



HAL
open science

Estimation via length-constrained generalized empirical principal curves under small noise

Sylvain Delattre, Aurélie Fischer

► **To cite this version:**

Sylvain Delattre, Aurélie Fischer. Estimation via length-constrained generalized empirical principal curves under small noise. 2019. hal-02361315

HAL Id: hal-02361315

<https://hal.science/hal-02361315>

Preprint submitted on 13 Nov 2019

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

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

Estimation via length-constrained generalized empirical principal curves under small noise

Sylvain Delattre & Aurélie Fischer*

November 13, 2019

Abstract

Let $g : [0, 1] \rightarrow \mathbb{R}^d$ be a rectifiable curve, and, for $n \geq 1$, let U_i^n , $i = 1, \dots, n$, denote independent random variables taking their values in $[0, 1]$, with full support. Consider the model

$$X_i^n = g(U_i^n) + \varepsilon_i^n, \quad i = 1, \dots, n,$$

where g is unknown and the noise tends to 0 in probability (in a way to be specified). We are interested in the estimation of the image of g .

Given a square integrable random vector X , let $f : [0, 1] \rightarrow \mathbb{R}^d$ be a minimizer of

$$\Delta(f) = E \left[\min_{t \in [0, 1]} V(|X - f(t)|) \right]$$

over all curves with length not greater than a certain threshold. Here, $V : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is a lower semi-continuous strictly increasing function. For instance, $V(x) = x^p$, where $p > 0$, or $V(x) = \frac{x}{1+x}$.

Similarly, an empirical optimal curve associated to X_1^n, \dots, X_n^n may be defined as a minimizer, under length-constraint, of the criterion

$$\Delta_n(f) = \frac{1}{n} \sum_{i=1}^n \min_{t \in [0, 1]} V(|X_i^n - f(t)|).$$

In this paper, we propose a method to build a sequence of generalized empirical principal curves, with selected length, so that, in Hausdorff distance, the images of the estimating principal curves converge in probability to the image of g .

Keywords – Curve estimation, additive noise, principal curves, length constraint.

2010 Mathematics Subject Classification: 62G05.

1 Introduction

1.1 Preliminary picture of the estimation result

Let $n \geq 1$. We observe random vectors X_i^n , given by

$$X_i^n = g(U_i^n) + \varepsilon_i^n, \quad i = 1, \dots, n, \tag{1.1}$$

*The research work of this author has been partially supported by ANR-18-IDEX-0001, IdEx Université de Paris.

where the unknown function $g : [0, 1] \rightarrow \mathbb{R}^d$ is continuous. Moreover, g is assumed to have finite length equal to its 1-dimensional Hausdorff measure and to have constant speed. Here, the random variables U_i^n , $i = 1, \dots, n$, taking their values in $[0, 1]$, are independent, and belong to a class of distributions with full support, enclosing for instance the uniform distribution as a particular case.

We study an asymptotic context, where the noise tends in probability to 0 (in a sense that will be specified below) when the number of observations n tends to infinity.

The main result of this paper is the construction, relying on the principal curve notion, of an estimator \hat{f}_n , which converges to the unknown curve g in Hausdorff distance, in the sense that the Hausdorff distance between $\text{Im}\hat{f}_n$ and $\text{Im}g$ converges in probability to 0.

1.2 Related work

The problem of estimating the image of g may be cast into the general context of filament or manifold estimation from observations sampled on or near the unknown shape.

The literature mainly focuses on shapes with a reach bounded away from zero. The reach ρ , characterizing the regularity of the shape, is the maximal radius of a ball rolling on it (see Federer (1969)). In Genovese et al. (2012a), an additive noise model of the form (1.1) is studied. The curve g is parameterized by arc-length, normalized to $[0, 1]$. The authors assume that the U_i , $i = 1, \dots, n$, have a common density with respect to the Lebesgue measure on $[0, 1]$, bounded and bounded away from zero. The noise has support in a ball $B(0, \sigma)$, with $\sigma < \rho(g)$, and admits a bounded density with respect to the Lebesgue measure, which is continuous on $\overset{\circ}{B}(0, \sigma)$, nondecreasing and symmetric, with a regularity condition on the boundary of the support. For an open curve (with endpoints), in addition, $|f(1) - f(0)|/2 > \sigma$. In the plane \mathbb{R}^2 , the assumptions made allow to estimate the support S of the distribution of the observations, the boundary of this set S , in order to find its medial axis, which is the closure of the set of points in S that have at least two closest points in the boundary ∂S . In the same article, the authors also consider clutter noise, corresponding to the situation where one observes points sampled from a mixture density $(1 - \eta)u(x) + \eta h(x)$, where u is the uniform density over some compact set, and h is the density of points on the shape. Another additive model is investigated in Genovese et al. (2012b), for the estimation of manifolds without boundary, with dimension lower than the dimension of the ambient space, contained in a compact set. The model may be written

$$X_i = G_i + \varepsilon_i, \quad i = 1, \dots, n, \quad (1.2)$$

where the random vectors G_i are drawn uniformly on the shape M , and the noise is drawn uniformly on the normal to the manifold, at distance at most $\sigma < \rho(M)$. The article Genovese et al. (2012c) is also dedicated to manifold estimation, under reach condition, first in a noiseless model, where the observations are exactly sampled on the manifold, according to some density with respect to the uniform distribution on the manifold, and then in the presence of clutter noise. An additive noise model, with known Gaussian noise, is examined as well. This latter case is related to density deconvolution. Estimating manifolds without boundary, with low dimension and a lower bound on the reach, is also the purpose of Aamari and Levrard (2018, 2019). The points sampled on the manifold have a common density with respect to the d -dimensional Hausdorff measure of the manifold, which is bounded and bounded away from zero. In Aamari and Levrard (2018), estimation relies on Tangential Delaunay Complexes. It is performed in the noiseless case, with additive noise, bounded by σ , and under clutter noise. Aamari and Levrard (2019) deal

with compact manifolds belonging to particular regularity classes. The authors examine the noiseless situation, as well as centered bounded noise perpendicular to the manifold. Estimators based on local polynomials are proposed.

To sum up, all these models involve strong conditions on the noise, which is either bounded, or of type clutter noise. Such assumptions allow the authors to derive rates of convergence. Here, we investigate a different situation, with a weak assumption on the noise. In particular, the noise does not need to be bounded. Regarding the regularity of the curve g , which has constant speed, there is no reach assumption, and g is not required to be injective. Although rates of convergence cannot be expected here, this weak framework is worth studying, since it is not obvious at first sight that it is even possible to build a convergent estimator without knowledge of either length or noise.

The estimation strategy relies on generalized empirical principal curves.

1.3 Extension of the notion of length-constrained principal curve

The notion of principal curve with length constraint has been proposed by Kégl et al. (2000). According to this definition, if X denotes a random vector with finite second moment, a principal curve is a continuous map $f^* : [0, 1] \rightarrow \mathbb{R}^d$ minimizing under a length constraint the quantity

$$E \left[\min_{t \in [0,1]} |X - f(t)|^2 \right] = E \left[d(X, \text{Im}f)^2 \right], \quad (1.3)$$

where $|\cdot|$ denotes the Euclidean norm and d stands for the Euclidean distance to a set. This optimization problem may also be seen as a version of the “average distance problem” studied in the calculus of variations community (see, e.g., Buttazzo and Stepanov (2003); Buttazzo et al. (2002)). Originally, a principal curve was defined by Hastie and Stuetzle (1989) as a self-consistent curve, that is, a curve f satisfying $f(t_f(X)) = \mathbb{E}[X | t_f(X)]$ a.s., with t_f given by $t_f(x) = \max \arg \min_t |x - f(t)|$. In addition to self-consistency, smoothness conditions were required: the principal curve has to be of class C^∞ , it does not intersect itself, and has finite length inside any ball in \mathbb{R}^d . Tibshirani (1992) revisited the problem as a mixture model, which forces the curve g in models of the form (1.1) to be a principal curve. The point of view by Kégl et al. (2000), where no smoothness assumption is made, was motivated in particular by the fact that the existence of principal curves defined in terms of self-consistency was only proved for a few particular examples (see Duchamp and Stuetzle (1996a,b)). Note that principal curves introduced by Kégl et al. (2000) include polygonal lines.

As stated in the next lemma, shown in Section A.1, existence of optimal curves is still guaranteed when replacing the squared Euclidean distance in the definition (1.3) by more general distortion measures.

Lemma 1.1. *Let $V : [0, \infty) \rightarrow [0, \infty)$ be a lower semi-continuous, strictly increasing function, continuous at 0, and such that $V(0) = 0$. Let X denote a random vector such that $E[V(|X|)] < \infty$. Then, for any finite length L , there exists a curve $f_L^* : [0, 1] \rightarrow \mathbb{R}^d$ with length $\mathcal{L}(f_L^*) \leq L$ minimizing over all curves with length at most L the criterion*

$$\Delta(f) = E \left[V(d(X, \text{Im}f)) \right].$$

The motivation for introducing this generalized notion of principal curves is that this allows for greater flexibility in the way we measure distances. This framework encloses

for instance as particular cases the power functions $V(x) = x^p$, $p > 0$. An appropriate choice of V may enhance robustness. A typical example in this regard is the function defined by $V(x) = \frac{x}{1+x}$.

In a statistical context, one has at hand independent observations X_1, \dots, X_n , and an empirical principal curve is defined as a minimizer, under a length constraint, of the criterion

$$\frac{1}{n} \sum_{i=1}^n d(X_i, \text{Im}f)^2.$$

Similarly, a generalized empirical principal curve may be obtained by minimizing

$$\frac{1}{n} \sum_{i=1}^n V(d(X_i, \text{Im}f)).$$

Observe that, in this case, existence of a minimizer is more straightforward since the empirical measure is compactly supported.

1.4 Organization of the paper

The manuscript is organized as follows. In Section 2, we set up notation and introduce more formally the model. In Section 3, we state and prove the main result: we build a sequence of generalized empirical principal curves converging to the curve to be estimated in Hausdorff distance. The proof is structured in two subsections. The first one gathers results around the Cauchy-Crofton formula, which allows to show a useful fact about the considered class of sampling distributions on $[0, 1]$. The proof of the existence Lemma 1.1, as well as a technical measurability result, are collected in Appendix A.

2 Definitions and notation

2.1 Notation

We consider the space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), |\cdot|)$, equipped with the standard Euclidean norm, associated to the inner product $\langle \cdot, \cdot \rangle$. Here, $\mathcal{B}(E)$ denotes the Borel sigma-algebra of a space E .

Let \mathcal{H}^1 denotes the 1-dimensional Hausdorff measure in \mathbb{R}^d .

In the sequel, for a compact set A , $\text{diam}(A)$ stands for the diameter of a set A and $d(x, A)$ for the distance from the point x to the set A , that is

$$\text{diam}(A) = \max_{x, y \in A} |x - y|, \quad d(x, A) = \min_{y \in A} |x - y|.$$

We denote by $d_H(A, B)$ the Hausdorff distance between two sets A and B , given by

$$d_H(A, B) = \sup_{a \in A} d(a, B) \vee \sup_{b \in B} d(b, A).$$

Let λ stand for the Lebesgue measure and δ_x for the Dirac measure at x .

Throughout, an interval (a, b) will denote an open interval of $[0, 1]$ equipped with the induced topology.

Denote by \mathcal{D} a metric associated to weak convergence. For a probability measure μ and a closed set of probability measures \mathcal{M} , let $\mathcal{D}(\mu, \mathcal{M}) = \min_{\mu' \in \mathcal{M}} \mathcal{D}(\mu, \mu')$.

For two probability measures μ and μ' , we define the bounded Lipschitz metric between μ and μ' by

$$|\mu - \mu'|_{BL} = \sup \left\{ |\mu(h) - \mu'(h)| : |h|_\infty \leq 1, \sup_{x \neq y} \frac{|h(x) - h(y)|}{|x - y|} \leq 1 \right\}.$$

A continuous function from $[0, 1]$ to \mathbb{R}^d will be called a curve. If a curve f is rectifiable, its length will be denoted by $\mathcal{L}(f)$. Finally, we will denote by $C([0, 1])$ the metric space of continuous functions from $[0, 1]$ to \mathbb{R}^d , equipped with the topology of uniform convergence.

2.2 Description of the model

Let $g : [0, 1] \rightarrow \mathbb{R}^d$ be a curve with finite length and constant speed, such that the length equals the 1-dimensional Hausdorff distance.

Given $c > 0$, we define \mathcal{M}_c as the closed family of probability distributions μ on $[0, 1]$ satisfying $\mu \geq c\lambda$ on $[0, 1]$.

For $n \geq 1$, we observe a triangular array of random vectors X_i^n , given by the model

$$X_i^n = g(U_i^n) + \varepsilon_i^n, \quad i = 1, \dots, n, \quad (2.1)$$

where the U_i^n , $i = 1, \dots, n$, are independent and for every $i = 1, \dots, n$, the distribution μ_i^n of U_i^n belongs to \mathcal{M}_c .

Let $V : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be a lower semi-continuous, strictly increasing function, continuous at 0, and such that $V(0) = 0$. Moreover, we assume that V satisfies the following property: there exist a constant $C > 0$, such that, for every $(x, y) \in \mathbb{R}_+$

$$V(x + y) \leq C(V(x) + V(y)).$$

For a curve f , we define

$$\Delta_n(f) = \frac{1}{n} \sum_{i=1}^n V(d(X_i^n, \text{Im}f)).$$

Remark 1. If we set $V(x) = x^2$, we find the usual principal curve definition by Kégl et al. (2000).

We also define a function $T(f, \cdot) : \mathbb{R}^d \rightarrow [0, 1]$, by setting

$$T(f, x) = \max \arg \min_{t \in [0, 1]} |x - f(t)|.$$

For every $L > 0$, let

$$G_n(L) = \min_{\mathcal{L}(f) \leq L} \Delta_n(f),$$

and let $\hat{f}_{n,L}$ denote an empirically optimal curve with length at most L , that is a random variable taking its values in $C([0, 1])$ such that

$$\Delta_n(\hat{f}_{n,L}) = G_n(L).$$

Moreover, we choose $\hat{f}_{n,L}$ L -Lipschitz. We set $\Lambda_n := \inf\{L \geq 0, G_n(L) = 0\}$.

3 Main result

We consider the estimation of the curve g in Model (2.1) using a sequence of generalized empirical principal curves, that is a sequence of curves which are optimal with respect to the criterion Δ_n .

Theorem 3.1. *Let $g : [0, 1] \rightarrow \mathbb{R}^d$ be a curve, such that $\mathcal{L}(g) \leq \Lambda < \infty$, and $|g'(t)| = \mathcal{L}(g) dt$ -a.e.. Assume that $\mathcal{L}(g) = \mathcal{H}^1(\text{Img})$. We consider Model (2.1), with $\frac{1}{n} \sum_{i=1}^n V(|\varepsilon_i^n|)$ tending to 0 in probability as n tends to infinity. Let \hat{L}_n be defined by*

$$\hat{L}_n \in \arg \min_{L \in a_n \mathbb{N} \cap [0, \Lambda_n \wedge \Lambda]} \left[L^2 \mathcal{D} \left(\frac{1}{n} \sum_{i=1}^n \delta_{T(\hat{f}_{n,L}, X_i^n)}, \mathcal{M}_c \right) + \Delta_n(\hat{f}_{n,L}) \right],$$

where $a_n > 0$ for every $n \geq 1$ and $a_n \rightarrow 0$ as $n \rightarrow \infty$. Then, $d_H(\text{Im} \hat{f}_{n, \hat{L}_n}, \text{Img})$ converges in probability to 0 as n tends to infinity.

Some comments are in order.

First, let us discuss the assumptions. The requirement $\mathcal{L}(g) \leq \Lambda < \infty$ is technical. It allows, in the proof, to consider limit points of the constructed sequence of empirical principal curves. From an applied point of view, this is not a limitation of the procedure. Indeed, in practice, we will always consider a finite grid for the length. Moreover, with a fixed number of observations, the minimal length needed to join all points is a finite upper bound for the length. The condition $\mathcal{L}(g) = \mathcal{H}^1(\text{Img})$ ensures that the image of g is parameterized with minimal possible length. Indeed, there exist an infinite number of parameterizations, with infinite possibilities for the length. In words, generically, a portion of image of g cannot be traveled several times. The case where g is injective is a particular case. Nevertheless, here, an image with loops is allowed. We also require $|g'(t)| = \mathcal{L}(g) dt$ -a.e., which means that the image of g is parameterized with constant speed $\mathcal{L}(g)$. These assumptions about the parametrization allow to show a key relation between the distribution class \mathcal{M}_c and its image by g (see Lemma 4.4 below), the proof of which relies on the Cauchy-Crofton formula for the length of a rectifiable curve.

Observe that the main strength of the result is that it provides a convergent estimator in a very general framework. Neither the length, nor the noise level, converging to 0 in a very weak sense, is known. Intuitively, considering a practical situation with a fixed number of observations, the same data cloud could arise from several different generative curves, more or less long, in a model with more or less noise. This illustrates the benefit of an estimator construction which does not require the knowledge of any of the two parameters. Apart from the upper bound Λ , which does not really need calibration in practice, as already mentioned, the procedure only depends on a single parameter, namely the constant c characterizing the class of possible sampling distributions \mathcal{M}_c .

It should be noticed that the theorem does not guarantee that the procedure allows to recover the true underlying length. Nevertheless, the proof below shows that the selected length cannot be too short: for all $\varepsilon > 0$ one has $P(\hat{L}_n \leq \mathcal{L}(g) - \varepsilon) \rightarrow 0$.

If g is closed ($g(0) = g(1)$), then Theorem 3.1 still holds when $\hat{f}_{n,L}$ is chosen as a closed empirical optimal curve with length less than L .

As mentioned in the Introduction, the proof of Theorem 3.1 is split into two parts. First, we state and prove the Cauchy-Crofton formula, together with a related result, and we use them to establish an equivalence linking \mathcal{M}_c and its image by g (Section 4). The rest of the proof of the theorem, divided in several lemmas, is presented thereafter (Section 5).

4 Cauchy-Crofton formula and relation linking \mathcal{M}_c to its image

In the sequel, we will make use of the Cauchy-Crofton formula (Cauchy (1850); Crofton (1868)) for the length of a rectifiable curve f in \mathbb{R}^d (see, e.g., Ayari and Dubuc (1997)). We recall the formula in the next lemma, and give a proof for the sake of completeness. Let $\mathcal{S}^{d-1} = \{z \in \mathbb{R}^d, |z| = 1\}$. For $\theta \in \mathcal{S}^{d-1}$ and $r \in [0, \infty)$, let

$$D_{\theta,r} = \{z \in \mathbb{R}^d \mid \langle \theta, z \rangle = r\}.$$

Lemma 4.1 (Cauchy-Crofton formula). *The length of a rectifiable curve $f : [0, 1] \rightarrow \mathbb{R}^d$ is given by*

$$\mathcal{L}(f) = \frac{1}{c_d} \int_{\mathcal{S}^{d-1}} \int_0^\infty \text{Card}(\{t \in [0, 1], f(t) \in D_{\theta,r}\}) dr d\theta,$$

where $c_d > 0$ is a constant depending on the dimension d .

Remark 2. This result may also be written in the following equivalent form :

$$\mathcal{L}(f) = \frac{1}{c_d} \int_{\mathcal{S}^{d-1}} \int_0^\infty \sum_{y \in \text{Im}f \cap D_{\theta,r}} \text{Card}(f^{-1}(\{y\})) dr d\theta.$$

Proof. For $p \geq 1$, consider the polygonal line f_p defined by the segments $[f(\frac{i}{p}), f(\frac{i+1}{p})]$, $0 \leq i \leq p-1$. We define f_θ and $f_{p,\theta}$ respectively by $f_\theta(t) = \langle \theta, f(t) \rangle$ and $f_{p,\theta}(t) = \langle \theta, f_p(t) \rangle$. There exist $\theta_i \in \mathcal{S}^{(d-1)}$ and $\rho_i \in [0, \infty)$, $0 \leq i \leq p-1$, such that $f(\frac{i+1}{p}) - f(\frac{i}{p}) = \rho_i \theta_i$. Then, the variation of $f_{p,\theta}$ is $V(f_{p,\theta}) = \sum_{i=1}^{p-1} |\langle \theta, f(\frac{i+1}{p}) - f(\frac{i}{p}) \rangle| = \sum_{i=1}^{p-1} \rho_i |\langle \theta, \theta_i \rangle|$. Hence,

$$\int_{\mathcal{S}^{d-1}} V(f_{p,\theta}) d\theta = \sum_{i=1}^{p-1} \rho_i \int_{\mathcal{S}^{d-1}} |\langle \theta, \theta_i \rangle| d\theta := c_d \sum_{i=1}^{p-1} \rho_i = c_d \mathcal{L}(f_p).$$

We have $\lim_{p \rightarrow +\infty} \mathcal{L}(f_p) = \mathcal{L}(f)$ and $\lim_{p \rightarrow +\infty} V(f_{p,\theta}) = V(f_\theta)$ (Alexandrov and Reshetnyak, 1989, Corollary of Theorem 2.1.2). By the Cauchy-Schwarz inequality, $V(f_{p,\theta}) \leq \mathcal{L}(f_p)$, and by definition of the length, $\mathcal{L}(f_p) \leq \mathcal{L}(f)$. Thanks to Lebesgue's dominated convergence theorem,

$$\lim_{p \rightarrow +\infty} \int_{\mathcal{S}^{d-1}} V(f_{p,\theta}) d\theta = \int_{\mathcal{S}^{d-1}} V(f_\theta) d\theta.$$

We deduce that

$$\mathcal{L}(f) = \frac{1}{c_d} \int_{\mathcal{S}^{d-1}} V(f_\theta) d\theta.$$

Besides, according to Banach's formula (see Banach (1925)), we have

$$V(f_\theta) = \int_0^\infty \text{Card}(\{t \in [0, 1], \langle \theta, f(t) \rangle = r\}) dr.$$

Consequently, we get the Cauchy-Crofton formula:

$$\mathcal{L}(f) = \frac{1}{c_d} \int_{\mathcal{S}^{d-1}} \int_0^\infty \text{Card}(\{t \in [0, 1], \langle \theta, f(t) \rangle = r\}) dr d\theta.$$

□

The next equality, corresponding to the Cauchy-Crofton formula applied to open subset of Img , will be useful in the sequel.

Remark 3. Let $(a, b) \subset [0, 1]$. Then,

$$\mathcal{L}(f|_{(a,b)}) = \frac{1}{c_d} \int_{\mathcal{S}^{d-1}} \int_0^\infty \text{Card}(\{t \in (a, b), f(t) \in D_{\theta,r}\}) dr d\theta.$$

Since

$$\mathcal{L}(f|_{(a,b)}) = \int_0^1 \mathbf{1}_{(a,b)}(t) |f'(t)| dt,$$

we have

$$\int_0^1 \mathbf{1}_{(a,b)}(t) |f'(t)| dt = \frac{1}{c_d} \int_{\mathcal{S}^{d-1}} \int_0^\infty \sum_{t \in [0,1]} \mathbf{1}_{(a,b)}(t) \mathbf{1}_{\{f(t) \in D_{\theta,r}\}} dr d\theta.$$

Hence, by linearity, if (a_i, b_i) , $i \geq 1$, are pairwise disjoint open intervals of $[0, 1]$, we have

$$\int_0^1 \mathbf{1}_{\bigcup_{i \geq 1} (a_i, b_i)}(t) |f'(t)| dt = \frac{1}{c_d} \int_{\mathcal{S}^{d-1}} \int_0^\infty \sum_{t \in [0,1]} \mathbf{1}_{\bigcup_{i \geq 1} (a_i, b_i)}(t) \mathbf{1}_{\{f(t) \in D_{\theta,r}\}} dr d\theta. \quad (4.1)$$

In the sequel, we will also use a Cauchy-Crofton-type formula for g taking the form of an equality for measures.

This result relies on the following lemma.

Lemma 4.2. *Let $f : [0, 1] \rightarrow \mathbb{R}^d$ be a rectifiable curve. Then, the trace of \mathcal{H}^1 on $\text{Im}f$ satisfies $\mathcal{H}^1 \leq \gamma$, where γ is the measure defined on every Borel set $A \subset \text{Im}f$ by*

$$\gamma(A) = \frac{1}{c_d} \int_{\mathcal{S}^{d-1}} \int_0^\infty \text{Card}(A \cap D_{\theta,r}) dr d\theta.$$

As a preliminary result, the next lemma states the measurability of $(\theta, r) \mapsto \text{Card}(A \cap D_{\theta,r})$. The proof is postponed to the Appendix (Section A.2).

Lemma 4.3. *Let $f : [0, 1] \rightarrow \mathbb{R}^d$ be a rectifiable curve. For $\theta \in \mathcal{S}^{d-1}$, $r \in \mathbb{R}_+$, A a Borel subset of $\text{Im}f$, let $N_{\theta,r}(A) = \text{Card}(A \cap D_{\theta,r})$. Then, the function $(\theta, r) \rightarrow N_{\theta,r}(A)$ is measurable for the Lebesgue sigma-algebra.*

Proof of Lemma 4.2. An open subset of $\text{Im}f$ may be written

$$\mathcal{O} \cap \text{Im}f = f\left(\bigcup_{k \geq 1} (a_k, b_k)\right) = \bigcup_{k \geq 1} f((a_k, b_k)),$$

where \mathcal{O} is an open subset of \mathbb{R}^d , and (a_k, b_k) , $k \geq 1$, are pairwise disjoint open intervals of $[0, 1]$. Let $\mathcal{V} = \{f([\alpha, \beta]); \exists k \geq 1, a_k < \alpha < \beta < b_k, \}$. This set is a Vitali class for $\mathcal{O} \cap \text{Im}f$, that is, for every $y = f(x)$, where $x \in \bigcup_{k \geq 1} (a_k, b_k)$, and every $\delta > 0$, there exist $\alpha < \beta$, such that $y \in f([\alpha, \beta])$ and $0 < \text{diam}(f([\alpha, \beta])) < \delta$. According to Vitali's covering theorem, for every $\varepsilon > 0$, there exist intervals $[\alpha_i, \beta_i] \subset \bigcup_{k \geq 1} (a_k, b_k)$, $i \geq 1$, such that the sets $f([\alpha_i, \beta_i])$, $i \geq 1$, are pairwise disjoint, $\mathcal{H}^1(\bigcup_{i \geq 1} f([\alpha_i, \beta_i])) = \mathcal{H}^1(\mathcal{O} \cap \text{Im}f)$ and $\mathcal{H}^1(\mathcal{O} \cap \text{Im}f) \leq \sum_{i \geq 1} \text{diam}(f([\alpha_i, \beta_i])) + \varepsilon$ (see (Falconer, 1985, Theorem 1.10)). Hence, for every $i \geq 1$, there exist $x_i < y_i$, in $[\alpha_i, \beta_i]$, such that

$$\begin{aligned} \mathcal{H}^1(\mathcal{O} \cap \text{Im}f) &\leq \sum_{i \geq 1} |f(x_i) - f(y_i)| + \varepsilon \\ &= \sum_{i \geq 1} \frac{1}{c_d} \int_{\mathcal{S}^{d-1}} \int_0^\infty \text{Card}(t \in [0, 1], h_i(t) \in D_{\theta,r}) dr d\theta + \varepsilon, \end{aligned}$$

thanks to the Cauchy-Crofton formula applied to the functions $h_i : t \in [0, 1] \mapsto tf(x_i) + (1-t)f(y_i)$, for all $i \geq 1$. Observe that $\text{Card}(t \in [0, 1], h_i(t) \in D_{\theta,r}) \in \{0, 1\}$ $d\theta dr$ a.e.. By Lemma 4.3, the function $(\theta, r) \rightarrow \text{Card}(A \cap D_{\theta,r})$ is measurable for the Lebesgue sigma-algebra, for every Borel subset A of $\text{Im}f$. If $\text{Card}(t \in [0, 1], h_i(t) \in D_{\theta,r}) = 1$, then $\text{Card}(\{f([\alpha_i, \beta_i]) \cap D_{\theta,r}\}) \geq 1$. Thus,

$$\begin{aligned} \mathcal{H}^1(\mathcal{O} \cap \text{Im}f) &\leq \sum_{i \geq 1} \frac{1}{c_d} \int_{S^{d-1}} \int_0^\infty \text{Card}(\{f([\alpha_i, \beta_i]) \cap D_{\theta,r}\}) dr d\theta + \varepsilon \\ &= \frac{1}{c_d} \int_{S^{d-1}} \int_0^\infty \text{Card}(\{\mathcal{O} \cap \text{Im}f \cap D_{\theta,r}\}) dr d\theta + \varepsilon. \end{aligned}$$

As ε is arbitrary, $\mathcal{H}^1(\mathcal{O} \cap \text{Im}f) \leq \frac{1}{c_d} \int_{S^{d-1}} \int_0^\infty \text{Card}(\{\mathcal{O} \cap \text{Im}f \cap D_{\theta,r}\}) dr d\theta$. We define γ , for every Borel set $A \subset \text{Im}f$, by $\gamma(A) = \frac{1}{c_d} \int_{S^{d-1}} \int_0^\infty \text{Card}(A \cap D_{\theta,r}) dr d\theta$: γ is a measure, satisfying $\mathcal{H}^1(\mathcal{O} \cap \text{Im}f) \leq \gamma(\mathcal{O} \cap \text{Im}f)$. According to the Cauchy-Crofton formula, $\gamma(\text{Im}f) \leq \mathcal{L}(f) < \infty$, so that the measure γ is finite. By outer regularity of finite measures, the trace of \mathcal{H}^1 on $\text{Im}f$ is less than the measure γ . \square

Remark 4. For g such that $\mathcal{H}^1(\text{Im}g) = \mathcal{L}(g)$, $\mathcal{H}^1 = \gamma$. Indeed, since $\mathcal{H}^1 \leq \gamma$ by Lemma 4.2, it is sufficient to show that both measures have the same mass. Yet, on the one hand, $\mathcal{H}^1(\text{Im}g) \leq \gamma(\text{Im}g)$ by Lemma 4.2, and on the other hand, $\gamma(\text{Im}g) \leq \mathcal{L}(g)$ by the Cauchy-Crofton formula (Lemma 4.1), so that the assumption $\mathcal{H}^1(\text{Im}g) = \mathcal{L}(g)$ implies $\mathcal{H}^1(\text{Im}g) = \gamma(\text{Im}g)$.

Remark 5. For g such that $\mathcal{H}^1(\text{Im}g) = \mathcal{L}(g)$, $\text{Card}(g^{-1}(\{y\})) = 1$ for almost every y with respect to the trace of \mathcal{H}^1 on $\text{Im}g$. This fact follows from

$$\mathcal{H}^1(\text{Im}g) = \gamma(\text{Im}g) = \frac{1}{c_d} \int_{S^{d-1}} \int_0^\infty \sum_{y \in \text{Im}g \cap D_{\theta,r}} 1 dr d\theta,$$

together with the Cauchy-Crofton formula for g (see Remark 2):

$$\mathcal{L}(f) = \frac{1}{c_d} \int_{S^{d-1}} \int_0^\infty \sum_{y \in \text{Im}g \cap D_{\theta,r}} \text{Card}(g^{-1}(\{y\})) dr d\theta.$$

We are now in a position to state the next lemma, which characterizes the image by g of a distribution belonging to the class \mathcal{M}_c .

Lemma 4.4. *Let $g : [0, 1] \rightarrow \mathbb{R}^d$ be a curve such that $0 < \mathcal{L}(g) < \infty$, $|g'(t)| = \mathcal{L}(g)$ a.e., and $\mathcal{H}^1(\text{Im}g) = \mathcal{L}(g)$. Let μ be a probability distribution supported in $[0, 1]$, and let $c > 0$ denote a constant. Then,*

$$\mu \geq c\lambda \Leftrightarrow \forall A \subset \mathcal{B}(\mathbb{R}^d) \cap \text{Im}g, \mu \circ g^{-1}(A) \geq c \frac{\mathcal{H}^1(A)}{\mathcal{L}(g)}. \quad (4.2)$$

Let us denote by \mathcal{M}_c^g the family of probability distributions m on \mathbb{R}^d , with support $\text{Im}g$, such that $\forall A \subset \mathcal{B}(\mathbb{R}^d) \cap \text{Im}g, m(A) \geq c \frac{\mathcal{H}^1(A)}{\mathcal{L}(g)}$. Hence, the equivalence (4.2) means

$$\mu \in \mathcal{M}_c \Leftrightarrow \mu \circ g^{-1} \in \mathcal{M}_c^g.$$

In the proof of Lemma 4.4, we will use the fact that the property $\mathcal{H}^1(\text{Im}g) = \mathcal{L}(g)$ may be localized, as shown in the next lemma.

Lemma 4.5. *Let $g : [0, 1] \rightarrow \mathbb{R}^d$ be a curve such that $0 < \mathcal{L}(g) < \infty$, and $\mathcal{H}^1(\text{Img}) = \mathcal{L}(g)$. Considering a subdivision $a_0 = 0 < a_1 < \dots < a_n = 1$, we have, for every $1 \leq i \leq n$,*

$$\mathcal{L}(g|_{(a_{i-1}, a_i)}) = \mathcal{H}^1(g((a_{i-1}, a_i))).$$

Proof. If not, there exists $i_0 \in \{1, \dots, n\}$ such that $\mathcal{L}(g|_{(a_{i_0-1}, a_{i_0})}) > \mathcal{H}^1((a_{i_0-1}, a_{i_0}))$, which implies

$$\begin{aligned} \mathcal{H}^1(g([0, 1])) &\leq \sum_{i=1}^n \mathcal{H}^1(g((a_{i-1}, a_i))) \\ &< \sum_{i=1}^n \mathcal{L}(g|_{(a_{i-1}, a_i)}) = \mathcal{L}(g). \end{aligned}$$

□

Proof of Lemma 4.5. \Rightarrow Assume that $\mu \geq c\lambda$. An open subset of Img may be written

$$\mathcal{O} \cap \text{Img} = g \left(\bigcup_{i \geq 1} (a_i, b_i) \right) = \bigcup_{i \geq 1} g((a_i, b_i)),$$

where \mathcal{O} is an open subset of \mathbb{R}^d , and (a_i, b_i) , $i \geq 1$, are pairwise disjoint open intervals of $[0, 1]$.

Thanks to the assumption on μ , we have

$$\begin{aligned} \mu \circ g^{-1}(\mathcal{O} \cap \text{Img}) &\geq c\lambda(g^{-1}(\mathcal{O} \cap \text{Img})) \\ &\geq c\lambda \left(\bigcup_{i \geq 1} (a_i, b_i) \right) \\ &= c \sum_{i \geq 1} (b_i - a_i) \end{aligned} \tag{4.3}$$

$$= c \sum_{i \geq 1} \frac{\mathcal{L}(g|_{(a_i, b_i)})}{\mathcal{L}(g)} \tag{4.4}$$

$$\begin{aligned} &= c \sum_{i \geq 1} \frac{\mathcal{H}^1(g((a_i, b_i)))}{\mathcal{L}(g)} \\ &\geq c \frac{\mathcal{H}^1(\mathcal{O} \cap \text{Img})}{\mathcal{L}(g)}. \end{aligned} \tag{4.5}$$

For the equality (4.3), we used that the intervals (a_i, b_i) are disjoint, for (4.4), the property $|g'(t)| = \mathcal{L}(g)$ a.e., and then for (4.5), the localized version of the equality $\mathcal{H}^1(\text{Img}) = \mathcal{L}(g)$ (Lemma 4.5).

The result extends to every Borel subset of Img , using the outer regularity of probability measures.

Remark 6. Taking $c = 1$ and $\mu = \lambda$, we obtain that $\lambda \circ g^{-1}$ is the trace of $\frac{\mathcal{H}^1}{\mathcal{L}(g)}$ on Img , since both measures are probability measures.

\Leftarrow Assume that $\forall A \subset \mathcal{B}(\mathbb{R}^d) \cap \text{Img}, \mu \circ g^{-1}(A) \geq c \frac{\mathcal{H}^1(A)}{\mathcal{L}(g)}$. An open subset of $[0, 1]$, for the induced topology, has the form $\bigcup_{i \geq 1} (a_i, b_i)$, where $(a_i, b_i), i \geq 1$ are pairwise disjoint open intervals of $[0, 1]$. Let $\mathcal{O} \cap \text{Img} = g(\bigcup_{i \geq 1} (a_i, b_i))$. Using the assumption, the fact that $\mathcal{H}^1 = \gamma$ (Remark 4), and the property $\text{Card}(g^{-1}(\{y\})) = 1$ for a.e. y with respect to the trace of \mathcal{H}^1 on Img (Remark 5), we may write

$$\begin{aligned} \mu \circ g^{-1}(\mathcal{O} \cap \text{Img}) &\geq c \frac{\mathcal{H}^1(\mathcal{O} \cap \text{Img})}{\mathcal{L}(g)} \\ &= \frac{c}{\mathcal{L}(g)} \frac{1}{c_d} \int_{S^{d-1}} \int_0^\infty \text{Card}(\mathcal{O} \cap \text{Img} \cap D_{\theta,r}) dr d\theta \\ &= \frac{c}{\mathcal{L}(g)} \frac{1}{c_d} \int_{S^{d-1}} \int_0^\infty \sum_{y \in \mathcal{O} \cap \text{Img} \cap D_{\theta,r}} 1 dr d\theta \\ &= \frac{c}{\mathcal{L}(g)} \frac{1}{c_d} \int_{S^{d-1}} \int_0^\infty \sum_{y \in \mathcal{O} \cap \text{Img} \cap D_{\theta,r}} \text{Card}(g^{-1}(\{y\})) dr d\theta \\ &= \frac{c}{\mathcal{L}(g)} \frac{1}{c_d} \int_{S^{d-1}} \int_0^\infty \sum_{t \in [0,1]} \mathbf{1}_{\bigcup_{i \geq 1} (a_i, b_i)}(t) \mathbf{1}_{\{g(t) \in D_{\theta,r}\}} dr d\theta \end{aligned}$$

Thanks to the equality (4.1), and using $|g'(t)| = \mathcal{L}(g)$ a.e., we deduce that

$$\begin{aligned} \mu \circ g^{-1}(\mathcal{O} \cap \text{Img}) &\geq \frac{c}{\mathcal{L}(g)} \int_0^1 \mathbf{1}_{\bigcup_{i \geq 1} (a_i, b_i)}(t) |g'(t)| dt \\ &= c \sum_{i \geq 1} (b_i - a_i) \\ &= c \lambda \left(\bigcup_{i \geq 1} (a_i, b_i) \right). \end{aligned}$$

Let us show that $\{t \in [0, 1], \text{Card}(g^{-1}(\{g(t)\})) > 1\}$ is negligible for λ .

Let $A \subset \text{Img}$ be a negligible set for the trace of \mathcal{H}^1 on Img . Then, $g^{-1}(A)$ is negligible for λ . Indeed, there exists a Borel set N , such that $A \subset N$ and $\mathcal{H}^1(N) = 0$. Since $g^{-1}(A) \subset g^{-1}(N)$, $\lambda(g^{-1}(A)) \leq \lambda(g^{-1}(N))$. By Remark 6, $\lambda(g^{-1}(N)) = \frac{\mathcal{H}^1(N)}{\mathcal{L}(g)}$. Thus, $\lambda(g^{-1}(A)) = 0$.

Hence, the fact that $\{y \in \text{Img}, \text{Card}(g^{-1}(\{y\})) > 1\}$ is a negligible set for the trace of \mathcal{H}^1 on Img (Remark 5) implies that $\{t \in [0, 1], \text{Card}(g^{-1}(\{g(t)\})) > 1\}$ is negligible for λ .

Consequently,

$$\mu \left(\bigcup_{i \geq 1} (a_i, b_i) \right) = \mu \circ g^{-1}(\mathcal{O} \cap \text{Img}),$$

and, thus,

$$\mu \left(\bigcup_{i \geq 1} (a_i, b_i) \right) \geq c \lambda \left(\bigcup_{i \geq 1} (a_i, b_i) \right).$$

This extends to every Borel subset of $[0, 1]$ by outer regularity. Hence, $\mu \geq c\lambda$. \square

Now, equipped with Lemma 4.4, let us turn to the proof of Theorem 3.1 itself.

5 Proof of Theorem 3.1

Note that $\mathcal{L}(g) - a_n < a_n \left\lfloor \frac{\mathcal{L}(g)}{a_n} \right\rfloor \leq \mathcal{L}(g)$. We set, for every $n \geq 1$, $f_n^* := \hat{f}_{n, a_n \lfloor \frac{\mathcal{L}(g)}{a_n} \rfloor}$.

5.1 Step 1

We will first prove that $L^2\mathcal{D} \left(\frac{1}{n} \sum_{i=1}^n \delta_{T(f_n^*, X_i^n)}, \mathcal{M}_c \right) + \Delta_n(f_n^*)$ converges in probability to 0 as n goes to infinity.

To begin with, let us consider the term $\Delta_n(f_n^*)$.

Lemma 5.1. $\Delta_n(f_n^*)$ converges in probability to 0 as n tends to infinity.

Proof of Lemma 5.1. We write

$$\begin{aligned} \Delta_n(f_n^*) &\leq \Delta_n(g) \\ &= \frac{1}{n} \sum_{i=1}^n V(d(X_i^n, \text{Img})) = \frac{1}{n} \sum_{i=1}^n V(d(g(U_i^n) + \varepsilon_i^n, \text{Img})) \\ &\leq \frac{C}{n} \sum_{i=1}^n V(d(g(U_i^n), \text{Img})) + \frac{C}{n} \sum_{i=1}^n V(|\varepsilon_i^n|) = \frac{C}{n} \sum_{i=1}^n V(|\varepsilon_i^n|), \end{aligned}$$

which converges in probability to 0. \square

For $n \geq 1$, let $\mathcal{D}_n^*(\mathcal{M}_c) = \mathcal{D} \left(\frac{1}{n} \sum_{i=1}^n \delta_{T(f_n^*, X_i^n)}, \mathcal{M}_c \right)$. The remaining part of Step 1 is dedicated to the convergence of $(\mathcal{D}_n^*(\mathcal{M}_c))_{n \geq 1}$.

Proposition 5.1. If $\mathcal{L}(g) > 0$, the sequence $(\mathcal{D}_n^*(\mathcal{M}_c))_{n \geq 1}$ converges in probability to 0 as n tends to infinity.

Proof of Proposition 5.1

We will show that the sequence $(\mathcal{D}_n^*(\mathcal{M}_c))_{n \geq 1}$ is tight and that every limit point for the convergence in distribution is δ_0 . We set $\nu_n^* = \frac{1}{n} \sum_{i=1}^n \delta_{(T(f_n^*, X_i^n), X_i^n)}$ and $\nu_n^{*'} = \frac{1}{n} \sum_{i=1}^n \delta_{(T(f_n^*, X_i^n), g(U_i^n))}$.

Lemma 5.2. The difference $|\nu_n^* - \nu_n^{*'}|_{BL}$ converges in probability to 0 as n tends to infinity.

Proof of Lemma 5.2. We have

$$\begin{aligned} &|\nu_n^* - \nu_n^{*'}|_{BL} \\ &= \sup \left\{ |\nu_n^*(h) - \nu_n^{*'}(h)| : |h|_\infty \leq 1, \sup_{x \neq y} \frac{|h(x) - h(y)|}{|x - y|} \leq 1 \right\} \\ &\leq \sup \left\{ \frac{1}{n} \sum_{i=1}^n |h(T(f_n^*, X_i^n), X_i^n) - h(T(f_n^*, X_i^n), g(U_i^n))| : |h|_\infty \leq 1, \sup_{x \neq y} \frac{|h(x) - h(y)|}{|x - y|} \leq 1 \right\} \\ &\leq \frac{1}{n} \sum_{i=1}^n |\varepsilon_i^n| \\ &\leq \left(\frac{1}{n} \sum_{i=1}^n |\varepsilon_i^n|^2 \right)^{1/2}, \end{aligned} \tag{5.1}$$

which converges in probability to 0 as n tends to infinity. \square

Lemma 5.3. *The sequence $(f_n^*, \nu_n^{*'})_{n \geq 1}$ is tight.*

Proof of Lemma 5.3. Observe that, for every $n \geq 1$, $\nu_n^{*'}$ is supported in a compact set $[0, 1] \times \text{Img}$, and thus, belongs to a compact set of measures for the topology of weak convergence, so that the sequence $(\nu_n^{*'})_{n \geq 1}$ is tight.

Besides, for every $n \geq 1$, f_n^* is $\mathcal{L}(g)$ -Lipschitz. Thus, to obtain tightness of the sequence $(f_n^*)_{n \geq 1}$ in $C([0, 1])$, it suffices to show that the sequence $(f_n^*(0))_{n \geq 1}$ is tight (see, e.g., (Prokhorov, 1956, Lemma 2.1)). If it is not the case, there exists $\eta > 0$ such that, for every $A > 0$, there exists $n \geq 1$ such that $P(|f_n^*(0)| > A) \geq \eta > 0$, and hence, as the length $\mathcal{L}(g)$ is finite,

$$P(d(0, \text{Im} f_n^*) > A - \mathcal{L}(g)) \geq \eta > 0. \quad (5.2)$$

Observe that

$$\Delta_n(f_n^*) \leq \Delta_n(0) = \frac{1}{n} \sum_{i=1}^n V(|X_i^n|).$$

Hence, since the sequence $(\frac{1}{n} \sum_{i=1}^n V(|X_i^n|))_{n \geq 1}$ is tight, for every $\varepsilon > 0$, there exists M_ε such that for all $n \geq 1$,

$$P(\Delta_n(f_n^*) > M_\varepsilon) \leq P\left(\frac{1}{n} \sum_{i=1}^n V(|X_i^n|) > M_\varepsilon\right) \leq \varepsilon. \quad (5.3)$$

Moreover,

$$\begin{aligned} P(\Delta_n(f_n^*) > M_\varepsilon) &= P\left(\frac{1}{n} \sum_{i=1}^n V(d(X_i^n, \text{Im} f_n^*)) > M_\varepsilon\right) \\ &\geq P\left(\min_{t \in [0, 1]} \frac{V(|f_n^*(t)|)}{C} - \frac{1}{n} \sum_{i=1}^n V(|X_i^n|) > M_\varepsilon\right) \\ &\geq P(V(d(0, \text{Im} f_n^*)) > 2CM_\varepsilon) - P\left(\frac{1}{n} \sum_{i=1}^n V(|X_i^n|) > M_\varepsilon\right) \\ &\geq P(V(d(0, \text{Im} f_n^*)) > 2CM_\varepsilon) - \varepsilon, \end{aligned}$$

according to (5.3). Thus, thanks to (5.2), there exists n_ε such that

$$\eta - \varepsilon \leq P(\Delta_{n_\varepsilon}(f_{n_\varepsilon}^*) > M_\varepsilon) \leq \varepsilon,$$

which leads to a contradiction since ε is arbitrary. Consequently, $(f_n^*(0))_{n \geq 1}$ is tight. \square

From Lemma 5.3, the sequence $(\mathcal{D}_n^*(\mathcal{M}_c))_{n \geq 1}$ is tight, and thus, by Prohorov's theorem, $(\mathcal{D}_n^*(\mathcal{M}_c))_{n \geq 1}$ has a limit point for weak convergence. So, there exists a weakly convergent subsequence. Let us consider an arbitrary such convergent subsequence $(\mathcal{D}_{\sigma(n)}^*(\mathcal{M}_c))_{n \geq 1}$, where $\sigma : \mathbb{N} \rightarrow \mathbb{N}$ denotes a strictly increasing function, and show that the corresponding limit point is δ_0 . The sequence $(f_n^*, \nu_n^*)_{n \geq 1}$ is tight, by Lemma 5.2 and Lemma 5.3. So, by Prohorov's theorem, we may assume, up to extracting a further subsequence, that the sequence $(f_{\sigma(n)}^*, \nu_{\sigma(n)}^*)_{n \geq 1}$ converges in distribution to a tuple (φ^*, ν^*) , where φ^* is a random function, and ν^* a random probability measure on $[0, 1] \times \mathbb{R}^d$. According to Skorokhod's representation theorem, there exist a random sequence $(\tilde{f}_{\sigma(n)}^*, \tilde{\nu}_{\sigma(n)}^*)_{n \geq 1}$ and a tuple $(\tilde{\varphi}^*, \tilde{\nu}^*)$ with the same distribution as $(f_{\sigma(n)}^*, \nu_{\sigma(n)}^*)_{n \geq 1}$ and (φ^*, ν^*) respectively,

such that $(\tilde{f}_{\sigma(n)}^*, \tilde{\nu}_{\sigma(n)}^*)_{n \geq 1}$ converges almost surely to $(\tilde{\varphi}^*, \tilde{\nu}^*)$. Moreover, up to considering an extension of the probability space where this representation holds, there exist random vectors \tilde{X}_i^n , $i = 1, \dots, n$, such that $((\tilde{X}_i^{\sigma(n)})_{1 \leq i \leq \sigma(n)}, \tilde{f}_{\sigma(n)}^*)$ has the same joint distribution as $((X_i^{\sigma(n)})_{1 \leq i \leq \sigma(n)}, f_{\sigma(n)}^*)$. In the sequel, to lighten notation, the tilde will be omitted. The marginal distributions of ν^* will be denoted by $\nu^{*,i}$, $i = 1, 2$.

Lemma 5.4. *Assume that $\mathcal{L}(g) > 0$. Then, we have $\nu^{*,2} \in \mathcal{M}_c^g$ a.s.*

Proof of Lemma 5.4. First, recall that $\nu_{\sigma(n)}^*$ converges almost surely to ν^* . This implies that $\frac{1}{\sigma(n)} \sum_{i=1}^{\sigma(n)} \delta_{g(U_i^{\sigma(n)})}$ converges almost surely to $\nu^{*,2}$.

For $n \geq 1$, the measure $\frac{1}{n} \sum_{i=1}^n \mu_i^n \circ g^{-1}$ is supported in the compact set Img , so that $(\frac{1}{n} \sum_{i=1}^n \mu_i^n \circ g^{-1})_{n \geq 1}$ is tight. Thus, up to a further extraction, we may assume that $\frac{1}{\sigma(n)} \sum_{i=1}^{\sigma(n)} \mu_i^{\sigma(n)} \circ g^{-1}$ converges weakly to a measure m .

Besides, for every continuous and bounded function h , by the weak law of large numbers for triangular arrays, we have $\frac{1}{n} \sum_{i=1}^n h(g(U_i^n)) - \frac{1}{n} \sum_{i=1}^n E[h(g(U_i^n))] \rightarrow 0$ in probability almost surely. Consequently, for every $x \in \mathbb{R}^d$, $\int \exp(i\langle u, x \rangle) d\nu^{*,2}(u) = \int \exp(i\langle u, x \rangle) dm(u)$ almost surely. Hence, almost surely, for every $x \in \mathbb{Q}^d$,

$$\int \exp(i\langle u, x \rangle) d\nu^{*,2}(u) = \int \exp(i\langle u, x \rangle) dm(u).$$

We obtain that $\nu^{2,*} = m$ a.s..

By Lemma 4.4, for $i = 1, \dots, n$, $\mu_i^n \circ g^{-1} \in \mathcal{M}_c^g$. As \mathcal{M}_c^g is convex, for every $n \geq 1$, $\frac{1}{n} \sum_{i=1}^n \mu_i^n \circ g^{-1} \in \mathcal{M}_c^g$. Since \mathcal{M}_c^g is closed, $\nu^{2,*} \in \mathcal{M}_c^g$ a.s.. □

Lemma 5.5. *Denoting by (T, Z) the identity on $[0, 1] \times \mathbb{R}^d$, we have*

$$\nu^*(|Z - \varphi^*(T)| = \min_{t \in [0,1]} |Z - \varphi^*(t)|) = 1 \text{ a.s.}$$

Proof of Lemma 5.5. For a curve f , define, for $\varepsilon > 0$, the open set

$$\mathcal{O}_f^\varepsilon = \{|Z - f(T)| > \min_{t \in [0,1]} |Z - f(t)| + \varepsilon\}.$$

By the Portmanteau lemma, $\nu^*(\mathcal{O}_{\varphi^*}^\varepsilon) \leq \liminf \nu_{\sigma(n)}^*(\mathcal{O}_{\varphi^*}^\varepsilon)$ a.s., and by definition of $f_{\sigma(n)}^*$, for every $\delta > 0$, $\nu_{\sigma(n)}^*(\mathcal{O}_{f_{\sigma(n)}^*}^\delta) = 0$. Moreover, thanks to the convergence of $f_{\sigma(n)}^*$ to φ^* , $\mathcal{O}_{\varphi^*}^\varepsilon \subset \mathcal{O}_{f_{\sigma(n)}^*}^{\varepsilon/2}$ a.s. as soon as n is large enough. Thus, $\nu^*(\mathcal{O}_{\varphi^*}^\varepsilon) = 0$ a.s., and, letting ε tend to 0, we get

$$\nu^*(|Z - \varphi^*(T)| > \min_{t \in [0,1]} |Z - \varphi^*(t)|) = 0 \text{ a.s.,}$$

that is

$$\nu^*(|Z - \varphi^*(T)| = \min_{t \in [0,1]} |Z - \varphi^*(t)|) = 1 \text{ a.s.,}$$

as desired. □

Lemma 5.6. *We have $\int |z - \varphi^*(t)|^2 d\nu^*(t, z) = 0$ almost surely.*

Proof of Lemma 5.6. First, recall that $\Delta_n(f_n^*)$ converges in probability to 0 as n tends to infinity by Lemma 5.1. Let

$$\Delta'_n(f) = \frac{1}{n} \sum_{i=1}^n V(d(g(U_i^n), \text{Im} f)).$$

We have

$$\begin{aligned} \Delta'_n(f_n^*) &\leq \frac{C}{n} \sum_{i=1}^n V(d(X_i^n, \text{Im} f_n^*)) + \frac{C}{n} \sum_{i=1}^n V(|\varepsilon_i^n|) \\ &= C\Delta_n(f_n^*) + \frac{C}{n} \sum_{i=1}^n V(|\varepsilon_i^n|). \end{aligned}$$

Thus, $\Delta'_n(f_n^*)$ converges in probability to 0, thanks to the convergence of $\Delta_n(f_n^*)$ and the assumption on the ε_i^n . Moreover,

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^n V(|g(U_i^n) - \varphi^*(T(f_n^*, X_i^n))|) \\ &= \frac{1}{n} \sum_{i=1}^n V(|g(U_i^n) - f_n^*(T(f_n^*, X_i^n)) + f_n^*(T(f_n^*, X_i^n)) - \varphi^*(T(f_n^*, X_i^n))|) \\ &\leq \frac{C}{n} \sum_{i=1}^n V(d(g(U_i^n), \text{Im} f_n^*)) + \frac{C}{n} \sum_{i=1}^n V(|f_n^*(T(f_n^*, X_i^n)) - \varphi^*(T(f_n^*, X_i^n))|), \end{aligned}$$

so that $\frac{1}{\sigma(n)} \sum_{i=1}^{\sigma(n)} V(|g(U_i^{\sigma(n)}) - \varphi^*(T(f_{\sigma(n)}^*, X_i^{\sigma(n)}))|)$ converges in probability to 0, using the convergence of $\Delta'_n(f_n^*)$, the uniform convergence of $f_{\sigma(n)}^*$ to φ^* , and the fact that V is continuous, with $V(0) = 0$. The function ψ defined by $\psi(z, t) = |z - \varphi^*(t)|$ for $(z, t) \in \text{Im} g \times [0, 1]$ is lower semi-continuous. This implies that, almost surely,

$$\liminf_{n \rightarrow \infty} \frac{1}{\sigma(n)} \sum_{i=1}^{\sigma(n)} V(|g(U_i^{\sigma(n)}) - \varphi^*(T(f_{\sigma(n)}^*, X_i^{\sigma(n)}))|) \geq \int V(|z - \varphi^*(t)|) d\nu^*(t, z).$$

Consequently, $\int V(|z - \varphi^*(t)|) d\nu^*(t, z) = 0$ a.s. □

Note that $\mathcal{L}(\varphi^*) \leq \mathcal{L}(g)$ a.s., by the lower semi-continuity property of the length (see, e.g., Alexandrov and Reshetnyak (1989, Theorem 2.1.2)). Together with Lemma 5.6, this property allows to show the next result.

Lemma 5.7. *Assume $\mathcal{L}(g) > 0$. We have $\text{Im} g = \text{Im} \varphi^*$ a.s. and $\mathcal{L}(\varphi^*) = \mathcal{L}(g)$ a.s..*

Proof of Lemma 5.7. Since, by Lemma 5.6, $\int |z - \varphi^*(t)|^2 d\nu^*(t, z) = 0$ almost surely, $\nu^*(Z = \varphi^*(T)) = 1$ a.s., that is, by Lemma 5.5, $\nu^*(d(Z, \text{Im} \varphi^*) = 0) = 1$ a.s.. As $\nu^{*,2} \in \mathcal{M}_c^g$, $d(g(u), \text{Im} \varphi^*) = 0$ du -a.e., a.s.. By continuity, $d(g(t), \text{Im} \varphi^*) = 0$ for every $t \in [0, 1]$, a.s.. Thus, $\text{Im} g \subset \text{Im} \varphi^*$ a.s.. As $\mathcal{L}(\varphi^*) \leq \mathcal{L}(g)$ a.s., using the assumption $\mathcal{L}(g) = \mathcal{H}^1(\text{Im} g)$, we have in fact equality $\text{Im} g = \text{Im} \varphi^*$ a.s., and $\mathcal{L}(\varphi^*) = \mathcal{L}(g)$ a.s.. □

In particular, $\mathcal{L}(\varphi^*) = \mathcal{H}^1(\text{Im} \varphi^*)$ a.s.. Moreover, since f_n^* is $\mathcal{L}(g)$ -Lipschitz for every $n \geq 1$, passing to the limit for $f_{\sigma(n)}^*$, we obtain that φ^* is $\mathcal{L}(g)$ -Lipschitz, that is $\mathcal{L}(\varphi^*)$ -Lipschitz, a.s.. We deduce that $|\varphi^{*'}(t)| = \mathcal{L}(\varphi^*)$ dt -a.e., a.s.. So, φ^* satisfies the assumptions of Lemma 4.4.

Lemma 5.8. *Assume that $\mathcal{L}(g) > 0$. We have*

$$\mathcal{D}_{\sigma(n)}^*(\mathcal{M}_c) \rightarrow 0,$$

in probability.

Proof of Lemma 5.8. The distribution $\nu^{*,1} \circ (\varphi^*)^{-1}$ is $\nu^{*,2}$ a.s.. Since $\nu^{*,2}$ belongs to \mathcal{M}_c^g a.s., which is \mathcal{M}_c^φ a.s., Lemma 4.4 shows that $\nu^{*,1} \in \mathcal{M}_c$ a.s..

Hence,

$$\mathcal{D} \left(\frac{1}{\sigma(n)} \sum_{i=1}^{\sigma(n)} \delta_{T(f_{\sigma(n)}^*, X_i^{\sigma(n)})}, \mathcal{M}_c \right)$$

tends to 0 in probability as n tends to infinity. \square

Finally, $(\mathcal{D}_n^*(\mathcal{M}_c))_{n \geq 1}$ converges in probability to 0, which proves Proposition 5.1.

5.2 Step 2

Let $\hat{f}_n := \hat{f}_{n, \hat{L}_n}$. In the sequel, we will consider extractions of the sequence $(\hat{f}_n)_{n \geq 1}$ converging in distribution, and show that for every limit point φ of $(\hat{f}_n)_{n \geq 1}$, $d_H(\text{Im}\varphi, \text{Im}g) = 0$ a.s..

By definition of \hat{L}_n , we have

$$\hat{L}_n^2 \mathcal{D} \left(\frac{1}{n} \sum_{i=1}^n \delta_{T(\hat{f}_n, X_i^n)}, \mathcal{M}_c \right) + \Delta_n(\hat{f}_n) \leq \mathcal{L}^2(f_n^*) \mathcal{D} \left(\frac{1}{n} \sum_{i=1}^n \delta_{T(f_n^*, X_i^n)}, \mathcal{M}_c \right) + \Delta_n(f_n^*).$$

By Lemma 5.1 and Proposition 5.1, the right-hand term tends to 0, which implies the two following results.

Lemma 5.9. $\Delta_n(\hat{f}_n)$ converges in probability to 0 as n tends to infinity.

Lemma 5.10. We have

$$\hat{L}_n^2 \mathcal{D} \left(\frac{1}{n} \sum_{i=1}^n \delta_{T(\hat{f}_n, X_i^n)}, \mathcal{M}_c \right) \rightarrow 0,$$

in probability.

We set $\nu_n = \frac{1}{n} \sum_{i=1}^n \delta_{(T(\hat{f}_n, X_i^n), X_i^n)}$ and $\nu'_n = \frac{1}{n} \sum_{i=1}^n \delta_{(T(\hat{f}_n, X_i^n), g(U_i^n))}$.

The same arguments as in the proof of Lemma 5.2 and Lemma 5.3 lead to the following statements for ν_n , ν'_n , and \hat{f}_n .

Lemma 5.11. *The difference $|\nu_n - \nu'_n|_{BL}$ converges in probability to 0 as n tends to infinity.*

Lemma 5.12. *The sequence $(\hat{f}_n, \nu'_n)_{n \geq 1}$ is tight.*

To obtain this result, it suffices to notice that \hat{f}_n is \hat{L}_n -Lipschitz for every $n \geq 1$, and, thus, Λ -Lipschitz, and to show that the sequence $(\hat{f}_n(0))_{n \geq 1}$ is tight, by replacing $\mathcal{L}(g)$ by Λ in the proof of Lemma 5.3.

By Lemma 5.12, the sequence $(d_H(\text{Im}\hat{f}_n, \text{Img}))_{n \geq 1}$ is tight. Considering an arbitrary limit point for weak convergence, we let $\kappa : \mathbb{N} \rightarrow \mathbb{N}$ denote a strictly increasing function such that the subsequence $(d_H(\text{Im}\hat{f}_{\kappa(n)}, \text{Img}))_{n \geq 1}$ is weakly convergent. We will show that the considered limit point is δ_0 . By Lemma 5.11 and Lemma 5.12, the sequence $(\hat{f}_n, \nu_n)_{n \geq 1}$ is tight. Thus, by Prohorov's theorem, up to an extraction, we may assume that $(\hat{f}_{\kappa(n)}, \nu_{\kappa(n)})_{n \geq 1}$ converges in distribution to a tuple (φ, ν) , where φ is a random function and ν a random probability measure on $[0, 1] \times \mathbb{R}^d$. Thanks to Skorokhod's representation theorem, there exist a random sequence $(\tilde{f}_{\kappa(n)}, \tilde{\nu}_{\kappa(n)})_{n \geq 1}$ and a tuple $(\tilde{\varphi}, \tilde{\nu})$ with the same distribution as $(\hat{f}_{\kappa(n)}, \nu_{\kappa(n)})_{n \geq 1}$ and (φ, ν) respectively, such that $(\tilde{f}_{\kappa(n)}, \tilde{\nu}_{\kappa(n)})_{n \geq 1}$ converges almost surely to $(\tilde{\varphi}, \tilde{\nu})$. Again, the tilde will be omitted to lighten notation, and the marginal distributions of ν will be denoted by ν^i , $i = 1, 2$. As above, $\nu^2 \in \mathcal{M}_c^g$. Moreover, we have the next result, similar to Lemma 5.5.

Lemma 5.13. *Denoting by (T, Z) the identity on $[0, 1] \times \mathbb{R}^d$, we have*

$$\nu(|Z - \varphi(T)| = \min_{t \in [0, 1]} |Z - \varphi(t)|) = 1 \text{ a.s.} \quad (5.4)$$

The next lemma is obtained by mimicking the proof of Lemma 5.6.

Lemma 5.14. *We have $\int |z - \varphi(t)|^2 d\nu(t, z) = 0$ almost surely.*

Lemma 5.15. *We have $\text{Img} = \text{Im}\varphi$ almost surely.*

Proof of Lemma 5.15. Since, by Lemma 5.14, $\nu(Z = \varphi(T)) = 1$ almost surely, using Lemma 5.4 and the fact that $\nu^2 \in \mathcal{M}_c^g$, we get, as in the proof of Lemma 5.7, $\text{Img} \subset \text{Im}\varphi$ a.s.. On the event $\{\mathcal{L}(\varphi) = 0\}$, we get $\text{Img} = \text{Im}\varphi$ a.s.. From now on, we consider the event $\{\mathcal{L}(\varphi) > 0 \cap \text{Img} \neq \text{Im}\varphi\}$. It suffices to prove that it has probability zero. Yet, on this event, $\liminf \hat{L}_n \geq \mathcal{L}(\varphi) > 0$ a.s., therefore Lemma 5.10 yields the convergence in probability

$$\mathcal{D} \left(\frac{1}{n} \sum_{i=1}^n \delta_{T(\hat{f}_n, X_i^n)}, \mathcal{M}_c \right) \rightarrow 0,$$

which implies $\nu^1 \in \mathcal{M}_c$ a.s.. Moreover, there exists $t \in [0, 1]$ such that $\varphi(t) \notin \text{Img}$, that is $d(\varphi(t), \text{Img}) > 0$. By continuity, $d(\varphi(s), \text{Img}) > 0$ for every s in a non-empty open interval (a, b) . We get $\nu(\varphi(T) \notin \text{Img}) \geq \nu(T \in (a, b)) \geq c(b - a)$ since $\nu^1 \in \mathcal{M}_c$ a.s.. Besides, $\nu(Z = \varphi(T)) = 1$ and $\nu(Z \in \text{Img}) = 1$ a.s., leading to a contradiction whenever the event $\{\mathcal{L}(\varphi) > 0 \cap \text{Img} \neq \text{Im}\varphi\}$ has positive probability. So, $\text{Img} = \text{Im}\varphi$ a.s.. \square

Hence, $d_H(\text{Im}\varphi, \text{Img}) = 0$ a.s., as desired. In other words, every limit point of the sequence $(d_H(\text{Im}\hat{f}_n, \text{Img}))_{n \geq 1}$ is δ_0 , and, thus, Theorem 3.1 is proved.

A Appendix

A.1 Proof of Lemma 1.1

First, we show that Δ is lower semi-continuous. Let $(f_n)_{n \geq 1}$ be a sequence of curves converging uniformly to a curve f . Then, for every $x \in \mathbb{R}^d$, $d(x, \text{Im}f_n)$ converges to $d(x, \text{Im}f)$, since $|d(x, \text{Im}f_n) - d(x, \text{Im}f)| \leq \sup_t |f(t) - f_n(t)|$. Then, by lower semi-continuity of V , for every $x \in \mathbb{R}^d$,

$$\liminf_{n \rightarrow \infty} V(d(x, \text{Im}f_n)) \geq V(d(x, \text{Im}f)). \quad (\text{A.1})$$

By Fatou's Lemma,

$$\liminf_{n \rightarrow \infty} \Delta(f_n) = \liminf_{n \rightarrow \infty} E[V(d(x, \text{Im}f_n))] \geq E \left[\liminf_{n \rightarrow \infty} V(d(x, \text{Im}f_n)) \right].$$

Using (A.1), we obtain

$$\liminf_{n \rightarrow \infty} \Delta(f_n) \geq E[V(d(x, \text{Im}f))] = \Delta(f).$$

So, Δ is lower semi-continuous.

Now, we prove that a minimizing sequence of Δ is relatively compact. Let $(f_n)_{n \geq 1}$ denote a sequence of curves with length at most L and constant speed, which is a minimizing sequence, that is

$$\lim_{n \rightarrow \infty} \Delta(f_n) = \inf_{f, \mathcal{L}(f) \leq L} \Delta(f).$$

For every $n \geq 1$, f_n is L -Lipschitz. Thus, the sequence is equi-uniformly continuous. Let us show that $(f_n(0))_{n \geq 1}$ is bounded. We may write, for every $n \geq 1$, $t \in [0, 1]$, $|f_n(t)| \geq |f_n(0)| - Lt \geq |f_n(0)| - L$. Here, the length L is finite. Thus, if there exists a strictly increasing function $\kappa : \mathbb{N} \rightarrow \mathbb{N}$ such that $|f_{\kappa(n)}(0)| \rightarrow \infty$, one has $\Delta(f_{\kappa(n)}) \rightarrow \sup V$, which is impossible since

$$\inf_{f, \mathcal{L}(f) \leq L} \Delta(f) \leq E[V(|X|)] < \sup V.$$

So, the sequence $(f_k(0))_{k \geq 1}$ is bounded.

A.2 Proof of Lemma 4.3

1. Let us first assume that A is compact. For $k \geq 1$, we set

$$N_{\theta, r}^k(A) = \inf_{\mathcal{C}} \sum_{S \in \mathcal{C}} \mathbf{1}_{\{D_{\theta, r} \cap S = \emptyset\}},$$

where the infimum is taken over all finite coverings \mathcal{C} of the compact set A with open sets S such that $\text{diam}(S) \leq \frac{1}{k}$. For every open set S , the function $(\theta, r) \mapsto \mathbf{1}_{\{D_{\theta, r} \cap S = \emptyset\}}$ is upper semi-continuous since $\{(\theta, r), D_{\theta, r} \cap S \neq \emptyset\}$ is open. Thus, $(\theta, r) \mapsto N_{\theta, r}^k(A)$ is also upper semi-continuous. Moreover,

$$N_{\theta, r}(A) = \lim_{k \rightarrow \infty} \uparrow N_{\theta, r}^k(A).$$

Consequently, $(\theta, r) \mapsto N_{\theta, r}(A)$ is measurable.

2. According to the Cauchy formula,

$$\begin{aligned} \mathcal{L}(f) &= \frac{1}{c_d} \int_{S^{d-1}} \int_0^\infty \text{Card}(\{t \in [0, 1], f(t) \in D_{\theta, r}\}) dr d\theta \\ &= \frac{1}{c_d} \int_{S^{d-1}} \int_0^\infty \sum_{y \in \text{Im}f \cap D_{\theta, r}} \text{Card}(f^{-1}(\{y\})) dr d\theta. \end{aligned}$$

Since $\mathcal{L}(f) < \infty$, the set $B = \{(\theta, r), N_{\theta, r}(\text{Im}f) = \infty\}$ is $d\theta dr$ negligible. Moreover, $\mathcal{M} = \{A \in \mathcal{B}(\mathbb{R}^d) \cap \text{Im}f, (\theta, r) \mapsto N_{\theta, r}(A) \mathbf{1}_{B^c}(\theta, r) \text{ is measurable}\}$ is a monotone class. Hence, $\mathcal{M} = \mathcal{B}(\mathbb{R}^d) \cap \text{Im}f$. Thus, $(\theta, r) \mapsto N_{\theta, r}(A)$ is measurable for the Lebesgue sigma-algebra for every $A \subset \mathcal{B}(\mathbb{R}^d) \cap \text{Im}f$.

References

- E. Aamari and C. Levrard. Stability and minimax optimality of tangential Delaunay complexes for manifold reconstruction. *Discrete & Computational Geometry*, 59(4): 923–971, 2018.
- E. Aamari and C. Levrard. Nonasymptotic rates for manifold, tangent space and curvature estimation. *The Annals of Statistics*, 47(1):177–204, 2019.
- A. D. Alexandrov and Y. G. Reshetnyak. *General Theory of Irregular Curves*. Mathematics and its Applications. Kluwer Academic Publishers, Dordrecht, 1989.
- S. Ayari and S. Dubuc. La formule de Cauchy sur la longueur d’une courbe. *Canadian Mathematical Bulletin*, 40:3–9, 1997.
- S. Banach. Sur les lignes rectifiables et les surfaces dont l’aire est finie. *Fundamenta Mathematicae*, 7:225–237, 1925.
- G. Buttazzo and E. Stepanov. Optimal transportation networks as free Dirichlet regions for the Monge-Kantorovich problem. *Ann. Sc. Norm. Sup. Pisa Cl. Sci.*, II(4):631–678, 2003.
- G. Buttazzo, E. Oudet, and E. Stepanov. Optimal transportation problems with free Dirichlet regions. *Progress in Nonlinear Diff. Equations and their Applications*, 51: 41–65, 2002.
- A. Cauchy. Mémoire sur la rectification des courbes et la quadrature des surfaces courbes. 1850.
- M. W. Crofton. On the theory of local probability, applied to straight lines at random in a plane. *Philosophical Transactions of the Royal Society*, 158:181–199, 1868.
- T. Duchamp and W. Stuetzle. Extremal properties of principal curves in the plane. *The Annals of Statistics*, 24:1511–1520, 1996a.
- T. Duchamp and W. Stuetzle. Geometric properties of principal curves in the plane. In H. Rieder, editor, *Robust Statistics, Data Analysis, and Computer Intensive Methods: in Honor of Peter Huber’s 60th Birthday*, volume 109 of *Lecture Notes in Statistics*, pages 135–152. Springer-Verlag, New York, 1996b.
- J. Falconer, K. *The Geometry of Fractal Sets*. Cambridge tracts in Mathematics. Cambridge University Press, 1985.
- H. Federer. *Geometric measure theory*, volume 153 of *Die Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, 1969.
- C. R. Genovese, M. Perone-Pacífico, I. Verdinelli, and L. Wasserman. The geometry of nonparametric filament estimation. *Journal of the American Statistical Association*, 107:788–799, 2012a.
- C. R. Genovese, M. Perone-Pacífico, I. Verdinelli, and L. Wasserman. Minimax manifold estimation. *Journal of Machine Learning Research*, 13:1263–1291, 2012b.

- C. R. Genovese, M. Perone-Pacifico, I. Verdinelli, and L. Wasserman. Manifold estimation and singular deconvolution under Hausdorff loss. *The Annals of Statistics*, 40(2):941–963, 2012c.
- T. Hastie and W. Stuetzle. Principal curves. *Journal of the American Statistical Association*, 84:502–516, 1989.
- B. Kégl, A. Krzyzak, T. Linder, and K. Zeger. Learning and design of principal curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:281–297, 2000.
- Y. Prokhorov. Convergence of random processes and limit theorems in probability theory. *Theory of Probability & Its Applications*, 1:157–214, 1956.
- R. Tibshirani. Principal curves revisited. *Statistics and Computing*, 2:183–190, 1992.