 19 Let $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$ and let μ_n be the empirical average $\sum_{i=1}^n \delta_{\mathbf{v}_i}$. Then:

$$\text{rank} \left((\kappa_{\mu,d}(\mathbf{v}_i, \mathbf{v}_j))_{i,j=1\dots n} \right) = \text{rank} (M_{\mu_n,d}).$$

Proof This is a consequence of equation (10). Let $\{P_i\}_{i=1}^{N(d)}$ be an orthonormal basis of $L_d^2(\mu)$. For each $i = 1, \dots, N(d)$ we choose $Q_i \in \mathbb{R}_d[X]$ to be one element in the equivalence class of P_i . The family $\{Q_i\}_{i=1}^{N(d)}$ must be independent in $\mathbb{R}_d[X]$ otherwise this would contradict independence of $\{P_i\}_{i=1}^{N(d)}$ in $L_d^2(\mu)$. This independent family can be extended to a basis of $\mathbb{R}_d[X]$ which we denote by $\{Q_i\}_{i=1}^{s(d)}$. Consider the matrix $D = (Q_j(\mathbf{v}_i))_{i=1\dots n, j=1\dots s(d)}$. Then $\text{rank} \left((K(\mathbf{v}_i, \mathbf{v}_j))_{i,j=1\dots n} \right) = \text{rank}(DD^T)$ from (10) and $\text{rank}(M_{\mu_n,d}) = \text{rank}(D^T D)$ since the rank of the moment matrix does not depend on the choice of polynomial basis in $\mathbb{R}_d[X]$. Both ranks are the same which is the desired result. \square

Now let $V^{N(d)}$ be the cartesian product space $\underbrace{V \times \dots \times V}_{N(d) \text{ times}}$. It is irreducible [7, Theorem 2.8.3] and its area

measure is the product measure $\sigma = \otimes_{i=1}^{N(d)} \sigma_V$. The determinantal function:

$$F: V^{N(d)} \mapsto \mathbb{R}$$

$$(\mathbf{x}_1, \dots, \mathbf{x}_n) \mapsto \det \left((K(\mathbf{x}_i, \mathbf{x}_j))_{i,j=1,\dots,N(d)} \right)$$

is a polynomial function which is not identically zero on $V^{N(d)}$, since by (9), $F(\mathbf{e}_1, \dots, \mathbf{e}_{N(d)}) > 0$. The following observation, of some independent interest, is in order.

Lemma 20 Let W be an irreducible real algebraic set with corresponding area measure σ_W and P be a polynomial on W . The following are equivalent

- (i) $\sigma_W(\{\mathbf{z} \in W : P(\mathbf{z}) = 0\}) > 0$.
- (ii) P vanishes on W .

Proof The implication (ii) to (i) is trivial since σ_W is supported on W . Let us assume that (i) is true and deduce (ii). This is classically formulated in the language of sheafs, we adopt a more elementary language and describe details for completeness. The main idea of the proof is to reduce the statement to Rückert's complex analytic Nullstellensatz [30, Proposition 1.1.29] which characterizes the class of analytic functions vanishing locally on the zero set of a family of analytic functions. The first point is precisely of local nature and the Nullstellensatz combined with properties of analytic functions allow to deduce properties of P which extend globally [47]. The irreducibility hypothesis and the definition of the area measure allow precisely to switch from real to complex variables.

First, let P_1, \dots, P_k generate the ideal I of polynomials vanishing on W . Since W is irreducible, I is a prime ideal and $W = \{\mathbf{x} \in \mathbb{R}^p, P_i(\mathbf{x}) = 0, i = 1, \dots, k\}$ (see propositions 3.3.14 and 3.3.16 in [7]). Point (i) entails that there exists $\mathbf{x}_0 \in W$ and U_1 an Euclidean neighborhood of \mathbf{x}_0 in \mathbb{R}^p such that $W \cap U_1$ is an analytic submanifold, or more precisely the Jacobian matrix $\left(\frac{\partial P_i}{\partial X_j}\right)_{i=1 \dots k, j=1 \dots p}$ has rank k on U_1 , and furthermore, $P(\mathbf{x}) = 0$ for all $\mathbf{x} \in W \cap U_1$.

Consider the analytic complex manifold $Z = \{\mathbf{z} \in \mathbb{C}^p, P_i(\mathbf{z}) = 0, i = 1, \dots, k\}$ and the polynomial map

$$G: (X_1, \dots, X_p) \mapsto (P_1(X_1, \dots, X_p), \dots, P_k(X_1, \dots, X_p), X_{k+1}, \dots, X_p).$$

This map is locally invertible around \mathbf{x}_0 in \mathbb{C}^p and its inverse is analytic. The function $(H: X_{k+1}, \dots, X_p) \mapsto P(G^{-1}(0, \dots, 0, X_{k+1}, \dots, X_p))$ is analytic and vanishes in an Euclidean neighborhood, of $(x_{0,k+1}, \dots, x_{0,p})$, the last k coordinates of \mathbf{x}_0 , in \mathbb{R}^k . Hence, it can be identified with the constant null function on U_1 . This shows that H vanishes in an Euclidean neighborhood of $(x_{0,k+1}, \dots, x_{0,p})$ in \mathbb{C}^k .

This proves that there exists a Euclidean neighborhood of \mathbf{x}_0 in \mathbb{C}^p , U_2 , such that P vanishes on $Z \cap U_2$. At this point we can invoke the complex analytic Nullstellensatz [30, Proposition 1.1.29], to obtain k analytic functions, O_1, \dots, O_k , on U_2 , and an integer $m \geq 1$, such that, for all $\mathbf{z} \in U_2$, we have

$$(P(\mathbf{z}))^m = \sum_{i=1}^k P_i(\mathbf{z})O_i(\mathbf{z}). \quad (11)$$

We can now use powerful result related to completion of local rings. Combining Corollary 1 of Proposition 4 and Proposition 22 in [47] we obtain that identity (11) still holds with the constraint that O_i , $i = 1, \dots, k$, are rational functions. In other words, reducing to common denominator, there exists an Euclidean neighbourhood of \mathbf{x}_0 , U_3 , and k complex polynomials Q_1, \dots, Q_k and a complex polynomial Q such that Q does not vanish on U_3 , and

$$Q(\mathbf{z})(P(\mathbf{z}))^m = \sum_{i=1}^k P_i(\mathbf{z})Q_i(\mathbf{z}). \quad (12)$$

From this we deduce that identity (12) still holds by restricting to real variables and real polynomials. Now since the ideal I is prime and $Q \notin I$ (since $Q(\mathbf{x}_0) \neq 0$), we deduce that $P \in I$, that is, P vanishes on W . This is what we wanted to prove. \square

We may now proceed to the proof of the rank stabilization result from Theorem 18. First, since V is irreducible, $V^{N(d)}$ is also irreducible [7, Theorem 2.8.3]. Since F is a polynomial which does not vanish everywhere on $V^{N(d)}$, we deduce from Lemma 20 that

$$\begin{aligned} & \sigma \left(\left\{ \mathbf{x}_1, \dots, \mathbf{x}_{N(d)} \in V^{N(d)} : F(\mathbf{x}_1, \dots, \mathbf{x}_{N(d)}) = 0 \right\} \right) \\ &= \int_{V^{N(d)}} \mathbb{I} [F(\mathbf{x}_1, \dots, \mathbf{x}_{N(d)}) = 0] d\sigma(\mathbf{x}_1, \dots, \mathbf{x}_{N(d)}) \\ &= 0. \end{aligned} \quad (13)$$

Noticing that $\text{rank}(M_{\mu,d}) = N(d)$, the result follows because μ is absolutely continuous with respect to σ_V and hence $\otimes_{i=1}^{N(d)} \mu$ is absolutely continuous with respect to σ so that, combining Lemma 19 and (13) with

the i.i.d. assumption, yields:

$$\begin{aligned}\mathbb{P}(\text{rank}(M_{\mu_n, d} < N(d))) &= \mathbb{P}(F(x_1, \dots, x_{N(d)}) = 0) \\ &= \int_{V^{N(d)}} \mathbb{I}[F(x_1, \dots, x_{N(d)}) = 0] d_\mu(x_1) \dots d_\mu(x_{N(d)}) = 0.\end{aligned}$$

This proves the first point. The second point is a consequence of the first point since the existence of a polynomial vanishing on the sample but not on the whole V would induce a rank deficiency in the empirical moment matrix and this occurs with zero probability.

C.5 Reference measure with uniform asymptotic behavior

Although the proof is mostly adapted from [33, Theorem 1.1], we provide explicit details and use a simplified set of assumptions which allows us to deal with singular measures. Our main hypothesis is related to a special property of a reference measure whose existence is assumed.

Assumption 21 *Let Z be a compact subset of \mathbb{R}^p and assume that there exists a reference Borel probability measure λ whose support is Z and a polynomial function $N: \mathbb{R}_+ \mapsto \mathbb{R}_+^*$ such that*

$$\limsup_{d \rightarrow \infty} \sup_{z \in Z} |N(d)\Lambda_{\lambda, d}(z) - 1| = 0.$$

Theorem 22 *Let Z and λ satisfy Assumption 21. Let μ be a borel probability measure on \mathbb{R}^p , absolutely continuous with respect to λ , with density $f: Z \rightarrow \mathbb{R}_+^*$ which is continuous and positive. Then:*

$$\limsup_{d \rightarrow \infty} \sup_{\mathbf{x} \in Z} |N(d)\Lambda_{\mu, d}(\mathbf{x}) - f(\mathbf{x})| = 0.$$

Proof Our arguments, mostly inspired from [33], are adapted to our setting. In addition, we provide some novel quantitative details. We split the proof into two parts. The second part is essentially a repetition of arguments similar to those used in the first part.

One direction By Assumption 21, Z is compact and so denote by D the finite diameter of Z . Fix an arbitrary $d \in \mathbb{N}^*$ and set:

$$\epsilon_d = \sup_{\mathbf{x}, \mathbf{y} \in Z, \|\mathbf{x} - \mathbf{y}\| \leq D/\sqrt{d}} |f(\mathbf{x}) - f(\mathbf{y})|, \quad (14)$$

which is well-defined because f is continuous on the compact set Z and hence uniformly continuous on Z , which ensures that the supremum is finite. Furthermore, $\epsilon_d \rightarrow 0$ as $d \rightarrow \infty$.

Next, fix an arbitrary $\mathbf{x}_0 \in Z$ and choose $P_{d - \lfloor d^{3/4} \rfloor}$ realizing the infimum for $\Lambda_{\lambda, d - \lfloor d^{3/4} \rfloor}(\mathbf{x}_0)$ ($\lfloor \cdot \rfloor$ denotes the floor integer part). Set $Q_{2 \lfloor d^{3/4} / 2 \rfloor}: \mathbf{y} \mapsto Q((\mathbf{y} - \mathbf{x}_0)/D)$ where Q is given by Lemma 25 with $\delta = 1/\sqrt{d}$ and of degree at most $2 \lfloor d^{3/4} / 2 \rfloor$. Then:

$$2 \lfloor d^{3/4} / 2 \rfloor \leq \lfloor d^{3/4} \rfloor,$$

and hence, the polynomial $P = P_{d - \lfloor d^{3/4} \rfloor} Q_{2 \lfloor d^{3/4} / 2 \rfloor}$ is of degree at most d and satisfies

$$\begin{cases} P(\mathbf{x}_0) = 1 \\ |P| \leq |P_{d - \lfloor d^{3/4} \rfloor}| \text{ on } Z \\ P^2 \leq 2^{2 - \sqrt{d}/2} \sup_{\mathbf{x} \in Z} P_{d - \lfloor d^{3/4} \rfloor}(x)^2 \text{ on } Z \setminus B_{D/\sqrt{d}}(\mathbf{x}_0), \end{cases} \quad (15)$$

where the first identity is because P is a product of polynomial whose value at \mathbf{x}_0 is 1, the last two identities follow by maximizing both terms of the product, using Lemma 25 and the fact that $1 - \delta \lfloor d^{3/4} / 2 \rfloor \leq$

$1 - d^{3/4}/(2\sqrt{d}) + 1/\sqrt{d} \leq 2 - \sqrt{d}/2$. P is feasible and provide an upper bound for the computation of the Christoffel function of μ . We obtain:

$$\begin{aligned}
& \Lambda_{\mu,d}(\mathbf{x}_0) \\
& \stackrel{(i)}{\leq} \int_Z (P(\mathbf{x}))^2 d\mu(\mathbf{x}) \\
& \stackrel{(ii)}{\leq} \int_{Z \cap \mathbf{B}(\mathbf{x}_0, D/\sqrt{d})} (P_{d-\lfloor d^{3/4} \rfloor}(\mathbf{x}))^2 f d\lambda(\mathbf{x}) + 2^{2-\sqrt{d}/2} \sup_{\mathbf{x} \in Z} P_{d-\lfloor d^{3/4} \rfloor}(\mathbf{x})^2 \int_{Z \setminus \mathbf{B}(\mathbf{x}_0, D/\sqrt{d})} d\mu(\mathbf{x}) \\
& \stackrel{(iii)}{\leq} (f(\mathbf{x}_0) + \epsilon_d) \int_{Z \cap \mathbf{B}(\mathbf{x}_0, D/\sqrt{d})} (P_{d-\lfloor d^{3/4} \rfloor}(\mathbf{x}))^2 d\lambda(\mathbf{x}) \\
& \quad + 2^{2-\sqrt{d}/2} \sup_{\mathbf{z} \in Z} (\Lambda_{\lambda, d-\lfloor d^{3/4} \rfloor}(\mathbf{z}))^{-1} \int_Z (P_{d-\lfloor d^{3/4} \rfloor}(\mathbf{x}))^2 d\lambda(\mathbf{x}) \\
& \stackrel{(iv)}{\leq} \left(f(\mathbf{x}_0) + \epsilon_d + 2^{2-\sqrt{d}/2} \sup_{\mathbf{z} \in Z} (\Lambda_{\lambda, d-\lfloor d^{3/4} \rfloor}(\mathbf{z}))^{-1} \right) \int_Z (P_{d-\lfloor d^{3/4} \rfloor}(\mathbf{x}))^2 d\lambda(\mathbf{x}) \\
& = \left(f(\mathbf{x}_0) + \epsilon_d + 2^{2-\sqrt{d}/2} \sup_{\mathbf{z} \in Z} (\Lambda_{\lambda, d-\lfloor d^{3/4} \rfloor}(\mathbf{z}))^{-1} \right) \Lambda_{\lambda, d-\lfloor d^{3/4} \rfloor}(\mathbf{x}_0).
\end{aligned}$$

where (i) is because $P(\mathbf{x}_0) = 1$, (ii) follows by decomposition of the integral over two domains and the uniform bounds in (15), (iii) follows by combining (14), Lemma 24 and the fact that μ is a probability measure on Z , (iv) follows by extending the domain of the first integral and the last identity is due to the choice of $P_{d-\lfloor d^{3/4} \rfloor}$ and Lemma 24. Therefore:

$$\begin{aligned}
& \Lambda_{\mu,d}(\mathbf{x}_0)N(d) - f(\mathbf{x}_0) \\
& \leq \left(f(\mathbf{x}_0) + \epsilon_d + 2^{2-\sqrt{d}/2} \sup_{\mathbf{z} \in Z} (\Lambda_{\lambda, d-\lfloor d^{3/4} \rfloor}(\mathbf{z}))^{-1} \right) \Lambda_{\lambda, d-\lfloor d^{3/4} \rfloor}(\mathbf{x}_0)N(d) - f(\mathbf{x}_0) \\
& \leq \left(\epsilon_d + 2^{2-\sqrt{d}/2} \sup_{\mathbf{z} \in Z} (\Lambda_{\lambda, d-\lfloor d^{3/4} \rfloor}(\mathbf{z}))^{-1} \right) \Lambda_{\lambda, d-\lfloor d^{3/4} \rfloor}(\mathbf{x}_0)N(d - \lfloor d^{3/4} \rfloor) \frac{N(d)}{N(d - \lfloor d^{3/4} \rfloor/2)} \\
& \quad + f(\mathbf{x}_0) \left(\Lambda_{\lambda, d-\lfloor d^{3/4} \rfloor}(\mathbf{x}_0)N(d - \lfloor d^{3/4} \rfloor) \frac{N(d)}{N(d - \lfloor d^{3/4} \rfloor/2)} - 1 \right) \\
& \leq \left(\epsilon_d + 2^{2-\sqrt{d}/2} \sup_{\mathbf{z} \in Z} (\Lambda_{\lambda, d-\lfloor d^{3/4} \rfloor}(\mathbf{z}))^{-1} \right) \Lambda_{\lambda, d-\lfloor d^{3/4} \rfloor}(\mathbf{x}_0)N(d - \lfloor d^{3/4} \rfloor) \frac{N(d)}{N(d - \lfloor d^{3/4} \rfloor/2)} \\
& \quad + C \left(\Lambda_{\lambda, d-\lfloor d^{3/4} \rfloor}(\mathbf{x}_0)N(d - \lfloor d^{3/4} \rfloor) \frac{N(d)}{N(d - \lfloor d^{3/4} \rfloor/2)} - 1 \right)
\end{aligned} \tag{16}$$

where $C = \sup_{\mathbf{x} \in Z} f(\mathbf{x})$. Now the upper bound in (16) does not depend on \mathbf{x}_0 and goes to 0 as $d \rightarrow \infty$ since as $d \rightarrow \infty$,

$$\begin{aligned}
& \epsilon_d \rightarrow 0, \\
& 2^{2-\sqrt{d}/2} \sup_{\mathbf{z} \in Z} (\Lambda_{\lambda, d-\lfloor d^{3/4} \rfloor}(\mathbf{z}))^{-1} \sim 2^{2-\sqrt{d}/2} N(d - \lfloor d^{3/4} \rfloor) \rightarrow 0, \\
& \Lambda_{\lambda, d-\lfloor d^{3/4} \rfloor}(\mathbf{x}_0)N(d - \lfloor d^{3/4} \rfloor) \frac{N(d)}{N(d - \lfloor d^{3/4} \rfloor/2)} \rightarrow 1,
\end{aligned}$$

where the first limit is obtained by uniform continuity, the second limit comes from N being a polynomial (by Assumption 21), and the last one also follows from Assumption 21. As a result:

$$\limsup_{d \rightarrow \infty} \sup_{\mathbf{x} \in Z} \Lambda_{\mu,d}(\mathbf{x})N(d) - f(\mathbf{x}) \leq 0, \tag{17}$$

which concludes the first part of the proof.

The other direction: To obtain the opposite direction inequality, we permute the role of μ and λ which corresponds to a density $\tilde{f} = 1/f$ which remains positive and continuous on Z . We repeat similar arguments, fix an arbitrary $d \in \mathbb{N}^*$ and set

$$\epsilon_d = \sup_{\mathbf{x}, \mathbf{y} \in Z, \|\mathbf{x} - \mathbf{y}\| \leq D/\sqrt{d}} |1/f(\mathbf{x}) - 1/f(\mathbf{y})|, \quad (18)$$

which again is well-defined because f is positive and continuous on the compact set Z and so $1/f$ is uniformly continuous on Z , which ensures that the supremum is finite. Furthermore, $\epsilon_d \rightarrow 0$ as $d \rightarrow \infty$.

Fix an arbitrary $\mathbf{x}_0 \in Z$ and choose P_d realizing the infimum for $\Lambda_{\mu, d}(\mathbf{x}_0)$. The polynomial $Q_{2\lfloor d^{3/4}/2 \rfloor} : \mathbf{y} \mapsto Q((\mathbf{y} - \mathbf{x}_0)/D)$ is the same as in the first part of the proof. The polynomial $P = P_d Q_{2\lfloor d^{3/4}/2 \rfloor}$ is of degree at most $d + \lfloor d^{3/4} \rfloor$ and satisfies

$$\begin{cases} P(\mathbf{x}_0) = 1 \\ |P| \leq |P_d| \text{ on } Z \\ P^2 \leq 2^{2-\sqrt{d}/2} \sup_{\mathbf{x} \in Z} P_d(\mathbf{x})^2 \text{ on } Z \setminus B_{D/\sqrt{d}}(\mathbf{x}_0). \end{cases} \quad (19)$$

As P is feasible one may compute an upper bound for the Christoffel function associated with λ , by:

$$\begin{aligned} & \Lambda_{\lambda, d + \lfloor d^{3/4}/2 \rfloor}(\mathbf{x}_0) \\ & \stackrel{(i)}{\leq} \int_Z (P(\mathbf{x}))^2 d\lambda(\mathbf{x}) \\ & \stackrel{(ii)}{\leq} \int_{Z \cap \mathbf{B}(\mathbf{x}_0, D/\sqrt{d})} \frac{(P_d(\mathbf{x}))^2}{f(\mathbf{x})} d\mu(\mathbf{x}) + 2^{2-\sqrt{d}/2} \sup_{\mathbf{x} \in Z} P_d(\mathbf{x})^2 \int_{Z \setminus \mathbf{B}(\mathbf{x}_0, D/\sqrt{d})} d\lambda \\ & \stackrel{(iii)}{\leq} (1/f(\mathbf{x}_0) + \epsilon_d) \int_{Z \cap \mathbf{B}(\mathbf{x}_0, D/\sqrt{d})} (P_d(\mathbf{x}))^2 d\mu(\mathbf{x}) + 2^{2-\sqrt{d}/2} \sup_{\mathbf{z} \in Z} (\Lambda_{\mu, d}(\mathbf{z}))^{-1} \int_Z (P_d(\mathbf{x}))^2 d\mu(\mathbf{x}) \\ & \stackrel{(iv)}{\leq} \left(1/f(\mathbf{x}_0) + \epsilon_d + 2^{2-\sqrt{d}/2} \sup_{\mathbf{z} \in Z} (\Lambda_{\mu, d}(\mathbf{z}))^{-1} \right) \int_Z (P_d(\mathbf{x}))^2 d\mu(\mathbf{x}) \\ & = \left(1/f(\mathbf{x}_0) + \epsilon_d + 2^{2-\sqrt{d}/2} \sup_{\mathbf{z} \in Z} (\Lambda_{\mu, d}(\mathbf{z}))^{-1} \right) \Lambda_{\mu, d}(\mathbf{x}_0). \end{aligned}$$

The inequality (i) follows $P(\mathbf{x}_0) = 1$, (ii) follows by decomposition of the integral over two domains and the uniform bounds in (15), (iii) follows by combining (18), Lemma 24 and the fact that λ is a probability measure on Z , (iv) follows by extending the domain of the first integral and the last identity is due to the choice of P_d and Lemma 24.

Hence we have,

$$\begin{aligned} & \frac{1}{\Lambda_{\mu, d}(\mathbf{x}_0)N(d)} - \frac{1}{f(\mathbf{x}_0)} \\ & \leq \left(\frac{1}{f(\mathbf{x}_0)} + \epsilon_d + 2^{2-\sqrt{d}/2} \sup_{\mathbf{z} \in Z} (\Lambda_{\mu, d}(\mathbf{z}))^{-1} \right) \frac{1}{\Lambda_{\lambda, d + \lfloor d^{3/4} \rfloor}(\mathbf{x}_0)N(d)} - \frac{1}{f(\mathbf{x}_0)} \\ & \leq \left(\epsilon_d + 2^{2-\sqrt{d}/2} \sup_{\mathbf{z} \in Z} (\Lambda_{\mu, d}(\mathbf{z}))^{-1} \right) \frac{1}{\Lambda_{\lambda, d + \lfloor d^{3/4} \rfloor}(\mathbf{x}_0)N(d + \lfloor d^{3/4} \rfloor)} \frac{N(d + \lfloor d^{3/4} \rfloor)}{N(d)} \\ & \quad + \frac{1}{f(\mathbf{x}_0)} \left(\frac{1}{\Lambda_{\lambda, d + \lfloor d^{3/4} \rfloor}(\mathbf{x}_0)N(d + \lfloor d^{3/4} \rfloor)} \frac{N(d + \lfloor d^{3/4} \rfloor)}{N(d)} - 1 \right) \\ & \leq \left(\epsilon_d + C2^{2-\sqrt{d}/2} \sup_{\mathbf{z} \in Z} (\Lambda_{\lambda, d}(\mathbf{z}))^{-1} \right) \frac{1}{\Lambda_{\lambda, d + \lfloor d^{3/4} \rfloor}(\mathbf{x}_0)N(d + \lfloor d^{3/4} \rfloor)} \frac{N(d + \lfloor d^{3/4} \rfloor)}{N(d)} \\ & \quad + C \left(\frac{1}{\Lambda_{\lambda, d + \lfloor d^{3/4} \rfloor}(\mathbf{x}_0)N(d + \lfloor d^{3/4} \rfloor)} \frac{N(d + \lfloor d^{3/4} \rfloor)}{N(d)} - 1 \right) \end{aligned} \quad (20)$$

where $C = \sup_{\mathbf{z} \in Z} 1/f(\mathbf{z})$. We have used the fact that for all $\mathbf{x} \in Z$, $\inf_{\mathbf{z} \in Z} f(\mathbf{z})\Lambda_{\lambda,d}(\mathbf{x}) \leq \Lambda_{\mu,d}(\mathbf{x})$ and hence $\sup_{\mathbf{z} \in Z} (\Lambda_{\mu,d}(\mathbf{z}))^{-1} \leq C \sup_{\mathbf{z} \in Z} (\Lambda_{\lambda,d}(\mathbf{z}))^{-1}$, to obtain the last identity. Now the upper bound in (20) does not depend on \mathbf{x}_0 and goes to 0 as $d \rightarrow \infty$ since as $d \rightarrow \infty$,

$$\begin{aligned} \epsilon_d &\rightarrow 0, \\ 2^{2-\sqrt{d}/2} \sup_{\mathbf{z} \in Z} (\Lambda_{\lambda,d}(\mathbf{z}))^{-1} &\sim 2^{2-\sqrt{d}/2} N(d) \rightarrow 0, \\ \frac{1}{\Lambda_{\lambda,d+\lfloor d^{3/4} \rfloor}(\mathbf{x}_0) N(d+\lfloor d^{3/4} \rfloor)} \frac{N(d+\lfloor d^{3/4} \rfloor)}{N(d)} &\rightarrow 1. \end{aligned}$$

The first limit is obtained by uniform continuity, the second limit follows from N being polynomial (by Assumption 21) and the last one also follows from Assumption 21. Therefore:

$$\limsup_{d \rightarrow \infty} \sup_{\mathbf{x} \in Z} \frac{1}{\Lambda_{\mu,d}(\mathbf{x}) N(d)} - \frac{1}{f(\mathbf{x})} \leq 0, \quad (21)$$

from which we deduce that

$$\liminf_{d \rightarrow \infty} \inf_{\mathbf{x} \in Z} \Lambda_{\mu,d}(\mathbf{x}) N_d(Z) - f(\mathbf{x}) \geq 0. \quad (22)$$

Combining (17) and (22) concludes the proof. \square

Remark In the case of the sphere, Assumption 21 does not hold only in the limit and from the proof of Theorem 22 we are able to obtain explicit quantitative bounds and a convergence rate. Suppose $f: \mathbb{R}^p \rightarrow \mathbb{R}_+$ with $0 < c \leq f \leq C$ on \mathbb{S}^{p-1} and that μ has density f with respect to the uniform measure on the sphere. In addition, assume that f is L -Lipschitz on the unit ball for some $L > 0$. Then equation (16) simplifies to:

$$\Lambda_{\mu,d}(\mathbf{x}_0) N(d) - f(\mathbf{x}_0) \leq \left(\frac{2L}{\sqrt{d}} + 2^{2-\sqrt{d}/2} N(d) \right) \frac{N(d)}{N(d - \lfloor d^{3/4}/2 \rfloor)} + C \left(\frac{N(d)}{N(d - \lfloor d^{3/4}/2 \rfloor)} - 1 \right)$$

and equation (20)

$$\Lambda_{\mu,d}(\mathbf{x}_0) N(d) - f(\mathbf{x}_0) \geq -C^2 \left(\frac{2L}{\sqrt{d}} + \frac{2^{2-\sqrt{d}/2}}{c} N(d) \right) \frac{N(d + \lfloor d^{3/4} \rfloor)}{N(d)} + \frac{C^2}{c} \left(\frac{N(d + \lfloor d^{3/4} \rfloor)}{N(d)} - 1 \right).$$

As a result we obtain the overall convergence rate $O(d^{1/4})$, which to the best of our knowledge is the first estimate of this type in this context. We leave for future research the task of improving this rate.

D Numerical experiments

D.1 Rank of the moment matrix

We choose four different subsets of \mathbb{R}^3 :

- The unit cube $\{x, y, z, |x| \leq 1, |y| \leq 1, |z| \leq 1\}$.
- The 3 dimensional unit sphere $\{x, y, z, x^2 + y^2 + z^2 = 1\}$.
- The 3 dimensional TV screen $\{x, y, z, x^6 + y^6 + z^6 - 2x^2y^2z^2 = 1\}$.
- The 3 dimensional torus $\left\{x, y, z, \left(x^2 + y^2 + z^2 + \frac{9}{16} - \frac{1}{16}\right)^2 - \frac{9}{16}(x^2 + y^2) = 0\right\}$.

Among the above sets the first one is three dimensional while all the others are 2-dimensional. The 2-dimensional sets are displayed in Figure 5. For each set, we sample 20000 points on it and compute the rank of the empirical moment matrix for different values of the degree. To perform this computation, we threshold the singular values of the design matrix consisting in the expansion of each data point in the multivariate Tchebychev polynomial basis.

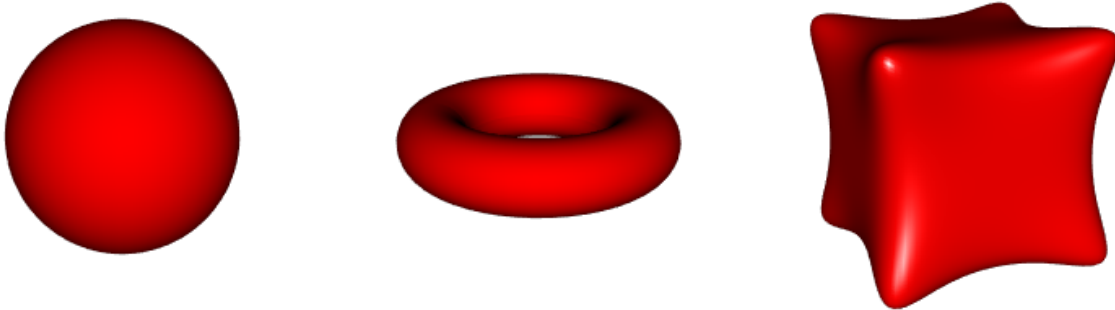


Figure 5: A visualization of the 2 dimensional surfaces considered in this example, the sphere, the torus and the TV screen.

D.2 Density estimation on algebraic sets

The following quantities are used in the litteratures as divergences combined with density estimation techniques for the examples treated in the main article.

- The quantity $\cos(\theta_1 - \theta_2)$ where θ_1 and θ_2 are angular coordinates of two points on the circle.
- The dot product on the sphere which generalizes the previous situation to larger dimensions, used in [15].
- The quantity $\cos\left(\sqrt{(\phi_1 - \phi_2)^2 + (\psi_1 - \psi_2)^2}\right)$ where $\phi_1, \phi_2, \psi_1, \psi_2$ are angles which correspond to points on the torus in \mathbb{R}^4 , used in [36].

E Perturbation of the moment matrix

The following Lemma describes the effect of small positive semidefinite perturbations of the moment matrix on the the Christoffel function.

Lemma 23 *For a given $m \in \mathbb{N}$, let $A, M \in \mathbb{R}^{m \times m}$ be symmetric semidefinite. Then, for any $\mathbf{v} \in \mathbb{R}^m$ (uniformly on any compact)*

$$\lim_{l \rightarrow 0_+} \inf_{\mathbf{x} \in \mathbb{R}^m} \{ \mathbf{x}^T (M + lA) \mathbf{x} \quad \text{s.t. } \mathbf{x} \in \mathbb{R}^p, \mathbf{x}^T \mathbf{v} = 1 \} = \inf \{ \mathbf{x}^T M \mathbf{x} \quad \text{s.t. } \mathbf{x} \in \mathbb{R}^m, \mathbf{x}^T \mathbf{v} = 1 \}.$$

Proof Since A is positive semidefinite, the left-hand side must be greater than the right-hand side. In addition, the infimum in the right-hand side is attained at some \mathbf{x}^* which can be used in the right-hand side to show the result. \square

An important example of application of Lemma 23 is the case of a convolution, which corresponds to the addidtion of noise. Consider the following random process

$$\mathbf{y} = \mathbf{x} + \epsilon \tag{23}$$

where \mathbf{x} is distributed according to μ and ϵ is some independent noise, assumed to be small. This corresponds to a convolution of the law of \mathbf{x} and that of ϵ . If $\|\epsilon\|_\infty$ is bounded by a small δ , for a given $d \in \mathbb{N}$, is it possible to measure the difference between moments of \mathbf{y} and moments of \mathbf{x} as a function of ϵ . The motivation for studying (23) is from *manifold learning* or *manifold identification*, where one considers a cloud of points which is not supported on an algebraic set because of additive small noise, but possibly very close to an algebraic set. For illustration purpose, in Figure 6 one observes how the level sets tend to concentrate as the perturbation gets smaller.

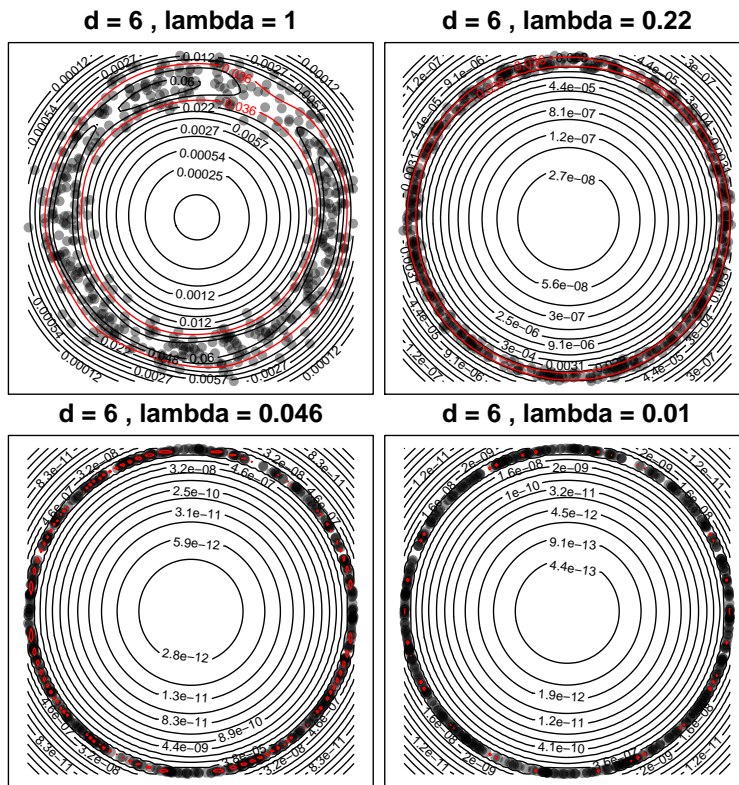


Figure 6: Small perturbation (23) of a set of points on the circle. The perturbation tends to 0 and the level sets of the Christoffel function tend to converge to the circle as predicted by Lemma 23.

F Appendix: Technical Lemmas

We begin with the following simple Lemma.

Lemma 24 *Let μ be a Borel probability measure on \mathbb{R}^p and S its support which is assumed to be a bounded subset of \mathbb{R}^p . Then for any $d \in \mathbb{N}^*$ and for any $P \in \mathbb{R}_d[X]$,*

$$\sup_{\mathbf{x} \in S} |P(\mathbf{x})|^2 \leq \sup_{\mathbf{z} \in S} \Lambda_{\mu,d}^{-1}(\mathbf{z}) \int (P(\mathbf{x}))^2 d\mu(\mathbf{x}).$$

Proof For any $\mathbf{x} \in S$ and $P \in \mathbb{R}_d[X]$, we have

$$\Lambda_{\mu,d}(\mathbf{x}) \leq \frac{\int (P(\mathbf{z}))^2 d\mu(\mathbf{z})}{P(\mathbf{x})^2}$$

and

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

The result follows by considering the supremum over S on both sides. \square

The following Lemma is a quantitative adaptation of [33, Lemma 2.1]

Lemma 25 *For any $d \in \mathbb{N}^*$ and any $\delta \in (0, 1)$, there exists a p -variate polynomial of degree $2d$, Q , such that*

$$Q(0) = 1; \quad -1 \leq Q \leq 1, \text{ on } \mathbf{B}; \quad |Q| \leq 2^{1-\delta d} \text{ on } \mathbf{B} \setminus \mathbf{B}_\delta(0).$$

Proof Let R be the univariate polynomial of degree $2d$, defined by

$$R: t \rightarrow \frac{T_d(1 + \delta^2 - t^2)}{T_d(1 + \delta^2)},$$

where T_d is the usual Chebyshev polynomial. We have

$$R(0) = 1. \tag{24}$$

Furthermore, for $t \in [-1, 1]$, we have $0 \leq 1 + \delta^2 - t^2 \leq 1 + \delta^2$. T_d has absolute value less than 1 on $[-1, 1]$ and is increasing on $[1, \infty)$ with $T_d(1) = 1$, so for $t \in [-1, 1]$,

$$-1 \leq R(t) \leq 1. \tag{25}$$

For $|t| \in [\delta, 1]$, we have $\delta^2 \leq 1 + \delta^2 - t^2 \leq 1$, so

$$|R(t)| \leq \frac{1}{T_d(1 + \delta^2)}. \tag{26}$$

Let us bound the last quantity. Recall that for $t \geq 1$, we have the following explicit expression

$$T_d(t) = \frac{1}{2} \left(\left(t + \sqrt{t^2 - 1} \right)^d + \left(t + \sqrt{t^2 - 1} \right)^{-d} \right).$$

We have $1 + \delta^2 + \sqrt{(1 + \delta^2)^2 - 1} \geq 1 + \sqrt{2}\delta$, which leads to

$$\begin{aligned} T_d(1 + \delta^2) &\geq \frac{1}{2} \left(1 + \sqrt{2}\delta \right)^d \\ &= \frac{1}{2} \exp \left(\log \left(1 + \sqrt{2}\delta \right) d \right) \\ &\geq \frac{1}{2} \exp \left(\log(1 + \sqrt{2}) \delta d \right) \\ &\geq 2^{\delta d - 1}, \end{aligned} \tag{27}$$

where we have used concavity of the log and the fact that $1 + \sqrt{2} \geq 2$. It follows by combining (24), (25), (26) and (27), that $Q: \mathbf{y} \rightarrow R(\|\mathbf{y}\|_2)$ satisfies the claimed properties. \square