1
Introduction

To get a glimpse of the main theme of the book, consider an arbitrary cloud of N points $\mathbf{x}_i = (x_i, y_i) \in \mathbb{R}^2$, $i = 1, \dots, N$, in the plane, dense enough to form some geometric shape. For instance in Figure 1.1 the shape looks like a rotated letter "T"; similarly in the frontispiece, the cloud of points is concentrated on the letters "C" and "D" (for Christoffel and Darboux). Then we invite the reader to perform the following simple operations on the preferred cloud of points:

1. Fix $n \in \mathbb{N}$ (for instance $n = 2$) and let $s(n) = \binom{n+2}{2}$.
2. Let $\mathbf{v}_n(\mathbf{x}) = (1, x, y, x^2, x\,y, \dots, x\,y^{n-1}, y^n)$ be the vector of all monomials $x^i y^j$ of total degree $i + j \leq n$.
3. Form $\mathbf{X}_n \in \mathbb{R}^{n \times s(n)}$, the design matrix whose ith row is $\mathbf{v}(\mathbf{x}_i)$, and the real symmetric matrix $\mathbf{M}_n \in \mathbb{R}^{s(n) \times s(n)}$ with rows and columns indexed by monomials such that
$$\mathbf{M}_n := \frac{1}{N}\mathbf{X}_n^T \mathbf{X}_n.$$
4. Form the polynomial
$$\mathbf{x} \mapsto p_n(\mathbf{x}) := \mathbf{v}_n(\mathbf{x})^T \mathbf{M}_n^{-1}\,\mathbf{v}_n(\mathbf{x}).$$
5. Plot the level sets $S_\gamma := \{\mathbf{x} \in \mathbb{R}^2 : \; p_n(\mathbf{x}) = \gamma\}$ for some values of γ, and in red for the particular value $\gamma = \binom{2+n}{2}$.

As the reader can observe in Figure 1.1, the various level sets (and in particular the red one) capture quite accurately the shape of the cloud of points.

The above polynomial p_n is associated with the cloud of points $(\mathbf{x}_i)_{i \leq N}$ only via the real symmetric matrix $\mathbf{M}_n$ in a conceptually simple manner, the main

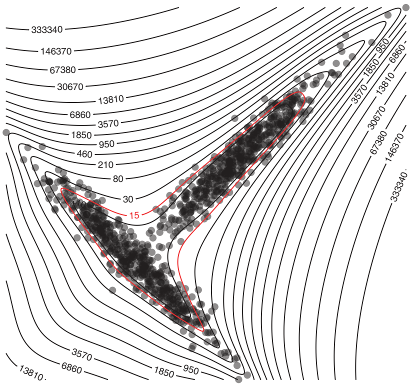

Figure 1.1 $n = 4$; $N = 1000$; level sets S_γ $\left(\text{in red for } \gamma = \binom{2+4}{2}\right)$.

computational step being matrix inversion. It turns out that $\mathbf{M}_n$ is called the *moment matrix* associated with the empirical probability measure

$$\mu_N := \frac{1}{N}\sum_{k=1}^{N}\delta_{\mathbf{x}_i}, \tag{1.1}$$

where $\delta_{\mathbf{x}}$ is the Dirac measure at the point $\mathbf{x}$, and $\mathbf{M}_n = \frac{1}{N}\sum_{i=1}^{N}\mathbf{v}_n(\mathbf{x}_i)\,\mathbf{v}_n(\mathbf{x}_i)^T$.

The reciprocal function $\mathbf{x} \mapsto \Lambda_n^{\mu_N}(\mathbf{x}) := p_n(\mathbf{x})^{-1}$ is called the *Christoffel function* (say of *degree* n as p_n is polynomial of degree $2n$) associated with the empirical measure μ_N. It depends only on the moments of μ_N, up to order $2n$.

In mathematical terms, the level sets S_γ in Figure 1.1 depict the shape of the *support* of the measure μ_N. This striking property is not an accident for this particular cloud. Indeed in Figure 1.2 we have displayed other clouds of two-dimensional points with various shapes and the corresponding level sets S_γ for various values of γ and n. Again, remarkably, the level sets S_γ approximate to high precision the shape of clouds even for a relatively small value of n. The same observation applies for the picture on the book cover where the shape of the "C" and "D" letters is very well approximated by level sets S_γ of the bivariate Cristoffel polynomial p_{10} of degree 20, and in particular by the one with boundary in red.

It turns out that in fact this property holds in the general framework of the Christoffel function Λ_n^μ associated with an abstract measure μ on a compact set $\mathbf{\Omega} \subset \mathbb{R}^d$. This time one replaces the empirical moment matrix $\mathbf{M}_n$ with

Cambridge University Press
978-1-108-83806-1 — The Christoffel–Darboux Kernel for Data Analysis
Jean Bernard Lasserre , Edouard Pauwels , Mihai Putinar
Excerpt
More Information

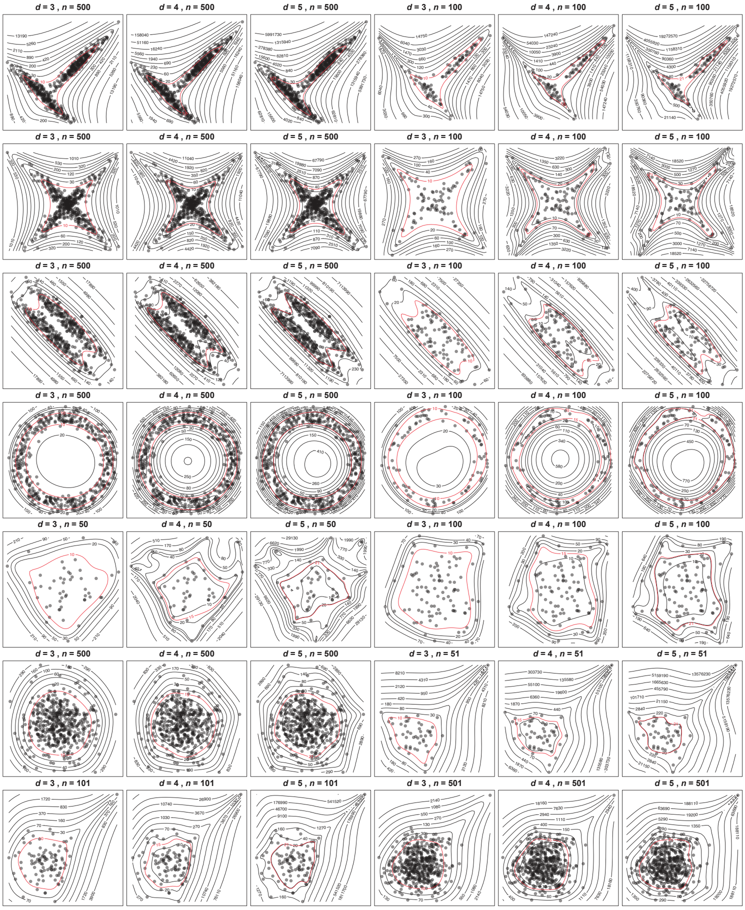

Figure 1.2 Level sets S_γ (in red for $\gamma = \binom{2+n}{2}$) for various clouds and various values of n.

the *exact* moment matrix $\mathbf{M}_n(\mu)$ associated with μ. For instance for $\boldsymbol{\Omega} \subset \mathbb{R}^2$, $\mathbf{M}_n(\mu)$ now reads

$$\mathbf{M}_n(\mu) := \int_{\boldsymbol{\Omega}} \mathbf{v}_n(\mathbf{x}_i)\, \mathbf{v}_n(\mathbf{x}_i)^T \, d\mu(\mathbf{x}),$$

where the integral is understood coordinate-wise. Notice that the empirical probability measure μ_N in (1.1) could have been obtained from a sample of N points $(\mathbf{x}_i)_{i \le N} \subset \boldsymbol{\Omega}$, drawn from some probability distribution μ on $\boldsymbol{\Omega}$.

This property has been known for a long time in various related domains of mathematics, for example approximation theory, orthogonal polynomials,

Cambridge University Press
978-1-108-83806-1 — The Christoffel–Darboux Kernel for Data Analysis
Jean Bernard Lasserre , Edouard Pauwels , Mihai Putinar
Excerpt
More Information

potential theory, reproducing kernel Hilbert spaces, function theory, spectral analysis, statistical mechanics, to cite a few. However, most works in these areas have been concerned with the asymptotic analysis of Λ_n^μ (appropriately scaled) when $n \to \infty$, and proportionally much less attention has been paid to the study of $\Lambda_n^{\mu_N}$ for fixed n, when μ_N is some empirical measure μ_N supported on a cloud of points, not to mention the clear benefits and potential applications in data analysis.

We know sketch out the structure of the book. It is divided into three rather distinct parts:

Part ONE consists of four chapters and focuses on historical and theoretical background. After introducing the key concept of a *reproducing kernel Hilbert space* (RKHS), classical results pertaining to Christoffel–Darboux kernel in the univariate case and separately the more involved multivariate case are recorded. More specifically:

Chapter 3 recalls classical results referring to the univariate Christoffel function (either in complex setting $\mathbb{C}$ or in the real line). Familiar results, sometimes several decades old, offer a necessary comparison basis for multivariate analogs. The latter are sometimes much more involved, still under investigation, or simply do not exist.

Chapter 4 focuses on the real multivariate Christoffel function for a measure on a compact set $\mathbf{\Omega} \subset \mathbb{R}^d$ and introduces both qualitative and quantitative asymptotics results. Several key theorems are stated without proofs due to intricate ingredients or necessary vast preliminaries (such as, for instance, pluripotential theory). Such important details fall beyond the scope of the present book. For the interested reader we provide some historical notes, technical statements and their sources. The level of depth and sophistication of recent advances in the multivariate theory of Christoffel–Darboux is barely suggested by our brief comments.

Chapter 5 is concerned with CD kernels associated with measures supported by a real algebraic variety. Think for instance of data points located by their very nature on a subset of the Euclidean sphere or a torus. In the singular support situation we fully exploit the concept of localized Hilbert function spaces. In spite of its theoretical flavor this is the natural framework for manipulating a structured moment matrix. To be more precise, it is quite remarkable that the degenerate moment matrix $\mathbf{M}_n$ (already described in the introduction above in dimension 2 and easily accessible from observed data) encodes profound analytical and geometrical characteristics of the generating measure.

Part TWO develops the motivation of this book, namely the utilization of the Christoffel–Darboux kernel associated with the empirical measure supported on a cloud of data points as the central carrier of structural information. How to decode this information into qualitative geometric, analytic or probabilistic features is our main task.

Chapter 6. As is typical in statistics and data analysis, a finite sample of size N is drawn independently from some unknown distribution μ with compact support $S \subset \mathbb{R}^d$. As detailed in the introduction, this provides an empirical version of the Christoffel function, related to the empirical measure supported on the finite sample. As expected, a generalization trade-off occurs at this point. For small values of n and large N the empirical Christoffel function is close to its population counterpart related to μ. On the other hand, for large n, the empirical Christoffel function has a trivial behavior which does not depend at all on μ. As described in Chapter 4, the population Christoffel function captures information on the underlying measure μ and its support S provided that the degree n increases. Therefore, in order to benefit from this phenomenon and avoid the trivial behavior of the empirical Christoffel function, it is crucial to relate the degree n of the Christoffel function. This chapter exposes recent results about statistical concentration for the Christoffel function and joint asymptotics in (n, N) under certain restrictions related to relative growth of n and N. Furthermore, in the context of singularly supported population measure μ with support contained in an algebraic set (which can be described by polynomial equations) we describe a finite-sample convergence phenomenon. Namely, under technical assumptions, the intrinsic rigidity of algebraic sets allows us to prove that, beyond a certain sample size N_0, with probability 1 the information contained in a finite sample is sufficient to fully characterize through the moment matrix the underlying algebraic set, that is, the set of equations describing it.

Chapter 7 illustrates the theory and expands on occurrences of CD kernels and Christoffel functions in statistics, for example the empirical CD kernel constitutes a higher-degree generalization of the well-known Mahalanobis distance. First in a parameteric regression setting, it turns out that the CD kernel has a natural interpretation in terms of predictive variance. This view allows us to make a direct connection with well-established quantities, such as leverage scores, and to discuss the problem of optimal design of experiments through the prism of the Christoffel function. Beyond parametric regression, the statistical results developed in Chapter 6 are illustrated on support inference problems for which we provide Christoffel function based estimators which benefit from the conceptual simplicity of the CD kernel, the main computational step being matrix inversion. This chapter also reports empirical results which were obtained for singularly supported densities on an algebraic set (sphere, torus), a situation that occurs for certain types of data (orientation, angles, positions on earth). We also provide results related to the motivating example of this introduction with outlier detection problems where we consider unsupervised network intrusion detection.

Part THREE contains a representative selection of complementary topics.

Chapter 8 focuses on two applications: one in basic approximation theory of nonsmooth functions, the other in the spectral analysis of certain ergodic dynamical systems. The clear advantages of treating such fundamental questions of mathematical analysis with techniques originating in the study of a CD kernel are simply stunning.

Chapter 9 deals with recent advances of immediate relevance to the topics of the book: stability under perturbations of Christoffel–Darboux kernels and a noncommutative, matrix analysis scheme of isolating the dense cloud from scattered and possibly *embedded* outliers of a 2D point distribution.

Chapter 10 is concerned with some spectral characterization as well as extensions of the Christoffel function. A first extension is to depart from the standard and classical $L^2(\mu)$ Hilbert space associated with the underlying measure μ and rather consider standard $L^p(\mu)$ Banach spaces. Another extension is to consider some natural convex cones of polynomials (positive on the support of μ) rather than "squares" in the variational $L^2(\mu)$ formulation of the Christoffel function. This yields alternative Christoffel-like functions with their own properties. Finally, when viewing the standard Christoffel function as single-point *interpolation*, it is also natural to investigate its natural *multi-point* extension.

The year appearing in cross references, for instance Hilbert (1953) or Marcel Riesz (2013), does not reflect the date of the original publication, but rather the year of a reprint edition.