ESAIM: PS 26 (2022) 171–207
https://doi.org/10.1051/ps/2022003

**ESAIM: Probability and Statistics**
www.esaim-ps.org

# RATE OF CONVERGENCE FOR GEOMETRIC INFERENCE BASED ON THE EMPIRICAL CHRISTOFFEL FUNCTION

Mai Trang Vu[1], François Bachoc[1,*] and Edouard Pauwels[2]

**Abstract.** We consider the problem of estimating the support of a measure from a finite, independent, sample. The estimators which are considered are constructed based on the empirical Christoffel function. Such estimators have been proposed for the problem of set estimation with heuristic justifications. We carry out a detailed finite sample analysis, that allows us to select the threshold and degree parameters as a function of the sample size. We provide a convergence rate analysis of the resulting support estimation procedure. Our analysis establishes that we may obtain finite sample bounds which are comparable to existing rates for different set estimation procedures. Our results rely on concentration inequalities for the empirical Christoffel function and on estimates of the supremum of the Christoffel-Darboux kernel on sets with smooth boundaries, that can be considered of independent interest.

**Mathematics Subject Classification.** 62G05, 42C05.

Received March 8, 2021. Accepted February 8, 2022.

## 1. Introduction

Given a measure $\nu$ on $\mathbb{R}^p$ and under appropriate assumption, the Christoffel function with degree bound $d \in \mathbb{N}$ can be defined on $\mathbb{R}^p$ as

$$\Lambda_{\nu,d} \colon z \quad \mapsto \quad \min_{\deg P \leq d,\, P(z)=1} \quad \int P^2 \mathrm{d}\nu,$$

where the infimum is over all polynomials of degree at most $d$. The empirical Christoffel function $\Lambda_{\mu_n,d}$ is associated to an input measure $\mu_n$, which is a scaled counting measure uniformly supported on a cloud of data-points. When $\mu_n$ is the empirical measure of an *iid* sample from $\mu$, $\Lambda_{\mu_n,d}$ can be seen as an estimation of the population Christoffel function $\Lambda_{\mu,d}$ (see [24]). Throughout this work the population measure $\mu$ has density $w$ on an unknown input set $S \subset \mathbb{R}^p$.

The (population) Christoffel function $\Lambda_{\mu,d}$ itself has a long history of research in the mathematical analysis literature. Its construction is based on multivariate polynomials of degree at most $d$ and it has strong links to the theory of orthogonal polynomials. Especially, the asymptotic behavior of the Christoffel function as the degree $d$ increases provides useful information regarding the support and density of the associated input measure $\mu$.

*Keywords and phrases:* Support estimation, christoffel function, concentration, finite sample, convergence rate.

[1] Institut de Mathématiques de Toulouse, UMR5219. Université de Toulouse, CNRS. UT3, 31062 Toulouse, France.
[2] Institut de Recherche en Informatique de Toulouse. Universite de Toulouse, CNRS. UT3, 31062 Toulouse, France.

[*] Corresponding author: francois.bachoc@math.univ-toulouse.fr

Important references in multivariate settings include [8, 9, 22, 23, 44], which concern specific cases of the input measure $\mu$ and set $S$. These works not only provide valuable information on the asymptotics of the population Christoffel function as $d$ goes to infinity, but also motivate the usage of this function in statistical contexts, especially in support recovery. Indeed, [24] provides a thresholding scheme using the Christoffel function which approximates the compact support $S$ of the measure $\mu$ with strong asymptotic guarantees. More precisely, [24] considers a family of polynomial sublevel sets $S_k = \{x \in \mathbb{R}^p : \Lambda_{\mu,d_k}(x) \geq \gamma_k\}$ with $k \in \mathbb{N}$, where the degree $d_k$ increases with $k$ and where the threshold $\gamma_k$ is well-chosen between a lower bound of the Christoffel function inside $S$ and an upper bound outside $S$. Another thresholding scheme can be found in [26], which provides useful results on the relation between $S$ and its estimator $S_k$. The topic of set estimation based on the population Christoffel function is thus currently a subject of active interest with a large range of applications in machine learning (see [24, 29]).

In a statistical context, the population Christoffel function $\Lambda_{\mu,d}$ is not available and only the empirical Christoffel function $\Lambda_{\mu_n,d}$ is, based on the observed empirical measure $\mu_n$. Let us detail results and discussions presented in [24]. Statistical procedures based on the empirical Christoffel function have three important features: (i) computations are remarkably simple and involve no optimization procedures, (ii) they scale efficiently with the number of observations and (iii) the procedures are affine equivariant (affine transformations of their input data result in affine transformations of their outputs, see Prop. 2.10). Furthermore, when considering a compactly supported population measure $\mu$ as well as its empirical counterpart $\mu_n$ supported on a sample of $n$ vectors in $\mathbb{R}^p$, drawn independently from $\mu$ and when the degree $d$ is fixed, the empirical Christoffel function $\Lambda_{\mu_n,d}$ converges uniformly to $\Lambda_{\mu,d}$, almost surely with respect to the draw of the random sample. This asymptotic result is appealing given the strong connections between $\Lambda_{\mu,d}$ and the support of $\mu$, which suggest that $\Lambda_{\mu_n,d}$ could be used for inferring the support of the population measure $\mu$. Yet more precise quantifications on the relation between sample size $n$ and the degree bound $d$ are required, but [24] does not provide any explicit way to choose the degree $d$ as a function of $n$, and does not provide any convergence guaranty for the full plugin approach based on the empirical Christoffel function $\Lambda_{\mu_n,d}$, when $d$ depends on $n$. These shortcomings constitute one of the main motivations for the present work.

## 1.1. Contribution

Our contribution is twofold:

1. We adapt the thresholding scheme in [24], using the empirical Christoffel function, by a careful tuning of the degree $d$ and the threshold $\gamma$ in the limit of large sample size. This scheme allows to estimate the compact support $S$ of a measure. Our results include, under regularity assumptions on $\mu$, a quantitative rate of convergence analysis which was unknown for this estimator. More precisely, we consider the Hausdorff distance between the original set $S$ and its estimator $S_n$ and between their respective boundaries, as well as the Lebesgue measure of their symmetric difference. These results rigorously establish the property that, when $n$ is large enough, these distances decrease to zero with an explicit rate.
2. This analysis relies on results which could be considered of independent interest. First, we provide a quantitative concentration result regarding the convergence of the empirical Christoffel function to its population counterpart. Second, this concentration relies on an estimate of the supremum of the Christoffel Darboux kernel on the support of the underlying measure. We prove that, for a large class of slowly decaying densities with smooth support boundary, this supremum is at most polynomial in the degree $d$. This shows that the considered class of measures is regular in the sense of the Bernstein-Markov property, see [31] and references therein.

## 1.2. Comparison with the existing literature on set estimation

Support inference (more generally set estimation) has been a topic of research in the statistics literature for more than half a century. The main subject of interest is to infer a set (support of an input measure, level set of an input density function,...) based on samples that are drawn independently from an unknown distribution.

Introduction and first results on this subject can be found in [20, 33], which motivate a subsequent analysis of estimators based on convex hulls for convex domains [10] or unions of balls for non-convex sets [16]. More involved estimators follow, such as the excess mass estimator [32], the plug-in approach based on the use of density estimators [12, 14, 27] or the $R$-convex hull of the samples, $R$ being a radius, [35].

Those works also motivated the development of minimax statistical analysis for the set estimation problem. We might find minimax results for the recovery of sets with (piecewise) smooth boundaries in [25], for the estimation of smooth or convex density level sets in [40] and for the plug-in approach in [34]. More current works related to set estimation include local convex hull estimators [1] and cone-convex hulls [11].

We obtain convergence rates both in terms of symmetric difference measure, and Hausdorff distance, which can be arbitrarily close to $n^{-1/(p+2r+2)}$ where $n$ is the sample size, $p$ is the ambient dimension and $r \geq 0$ measures the speed of decrease of the population density around the boundary of the support ($r = 0$ corresponds to a density which is uniformly bounded away from 0).

Our convergence rates hold under the assumption of a ball of fixed radius $R$ rolling inside and outside the support $S$. The rolling ball assumption is common [13, 42, 43]. Under this assumption, and in the case $r = 0$, [35] showed that the $R$-convex hull of the samples achieves the rate of convergence $n^{-2/(p+1)}$, for the Hausdorff distance[1] and the symmetric difference measure.

In [15], the Devroye and Wise estimator is shown to have a convergence rate of order $(\log(n)/n)^{1/p}$ in Hausdorff distance, under similar geometric assumptions as ours corresponding to the choice $r = 0$. Later on, [7] proved for the same estimator, under similar assumptions as ours, a rate which can be arbitrarily close to $n^{-1/(p+r)}$ for the measure of the symmetric difference for $r = 1$ and $r = 2$. Similarly, [37] obtained the same rate for an histogram based estimator in the context of density level set estimation. Earlier work presented in [25] proved that $n^{-1/p}$ is minimax optimal for the symmetric difference measure for a special class of piece-wise $C^1$ boundaries. Recently [28] proved a minimax lower bound on the convergence rate for symmetric difference, of order $n^{-1/(p+r)}$ for adaptive estimators to unknown $r \leq 2$.

Although the rates which we obtain are not optimal, for instance when compared to [35] in the case $r = 0$, the dependency in the dimension and speed of decrease of the density seem reasonable in comparison to existing rates. Let us insist on the fact that our analysis allows to cover a wide range of density decrease regimes and a variety of divergence measures between sets for which the results for other estimates are not known. A detailed comparison between all geometric conditions on the support, its boundary and different notions of divergence between sets is out of reach given the diversity of assumptions in the literature, and as such we only consider a high level general discussion based on orders of magnitude here.

From a computational point of view, our approach using the empirical Christoffel function has important advantages. The most important one is that this approach estimates the support of $\mu$ by a polynomial sublevel set, which is conceptually simple to manipulate. As an important illustration example, consider the situation when one is interested in performing numerical optimization over the estimated support. This situation can arise when a criterion is to be optimized over a feasible domain, which needs to be estimated from data. In this optimization case, the fact that the estimated support is a polynomial sublevel set is beneficial, for instance one can use nonlinear optimization techniques such as Sequential Quadratic Programming (SQP) or barrier functions. If the support is estimated by an union of balls centered at the observations [16], the estimated support may be less amenable to numerical optimization. Similarly, the $R$-convex hull estimator [35] is a set over which optimization may be challenging.

In terms of numerical implementation, our approach requires to compute and store the inverse of a matrix of size $s(d_n) = o(n)$ (see Sects. 2 and 3) where $d_n$ is the selected degree for the sample size $n$. Then, each input point can be tested to belong to the estimated support or not, with the cost of evaluating a quadratic form of size $s(d_n)$ and of computing $s(d_n)$ monomials in dimension $p$. In practice, $s(d_n)$ is smaller than $n$ (to avoid rank deficiencies), and in our asymptotic results, $d_n$ is selected such that $s(d_n) = o(n)$. Hence our approach relies on reasonably simple and classical computations. In comparison, for instance, computing the $R$-convex hull estimator [35] in general dimension $p$ may turn out to be challenging. In dimension $p = 2$, a point can be

[1] Similarly as we do, they consider Hausdorff distances both between sets and between their boundaries.

tested to belong to this set with computational cost $O(n \log n)$ [19], but we are not aware of similar efficient algorithms for larger $p$.

### 1.3. Organisation of the paper

Section 2 introduces the notation and definitions which will be used throughout the text, especially the definition of the population and empirical Christoffel functions and their known properties. In Section 3, we present our main assumptions as well as our results on support estimation and convergence of the empirical Christoffel function to the population one. Numerical illustration of the method appears in Section 4 for synthetic data in dimension 2 and 3 and an outlier detection benchmark in dimension 6. There, we also provide a fully data driven procedure for choosing the degree $d$ and the threshold $\gamma$. Concluding remarks are provided in Section 5. The proofs are postponed to the appendix. The appendix also contains additional results of interest on upper and lower bounds on the Christoffel function, outside and inside the support.

## 2. Preliminaries

### 2.1. General notation

When $\mu$ is a measure on $\mathbb{R}^p$, we denote by $\operatorname{supp}\mu$ the support of $\mu$. Let $f$ be a measurable function from $\mathbb{R}^p$ to $\mathbb{R}^p$. The push-forward measure of $\mu$ by $f$, denoted by $\mu_{\# f}$, is a measure on $\mathbb{R}^p$ defined by: $\mu_{\# f}(K) = \mu(f^{-1}(K))$ for all Borel sets $K$ of $\mathbb{R}^p$. Given an arbitrary (measurable) set $S \subset \mathbb{R}^p$, we denote by $\operatorname{Int} S$ the interior of $S$, $\partial S$ the boundary of $S$, $S^c$ the complement of $S$, $\lambda(S)$ the Lebesgue measure of $S$, $\operatorname{diam}(S)$ the diameter of $S$ (with respect to the Euclidean distance), $\lambda_S$ the Lebesgue measure restricted on $S$ and $\mu_S$ the uniform measure on $S$ (when $\lambda(S) > 0$).

When $M$ is a square matrix, we denote by $\|M\|$ the operator norm of $M$, *i.e.*

$$\|M\| = \sup_{x \neq 0} \frac{\|Mx\|_2}{\|x\|_2} = \sup_{\|x\|_2 = 1} \|Mx\|_2.$$

If in addition, $M$ is symmetric and positive definite, we can define its inverse $M^{-1}$ and its unique square root $M^{1/2}$ which are also symmetric positive definite matrices. We denote by $M^{-1/2}$ the inverse of the square root of $M$, which is also symmetric and positive definite.

For $x, y \in \mathbb{R}^p$, we let $d(x,y)$ be the Euclidean norm between $x$ and $y$. For $A \subset \mathbb{R}^p$ and $x \in \mathbb{R}^p$, let $d(x,A) = \inf_{y \in A} d(x,y)$.

We also denote by $B_r(x)$ the open Euclidean ball of radius $r > 0$ and centered at $x \in \mathbb{R}^p$ while $\overline{B}_r(x)$ is the associated closed ball. In particular, $B$ denotes the unit Euclidean ball $\overline{B}_1(0)$.

We denote by

$$\omega_p := \frac{2\pi^{\frac{p+1}{2}}}{\Gamma\left(\frac{p}{2}+1\right)} \tag{2.1}$$

the surface area of the $p$-dimensional unit sphere in $\mathbb{R}^{p+1}$. We denote by

$$c_r := \frac{\Gamma(p/2 + r + 1)}{\pi^{p/2}\Gamma(r+1)} \tag{2.2}$$

the normalization constant of the measure $\nu_r$ whose density is $(1 - \|z\|_2^2)^r$ on the unit ball $B$ (see *e.g.* [45], page 2441, (2.2)). For $\alpha \in \mathbb{R}$ and $k \in \mathbb{N}$, the associated binomial coefficient is defined as follows:

$$\binom{\alpha}{k} := \frac{\Gamma(\alpha+1)}{\Gamma(k+1)\Gamma(\alpha-k+1)} = \frac{\alpha(\alpha-1)(\alpha-2)\dots(\alpha-k+1)}{k!}.$$

Finally, for two positive quantities $a_n$ and $b_n$ depending on the sample size $n$ (and potentially depending on other constant quantities), we write $a_n \propto b_n$ when both quantities $a_n/b_n$ and $b_n/a_n$ are upper bounded uniformly in $n$. We proceed similarly, for two positive quantities $a_d$ and $b_d$ depending on a degree bound $d$ but not on the sample size $n$ (and potentially depending on other constant quantities).

## 2.2. Problem setting

The following notation and assumptions will be standing throughout the text.

**Assumption 2.1.**

1. $\mu$ is a Borel probability measure on $\mathbb{R}^p$ and its support $S := \operatorname{supp}(\mu)$ is compact with nonempty interior.
2. $n \in \mathbb{N}$, $n > 0$ is fixed and $X_1, \ldots, X_n$ are independent and identically distributed random vectors with distribution $\mu$. The corresponding empirical measure is denoted by

$$\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}, \tag{2.3}$$

where $\delta_x$ is the dirac measure at $x \in \mathbb{R}^p$.

Using the notation of Assumption 2.1, given the sample $(X_i)_{i=1}^n$ our goal is to build an estimator $S_n(X_1, \ldots, X_n) \subset \mathbb{R}^p$ in order to approximate $S$. We construct a specific kind of estimator $S_n$ based on the empirical Christoffel function. The rest of this section is dedicated to the presentation of further background needed to define our estimator. Convergence of our estimator to $S$ using different criteria is described next in Section 3.

## 2.3. The Christoffel function

### *2.3.1. Multivariate polynomials*

Polynomials of $p$ variables are indexed by the set $\mathbb{N}^p$ of multi-indices. For example, given a set of $p$ variables $x = (x_1, \ldots, x_p)$ and a multi-index $\alpha = (\alpha_1, \ldots, \alpha_p) \in \mathbb{N}^p$, the *monomial* $x^\alpha$ is given by $x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_p^{\alpha_p}$ which *degree* is

$$\deg x^\alpha = |\alpha| = \sum_{i=1}^{p} \alpha_i.$$

The *space of polynomials of degree at most d* is the linear span of monomials of degree up to $d$:

$$\Pi_d^p := \operatorname{span}\{x^\alpha : \alpha \in \mathbb{N}^p, |\alpha| \leq d\}.$$

The *space of polynomials of p variables* is

$$\Pi^p := \bigcup_{d \in \mathbb{N}} \Pi_d^p.$$

The *degree* of a polynomial $P \in \Pi^p$, denoted by $\deg P$, is the maximum degree of its monomial associated to a nonzero coefficient (the null polynomial has degree 0). Note that $\dim \Pi_d^p = \binom{d+p}{d}$. We denote by $s(d)$ the quantity $\binom{d+p}{d}$ throughout the text.

### *2.3.2. Orthonormal polynomials*

Since $\mu$ satisfies Assumption 2.1 (see also Remark 2.3), we have the following inner product:

$$\langle P, Q\rangle_\mu = \int\limits_{\mathbb{R}^p} P(x)Q(x)\mathrm{d}\mu(x),$$

where $P, Q$ are polynomials. A sequence of *orthonormal polynomials with respect to* $\mu$ is a sequence of polynomials $\{P_\alpha : \alpha \in I\}$ in $\Pi^p$ such that $\langle P_\alpha, P_\beta\rangle_\mu = \delta(\alpha, \beta)$ [2] for all $\alpha, \beta \in I$. The Gram-Schmidt orthonormalization process guarantees the existence of such an orthonormal sequence. Restricting the degree up to $d \in \mathbb{N}$, we obtain a sequence of orthonormal polynomials $\{P_j : 1 \leq j \leq s(d)\}$, which is also a basis of $\Pi^p_d$.

### *2.3.3. Moment matrix*

Now, let $\{P_j : 1 \leq j \leq s(d)\}$ be a basis of $\Pi^p_d$ (not necessarily orthonormal). We denote

$$\begin{aligned} v_d: \quad \mathbb{R}^p &\longrightarrow \mathbb{R}^{s(d)} \\ x &\longmapsto (P_1(x), P_2(x), \ldots, P_{s(d)}(x))^T. \end{aligned}$$

The *moment matrix of* $\mu$ *with respect to the basis* $\{P_j\}_{j=1}^{s(d)}$ is a square matrix of dimension $s(d)$ which is defined by

$$M_{\mu,d} = \int\limits_{\mathbb{R}^p} v_d(x)v_d(x)^T\mathrm{d}\mu(x), \tag{2.4}$$

where the integral is taken entry-wise. We have the following property of the moment matrix which is useful in the sequel.

**Lemma 2.2.** *Let* $P, Q \in \Pi^p_d$ *have representations with respect to the basis* $\{P_j : 1 \leq j \leq s(d)\}$ *of the form:*

$$P = \sum_{j=1}^{s(d)} (c_P)_j P_j = c_P^T v_d, \quad Q = \sum_{j=1}^{s(d)} (c_Q)_j P_j = c_Q^T v_d,$$

*where* $c_P, c_Q \in \mathbb{R}^{s(d)}$. *Then*

$$\int\limits_{\mathbb{R}^p} P(x)Q(x)\mathrm{d}\mu(x) = c_P^T M_{\mu,d} c_Q.$$

**Remark 2.3.** $M_{\mu,d}$ is a symmetric, positive definite square matrix of dimension $s(d)$. In fact, for any $c \in \mathbb{R}^{s(d)}$, we set $P_c = \sum\limits_{j=1}^{s(d)} c_j P_j$ and by Lemma 2.2, we have

$$c^T M_{\mu,d} c = \int\limits_{\mathbb{R}^p} P_c(x)^2\mathrm{d}\mu(x) \geq 0. \tag{2.5}$$

[2] $\delta(\alpha, \beta)$ is 1 if $\alpha = \beta$, 0 otherwise.

Since $\operatorname{Int} S \neq \emptyset$, $S$ is polynomial determining, that is, the equality of two polynomials is implied from their equality on the support. By combining this fact with (2.5), we obtain that $M_{\mu,d}$ is positive definite.

### *2.3.4. The Christoffel – Darboux kernel*

The space of polynomials of degree at most $d$ along with the inner product defined by $\mu$ $(\Pi_d^p, \langle . , . \rangle_\mu)$ is then a finite-dimensional Hilbert space of functions from $\mathbb{R}^p$ to $\mathbb{R}$ and $\dim \Pi_d^p = s(d)$. Moreover, $(\Pi_d^p, \langle . , . \rangle_\mu)$ is a reproducing kernel Hilbert space (RKHS) (see for example [3]). Indeed, we notice that the function $P \mapsto P(x)$ is linear on the space of polynomials and $\Pi_d^p$ is finite-dimensional (hence all norms are equivalent), therefore we obtain the continuity of this function on $(\Pi_d^p, \langle . , . \rangle_\mu)$ for any $x \in \mathbb{R}^p$. This property of $(\Pi_d^p, \langle . , . \rangle_\mu)$ guarantees the existence and uniqueness of a reproducing kernel which is defined as follows.

**Definition 2.4.** The *Christoffel – Darboux kernel*, denoted by $\kappa_{\mu,d}$, is the reproducing kernel of the RKHS $(\Pi_d^p, \langle . , . \rangle_\mu)$, *i.e.* for all $x \in \mathbb{R}^p$ and $P \in \Pi_d^p$, we have $\kappa_{\mu,d}(x, .) \in \Pi_d^p$ and

$$\langle P, \kappa_{\mu,d}(x, .) \rangle_\mu = \int_{\mathbb{R}^p} P(y)\kappa_{\mu,d}(x, y)\mathrm{d}\mu(y) = P(x).$$

The two following propositions are explicit formulas for the Christoffel – Darboux kernel. The first one is its expression as a sum of squares of orthonormal polynomials, while the other is a computation based on the moment matrix (and does not require an orthonormal basis).

**Proposition 2.5** (see *e.g.* [5], page 7 or [18], page 97, (3.6.3))**.** *Let* $\{P_j\}_{j=1}^{s(d)}$ *be an orthonormal basis of* $\Pi_d^p$ *with respect to* $\mu$*. Then for all* $x, y \in \mathbb{R}^p$

$$\kappa_{\mu,d}(x, y) = \sum_{j=1}^{s(d)} P_j(x)P_j(y).$$

**Proposition 2.6** (see *e.g.* [24], page 7, (3.1))**.** *Let* $v_d = (P_1, P_2, \ldots, P_{s(d)})^T$ *be a basis of* $\Pi_d^p$ *and* $M_{\mu,d}$ *be the corresponding moment matrix (see* (2.4)*). For all* $x, y \in \mathbb{R}^p$*, we have*

$$\kappa_{\mu,d}(x, y) = v_d(x)^T M_{\mu,d}^{-1} v_d(y).$$

**Remark 2.7.** By Proposition 2.5,

$$\kappa_{\mu,d}(x, x) = \sum_{j=1}^{s(d)} P_j(x)^2 \geq 0,$$

where $\{P_j : 1 \leq j \leq s(d)\}$ is an orthonormal basis of $\Pi_d^p$. Moreover, the $P_j(x)$ cannot be all 0 since otherwise, the polynomial 1 will be 0 at point $x$, which is impossible. So $\kappa_{\mu,d}(x, x) > 0$ for all $x \in \mathbb{R}^p$.

### *2.3.5. The Christoffel function*

Now, we will define the (population) Christoffel function and provide some of its properties which are useful for the sequel. The fact that the min exists in the next definition follows from *e.g.* [18], Theorem 3.6.6.

**Definition 2.8.** Let $d \in \mathbb{N}$. The *Christoffel function* associated to $\mu$ and $d$ is the function

$$\Lambda_{\mu,d} : \mathbb{R}^p \longrightarrow \mathbb{R}_+$$

$$z \longmapsto \min\left\{\int_{\mathbb{R}^p} P^2 \mathrm{d}\mu : P \in \Pi_d^p, P(z) = 1\right\}.$$

The following proposition is an equivalent definition of the Christoffel function, relying on the Christoffel-Darboux kernel.

**Proposition 2.9** (see *e.g.* [18], Thm. 3.6.6)**.**

$$\Lambda_{\mu,d}(z) = \frac{1}{\kappa_{\mu,d}(z,z)}.$$

Above, recall that the Christoffel – Darboux kernel is positive. We now highlight the following properties of the Christoffel function which will be useful in the sequel. The following proposition guarantees the equivariance of the Christoffel function by affine transformations.

**Proposition 2.10** (see *e.g.* [29], Lem. 1)**.** *Let $A$ be an invertible affine map from $\mathbb{R}^p$ to $\mathbb{R}^p$. Recall that $\mu_{\#A}$ is the push-forward measure of $\mu$ by $A$. Then for all $x \in \mathbb{R}^p$,*

$$\Lambda_{\mu,d}(x) = \Lambda_{\mu_{\#A},d}(Ax).$$

The next proposition expresses the monotonicity property of the Christoffel function. It is a direct consequence of Definition 2.8.

**Proposition 2.11.** *If $\nu$ is a Borel measure on $\mathbb{R}^p$, such that $\nu \leq \mu$, in the sense that $\nu(K) \leq \mu(K)$ for all Borel sets $K$, then for all $x \in \mathbb{R}^p$,*

$$\Lambda_{\nu,d}(x) \leq \Lambda_{\mu,d}(x).$$

**Remark 2.12.** All the previous definitions and results extend straightforwardly to the case when $\mu$ does not have unit mass (see Asm. 2.1). This extension is a simple scaling. This is a very slight abuse of notation that we will some times do without mentioning it (for instance in Prop. 2.11).

## 2.4. The empirical Christoffel function

The Christoffel function associated to $\mu_n$ (see Asm. 2.1), $\Lambda_{\mu_n,d}$ is called the *empirical Christoffel function*. It is to be compared to the *population Christoffel function* $\Lambda_{\mu,d}$. The convergence of the empirical Christoffel function towards its population counterpart as $n \to \infty$ and for a fixed $d$ has been shown in [24]. Furthermore, [24] show that, as $d \to \infty$, $\Lambda_{\mu,d}$ takes asymptotically much larger values in the support $S$ (polynomial decay) than outside (exponential decay).

Since $\Lambda_{\mu,d}$ is unknown but $\Lambda_{\mu_n,d}$ is observed, our objective is thus, with a careful choice of threshold $\gamma > 0$ and degree $d \in \mathbb{N}$, as functions of $n$, to construct a sequence of polynomial sublevel sets

$$\{x \in \mathbb{R}^p : \Lambda_{\mu_n,d}(x) \geq \gamma\}$$

which estimate the support $S$. It is worth mentioning that the empirical Christoffel function $\Lambda_{\mu_n,d}$ can be computed using the inversion of a square matrix of size $s(d)$ thanks to Proposition 2.6.

Assuming that $n \geq s(d)$, and that the empirical moment matrix $M_{\mu_n,d}$ (see (2.4) with $\mu$ replaced by $\mu_n$) is invertible (this being true with probability one for example if $\mu$ has a density, see for example [30]), by Propositions 2.5 and 2.9, we have

$$\int_{\mathbb{R}^p} \frac{1}{\Lambda_{\mu_n,d}(z)} \mathrm{d}\mu_n(z) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\Lambda_{\mu_n,d}(X_i)} = \frac{1}{n}\sum_{i=1}^{n} \kappa_{\mu_n,d}(X_i, X_i) = s(d).$$

By Jensen's inequality, we deduce that in this case $\frac{1}{n}\sum_{i=1}^n s(d)\Lambda_{\mu_n,d}(X_i) \geq 1$. This holds with equality for two special values.

- If $d = 0$, then $s(d) = 1$ and following Definition 2.8, we have $\Lambda_{\mu_n,d}(z) = 1$ for all $z$ and so $\frac{1}{n}\sum_{i=1}^n s(d)\Lambda_{\mu_n,d}(X_i) = 1$.
- If $s(d) = n$, for each $i = 1,\dots,n$, let $M_{\mu_{n,-i},d}$ be the empirical moment matrix obtained from (2.4) with $\mu$ replaced by $\mu_{n,-i}$, the empirical distribution with $X_i$ removed from $\mu_n$. The matrix $M_{\mu_{n,-i},d}$ is not full rank so from (2.5) (with $\mu$ replaced by $\mu_{n,-i}$), there exists a non-zero $P \in \Pi_d^p$ such that $P(X_j) = 0$ for $j \neq i$. We have $P(X_i) \neq 0$ because $M_{\mu_n,d}$ is invertible and again from (2.5). Then by renormalization of $P$, using Definition 2.8 we have $\Lambda_{\mu_n,d}(X_i) = \frac{1}{n} = \frac{1}{s(d)}$ so that $\frac{1}{n}\sum_{i=1}^n s(d)\Lambda_{\mu_n,d}(X_i) = 1$.

So the typical value of $\Lambda_{\mu_n,d}(X_i)$ is greater than $1/s(d)$ with equality in two extreme cases of very small and very large degree for $n$ fixed. The Christoffel function scaled by $s(d)$ is indeed a key quantity. In typical examples, $s(d)\Lambda_{\mu,d}(z)$ converges as $d \to \infty$ to a product of two terms, one accounting for the geometry of the support (the so called equilibrium measure from potential theory), one accounting for the density of $\mu$ [9, 39, 44]. The most general multivariate description of this phenomenon is found in [6], the equilibrium measure and density argument is found in [22]. As discussed above, the limit is zero outside the support of $\mu$.

# 3. Main results

## 3.1. Overview

From now on, we consider the case where the probability measure $\mu$ has density $w$ with respect to Lebesgue measure. Our main result is that for a large enough number of observations $n$, by choosing pertinently a degree $d_n \in \mathbb{N}$ and a threshold $\gamma_n > 0$ for the empirical Christoffel function $\Lambda_{\mu_n,d_n}$, we obtain a sequence of polynomial sublevel sets

$$S_n := \{x \in \mathbb{R}^d : \Lambda_{\mu_n,d_n}(x) \geq \gamma_n\}$$

which approximates the support of $\mu$. More explicitly, we show that under smoothness assumptions on $S$, $S_n$ is close to $S$ both in Hausdorff distance and Lebesgue measure of their symmetric difference. For any $\epsilon \in (0,1)$, we obtain an explicit convergence rate of order

$$n^{-\frac{1-\epsilon}{p+2r+2}}, \tag{3.1}$$

where $r$ measures the speed of decrease of the density of $\mu$, $w$, close to $\partial S$, see Assumption 3.5.

Those results are obtained from the following materials:

1. Properties of the population Christoffel function. We provide a lower bound on the Christoffel function $\Lambda_{\mu,d}$ in the interior of the support $S$ and an upper bound in the exterior of $S$. We also provide a bound on the supremum of the Christoffel-Darboux kernel $\kappa_{\mu,d}$ on $S$. Those results will be discussed in Appendices B and C.

2. Concentration results for the speed of convergence of the empirical Christoffel function $\Lambda_{\mu_n,d}$ to its population counterpart $\Lambda_{\mu,d}$. This part requires the above mentioned bound on the supremum of the Christoffel – Darboux kernel. Those results could be of independent interest and will be discussed in Section 3.4 with all the proofs in Appendix D.

3. We introduce a thresholding scheme using the empirical Christoffel function $\Lambda_{\mu_n,d_n}$ as in (3.3) by a careful tuning of the degree $d$ and the threshold $\gamma$ in the limit of large sample size $n$. With this thresholding scheme, we prove the desired results described in (3.1). The details will be in Section 3.3 with proofs postponed to Appendix E.

## 3.2. Conditions on the support and the density

Throughout the text, we consider a probability measure $\mu$ which is supported on $S \subset \mathbb{R}^p$ and has density $w \geq 0$.

### *3.2.1. Assumptions on the support S*

We first introduce the following definitions, notation and assumptions.

**Definition 3.1.** Consider a closed set $F \subset \mathbb{R}^p$ and a constant $R > 0$. We say that a ball of radius $R$ rolls inside $F$ if for any $x \in F$, there exists a ball $B_x$ centered at $z_x$ of radius $R$ such that $x \in B_x \subset F$. If a ball of radius $R$ rolls inside $\overline{F^c}$, then we say that a ball of radius $R$ rolls outside $F$.

**Definition 3.2.** Consider a closed set $F \subset \mathbb{R}^p$. Denote by $F^\epsilon$ the $\epsilon$-extension of $F$, defined as

$$F^\epsilon = \{x \in \mathbb{R}^p : d(x, F) \leq \epsilon\}.$$

We also define the volume function

$$\begin{aligned} V_F : \mathbb{R}_+ &\longrightarrow \mathbb{R}_+ \\ \epsilon &\longmapsto \lambda(F^\epsilon), \end{aligned}$$

where we recall that $\lambda(.)$ denotes the Lebesgue measure of a set.

**Assumption 3.3.** $S \subset \mathbb{R}^p$ is a compact set with nonempty interior. Furthermore, there exists $R > 0$ such that a ball of radius $R$ rolls inside and outside $S$.

We will rely on Assumption 3.3, in particular the rolling ball part, for our results and proofs. This latter assumption is made relatively frequently in the support inference literature, see for instance [15]. It is interpreted as meaning that the boundary of $S$ is smooth. In particular, it prevents corners in the boundary of $S$. The case of sets $S$ with non-smooth boundaries is a future research topic of interest that is not addressed here for the sake of concision. The following result will be needed when working with the Lebesgue measure of the symmetric difference, the proof is given in Appendix F.

**Lemma 3.4.** *Let $S \subset \mathbb{R}^p$ satisfy Assumption 3.3 and let $c > 0$. Then, there is $C_S > 0$ such that for all $0 < \epsilon < c$,*

$$V_{\partial S}(\epsilon) \leq \epsilon\, C_S.$$

Following [43, Theorem 1], there is a geometric sufficient condition for Assumption 3.3. If $S \subset \mathbb{R}^p$ has finitely many path-connected components, each with non-empty interior, and is compact, then it satisfies Assumption 3.3 if and only if its boundary $\partial S$ is a $C^1$ submanifold of $\mathbb{R}^p$, of dimension $p-1$ and its unit outer pointing normal vector, $n \colon \partial S \to \mathbb{R}^p$ is globally Lipschitz on $\partial S$, with constant $1/R$, that is

$$\|n(x) - n(y)\| \leq \frac{1}{R}\|x - y\|, \quad \forall x, y \in \partial S.$$

Note that the norm is that of $\mathbb{R}^p$, the proof of Lemma 3.4 is built on this construction.

### *3.2.2. Assumption on the density w*

Now, for $\delta > 0$, we set

$$L(\delta) := \inf\{w(x) : x \in S, d(x, \partial S) \geq \delta\}.$$

The next assumption concerns the rate of decay of the density of $\mu$ at the boundary of the support $S$.

**Assumption 3.5.** The density $w : S \to \mathbb{R}$ is such that for all $\delta \geq 0$, we have

$$L(\delta) \geq C\delta^r,$$

where $C > 0$ and $r \geq 0$ are fixed constants (depending only on $\mu$).

## 3.3. Main results for support estimation

### *3.3.1. Thresholding scheme*

First, we design our $n$-dependent thresholding scheme using the empirical Christoffel function $\Lambda_{\mu_n,d}$. This thresholding scheme depends on the constants $R, C, r$ given by the assumptions on $\mu$ (Asms. 3.3 and 3.5). It also depends on a constant $\epsilon \in (0,1)$ which can be made arbitrarily small (a smaller $\epsilon$ leads to a better rate of convergence, but possibly worse constants), and on a constant $\alpha \in (0,1)$ which is in principle a small risk threshold, such that our results hold with probability $1-\alpha$.

The $n$-dependent thresholding scheme relies on a sequence of degrees $d_n$ of order

$$d_n \propto n^{\frac{1}{p+2r+2}}, \tag{3.2}$$

which full expression is given in (A.3) in Appendix A. Recall that $\propto$ is defined in Section 2.1. Based on $d_n$, we define a sequence of thresholds $\gamma_n$ of order

$$\gamma_n \propto n^{-\frac{p(2-\epsilon)+(1-\epsilon)r}{p+2r+2}},$$

which full expression is given in (A.4) in Appendix A. The thresholding scheme, for support estimation, is then

$$S_n := \{x \in \mathbb{R}^p : \Lambda_{\mu_n,d_n}(x) \geq \gamma_n\}. \tag{3.3}$$

The explicit results for this thresholding scheme will be presented in the next subsections.

### *3.3.2. Result for the Hausdorff distance between two sets and two boundaries*

Recall the definition of the Hausdorff distance between two subsets $A, B$ of $\mathbb{R}^p$:

$$d_H(A,B) = \max\left(\sup_{x\in A} d(x,B), \sup_{y\in B} d(y,A)\right).$$

The following result provides an explicit quantitative rate of convergence for the estimation of $S$ using the thresholding scheme (3.3) based on the empirical Christoffel function. More explicitly, this estimation of $S$ by $S_n$ is measured by the Hausdorff distance between them and between their boundaries. Thus, this theorem is one of the most important results of this paper.

**Theorem 3.6.** *Let $S \subset \mathbb{R}^p$ satisfy Assumption 3.3 with radius $R > 0$ and $w : S \longrightarrow \mathbb{R}$ satisfy Assumption 3.5 with two constants $C > 0$ and $r \geq 0$. Let $\mu$ be the measure supported on $S$ with density $w$ with respect to Lebesgue measure. Then, there is a constant $n_0 \in \mathbb{N}$ (with full expression given in (A.2) in Appendix A) such that for $n \geq n_0$, the thresholding scheme (3.3) satisfies with probability at least $1-\alpha$ that*

$$d_H(S, S_n) \leq \delta_n$$

*and*

$$d_H(\partial S, \partial S_n) \leq \delta_n,$$

*where*

$$\delta_n \propto n^{-\frac{1-\epsilon}{p+2r+2}},$$

*with full expression given in* (A.5) *in Appendix A.*

### *3.3.3. Result for the Lebesgue measure of the symmetric difference between two sets*

Recall the definition of the symmetric difference between two subsets $A, B$ of $\mathbb{R}^p$:

$$A \triangle B = (A \setminus B) \cup (B \setminus A).$$

In this section, in order to measure the convergence of the estimator $S_n$ to the true set $S$, we will use the Lebesgue measure of their symmetric difference:

$$(A, B) \longmapsto \lambda(A \triangle B).$$

The following result, which is a counterpart of Theorem 3.6 for the Lebesgue measure of the symmetric difference, is the second main result of this paper.

**Corollary 3.7.** *Let $S \subset \mathbb{R}^p$ satisfy Assumption 3.3 with radius $R > 0$ and $w : S \longrightarrow \mathbb{R}$ satisfy Assumption 3.5 with two constants $C > 0$ and $r \geq 0$. Let $\mu$ be the measure supported on $S$ with density $w$ with respect to Lebesgue measure. Consider then $\delta_n \propto n^{-\frac{1-\epsilon}{p+2r+2}}$ as defined in Theorem 3.6. Let $C_S > 0$ be given by Lemma 3.4 with some $c$ such that $c > 2 \max_{n \in \mathbb{N}} \delta_n$. Then, for $n \geq n_0$, with $n_0$ as in Theorem 3.6, the thresholding scheme* (3.3) *satisfies with probability at least $1 - \alpha$ that*

$$\lambda(S \triangle S_n) \leq 2 C_S \delta_n.$$

**Remark 3.8.** The order of magnitude of the error for the thresholding scheme (3.3) is $n^{-\frac{1-\epsilon}{p+2r+2}}$ for both the Hausdorff distance between two sets and between their boundaries as well as the Lebesgue measure of their symmetric difference. Since $\epsilon \in (0, 1)$ can be taken arbitrarily small, the rate of convergence is essentially $n^{-\frac{1}{p+2r+2}}$.

**Remark 3.9.** The tuning of $d_n$ and $\gamma_n$ (see (A.3) and (A.4) in Appendix A) depends on the constants $C$ and $r$ from Assumption 3.5 and on the constant $R$ from Assumption 3.3. In practice, these constants are typically unknown. In the numerical simulations of Section 4, once $d_n$ is selected, $\gamma_n$ is taken as large as possible, under the constraint that the resulting $S_n$ contains all the observed points. We would recommend this choice for $\gamma_n$ in general. In Section 4, we also suggest a simple fully data driven heuristic to select $d_n$, based on considering the normalized Christoffel function $s(d)\Lambda_{\mu,d}$ as a proxy to a "density" (see also Sect. 2.4) and on taking the value of $d$ which associates the most "mass" to the observations. While the numerical performances of this heuristic are encouraging, we leave further empirical validation of this selection procedure, and the development of associated theory, as a topic of future research.

On a theoretical level, the main aim of this paper is to show that it is possible to obtain rates of convergence, by selecting $d_n$ and $\gamma_n$ according to the constants $C$, $r$ and $R$. For the sake of concision, the situation where $C$, $r$ and $R$ are estimated from data is not studied in this paper. Let us nevertheless discuss it briefly here. First, we remark that if Assumptions 3.5 and 3.3 hold with constants $C$, $r$ and $R$, then they hold a fortiori with constants $C' < C$, $r' > r$ and $R' < R$. Hence, in order to obtain rates of convergence, it is sufficient to tune $d_n$ and $\gamma_n$ based on conservative values of $C$ and $R$ that are overly small and of $r$ that are overly large, such that Assumptions 3.5 and 3.3 hold. Obtaining conservative values is statistically easier than obtaining the sharpest possible values of $C$, $r$ and $R$ such that Assumptions 3.5 and 3.3 hold. Another important question is adaptivity: obtaining a procedure based on the Christoffel function, with no knowledge of the values of $C$, $r$ and $R$ such that Assumptions 3.5 and 3.3 hold, and which yields the same rates of convergence as when knowing the sharpest values of $C$, $r$ and $R$ such that Assumptions 3.5 and 3.3 hold.

### *3.3.4. Sketch of proof of Theorem 3.6*

First, we suppose that the estimation of the population Christoffel function by its empirical counterpart can be controlled. More explicitly, we assume that there exists a constant $\beta < 1$ such that for all $x \in \mathbb{R}^p$,

$$|\Lambda_{\mu,d}(x) - \Lambda_{\mu_n,d}(x)| \leq \Lambda_{\mu,d}(x)\beta, \tag{3.4}$$

or equivalently

$$(1-\beta)\Lambda_{\mu,d}(x) \leq \Lambda_{\mu_n,d}(x) \leq (1+\beta)\Lambda_{\mu,d}(x). \tag{3.5}$$

Now we introduce a sequence of polynomial sublevel sets which estimates the support $S$ using the empirical function $\Lambda_{\mu_n,d}$ where $d$ does not depend on $n$. For $0 < \epsilon < 1$ fixed and for $d \in \mathbb{N}$, we define

$$\gamma_d \propto \frac{1}{d^{p(2-\epsilon)+(1-\epsilon)r}},$$

with full expression given in (A.9) in Appendix A. We then let

$$S_{d,n} := \{x \in \mathbb{R}^p : \Lambda_{\mu_n,d}(x) \geq \gamma_d\}. \tag{3.6}$$

The idea of this estimator $S_{d,n}$ comes from [26], Section 4.1. The difference is that we let $0 < \epsilon < 1$ arbitrarily small for a better rate of convergence (instead of setting $\epsilon = 1/2$ like in [26]). Moreover, by choosing carefully the threshold, we obtain an estimator $S_{d,n}$ such that not only $S_{d,n}$ is contained in a small enlargement of $S$ (which has been shown in [26]), but we also have a small enlargement of $S_{d,n}$ that contains $S$. The explicit result is as follows.

**Lemma 3.10.** *Let $S$ be a compact set with non-empty interior, $w : S \longrightarrow \mathbb{R}$ satisfy Assumption 3.5 with two constants $C > 0$, $r \geq 0$ and $\mu$ be the measure supported on $S$ with density $w$. Assume that there exists a constant $\beta < 1$ for which (3.4) holds. We define*

$$S^1 := \{x \in \mathbb{R}^p : d(x, S) \leq \delta_1(d)\}$$

*and*

$$S^2 := \{x \in S : d(x, \partial S) \geq \delta_2(d, \beta)\},$$

*where $\delta_1(d) \propto d^{-(1-\epsilon)}$, $\delta_2(d,\beta) \propto d^{-(1-\epsilon)}$ are defined in (A.6) and (A.7) respectively, in Appendix A. Then the thresholding scheme (3.6) satisfies that*

$$S^2 \subset S_{d,n} \subset S^1.$$

This above relation between $S$ and $S_{d,n}$ is important since it implies that the difference between $S$ and $S_{d,n}$ is controlled by

$$\delta_d = \max\left(\delta_1(d), \delta_2(d,\beta)\right) \propto d^{-(1-\epsilon)}. \tag{3.7}$$

Now, under Assumptions 3.3 and 3.5 and thanks to the concentration results in Section 3.4, we can select $d_n$ such that (3.4) holds with high probability with $\beta = 1/2$. Subsequently, we can select a threshold $\gamma_n$ that will optimize the convergence rate of $S_n$ to $S$. We obtain now the thresholding scheme (3.3) and the result regarding the Hausdorff distance.

All the proofs' details are postponed to Appendix E for the sake of clarity.

### 3.4. A concentration result for the approximation of the Christoffel function by its empirical counterpart

Let $\mu$ be a measure which satisfies Assumption 2.1 and $\mu_n$ be the corresponding empirical measure. We consider now the speed of convergence of the empirical Christoffel function $\Lambda_{\mu_n,d}$ towards $\Lambda_{\mu,d}$. All the proofs of the following results will be postponed to Appendix D.

First, we state below a technical lemma which bounds uniformly the quantity $|\Lambda_{\mu_n,d} - \Lambda_{\mu,d}|/\Lambda_{\mu,d}$ by the operator norm of a moment-based random matrix.

**Lemma 3.11.** *Let $v_d = \{P_j : 1 \le j \le s(d)\}$ be a basis of orthonormal polynomials with respect to $\mu$. Denote by $M_{\mu_n,d}$ the moment matrix of $\mu_n$ with respect to the basis $v_d$ (see Sect. 2.3). Then for all $x \in \mathbb{R}^p$, we have*

$$\left|\Lambda_{\mu,d}(x) - \Lambda_{\mu_n,d}(x)\right| \le \Lambda_{\mu,d}(x)\,\|M_{\mu_n,d} - I_{s(d)}\|$$

*where we recall that the norm of $s(d) \times s(d)$ matrices is the operator norm.*

Note that $I_{s(d)}$ is actually the associated moment matrix of $\mu$ with respect to the basis $v_d$. Now, to control the operator norm of the random matrix $M_{\mu_n,d} - I_{s(d)}$, we rely on Theorem 5.44 from [41]. The following theorem makes use of this random matrix result and of Lemma 3.11 to obtain an upper bound for the quantity $|\Lambda_{\mu_n,d} - \Lambda_{\mu,d}|/\Lambda_{\mu,d}$ with high probability.

**Theorem 3.12.** *Let $\mu$ be a measure which satisfies Assumption 2.1 and $\mu_n$ be the corresponding empirical measure. Then for all $x \in \mathbb{R}^p$ and $\alpha > 0$, we have*

$$\left|\Lambda_{\mu,d}(x) - \Lambda_{\mu_n,d}(x)\right| \le \Lambda_{\mu,d}(x)\,\max\left(\sqrt{\frac{16m}{3n}\log\frac{s(d)}{\alpha}}, \frac{16m}{3n}\log\frac{s(d)}{\alpha}\right)$$

*with probability at least $1-\alpha$, where*

$$m = \sup_{x\in\operatorname{supp}\mu} \kappa_{\mu,d}(x,x).$$

Note that in our case, the supremum of the Christoffel – Darboux kernel $m$ has a quantitative upper bound of order $d^{p+2r+1}$ which is of independent interest and will be provided in Appendix C. The following corollary is a consequence of Theorem 3.12 combined with Theorem C.3, and is useful in the tuning of $d_n$ for the thresholding scheme (3.3).

**Corollary 3.13.** *Let $S \subset \mathbb{R}^p$ satisfy Assumption 3.3 with radius $R > 0$ and $w : S \longrightarrow \mathbb{R}$ satisfy Assumption 3.5 with two constants $C > 0$ and $r \ge 0$. Let $\mu$ be the measure supported on $S$ with density $w$ with respect to Lebesgue* *measure and $\mu_n$ be the corresponding empirical measure. Then for all $d \geq 2$, $x \in \mathbb{R}^p$ and $\alpha > 0$, we have*

$$\begin{aligned} & |\Lambda_{\mu,d}(x) - \Lambda_{\mu_n,d}(x)| \\ \leq \quad & \Lambda_{\mu,d}(x) \max \left( \sqrt{\frac{16\, m(d,p,S,w)}{3n} \log \frac{s(d)}{\alpha}}, \frac{16\, m(d,p,S,w)}{3n} \log \frac{s(d)}{\alpha} \right) \end{aligned}$$

*with probability at least $1-\alpha$, where*

$$m(d,p,S,w) \propto d^{p+2r+1},$$

*with full expression given in* (A.10) *in Appendix A.*

**Remark 3.14.** When the dimension $p$ is fixed and $n$ is large, this uniform upper bound in high probability of $|\Lambda_{\mu,d} - \Lambda_{\mu_n,d}|/\Lambda_{\mu,d}$ is of order $\sqrt{d^{p+2r+1}/n}$, up to multiplicative $\log(d)$ factors.

## 4. Numerical illustration

### 4.1. A heuristic to tune $d$ and $\gamma$

For a given fixed $d$, an easy heuristic is to select $\gamma > 0$ such that $\Lambda_{\mu_n,d}(X_i) \geq \gamma$ for all $i = 1, \ldots, n$. Indeed we have $X_i \in \mathrm{supp}(\mu)$ for $i = 1, \ldots, n$ almost surely so that if this constraint is not satisfied, our estimate will inevitably miss a portion of the support. The proposed heuristic for $\gamma$ is to choose the largest such $\gamma$ which is the minimal value of $\Lambda_{\mu_n,d}(X_i)$ for $i = 1 \ldots n$.

As for the choice of $d$, following the discussion in Section 2.4, assuming that $n \geq s(d)$, we have

- $\frac{1}{n}\sum_{i=1}^n s(d)\Lambda_{\mu_n,d}(X_i) \geq 1$,
- the inequality holds with equality in two extreme cases, when $d = 0$ or $d$ is such that $s(d) = n$.

Choosing values of $d$ such that $s(d) > n$ would result in lack of invertibility for the moment matrix so we consider $d_{\min} = 1$ and $d_{\max}$ the largest value of $d$ such that $s(d) < n$. We will restrict the choice of $d$ in $\{d_{\min}, \ldots, d_{\max}\}$.

As described in Section 2.4, the term $s(d)\Lambda_{\mu,d}(\cdot)$ has asymptotically positive values on the support of $\mu$ and zero value outside. Our heuristic is to consider this term as a proxy to a "density" and take the value of $d$ which associates the most "mass" to the observations. All in all, we have the following

$$\begin{aligned} \hat{d}_n = \quad & \operatorname*{arg\,max}_{s(d)<n} \quad \frac{1}{n}\sum_{i=1}^n s(d)\Lambda_{\mu_n,d}(X_i) \\ \hat{\gamma}_n = \quad & \min_{i=1,\ldots,n} \quad \Lambda_{\mu_n,\hat{d}_n}(X_i) \\ \hat{S}_n = \quad & \{x \in \mathbb{R}^p : \Lambda_{\mu_n,\hat{d}_n}(x) \geq \hat{\gamma}_n\}. \end{aligned}$$

This procedure has the advantage to be fully data driven. It provides a reasonable choice for the parameters as illustrated with numerical simulations in the present section. A detailed study of this heuristic is beyond the scope of the present paper but constitutes an interesting question for future research.

### 4.2. Empirical comparison with the Devroye-Wise estimator

#### *4.2.1. Devroye-Wise estimator*

The estimator proposed by [16], which we will denote by DW, is simply a union of balls, given $r > 0$,

$$\hat{S}_{n,r} = \cup_{i=1}^n \bar{B}_r(X_i).$$

In order to define this estimator, one needs to choose the radius $r$. We will consider a data driven heuristic for this purpose: choose $r$ by leave-one-out. This corresponds to choose the smallest $r > 0$ such that for all $i$, $X_i$ is contained in the DW estimator evaluated on the same dataset with $X_i$ removed. This yields

$$\hat{r}_n = \max_{i=1,\ldots,n} \min_{j \neq i} \|X_j - X_i\|.$$

Finally, for completeness, we will also consider an optimal choice of $r$, given a target support $S$, we will estimate

$$r \in \quad \arg\min_t \quad \mathrm{vol}(\hat{S}_{n,t} \Delta S),$$

where the symmetric difference volume is estimated by Monte-Carlo simulation and the minimum is found by discrete search. Note that this procedure requires to know the target set $S$ so it is not implementable in practice, we will consider it for the sake of investigation only. This allows to have an estimate of the best performance achievable by the DW estimator for a given sample dataset. We will denote by DW(LO) the Devroye-Wise estimator with leave-one-out estimated $r = \hat{r}_n$ and by DW(OPT) the same estimator with optimal $r$.

#### *4.2.2. Quantitative comparison in dimension 2 and 3*

Given the dimension $p$, we consider the polynomial

$$P(x) = \sum_{j=1}^{p} x_j^4 - \frac{2}{3} \sum_{j=1}^{p} x_j^2,$$

and choose as targets some sublevel sets of $P$ with different thresholds, resulting in different shapes contained in the unit box $[-1, 1]^p$. The samples will be obtained from the uniform distribution over these target sets. We compare the performances of the proposed empirical Christoffel method with data driven parameter estimates, referred to as CD, and the performances of DW(LO) and DW(OPT), the Devroye-Wise estimator with leave-one-out or optimized choice of radius $r$. For each method we measure the volume of the symmetric difference with the target set using a Monte-Carlo sample of size $10^5$. We vary $n$ and consider dimensions $p = 2$ and $p = 3$ and different sublevel set thresholds for $P$.

The obtained results are shown in Figure 1. The boxplots represent ten repetitions of the same experiment (sampling uniformly from the considered sublevel set). Two comments are in order regarding these results. First it is clear that, in this experimental setting, the proposed CD estimator consistently outperforms the DW(LO) estimator, and empirically leads to smaller symmetric difference volumes. Note that both approaches are fully data driven. Second, the estimator DW(OPT) corresponds to the best possible choice of radius $r$ for the Devroye-Wise estimator. This shows that there is a good margin of improvement regarding the estimation of the radius $r$, beyond our proposed leave-one-out approach. Yet, the CD estimator (which is purely data driven) is consistently performing at least as good as DW(OPT), and in some situations outperforms DW(OPT) by a significant margin.

These results suggest that the proposed CD estimator constitutes a competitive alternative compared to the Devroye-Wise estimator independently of parameter estimation. Indeed, we propose a fully data driven procedure which does not perform worse than DW(OPT) which is an empirical estimate of the best result achievable by the Devroye-Wise estimator on this problem. The next section illustrates the sublevel set estimates obtained in the bivariate case.

#### *4.2.3. Representation in the plane*

The target sublevel set and estimated set in dimension 2 are depicted in Figures 2 for the leave-one-out radius estimate and in Figure 3 for the optimal radius. In both cases the CD estimated set is also shown. The densities considered are uniform on the chosen polynomial sublevel set which has a smooth boundary. Recall that CD

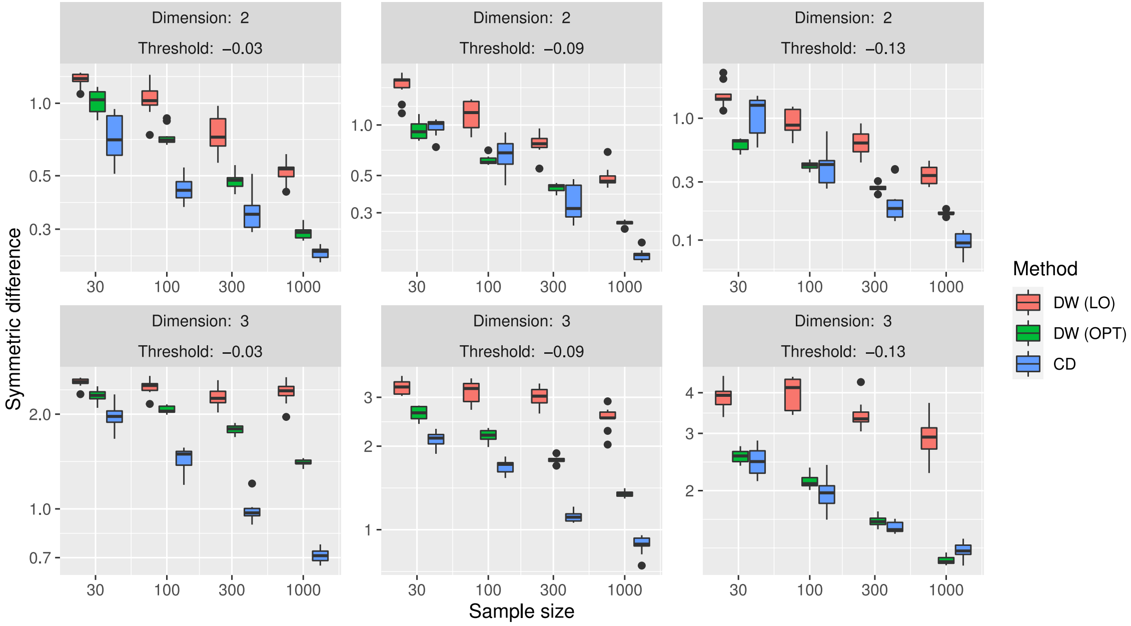

FIGURE 1. Comparison of [16] (DW) and the proposed Christoffel-Darboux (CD) estimators. Note that the first line corresponds to the setting of Figures 2 and 3. The radius is estimated by leave-one-out for DW(LO) or chosen as the optimal value for DW(OPT) (see main text). The Figure represents a Monte-Carlo estimate of the volume of symmetric difference ($10^5$ samples). Boxplots represent 10 repetitions.

and DW(LO) are fully data driven while DW(OPT) requires knowledge of the target set which is intractable in practice.

The results presented in Figures 2 and 3 correspond to the first line of Figure 1 in dimension 2. These results illustrate the fact that both CD and DW(LO) are able to identify the support set as well as its boundary and topological features correctly for large enough sample sizes. For DW(OPT) in Figure 3, although the radius minimizes the volume of the symmetric difference with the target set, the estimator contains many spurious holes (more than for DW(LO)), even for large values of $n$, illustrating the limitations of the symmetric difference volume as metric of quality for set estimation.

## 4.3. Empirical convergence rate estimation

We consider the experiment reported in Figure 1 and limit ourselves to dimension 2 because the three dimensional case is much harder especially for DW(LO). For each threshold level, we empirically estimate the slope of decrease of the volume of the symmetric difference as a function of $n$ in logarithmic space (by taking the slope between the medians for the two largest values of $n$ in Fig. 1). Our theory predicts that for CD, this slope should be asymptotically at least $-0.25$. Furthermore, the slope predicted for the DW estimator is of order $-0.5$, which is $-1/p$. We obtain the following results.

| Threshold | -0.03 | -0.09 | -0.13 |
|---|---|---|---|
| CD | -0.31 | - 0.55 | -0.56 |
| DW(LO) | 0.26 | -0.45 | -0.53 |
| DW(OPT) | -0.43 | -0.43 | -0.41 |

Except for the first threshold which is the hardest (it has a tiny hole), all slopes are close to $-0.5$, DW(OPT) being the most consistent. This experiment thus suggests that the additional term $+2$ in the denominator for

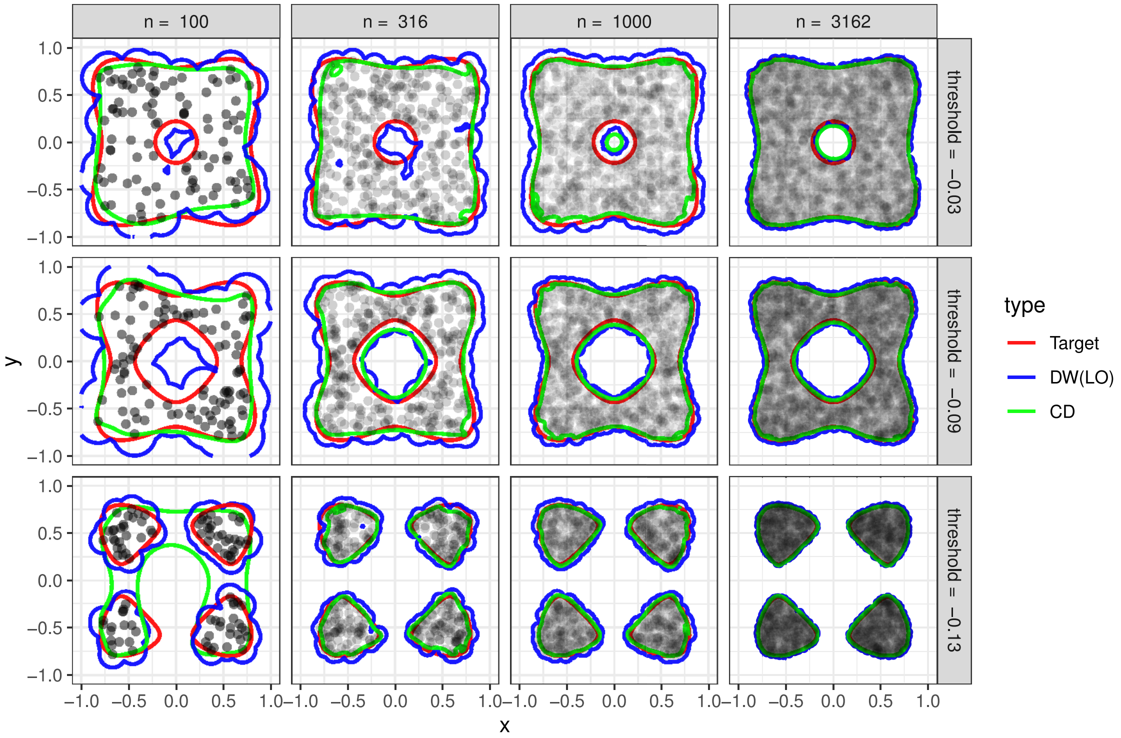

FIGURE 2. Experiment on synthetic data in the plane. The sample points are drawn uniformly on a polynomial sublevel set (Target in red). We represent the Devroye-Wise estimator with radius estimated by leave-one-out, DW(LO), and the proposed CD method. This corresponds to the first line of Figure 1.

our estimator (Thm. 3.6 and Cor. 3.7) could possibly be removed. We believe that this requires to completely change our proof strategy, switching from a worst case analysis (Appendix C and proof of Thm. 3.12, for which tools from orthogonal polynomials and approximation theory allow to have good estimates) to an average case analysis, with possibly improved rates, but for which approximation theoretic tools are not available, to the best of our knowledge. We leave this investigation for future research.

## 4.4. Outlier detection on benchmark dataset

We consider a thyroid disease dataset obtained from UCI repository [17]. This is a classification benchmark which contains 3772 examples with three classes, normal, hyperfunctioning and subnormal classes. The hyperfunctioning class contains 93 examples considered as outliers. Each example has 6 numerical descriptors so the effective dimension is 6. This dataset was used in [2, 21] to benchmark outlier detection methods.

We adopt the following procedure, for each concurrent method.

- Split the dataset randomly into a training set of normal examples and a test set with half malfunctioning cases and half normal examples.
- Estimate the support of the training set. This is done by computing a function for which a sublevel set represents the support, for example the Christoffel function or a kernel density estimate, and thresholding to a chosen value to obtain a set.
- On the test set predict outlyingness for half of the data for which the estimated function is most below the chosen threshold value on the support. Not that in this case, the threshold value is not important, only the order and rank of function value on the test set matters.

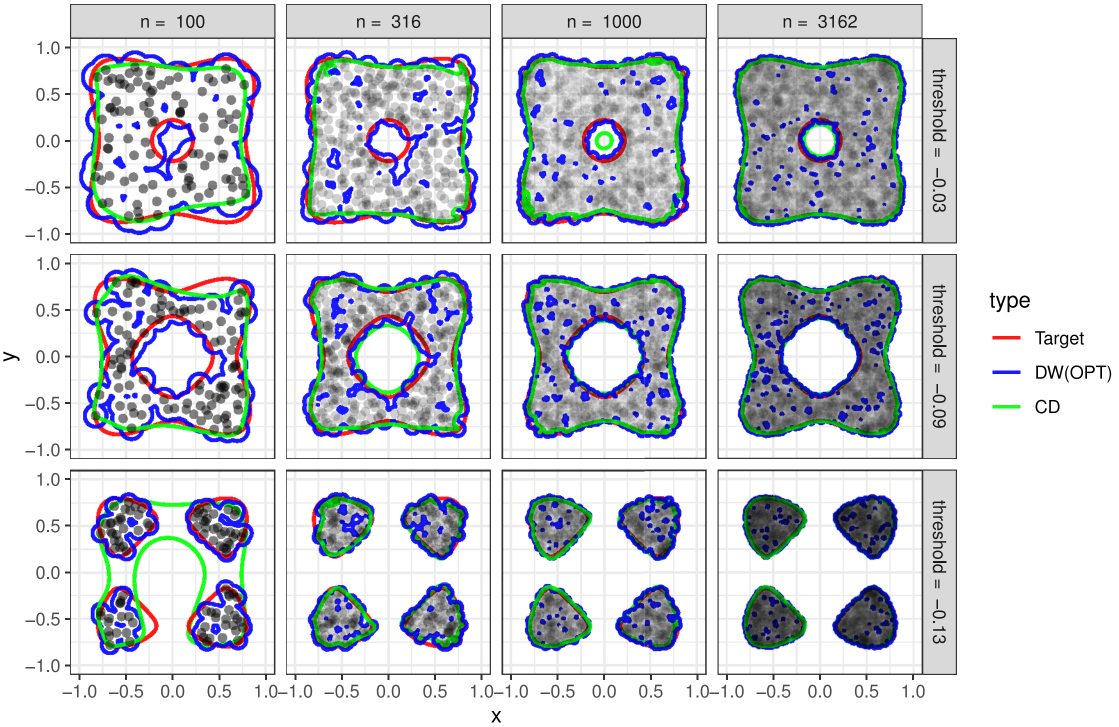

FIGURE 3. Same as Figure 2, but the DW estimator is considered with radius minimising the volume of the symmetric difference with the target set. Although this is optimal, the estimated set DW(OPT) contains many spurious holes, even for larger values of $n$.

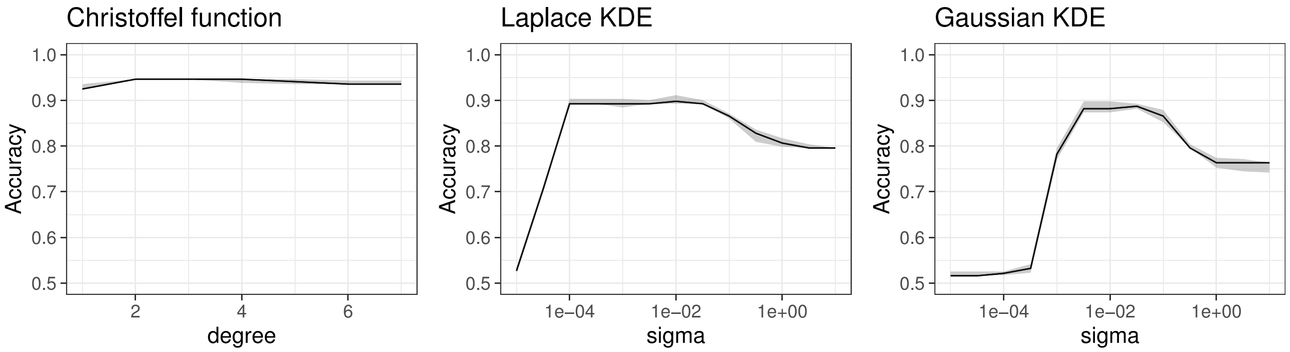

FIGURE 4. Experiment on the thyroid dataset. The dataset and experiments are described in the main text. We compare Christoffel function and kernel density estimation (KDE) to detect malfunctioning cases in a test set, based on a training set containing only normal cases. For each method, the middle line shows the median and the ribbons show quantiles for 10 random train test splits.

The results are displayed in Figure 4. We compare the Christoffel function with varying degree to kernel density estimators using Laplace or Gaussian kernels with various bandwidth. These results suggest that on this benchmark, the Christoffel function performs favorably and is more stable with respect to the choice of tuning parameter compared to kernel density estimators.

## 5. Conclusion

We have provided a detailed quantitative finite sample analysis of support estimation based on the empirical Christoffel function. We have obtained a sample-size-dependent choice of the degree $d$, together with errors bounds for the corresponding support inference procedure. An interest of our results is that support inference based on the empirical Christoffel function is computationally and conceptually attractive, as we illustrate in Section 4. These procedures have recently been subject to active developments, but there are only weak theoretical guarantees.

Our error rates are, generally speaking, slightly worse than convergence rates obtained by concurrent support inference methods. Differences in rates relate to the fact that our proofs are based on tools and developments from different fields, in particular matrix concentration inequalities, non-parametric statistics, geometry and orthonormal polynomials. Furthermore, our setting is quite general, in terms of assumptions on the unknown support and on the divergences between sets. In future work, it would be interesting to see if our proofs could be refined to obtain slightly sharper bounds, potentially in more specific settings (see also Sect. 4.3). Alternatively, it would be interesting to see if lower bounds can be provided specifically for estimation procedures based on the empirical Christoffel function, paving the way to a minimax theory for this approach. Finally it would be relevant to investigate theoretical performance of the data-driven degree bound selection method proposed in the numerical section.

Other problems of interest remain open. In particular, it would be interesting to extend our results to the case of supports with non-smooth boundaries. It would also be valuable to provide a quantitative analysis of the case where the underlying measure is supported on a manifold with smaller dimension than the ambient space.

## Appendix A. Full expressions of quantities in the main text

Below, recall that $R$ is given in Assumption 3.3, $C$, $r$ are given in Assumption 3.5 and $c_r$, $\omega_p$ are defined in (2.2) and (2.1) respectively. We let

$$\begin{aligned} C_{p,r,\alpha} := \frac{4^{r+2}}{3}\Bigg[ & 2^{p+1}c_r\left(\frac{e}{p+2r+1}\right)^{p+2r+1}\exp((p+2r+1)^2) \\ & +\frac{4^p(p+2)(p+3)(p+8)}{24\omega_p}\left(\frac{e}{p}\right)^p\exp(p^2)\Bigg]\left(p+p(1-\log p)+p^2-\log\alpha\right). \end{aligned} \tag{A.1}$$

We use the notation

$$\begin{aligned} & E_{p,r,\epsilon}\left(d,\beta\right) \\ & := \left(\frac{(1+\beta)(p+2)(p+3)(p+8)}{3C(1-\beta)\omega_p}\right)^{\frac{1}{p+r}}\left(\frac{3p(2-\epsilon)+3(1-\epsilon)r}{2\epsilon e}\right)^{\frac{p(2-\epsilon)+(1-\epsilon)r}{\epsilon(p+r)}}\left(\frac{e^{1+\frac{p}{d}}}{p}\right)^{\frac{p}{p+r}}, \end{aligned}$$

for all $d\in\mathbb{N}^*$ and $\beta\in[0,1($. Note that $E_{p,r,\epsilon}(d,\beta)$ is a bounded and decreasing function of $d$. We then write

$$D_{p,S,w,\epsilon} := \max\left(2,\left(\frac{\operatorname{diam}(S)}{R}+1\right)^{\frac{1}{1-\epsilon}},\left(\frac{2}{R}E_{p,r,\epsilon}\left(1,\frac{1}{2}\right)\right)^{\frac{1}{1-\epsilon}}\right).$$

Then $n_0$ in Theorem 3.6 is defined as

$$n_0 := \frac{4(D_{p,S,w,\epsilon}+1)^{p+2r+2}C_{p,r,\alpha}}{CR^{p+r}}. \tag{A.2}$$

The sequence of degrees $d_n$ in Section 3.3.1 is given by

$$d_n := \left\lfloor \left( \frac{CR^{p+r}}{4C_{p,r,\alpha}} n \right)^{\frac{1}{p+2r+2}} \right\rfloor \tag{A.3}$$

and the sequence of thresholds $\gamma_n$ in Section 3.3.1 is given by

$$\gamma_n := 12 \left( \frac{3p(2-\epsilon)+3(1-\epsilon)r}{2\epsilon e} \right)^{\frac{p(2-\epsilon)+(1-\epsilon)r}{\epsilon}} \frac{1}{d_n^{p(2-\epsilon)+(1-\epsilon)r}}. \tag{A.4}$$

In Theorem 3.6, $\delta_n$ is given by

$$\begin{aligned} \delta_n := \max \Bigg( \frac{\operatorname{diam}(S)}{d_n^{1-\epsilon}-1}, \frac{2}{d_n^{1-\epsilon}} \left( \frac{(p+2)(p+3)(p+8)}{C\omega_p} \right)^{\frac{1}{p+r}} \\ \times \left( \frac{3p(2-\epsilon)+3(1-\epsilon)r}{2\epsilon e} \right)^{\frac{p(2-\epsilon)+(1-\epsilon)r}{\epsilon(p+r)}} \left( \frac{e^{1+\frac{p}{d_n}}}{p} \right)^{\frac{p}{p+r}} \Bigg). \end{aligned} \tag{A.5}$$

In Lemma 3.10, $\delta_1(d)$ and $\delta_2(d,\beta)$ are given by

$$\delta_1(d) := \frac{\operatorname{diam}(S)}{d^{1-\epsilon}-1} \tag{A.6}$$

and

$$\delta_2(d,\beta) := \frac{2}{d^{1-\epsilon}} E_{p,r,\epsilon}(d,\beta). \tag{A.7}$$

Observe that we can rewrite $\delta_n$ as:

$$\delta_n = \max \left( \delta_1(d_n), \delta_2 \left( d_n, \frac{1}{2} \right) \right). \tag{A.8}$$

In Section 3.3.4, $\gamma_d$ is given by

$$\gamma_d := 8(1+\beta) \left( \frac{3p(2-\epsilon)+3(1-\epsilon)r}{2\epsilon e} \right)^{\frac{p(2-\epsilon)+(1-\epsilon)r}{\epsilon}} \frac{1}{d^{p(2-\epsilon)+(1-\epsilon)r}}. \tag{A.9}$$

Finally, in Corollary 3.13, $m(d,p,S,w)$ is given by

$$\begin{aligned} m(d,p,S,w) := \frac{4^{p+r}s(d)}{C\omega_p R^{p+r}} \frac{(d+p+1)(d+p+2)(2d+p+6)}{(d+1)(d+2)(d+3)} \\ + \frac{2^{p+2r}c_r}{CR^{p+r}} \left[ 2\binom{p+d+2r+1}{d} - \binom{p+d+2r}{d} \right]. \end{aligned} \tag{A.10}$$

## Appendix B. Bounds on the Christoffel function

The following results provide a lower bound on the Christoffel function $\Lambda_{\mu,d}$ inside the support $S$ and an upper bound outside $S$. These bounds are similar to those in Sections 6.3.1 and 6.3.2 of [24] and will be useful in the next proofs.

## B.1 Upper bound on the Christoffel function outside $S$

In this section, we consider a probability measure $\mu$ which satisfies Assumption 2.1. Now, to exhibit an upper bound on the Christoffel function outside $S$, we first provide a refinement of the "needle polynomial" which has been introduced in [22].

**Lemma B.1** (see *e.g.* [24], Lem. 6.3)**.** *For any $d \in \mathbb{N}^*$ and $\delta \in (0,1)$, there exists a $p$-variate polynomial $Q$ of degree $2d$ such that $Q(0) = 1$, $|Q| \leq 1$ on the unit ball $B$ and $|Q| \leq 2^{1-\delta d}$ on $B \setminus \overline{B}_\delta(0)$.*

**Lemma B.2.** *Let $\delta > 0$ and $x \notin S$ such that $d(x,S) \geq \delta$. Then, for any $d \in \mathbb{N}^*$ we have*

$$\Lambda_{\mu,d}(x) \leq 2^{3-\frac{\delta d}{\delta+\operatorname{diam}(S)}}.$$

*Proof.* First, we will prove Lemma B.2 with $x \notin S$ such that $d(x,S) = \delta$. In this case, $S \subset T := \overline{B}_{\delta+\operatorname{diam}(S)}(x) \setminus B_\delta(x)$. Indeed, for any $y \in S$, $d(x,y) \leq d(x,S) + \operatorname{diam}(S) = \delta + \operatorname{diam}(S)$. On the other hand, if $y \in B_\delta(x)$, then $d(x,S) \leq d(x,y) < \delta$ which is a contradiction.
Now, let $A$ be the affine transformation which maps $\overline{B}_{\delta+\operatorname{diam}(S)}(x)$ to the unit ball $B$ and $\mu_{\#A}$ be the push-forward measure of $\mu$ by $A$. Then $\operatorname{supp}\mu_{\#A} = A(S) \subset A(T) = B \setminus B_{\delta'}(0)$ where $\delta' = \frac{\delta}{\delta+\operatorname{diam}(S)}$ and by Proposition 2.10, we have

$$\Lambda_{\mu,d}(x) = \Lambda_{\mu_{\#A},d}(0).$$

Next, we apply Lemma B.1 to $k \in \mathbb{N}^*$ and $\delta' \in (0,1)$, we obtain a polynomial $Q$ of degree $2k$ such that $Q(0) = 1$ and $|Q| \leq 2^{1-\delta' k}$ on $B \setminus \overline{B}_{\delta'}(0)$, which implies that $|Q| \leq 2^{1-\delta' k}$ on $\operatorname{supp}\mu_{\#A}$. Thus

$$\begin{aligned}\Lambda_{\mu_{\#A},2k}(0) &= \min\left\{\int\limits_{\mathbb{R}^p} P^2 \mathrm{d}\mu_{\#A} : P \in \Pi^p_{2k}, P(z) = 1\right\}\\ &\leq \int\limits_{\mathbb{R}^p} Q^2 \mathrm{d}\mu_{\#A} = \int\limits_{\operatorname{supp}\mu_{\#A}} Q^2 \mathrm{d}\mu_{\#A}\\ &\leq 2^{2(1-\delta' k)} \leq 2^{3-\delta' 2k}.\end{aligned}$$

Then we have for any $k \in \mathbb{N}^*$

$$\Lambda_{\mu,2k}(x) \leq 2^{3-\delta' 2k}.$$

Now, the definition of the Christoffel function in Definition 2.8 makes sure that

$$\Lambda_{\mu,2k+1}(x) \leq \Lambda_{\mu_S,2k}(x) \leq 2^{2(1-\delta' k)} \leq 2^{2(1-\delta' k)+1-\delta'} \leq 2^{3-\delta'(2k+1)}.$$

By combining both cases $d = 2k$ and $d = 2k+1$, we have

$$\Lambda_{\mu,d}(x) \leq 2^{3-\delta' d}.$$

Finally, since $2^{3-\delta' d} = 2^{3-\frac{\delta d}{\delta+\operatorname{diam}(S)}}$ is a decreasing function of $\delta$, we have for all $x$ such that $d(x,S) \geq \delta$,

$$\Lambda_{\mu,d}(x) \leq 2^{3-\frac{d(x,S)d}{d(x,S)+\operatorname{diam}(S)}} \leq 2^{3-\frac{\delta d}{\delta+\operatorname{diam}(S)}}.$$

□

## B.2 Lower bound on the Christoffel function inside $S$

We now consider a compact set $S$ with non-empty interior, a density $w$ satisfying Assumption 3.5 with two constants $C > 0$, $r \geq 0$ and the measure $\mu$ supported on $S$ with density $w$.

**Lemma B.3.** *Let $\delta > 0$ and $x \in S$ such that $d(x, \partial S) \geq \delta$. Then for any $d \geq 2$ we have*

$$\Lambda_{\mu,d}(x) \geq \frac{C\omega_p\,\delta^{p+r}}{2^{p+r}}\frac{1}{s(d)}\frac{(d+1)(d+2)(d+3)}{(d+p+1)(d+p+2)(2d+p+6)}.$$

*Proof.* First, we will prove that the closed ball $\overline{B}_{\delta/2}(x) \subset \{x \in S : d(x, \partial S) \geq \delta/2\} \subset S$. Indeed, if $z \in \overline{B}_{\delta/2}(x)$, *i.e.* $\operatorname{dist}(x, z) \leq \delta/2$, then

$$d(z, \partial S) \geq d(x, \partial S) - d(x, z) \geq \delta - \delta/2 = \delta/2.$$

We have

$$\begin{aligned}
\Lambda_{\mu,d}(x) &= \min\left\{\int_{\mathbb{R}^p} P^2(z)\mathrm{d}\mu(z) : P \in \Pi_d^p, P(x) = 1\right\}\\
&= \min\left\{\int_{S} P^2(z)w(z)dz : P \in \Pi_d^p, P(x) = 1\right\}\\
&\geq \min\left\{\int_{\overline{B}_{\delta/2}(x)} P^2(z)\mathrm{d}\mu(z) : P \in \Pi_d^p, P(x) = 1\right\}\\
&\geq L\left(\frac{\delta}{2}\right)\min\left\{\int_{\overline{B}_{\delta/2}(x)} P^2(z)dz : P \in \Pi_d^p, P(x) = 1\right\}\\
&\geq C\left(\frac{\delta}{2}\right)^r\min\left\{\int_{\mathbb{R}^p} P^2(z)d\lambda_{\overline{B}_{\delta/2}(x)}(z) : P \in \Pi_d^p, P(x) = 1\right\}\\
&= C\left(\frac{\delta}{2}\right)^r\Lambda_{\lambda_{\overline{B}_{\delta/2}(x)},d}(x),
\end{aligned}$$

where the third inequality comes from Assumption 3.5. Now we have

$$\begin{aligned}
\Lambda_{\lambda_{\overline{B}_{\delta/2}(x)},d}(x) &= \lambda\left(\overline{B}_{\delta/2}(x)\right)\Lambda_{\mu_{\overline{B}_{\delta/2}(x)},d}(x) = \lambda\left(\overline{B}_{\delta/2}(x)\right)\Lambda_{\mu_B,d}(0)\\
&= \frac{\lambda\left(\overline{B}_{\delta/2}(x)\right)}{\lambda(B)}\Lambda_{\mu_B,d}(0) = \left(\frac{\delta}{2}\right)^p\Lambda_{\mu_B,d}(0)\\
&\geq \left(\frac{\delta}{2}\right)^p\frac{\omega_p}{s(d)}\frac{(d+1)(d+2)(d+3)}{(d+p+1)(d+p+2)(2d+p+6)},
\end{aligned}$$

where the first and third equality come from the monotonicity of the Christoffel function, the second equality comes from its affine equivariance and the last inequality is Lemma 6.1 in [24], which is obtained when $d \geq 2$. Then, by combining the above arguments, we have the lower bound result. □

## Appendix C. Supremum of the Christoffel – Darboux kernel on $S$

In this section, we consider a set $S$ satisfying Assumption 3.3 with radius $R > 0$ and a density $w$ satisfying Assumption 3.5 with two constants $C > 0$, $r \geq 0$. Let $\mu$ be the measure supported on $S$ with density $w$. First, we have the following upper bound of the Christoffel – Darboux kernel $\kappa_{\mu,d}$ inside the support $S$, which is a direct consequence of Lemma B.3.

**Corollary C.1.** *Let us consider $x \in S$ such that $d(x, \partial S) \geq R/2$. Then*

$$\kappa_{\mu,d}(x,x) \leq \frac{4^{p+r} s(d)}{C\omega_p R^{p+r}} \frac{(d+p+1)(d+p+2)(2d+p+6)}{(d+1)(d+2)(d+3)}.$$

*Proof.* We apply Lemma B.3 with $\delta = R/2$ and we use the fact that $\kappa_{\mu,d}(x,x) = 1/\Lambda_{\mu,d}(x)$. □

Next, for the points which stay near the boundary of $S$, we will rely on Theorem 3.1 from [45] which provides an explicit formula for the Christoffel – Darboux kernel associated to a measure with Jacobi-like weight on the unit Euclidean ball. The following lemma provides an upper bound near the boundary.

**Lemma C.2.** *Given $x \in S$ such that $d(x, \partial S) \leq R/2$, we have*

$$\kappa_{\mu,d}(x,x) \leq \frac{2^{p+2r} c_r}{CR^{p+r}} \left[ 2\binom{p+d+2r+1}{d} - \binom{p+d+2r}{d} \right],$$

*where $c_r$ is defined in (2.2).*

*Proof.* By Assumption 3.3 there exists a point $z_x \in S$ such that $x \in \overline{B}_R(z_x) \subset S$. We set $\epsilon = \|x - z_x\|_2$, then $\epsilon \leq R$ and

$$\epsilon \geq d(z_x, \partial S) - d(x, \partial S) \geq R - R/2 = R/2.$$

Evidently, $x$ is on the boundary of the closed ball $\overline{B}_\epsilon(z_x) \subset S$. Moreover, for all $y \in \overline{B}_\epsilon(z_x)$, we have

$$d(y, \partial S) \geq d(y, \partial \overline{B}_\epsilon(z_x)) = \epsilon - \|y - z_x\|_2 \geq 0.$$

Now, by Assumption 3.5, we have for all $y \in \overline{B}_\epsilon(z_x)$:

$$\begin{aligned} w(y) \geq C\left(\epsilon - \|y - z_x\|_2\right)^r &\geq C\left(\epsilon - \|y - z_x\|_2\right)^r \left(\frac{\epsilon + \|y - z_x\|_2}{2\epsilon}\right)^r \\ &= \frac{C}{(2\epsilon)^r}\left(\epsilon^2 - \|y - z_x\|_2^2\right)^r. \end{aligned}$$

We have

$$\begin{aligned} \Lambda_{\mu,d}(x) &= \min\left\{ \int_{\mathbb{R}^p} P^2(y) \mathrm{d}\mu(y) : P \in \Pi_d^p, P(x) = 1 \right\} \\ &= \min\left\{ \int_S P^2(y) w(y) dy : P \in \Pi_d^p, P(x) = 1 \right\} \end{aligned}$$

$$
\begin{aligned}
&\geq \min\left\{\int\limits_{\overline{B}_\epsilon(z_x)} P^2(y)w(y)dy : P\in\Pi_d^p, P(x)=1\right\}\\
&\geq \frac{C}{(2\epsilon)^r}\min\left\{\int\limits_{\overline{B}_\epsilon(z_x)} P^2(y)\left(\epsilon^2-\|y-z_x\|_2^2\right)^r dy : P\in\Pi_d^p, P(x)=1\right\}.
\end{aligned}
$$

Now, by changing the variable $z=\dfrac{y-z_x}{\epsilon}$ and setting $Q(z)=P(z_x+\epsilon z)$, we have

$$
\begin{aligned}
\int\limits_{\overline{B}_\epsilon(z_x)} P^2(y)\left(\epsilon^2-\|y-z_x\|_2^2\right)^r dy &= \int\limits_B P^2(z_x+\epsilon z)(\epsilon^2-\epsilon^2\|z\|_2^2)^r\epsilon^p dz\\
&= \epsilon^{p+2r}\int\limits_B Q^2(z)(1-\|z\|_2^2)^r dz,
\end{aligned}
$$

where $Q$ is a polynomial with degree at most $d$ and $Q\left(\frac{x-z_x}{\epsilon}\right)=P(x)=1$. We set $\tilde{x}=\frac{x-z_x}{\epsilon}\in\partial B$ since $x\in\partial\overline{B}_\epsilon(z_x)$. Now we have

$$
\begin{aligned}
\Lambda_{\mu,d}(x) &\geq \frac{C\epsilon^{p+2r}}{(2\epsilon)^r}\min\left\{\int\limits_B Q^2(z)(1-\|z\|_2^2)^r dz : Q\in\Pi_d^p, Q(\tilde{x})=1\right\}\\
&= \frac{C\epsilon^{p+r}}{2^r c_r}\min\left\{\int\limits_B Q^2(z)\,c_r(1-\|z\|_2^2)^r dz : Q\in\Pi_d^p, Q(\tilde{x})=1\right\}\\
&= \frac{C\epsilon^{p+r}}{2^r c_r}\Lambda_{\nu_r,d}(\tilde{x}),
\end{aligned}
$$

where we recall that $c_r=\dfrac{\Gamma(p/2+r+1)}{\pi^{p/2}\Gamma(r+1)}$ is the normalization constant of the measure $\nu_r$ which density is $(1-\|z\|_2^2)^r$ on the unit ball $B$. Then, by taking the inverse, we have

$$
\kappa_{\mu,d}(x,x)\leq\frac{2^r c_r}{C\epsilon^{p+r}}\kappa_{\nu_r,d}(\tilde{x},\tilde{x}).
$$

Now, to compute $\kappa_{\nu_r,d}(\tilde{x},\tilde{x})$, we use Theorem 3.1 in [45] with $\mu=r+\frac{1}{2}$ and we obtain that

$$
\begin{aligned}
\kappa_{\nu_r,d}(\tilde{x},\tilde{x}) &= \sum_{k=0}^d \frac{k+r+\frac{1}{2}+\frac{p-1}{2}}{r+\frac{1}{2}+\frac{p-1}{2}}\int\limits_0^\pi C_k^{\left(r+\frac{1}{2}+\frac{p-1}{2}\right)}\left(\langle\tilde{x},\tilde{x}\rangle+\sqrt{1-\|\tilde{x}\|_2^2}\sqrt{1-\|\tilde{x}\|_2^2}\cos\psi\right)\\
&\quad\times(\sin\psi)^{2\left(r+\frac{1}{2}\right)-1}d\psi\Big/\int\limits_0^\pi(\sin\psi)^{2\left(r+\frac{1}{2}\right)-1}d\psi\\
&= \sum_{k=0}^d\frac{k+\frac{p}{2}+r}{\frac{p}{2}+r}C_k^{\left(\frac{p}{2}+r\right)}(1),
\end{aligned}
$$

where the $C_k^{(\beta)}$'s are the classical Gegenbauer polynomials, which are orthogonal polynomials on $[-1,1]$ with respect to the weight function $(1-x^2)^{\beta-1/2}$. In particular, by [38, p.81, (4.7.3)], we have

$$C_k^{\left(\frac{p}{2}+r\right)}(1) = \binom{p+k+2r-1}{k},$$

where we recall that the binomial coefficient for $\alpha \in \mathbb{R}$ and $k \in \mathbb{N}$ is defined as:

$$\binom{\alpha}{k} := \frac{\Gamma(\alpha+1)}{\Gamma(k+1)\Gamma(\alpha-k+1)} = \frac{\alpha(\alpha-1)(\alpha-2)\dots(\alpha-k+1)}{k!}.$$

Hence

$$\begin{aligned}
\kappa_{\nu_r,d}(\tilde{x},\tilde{x}) &= \sum_{k=0}^{d} \frac{k+\frac{p}{2}+r}{\frac{p}{2}+r} C_k^{\left(\frac{p}{2}+r\right)}(1) = \sum_{k=0}^{d} \frac{k+\frac{p}{2}+r}{\frac{p}{2}+r} \binom{p+k+2r-1}{k} \\
&= \sum_{k=0}^{d} \frac{2k+p+2r}{p+2r} \frac{(p+k+2r-1)(p+k+2r-2)\dots(p+2r)}{k!} \\
&= \sum_{k=0}^{d} \left(\frac{2(k+p+2r)}{p+2r} - 1\right) \frac{(p+k+2r-1)(p+k+2r-2)\dots(p+2r)}{k!} \\
&= 2\sum_{k=0}^{d} \frac{(p+k+2r)(p+k+2r-1)\dots(p+2r+1)}{k!} - \binom{p+k+2r-1}{k} \\
&= 2\sum_{k=0}^{d} \binom{p+k+2r}{k} - \sum_{k=0}^{d} \binom{p+k+2r-1}{k} \\
&= 2\binom{p+d+2r+1}{d} - \binom{p+d+2r}{d}.
\end{aligned}$$

We finally have

$$\kappa_{\mu,d}(x,x) \leq \frac{2^r c_r}{C\epsilon^{p+r}} \left[2\binom{p+d+2r+1}{d} - \binom{p+d+2r}{d}\right],$$

and the result follows by using the fact that $\epsilon \geq R/2$. □

By combining Corollary C.1 and Lemma C.2, we have the following theorem regarding the supremum of the Christoffel – Darboux kernel.

**Theorem C.3.** *Let $S \subset \mathbb{R}^p$ satisfies Assumption 3.3 with radius $R > 0$ and $w : S \longrightarrow \mathbb{R}$ satisfies Assumption 3.5 with two constants $C > 0$ and $r \geq 0$. Let $\mu$ be the measure supported on $S$ with density $w$ with respect to Lebesgue measure. We have for any $d \geq 2$,*

$$\sup_{x\in S} \kappa_{\mu,d}(x,x) \leq m(d,p,S,w),$$

where $m(d,p,S,w)$ in (A.10) is of order $d^{p+2r+1}$ when $p$ is fixed.

## Appendix D. Proof of the concentration results

*Proof of Lemma 3.11.* For all $x \in \mathbb{R}^p$, we have

$$\begin{aligned}
\left|\Lambda_{\mu,d}(x) - \Lambda_{\mu_n,d}(x)\right| &= \Lambda_{\mu,d}(x)\Lambda_{\mu_n,d}(x)\left|\kappa_{\mu,d}(x,x) - \kappa_{\mu_n,d}(x,x)\right| \\
&= \Lambda_{\mu,d}(x)\Lambda_{\mu_n,d}(x)\left|v_d(x)^T\left(I_{s(d)} - M_{\mu_n,d}^{-1}\right)v_d(x)\right| \\
&= \Lambda_{\mu,d}(x)\Lambda_{\mu_n,d}(x)\left|v_d(x)^T\left(M_{\mu_n,d}^{-1/2}\right)^T\left(M_{\mu_n,d} - I_{s(d)}\right)M_{\mu_n,d}^{-1/2}v_d(x)\right| \\
&= \Lambda_{\mu,d}(x)\Lambda_{\mu_n,d}(x)\left|\left(M_{\mu_n,d}^{-1/2}v_d(x)\right)^T\left(M_{\mu_n,d} - I_{s(d)}\right)M_{\mu_n,d}^{-1/2}v_d(x)\right| \\
&\leq \Lambda_{\mu,d}(x)\Lambda_{\mu_n,d}(x)\left(M_{\mu_n,d}^{-1/2}v_d(x)\right)^T M_{\mu_n,d}^{-1/2}v_d(x)\left\|M_{\mu_n,d} - I_{s(d)}\right\| \\
&= \Lambda_{\mu,d}(x)\Lambda_{\mu_n,d}(x)v_d(x)^T M_{\mu_n,d}^{-1}v_d(x)\left\|M_{\mu_n,d} - I_{s(d)}\right\| \\
&= \Lambda_{\mu,d}(x)\Lambda_{\mu_n,d}(x)\kappa_{\mu_n,d}(x,x)\left\|M_{\mu_n,d} - I_{s(d)}\right\| \\
&= \Lambda_{\mu,d}(x)\left\|M_{\mu_n,d} - I_{s(d)}\right\|,
\end{aligned}$$

where the third equality comes from the fact that $M_{\mu_n,d}$ is symmetric and positive definite, which implies that $M_{\mu_n,d}^{-1/2}$ exists and is also symmetric; while the inequality can be seen as

$$\left|z^T A z\right| \leq z^T z\|A\|,$$

with $z = M_{\mu_n,d}^{-1/2}v_d(x) \in \mathbb{R}^{s(d)}$ and $A = M_{\mu_n,d} - I_{s(d)} \in \mathbb{R}^{s(d)\times s(d)}$. This inequality can be proved as below:

$$\left|z^T A z\right| = \langle z, Az\rangle \leq \|z\|_2\,\|Az\|_2 \leq \|z\|_2\|A\|\,\|z\|_2 = z^T z\|A\|,$$

where the first inequality is Cauchy-Schwarz and the second one comes from the definition of operator norm. □

*Proof of Theorem 3.12.* Let $v_d = \{P_j : 1 \leq j \leq s(d)\}$ be a system of orthonormal polynomials with respect to $\mu$ and $M_{\mu_n,d}$ be the moment matrix of $\mu_n$ with respect to $v_d$. We apply Theorem 5.44 in [41] to

$$A = \begin{bmatrix} P_1(X_1) & \dots & P_{s(d)}(X_1) \\ \dots & & \dots \\ P_1(X_n) & \dots & P_{s(d)}(X_n) \end{bmatrix},$$

which is a $n \times s(d)$ random matrix whose rows $A_k = \left(P_1(X_k), \dots, P_{s(d)}(X_k)\right)$ are independent random vectors in $\mathbb{R}^{s(d)}$ with the common second moment matrix $\Sigma = \mathbb{E}[A_k^T A_k] = I_{s(d)}$. We have

$$\frac{1}{n}A^T A = M_{\mu_n,d},$$

thus

$$\left\|\frac{1}{n}A^T A - \Sigma\right\| = \|M_{\mu_n,d} - I_{s(d)}\|.$$

If we can obtain an almost sure bound on the rows of $A_k$, then Theorem 5.44 in [41] provides an upper bound for $\|M_{\mu_n} - I_{s(d)}\|$, and then, by Lemma 3.11, an upper bound for $|\Lambda_{\mu,d}(x) - \Lambda_{\mu_n,d}(x)|$ with high probability.

Let us check the boundedness condition of the rows $A_k$. We have

$$\|A_k\|_2^2 = \sum_{j=1}^{s(d)} P_j(X_k)^2 = \kappa_{\mu,d}(X_k, X_k) = \frac{1}{\Lambda_{\mu,d}(X_k)}.$$

A natural upper bound for this will be

$$\sup_{x\in \operatorname{supp}\mu} \sum_{j=1}^{s(d)} P_j(x)^2 = \sup_{x\in \operatorname{supp}\mu} \kappa_{\mu,d}(x,x) := m,$$

which is finite since $x \mapsto \sum_{j=1}^{s(d)} P_j(x)^2$ is continuous and $\operatorname{supp}\mu$ is a compact set. Now, by Theorem 5.44 in [41], for all $t \geq 0$, with probability at least $1 - s(d).\exp(-3t^2/16)$, we have

$$\|M_{\mu_n,d} - I_{s(d)}\| \leq \max\left(t\sqrt{\frac{m}{n}}, \frac{t^2 m}{n}\right).$$

We choose $\alpha = s(d).\exp(-3t^2/16)$, which means $t = \sqrt{\frac{16}{3}\log\frac{s(d)}{\alpha}}$, and we have

$$\|M_{\mu_n,d} - I_{s(d)}\| \leq \max\left(\sqrt{\frac{16m}{3n}\log\frac{s(d)}{\alpha}}, \frac{16m}{3n}\log\frac{s(d)}{\alpha}\right)$$

with probability at least $1-\alpha$. Then by lemma 3.11,

$$\left|\Lambda_{\mu,d}(x) - \Lambda_{\mu_n,d}(x)\right| \leq \Lambda_{\mu,d}(x)\ \max\left(\sqrt{\frac{16m}{3n}\log\frac{s(d)}{\alpha}}, \frac{16m}{3n}\log\frac{s(d)}{\alpha}\right)$$

with probability at least $1-\alpha$. □

# Appendix E. Proofs of the main results regarding support estimation

## E.1 Proof of Lemma 3.10

First, we introduce some inequalities which will be useful in the proof of Lemma 3.10.

**Lemma E.1** (see *e.g.* [24], Lem. 6.5)**.** *For any $m, n \in \mathbb{N}^*$, we have*

$$\binom{m+n}{m} \leq m^n \left(\frac{e}{n}\right)^n \exp\left(\frac{n^2}{m}\right).$$

**Lemma E.2** (see e.g [26], Lem. 5)**.** *For any $q > 0$, we have*

$$\min_{x>0}\left[\log(2)x - 2q\log(x)\right] = 2q\left(1 - \log\left(\frac{2q}{\log(2)}\right)\right) \geq 2q(1-\log(3q)).$$

**Lemma E.3.** *For any $d \in \mathbb{N}$, $0 < \epsilon < 1$ and $q > 0$, we have*

$$2^{3-d^\epsilon} \leq \frac{8(3q)^{2q}}{e^{2q}d^{2q\epsilon}}.$$

*Proof.* We have

$$\begin{aligned}
&2^{3-d^\epsilon} \leq \frac{8(3q)^{2q}}{e^{2q}d^{2q\epsilon}}\\
\Leftrightarrow\ &(3-d^\epsilon)\log(2) \leq 3\log(2) + 2q\log(3q) - 2q - 2q\epsilon\log(d)\\
\Leftrightarrow\ &2q(1-\log(3q)) \leq \log(2)d^\epsilon - 2q\log(d^\epsilon),
\end{aligned}$$

which holds true by applying Lemma E.2 to $x = d^\epsilon > 0$. □

*Proof of Lemma 3.10.* The inclusion $S_{d,n} \subset S^1$ can be proved for a more general thresholding scheme, where we define for any $d \in \mathbb{N}$ and any $q \in \mathbb{N}$ such that $2q\epsilon > p$:

$$\begin{cases} \gamma_d := \dfrac{8(1+\beta)(3q)^{2q}}{e^{2q}d^{2q\epsilon}} \\ S_{d,n} := \{x \in \mathbb{R}^p : \Lambda_{\mu_n,d}(x) \geq \gamma_d\}. \end{cases} \tag{E.1}$$

Indeed, if $x \notin S^1$, *i.e.* $d(x,S) > \delta_1(d) = \dfrac{\operatorname{diam}(S)}{d^{1-\epsilon}-1}$, we apply Lemma B.2, then Lemma E.3 and we have

$$\Lambda_{\mu,d}(x) < 2^{3-\frac{\delta_1(d)d}{\delta_1(d)+\operatorname{diam}(S)}} = 2^{3-d^\epsilon} \leq \frac{8(3q)^{2q}}{e^{2q}d^{2q\epsilon}}.$$

Since $\Lambda_{\mu_n,d}(x) \leq (1+\beta)\Lambda_{\mu,d}(x)$ by (3.5), we obtain that

$$\Lambda_{\mu_n,d}(x) \leq \frac{8(1+\beta)(3q)^{2q}}{e^{2q}d^{2q\epsilon}} = \gamma_d,$$

which means that $x \notin S_{d,n}$ and we can deduce the result by contraposition.

By choosing $q = \frac{p(2-\epsilon)+(1-\epsilon)r}{2\epsilon}$ in the scheme (E.1), we obtain our thresholding scheme (3.6) and the result $S_{d,n} \subset S^1$ follows.

Now, if $x \in S^2$, *i.e.* $x \in S$ and $d(x, \partial S) \geq \delta_2(d,\beta)$, then by Lemmas B.3 and E.1, we have

$$\begin{aligned}
\Lambda_{\mu,d}(x) &\geq \frac{C\omega_p\,(\delta_2(d,\beta))^{p+r}}{2^{p+r}}\frac{1}{s(d)}\frac{(d+1)(d+2)(d+3)}{(d+p+1)(d+p+2)(2d+p+6)}\\
&\geq \frac{C\omega_p(\delta_2(d,\beta))^{p+r}}{2^{p+r}d^p\left(\dfrac{e}{p}\right)^p\exp\left(\dfrac{p^2}{d}\right)}\frac{24}{(p+2)(p+3)(p+8)}\\
&= \frac{C\omega_p}{2^{p+r}d^p\left(\dfrac{e}{p}\right)^p\exp\left(\dfrac{p^2}{d}\right)}\frac{24}{(p+2)(p+3)(p+8)}\frac{2^{p+r}}{d^{(1-\epsilon)(p+r)}}\\
&\times\frac{(1+\beta)(p+2)(p+3)(p+8)}{3C(1-\beta)\omega_p}\left(\frac{3p(2-\epsilon)+3(1-\epsilon)r}{2\epsilon e}\right)^{\frac{p(2-\epsilon)+(1-\epsilon)r}{\epsilon}}\left(\frac{e^{1+\frac{p}{d}}}{p}\right)^p
\end{aligned}$$

$$= \frac{8(1+\beta)}{1-\beta}\left(\frac{3p(2-\epsilon)+3(1-\epsilon)r}{2\epsilon e}\right)^{\frac{p(2-\epsilon)+(1-\epsilon)r}{\epsilon}} \frac{1}{d^{p(2-\epsilon)+(1-\epsilon)r}}.$$

Since $\Lambda_{\mu_n,d}(x) \geq (1-\beta)\Lambda_{\mu,d}(x)$ by (3.5), we have $\Lambda_{\mu_n,d}(x) \geq \gamma_d$, which means that $x \in S_{d,n}$. □

## E.2 Proof of Theorem 3.6

First, we have the following lemma which highlights an important property of a set $S$ which satisfies Assumption 3.3.

**Lemma E.4.** *Let $S \subset \mathbb{R}^p$ satisfies Assumption 3.3 with radius $R > 0$. Given $\delta_1, \delta_2 > 0$, we set*

$$S^1 := \{x \in \mathbb{R}^p : d(x, S) \leq \delta_1\}$$

*and*

$$S^2 := \{x \in S : d(x, \partial S) \geq \delta_2\}.$$

*Suppose that there exists a closed set $\tilde{S} \subset \mathbb{R}^p$ such that*

$$S^2 \subset \tilde{S} \subset S^1. \tag{E.2}$$

*If we have in addition that $\delta := \max(\delta_1, \delta_2) \leq R$, then*

$$d_H(S, \tilde{S}) \leq \delta$$

*and*

$$d_H(\partial S, \partial \tilde{S}) \leq \delta.$$

*Proof.* We will begin with the Hausdorff distance between two sets $S$ and $\tilde{S}$:

$$d_H(S, \tilde{S}) = \max\left(\sup_{x\in\tilde{S}} d(x, S), \sup_{x\in S} d(x, \tilde{S})\right).$$

Given $x \in \tilde{S}$, then by (E.2), $x \in S^1$, *i.e.* $d(x, S) \leq \delta_1 \leq \delta$. Hence

$$\sup_{x\in\tilde{S}} d(x, S) \leq \delta. \tag{E.3}$$

Given now $x \in S$. Since $S^2 \subset \tilde{S}$, we have $d(x, \tilde{S}) \leq d(x, S^2)$. By Assumption 3.3 and since $\delta_2 \leq R$, there exists $z_x \in S$ such that $x \in \overline{B}_{\delta_2}(z_x) \subset S$. Hence $\|x - z_x\|_2 \leq \delta_2$ and $d(z_x, \partial S) \geq \delta_2$. We have now $z_x \in S^2$ and $d(x, S^2) \leq \|x - z_x\|_2 \leq \delta_2$. Then $d(x, \tilde{S}) \leq \delta_2 \leq \delta$ and

$$\sup_{x\in S} d(x, \tilde{S}) \leq \delta. \tag{E.4}$$

By combining (E.3) and (E.4), we have $d_H(S,\tilde{S}) \le \delta$. Now, we continue with the Hausdorff distance between two boundaries:

$$d_H(\partial S, \partial\tilde{S}) = \max\left(\sup_{x\in\partial\tilde{S}} d(x,\partial S), \sup_{x\in\partial S} d(x,\partial\tilde{S})\right).$$

Consider $x \in \partial\tilde{S}$. We will consider two cases where $x \in S$ and $x \notin S$ separately. If $x \notin S$, then $d(x,\partial S) = d(x,S) \le \delta$ by (E.3) since $x \in \partial\tilde{S} \subset \tilde{S}$. Now, consider $x \in S \cap \partial\tilde{S}$. Note that since $S^2 \subset \tilde{S}$, $\operatorname{Int} S^2 \subset \operatorname{Int}\tilde{S}$, which implies that $\operatorname{Int} S^2 \cap \partial\tilde{S} = \emptyset$. Now when $x \in \partial\tilde{S}$, we have $x \notin \operatorname{Int} S^2 = \{x \in S : d(x,\partial S) > \delta_2\}$. By combining with the fact that $x \in S$, we have $d(x,\partial S) \le \delta_2 \le \delta$. Hence

$$\sup_{x\in\partial\tilde{S}} d(x,\partial S) \le \delta. \tag{E.5}$$

Consider now $x \in \partial S$. We also consider two cases where $x \in \tilde{S}$ and $x \notin \tilde{S}$ separately. If $x \notin \tilde{S}$, then $d(x,\partial\tilde{S}) = d(x,\tilde{S}) \le \delta$ by (E.4) since $x \in \partial S \subset S$. Given now $x \in \tilde{S} \cap \partial S$. Note that since $\tilde{S} \subset S^1$ and $x \in \tilde{S}$, we have $d(x,\partial\tilde{S}) \le d(x,\partial S^1) = d(x,\overline{(S^1)^c})$. When $x \in \partial S$, $x \in \overline{S^c}$. By Assumption 3.3 and since $\delta_1 \le R$, there exists $y_x \in \overline{S^c}$ such that $x \in \overline{B}_{\delta_1}(y_x) \subset \overline{S^c}$. Hence $\|x - y_x\|_2 \le \delta_1$ and $d(y_x,\partial S) = d(y_x,\partial\overline{S^c}) \ge \delta_1$. We have now $y_x \in \overline{(S^1)^c}$ and $d(x,\overline{(S^1)^c}) \le \|x - y_x\|_2 \le \delta_1$. Then $d(x,\partial\tilde{S}) \le \delta_1 \le \delta$. Now we have

$$\sup_{x\in\partial S} d(x,\partial\tilde{S}) \le \delta. \tag{E.6}$$

By combining (E.5) and (E.6), we obtain that $d_H(\partial S,\partial\tilde{S}) \le \delta$. □

Now, by combining the bounds on $S$ which have been shown in Lemma 3.10 with the previous property of $S$, we obtain a result concerning the Hausdorff distance between two sets and two boundaries for the thresholding scheme (3.6).

**Lemma E.5.** *Under the assumptions and definitions of Lemma 3.10, we suppose in addition that $S$ satisfies Assumption 3.3 with radius $R > 0$. Recall $\delta_d$ in (3.7). For any $d > 1$ large enough such that $\delta_d \le R$, the thresholding scheme (3.6) satisfies that*

$$d_H(S, S_{d,n}) \le \delta_d$$

*and*

$$d_H(\partial S, \partial S_{d,n}) \le \delta_d.$$

*Proof.* We use Lemma 3.10, then apply Lemma E.4 with $\tilde{S} = S_{d,n}$, $\delta_1 = \delta_1(d)$, $\delta_2 = \delta_2(d,\beta)$ under the assumption that $\delta_d = \max\big(\delta_1(d), \delta_2(d,\beta)\big) \le R$. □

The proof of Theorem 3.6 follows by combining Lemma E.5 with the concentration result in Corollary 3.13.

*Proof of Theorem 3.6.* By Corollary 3.13, we have with probability at least $1-\alpha$,

$$|\Lambda_{\mu,d_n}(x) - \Lambda_{\mu_n,d_n}(x)| \le \Lambda_{\mu,d_n}(x)\max\left(\sqrt{\frac{16\,m(d_n,p,S,w)}{3n}\log\frac{s(d_n)}{\alpha}},\right.$$
$$\left.\frac{16\,m(d_n,p,S,w)}{3n}\log\frac{s(d_n)}{\alpha}\right),$$

where

$$m(d_n, p, S, w) = \frac{4^{p+r} s(d_n)}{C\omega_p R^{p+r}} \frac{(d_n+p+1)(d_n+p+2)(2d_n+p+6)}{(d_n+1)(d_n+2)(d_n+3)} + \frac{2^{p+2r} c_r}{CR^{p+r}} \left[2\binom{p+d_n+2r+1}{p+2r+1} - \binom{p+d_n+2r}{p+2r}\right].$$

Note that

$$\frac{2^{p+2r} c_r}{CR^{p+r}} \left[2\binom{p+d_n+2r+1}{p+2r+1} - \binom{p+d_n+2r}{p+2r}\right] \le \frac{2^{p+2r+1} c_r}{CR^{p+r}} \binom{p+d_n+2r+1}{p+2r+1}$$
$$\le \frac{2^{p+2r+1} c_r}{CR^{p+r}} d_n^{p+2r+1} \left(\frac{e}{p+2r+1}\right)^{p+2r+1} \exp((p+2r+1)^2)$$

and

$$\begin{aligned}&\frac{4^{p+r} s(d_n)}{C\omega_p R^{p+r}} \frac{(d_n+p+1)(d_n+p+2)(2d_n+p+6)}{(d_n+1)(d_n+2)(d_n+3)}\\ &\le \frac{4^{p+r}}{C\omega_p R^{p+r}} d_n^p \left(\frac{e}{p}\right)^p \exp(p^2) \frac{(p+2)(p+3)(p+8)}{24}\\ &\le \frac{4^{p+r}}{C\omega_p R^{p+r}} d_n^{p+2r+1} \left(\frac{e}{p}\right)^p \exp(p^2) \frac{(p+2)(p+3)(p+8)}{24},\end{aligned}$$

where the inequalities come from Lemma E.1 and the fact that the function

$$d \mapsto \frac{(d+p+1)(d+p+2)(2d+p+6)}{(d+1)(d+2)(d+3)}$$

is decreasing. Hence

$$\frac{16\, m(d_n, p, S, w)}{3n} \le \frac{4^{r+2}}{3nCR^{p+r}} d_n^{p+2r+1} \Bigg[2^{p+1} c_r \left(\frac{e}{p+2r+1}\right)^{p+2r+1} \exp((p+2r+1)^2) + \frac{4^p (p+2)(p+3)(p+8)}{24\omega_p} \left(\frac{e}{p}\right)^p \exp(p^2)\Bigg].$$

On the other hand,

$$\begin{aligned}\log \frac{s(d_n)}{\alpha} &\le \log\left[d_n^p \left(\frac{e}{p}\right)^p \exp\left(\frac{p^2}{d_n}\right)\right] - \log\alpha = p\log d_n + p(1-\log p) + \frac{p^2}{d_n} - \log\alpha\\ &\le p d_n + p(1-\log p) + p^2 - \log\alpha\\ &\le d_n \left(p + p(1-\log p) + p^2 - \log\alpha\right).\end{aligned}$$

Then

$$\frac{16\, m(d_n, p, S, w)}{3n} \log \frac{s(d_n)}{\alpha} \le \frac{d_n^{p+2r+2}}{nCR^{p+r}} C_{p,r,\alpha},$$

where $C_{p,r,\alpha}$ is in (A.1). Since $d_n = \left\lfloor \left( \frac{CR^{p+r}}{4C_{p,r,\alpha}} n \right)^{\frac{1}{p+2r+2}} \right\rfloor$, then $d_n \leq \left( \frac{CR^{p+r}}{4C_{p,r,\alpha}} n \right)^{\frac{1}{p+2r+2}}$, which implies that

$$\frac{d_n^{p+2r+2}}{nCR^{p+r}} C_{p,r,\alpha} \leq \frac{1}{4}$$

and

$$\max\left( \sqrt{\frac{16\, m(d_n, p, S, w)}{3n} \log \frac{s(d_n)}{\alpha}}, \frac{16\, m(d_n, p, S, w)}{3n} \log \frac{s(d_n)}{\alpha} \right) \leq \frac{1}{2}.$$

Now we obtain with probability at least $1 - \alpha$,

$$|\Lambda_{\mu,d_n}(x) - \Lambda_{\mu_n,d_n}(x)| \leq \frac{1}{2} \Lambda_{\mu,d_n}(x).$$

We set

$$\delta_n^1 := \delta_1(d_n),$$

$$\delta_n^2 := \delta_2\left(d_n, \frac{1}{2}\right)$$

and thus we have from (A.8)

$$\delta_n = \max(\delta_n^1, \delta_n^2).$$

Recall that we have set

$$D_{p,S,w,\epsilon} = \max\left( 2, \left( \frac{\operatorname{diam}(S)}{R} + 1 \right)^{\frac{1}{1-\epsilon}}, \left( \frac{2}{R} E_{p,r,\epsilon}\left(1, \frac{1}{2}\right) \right)^{\frac{1}{1-\epsilon}} \right).$$

Then, for $n \geq n_0 = \frac{4(D_{p,S,w,\epsilon}+1)^{p+2r+2} C_{p,r,\alpha}}{CR^{p+r}}$, $d_n = \left\lfloor \left( \frac{CR^{p+r}}{4C_{p,r,\alpha}} n \right)^{\frac{1}{p+2r+2}} \right\rfloor \geq D_{p,S,w,\epsilon}$. Hence $d_n \geq 2 > 1$. Furthermore,

$$\delta_n^1 = \frac{\operatorname{diam}(S)}{d_n^{1-\epsilon} - 1} \leq \frac{\operatorname{diam}(S)}{D_{p,S,w,\epsilon}^{1-\epsilon} - 1} \leq \frac{\operatorname{diam}(S)}{\frac{\operatorname{diam}(S)}{R} + 1 - 1} \leq R$$

and

$$\begin{aligned} \delta_n^2 &= \frac{2}{d_n^{1-\epsilon}} E_{p,r,\epsilon}\left(d_n, \frac{1}{2}\right) \leq \frac{2}{D_{p,S,w,\epsilon}^{1-\epsilon}} E_{p,r,\epsilon}\left(d_n, \frac{1}{2}\right) \\ &\leq \frac{2}{\frac{2}{R} E_{p,r,\epsilon}\left(1, \frac{1}{2}\right)} E_{p,r,\epsilon}\left(d_n, \frac{1}{2}\right) \leq R, \end{aligned}$$

which implies that $\delta_n = \max(\delta_n^1, \delta_n^2) \leq R$.

Now, all the assumptions of Lemma E.5 hold with probability at least $1-\alpha$ and we obtain the threshold $\gamma_n$ along with the estimator $S_n$ such that

$$d_H(S, S_n) \leq \delta_n$$

and

$$d_H(\partial S, \partial S_n) \leq \delta_n.$$

□

### E.3 Proof of Corollary 3.7

*Proof of Corollary 3.7.* For any two sets $A$ and $B$ of $\mathbb{R}^p$, it is well-known that

$$A \triangle B \subset (\partial A)^{d_H(A,B)} \cup (\partial B)^{d_H(A,B)}.$$

Thus we have, from Theorem 3.6, for $n \geq n_0$ and with probability at least $1-\alpha$,

$$\begin{aligned} S \triangle S_n &\subset (\partial S)^{d_H(S,S_n)} \cup (\partial S_n)^{d_H(S,S_n)} \\ &\subset (\partial S)^{\delta_n} \cup (\partial S_n)^{\delta_n} \\ &\subset (\partial S)^{\delta_n} \cup (\partial S)^{d_H(\partial S, \partial S_n)+\delta_n} \\ &\subset (\partial S)^{\delta_n} \cup (\partial S)^{2\delta_n} \\ &= (\partial S)^{2\delta_n}. \end{aligned}$$

Hence from Lemma 3.4, we obtain

$$\lambda(S \triangle S_n) \leq 2 C_S \delta_n.$$

□

## Appendix F. Proof of Lemma 3.4

*Proof of Lemma 3.4.* First we remark that for a compact set $S$ if a ball of radius $R$ rolls freely inside $S$, then $S$ has finitely many path-connected components. Indeed $S$ is contained in a big ball, and each path-connected component contains at least a small ball so that the number of connected components is at most the volume ratio of the two balls.

Using [43, Theorem 1], we know that $\partial S$ is an embedded compact differentiable manifold of dimension $p-1$ with outer pointing unit normal $n$ being $1/R$ Lipschitz, for any $x, y \in \partial S$

$$\|n(x) - n(y)\| \leq \frac{\|x-y\|}{R}. \tag{F.1}$$

Indeed, [43] uses the notion of submanifold of Euclidean space in [4, Chapter 2] which corresponds to embedded differentiable manifolds. The manifold is compact as a compact subset of $\mathbb{R}^p$.

Fix $c > \epsilon > 0$. For any $z \in \mathbb{R}^p$, with $d(z, \partial S) \leq \epsilon$, $z$ is of the form $x + \alpha n(x)$ with $x$ realizing the minimal distance to $z$ on $\partial S$ and $|\alpha| = d(z, \partial S) \leq \epsilon$. Indeed, the function $x \mapsto \|x - z\|^2$ is smooth and its differential on

$\partial S$ (seen as a manifold), corresponds to the projection of the gradient on the tangent space. Minimizers have null differential, that is $x - z$ orthogonal to the tangent space $\partial S$, or co-linear to the normal vector $n(x)$.

Reciprocally, each $z \in \mathbb{R}^p$ of this form is at distance at most $\epsilon$ to $\partial S$ so that

$$\partial S^\epsilon = \{x + \alpha n(x),\, x \in \partial S,\, |\alpha| \leq \epsilon\}.$$

Since $\partial S$ is a compact embedded differentiable manifold of dimension $p-1$, there exists a family of bounded open subsets $U_1, \dots, U_l \subset \mathbb{R}^{p-1}$ and Lipschitz diffeomorphisms $\phi_1, \dots, \phi_l$ such that

$$\partial S = \cup_{i=1}^l \phi_i(U_i).$$

Indeed [4, Theorem 2.1.2] ensures that for every $x \in \partial S$, there is an open neighborhood $U$ of $\mathbb{R}^{p-1}$ and $\phi \colon U \to \mathbb{R}^p$ and an open ball $B \subset \mathbb{R}^p$ centered at $x$, such that $\phi(U) = B \cap \partial S$ and $\phi$ is a $C^1$ homeomorphism onto its image. Reducing $B$ if necessary, $\phi$ can be taken to be Lipschitz. By compactness, $\partial S$ can be covered by finitely many balls of this form which gives the desired family (Borel-Lebesgue property).

Set $\psi_i \colon U_i \times [-c, c] \to \partial S^c$, such that $\psi_i(u, \alpha) = \phi_i(u) + \alpha n(\phi_i(u))$. We have therefore that

$$\partial S^\epsilon = \cup_{i=1}^l \psi_i(U_i \times [-\epsilon, \epsilon]).$$

Each $\psi_i$ is Lipschitz on $U_i \times [c, c]$, say with constant $L$. Using $\lambda_m$ to denote the Lebesgue measure over $m$ dimensional Euclidean space, we obtain using Lemma F.1

$$\begin{aligned}\lambda_p(\partial S^\epsilon) &\leq \sum_{i=1}^l \lambda_p(\psi_i(U_i \times [-\epsilon, \epsilon])) \\ &\leq L^p \sum_{i=1}^l \lambda_p(U_i \times (-\epsilon, \epsilon)) \\ &= 2L^p \epsilon \sum_{i=1}^l \lambda_{p-1}(U_i),\end{aligned}$$

which is the desired result. □

**Lemma F.1.** *Let $F \colon \mathbb{R}^p \to \mathbb{R}^p$ be L-Lipschitz, then for any measurable set $A$,*

$$\lambda(F(A)) \leq L^p \lambda(A).$$

*Proof.* The measurability of $F(A)$ is given in [36, Proposition 18, Chapter 20].

Recall that for any measurable set $S \subset \mathbb{R}^p$,

$$\lambda(S) = \inf_{\{R_i\}_{i \in \mathbb{N}}} \sum_{i \in \mathbb{N}} \lambda(R_i) \tag{F.2}$$

where the infimum is taken over all covers of $S$ by hyper-rectangles [36, Section 20.2]. For any $\delta > 0$, any hyperrectangle $E$ can be covered by finitely many balls $\{B_i\}_{i \in J}$ such that $\sum_{i \in J} \lambda(B_i) \leq \lambda(E) + \delta$. This implies that the infimum in (F.2) can be taken over countable unions of balls.

As a consequence, for any $\epsilon > 0$, there exists a countable collection of balls $\{B_i\}_{i \in \mathbb{N}}$ covering $A$ such that

$$\lambda(A) + \epsilon \geq \sum_{i \in \mathbb{N}} \lambda(B_i).$$

The family $\{F(B_i)\}_{i\in\mathbb{N}}$ forms a covering of $F(A)$. The image of a ball of radius $r$ by an $L$-Lipschitz function is contained in a ball of radius $Lr$ and therefore for each $i \in \mathbb{N}$, we have $\lambda(F(B_i)) \leq L^p\lambda(B_i)$. Putting things together, we have

$$\lambda(F(A)) \leq \sum_{i\in\mathbb{N}} \lambda(F(B_i)) \leq L^p \sum_{i\in\mathbb{N}} \lambda(B_i) \leq L^p(\lambda(A) + \epsilon).$$

Since $\epsilon > 0$ was arbitrary, the result holds. □

*Acknowledgements.* This work was mostly carried out during the first author's MS final project which was funded by the CIMI labex (Centre International de Mathématiques et d'Informatique de Toulouse). The third author acknowledges support of ANR-3IA Artificial and Natural Intelligence Toulouse Institute, ANR MASDOL and Air Force Office of Scientific Research, Air Force Material Command, USAF, under grant number FA9550-19-1-7026/19IOE033 and FA9550-18-1-0226. The authors would like to thank Jonas Kahn and Sébastien Gerchinovitz for motivating preliminary discussions on the topic. The authors thank an Associate Editor and a Referee, for their comments and suggestions that led to an improvement of the manuscript.

## References

[1] C. Aaron and O. Bodart, Local convex hull support and boundary estimation. *J. Multivar. Anal.* **147** (2016) 82–101.

[2] C.C. Aggarwal and S. Sathe, Theoretical foundations and algorithms for outlier ensembles. *ACM Sigkdd Explor. Newsl.* **17** (2015) 24–47.

[3] N. Aronszajn, Theory of reproducing kernels. *Trans. Am. Math. Soc.* **68** (1950) 337–404.

[4] M. Berger and B. Gostiaux, Differential Geometry: Manifolds, Curves, and Surfaces: Manifolds, Curves, and Surfaces, Vol. 115, Springer Science & Business Media (2012).

[5] A. Berlinet and C. Thomas-Agnan, Reproducing kernel Hilbert spaces in probability and statistics. Springer Science & Business Media (2011).

[6] R. Berman, S. Boucksom and D.W. Nyström, Fekete points and convergence towards equilibrium measures on complex manifolds. *Acta Math.* **207** (2011) 1–27.

[7] G. Biau, B. Cadre and B. Pelletier, Exact rates in density support estimation. *J. Multivar. Anal.* **99** (2008) 2185–2207.

[8] L. Bos, Asymptotics for the Christoffel function for Jacobi like weights on a ball in $\mathbb{R}^m$. *New Zealand J. Math.* **23** (1994) 109.

[9] L. Bos, B. Della Vecchia and G. Mastroianni, On the asymptotics of Christoffel functions for centrally symmetric weight functions on the ball in $\mathbb{R}^d$. *Rend. Cir. Mat.* **52** (1998) 277–290.

[10] J. Chevalier, Estimation du support et du contour du support d'une loi de probabilité. *Ann. Inst. Henri Poincaré Stat.* **12** (1976) 339–364.

[11] A. Cholaquidis, A. Cuevas and R. Fraiman, On Poincaré cone property. *Ann. Stat.* **42** (2014) 255–284.

[12] A. Cuevas and R. Fraiman, A plug-in approach to support estimation. *Ann. Stat.* **25** (1997) 2300–2312.

[13] A. Cuevas, R. Fraiman and B. Pateiro-López, On statistical properties of sets fulfilling rolling-type conditions. *Adv. Appl. Prob.* **44** (2012) 311–329.

[14] A. Cuevas, W. González-Manteiga and A. Rodríguez-Casal, Plug-in estimation of general level sets. *Aust. New Zealand J. Stat.* **48** (2006) 7–19.

[15] A. Cuevas and A. Rodríguez-Casal, On boundary estimation. *Adv. Appl. Prob.* **36** (2004) 340–354.

[16] L. Devroye and G.L. Wise, Detection of abnormal behavior via nonparametric estimation of the support. *SIAM J. Appl. Math.* **38** (1980) 480–488.

[17] D. Dua and C. Graff, UCI Machine Learning Repository (2017).

[18] C.F. Dunkl and Y. Xu, Vol. 155 of Orthogonal polynomials of several variables. Cambridge University Press (2014).

[19] H. Edelsbrunner, D. Kirkpatrick and R. Seidel, On the shape of a set of points in the plane. *IEEE Trans. Inf. Theory* **29** (1983) 551–559.

[20] J. Geffroy, Sur un probleme d'estimation géométrique. *Publ. Inst. Statist. Univ. Paris* **13** (1964) 191–210.

[21] F. Keller, E. Muller and K. Bohm, HiCS: High contrast subspaces for density-based outlier ranking, in *2012 IEEE 28th international conference on data engineering, IEEE* (2012) 1037–1048.

[22] A. Kroó and D. Lubinsky, Christoffel functions and universality in the bulk for multivariate orthogonal polynomials. *Can. J. Math.* **65** (2013) 600–620.

[23] A. Kroó and D. Lubinsky, Christoffel functions and universality on the boundary of the ball. *Acta Math. Hung.* **140** (2013) 117–133.

[24] J.B. Lasserre and E. Pauwels, The empirical Christoffel function with applications in data analysis. *Adv. Comput. Math.* **45** (2019) 1439–1468.

[25] E. Mammen and A.B. Tsybakov, Asymptotical minimax recovery of sets with smooth boundaries. *Ann. Stat.* **23** (1995) 502–524.
[26] S. Marx, E. Pauwels, T. Weisser, D. Henrion and J. Lasserre, Tractable semi-algebraic approximation using Christoffel-Darboux kernel. Preprint arXiv:1904.01833 (2019).
[27] I.S. Molchanov, A limit theorem for solutions of inequalities. *Scand. J. Stat.* **25** (1998) 235–242.
[28] T. Patschkowski and A. Rohde, Adaptation to lowest density regions with application to support recovery. *Ann. Stat.* **44** (2016) 255–287.
[29] E. Pauwels and J.B. Lasserre, Sorting out typicality with the inverse moment matrix SOS polynomial, in Advances in Neural Information Processing Systems (2016) 190–198.
[30] E. Pauwels, M. Putinar and J.-B. Lasserre, Data analysis from empirical moments and the Christoffel function. *Found. Comput. Math.* **21** (2021) 243–273.
[31] F. Piazzon, *Bernstein Markov properties and applications*, Ph.D. thesis, Dipartimento di Matematica, Università degli Studi di Padova (2016).
[32] W. Polonik, Measuring mass concentrations and estimating density contour clusters-an excess mass approach. *Ann. Stat.* **23** (1995) 855–881.
[33] A. Rényi and R. Sulanke, Über die konvexe Hülle von n zufällig gewählten Punkten. *Prob. Theory Related Fields* **2** (1963) 75–84.
[34] P. Rigollet and R. Vert, Optimal rates for plug-in estimators of density level sets. *Bernoulli* **15** (2009) 1154–1178.
[35] A. Rodríguez Casal, Set estimation under convexity type assumptions. *Ann. Inst. Henri Poincaré, Prob. Stat.* **43** (2007) 763–774.
[36] H.L. Royden and P. Fitzpatrick, Real analysis, vol. 32. Macmillan New York (1988).
[37] A. Singh, C. Scott and R. Nowak, Adaptive Hausdorff estimation of density level sets. *Ann. Stat.* **37** (2009) 2760–2782.
[38] G. Szegö, Vol. 23 of Orthogonal polynomials. American Mathematical Soc. (1939).
[39] V. Totik, Asymptotics for Christoffel functions for general measures on the real line. *J. d'Anal. Math.* **81** (2000) 283–303.
[40] A.B. Tsybakov, On nonparametric estimation of density level sets. *Ann. Stat.* **25** (1997) 948–969.
[41] R. Vershynin, Introduction to the non-asymptotic analysis of random matrices. Preprint arXiv:1011.3027 (2010).
[42] G. Walther, Granulometric smoothing. *Ann. Stat.* **25** (1997) 2273–2299.
[43] G. Walther, On a generalization of Blaschke's rolling theorem and the smoothing of surfaces. *Math. Methods Appl. Sci.* **22** (1999) 301–316.
[44] Y. Xu, Asymptotics of the Christoffel functions on a simplex in $\mathbb{R}^d$. *J. Approx. Theory* **99** (1999) 122–133.
[45] Y. Xu, Summability of Fourier orthogonal series for Jacobi weight on a ball in $\mathbb{R}^d$. *Trans. Am. Math. Soc.* **351** (1999) 2439–2458.