



On principal curves with a length constraint

Sylvain Delattre, Aurélie Fischer

► To cite this version:

Sylvain Delattre, Aurélie Fischer. On principal curves with a length constraint. Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques, 2020, 56 (3), <10.1214/19-AIHP1030>. <hal-01555892v2>

HAL Id: hal-01555892

<https://hal.science/hal-01555892v2>

Submitted on 13 Oct 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.



HAL Authorization

On principal curves with a length constraint

Sylvain Delattre & Aurélie Fischer*

October 13, 2019

Abstract

In this paper, we are interested in the problem of finding a parametric curve f minimizing the quantity $\mathbb{E} [\min_{t \in [0,1]} \|X - f(t)\|^2]$, where X is a random variable, under a length constraint. This question is known in the probability and statistical learning context as length-constrained principal curves optimization, as introduced by Kégl et al. (2000), and it also corresponds to a version of the “average-distance problem” studied in the calculus of variation and shape optimization community (Buttazzo and Stepanov (2003); Buttazzo et al. (2002)).

We investigate the theoretical properties satisfied by a principal curve $f : [0,1] \rightarrow \mathbb{R}^d$ with length at most L associated to a probability distribution with second-order moment. We suppose that the probability distribution is not supported on the image of a curve with length L . Studying open as well as closed optimal curves, we show that they have finite curvature. We also derive a first order Euler-Lagrange equation. This equation is then used to show that a length-constrained principal curve in two dimension has no multiple point. Finally, some examples of optimal curves are presented.

Keywords – Principal curves, average-distance problem, quantization of probability measures, length constraint, finite curvature.

2000 Mathematics Subject Classification: Primary 60E99; Secondary 35B38, 49Q10, 49Q20.

1 Introduction

1.1 Context of the problem and motivation

We focus on the problem:

find a curve $f : [0,1] \rightarrow \mathbb{R}^d$ minimizing the quantity

$$\mathbb{E} [d(X, \text{Im} f)^2] = \int d(x, \text{Im} f)^2 d\mu(x), \quad (1)$$

over all curves with length $\mathcal{L}(f)$, such that $\mathcal{L}(f) \leq L$.

Here, $d(\cdot, \cdot)$ is the Euclidean distance from a point to a set, $\text{Im} f$ is the image of f , and X is some random vector with distribution μ , taking its values in \mathbb{R}^d . As an illustration, two examples of length-constrained principal curves, fitted via a stochastic gradient descent algorithm, are presented in Figure 1. This corresponds to principal curves with length constraint, as described by Kégl et al. (2000). These authors show that there exists indeed a minimizer whenever X is square integrable. Observe that such a length constraint makes perfectly sense in the empirical case, that is in the statistical framework, when the random vector is replaced by a data cloud. Indeed, from a practical point of view, it is essential to appropriately tune some parameter reflecting the complexity of the curve, in order to achieve a trade-off between a curve passing through all data points and a too rough one. The parameter selection issue was addressed in this statistical context for instance in Biau and Fischer (2012), Fischer (2013) and Gerber and Whitaker (2013).

*The research work of this author has been partially supported by the French National Research Agency via the TopData project ANR-13-BS01-0008

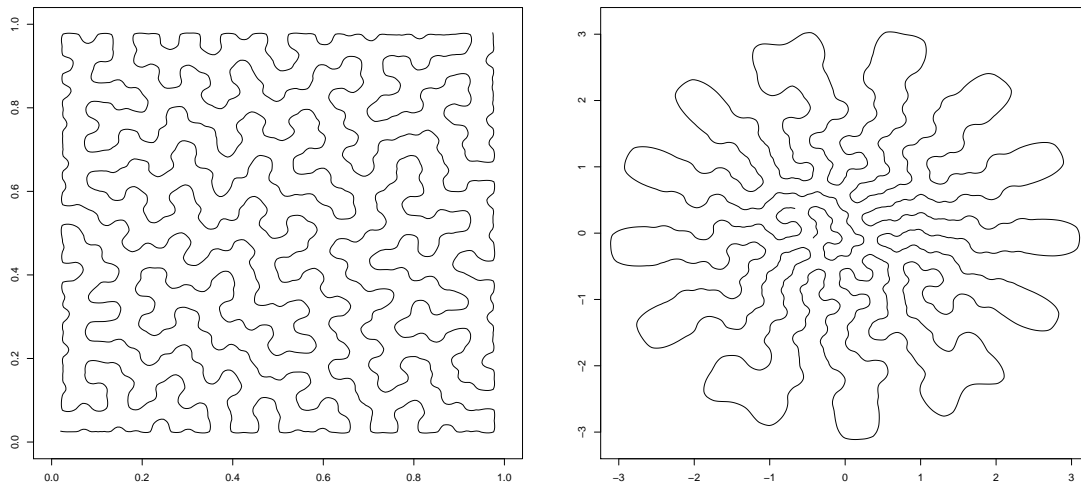


Figure 1: Two examples of principal curves with length constraint: (a) Uniform distribution over the square $[0, 1]^2$. (b) Standard Gaussian distribution.

Originally, principal curves were introduced by Hastie and Stuetzle (1989), with a different definition, based on the so-called self-consistency property. In this point of view, a curve f is said to be self-consistent for a random vector X with finite second moment if it satisfies:

$$f(t_f(X)) = \mathbb{E}[X|t_f(X)] \quad \text{a.s.},$$

where the projection index t_f is given by

$$t_f(x) = \max_t \arg \min \|x - f(t)\|.$$

The self-consistency property may be interpreted as follows: each point on the curve is the average of the mass of the probability distribution projecting there (for more details about the notion of self-consistency, see Tarpey and Flury (1996)). Some regularity assumptions are made in addition: the principal curve is required to be smooth (C^∞), it does not intersect itself, and has finite length inside any ball in \mathbb{R}^d . The existence of principal curves designed according to this definition cannot be proved in general (see Duchamp and Stuetzle (1996a), Duchamp and Stuetzle (1996b) for results obtained in the case of some particular distributions in two dimensions), which is the main motivation for the least-square minimization definition proposed by Kégl et al. (2000).

Note that several other principal curve definitions, as well as algorithms, were proposed in the literature (Tibshirani (1992), Verbeek et al. (2001), Delicado (2001), Sandilya and Kulkarni (2002), Einbeck et al. (2005a), Ozertem and Erdogmus (2011), Gerber and Whitaker (2013)). Note also that principal curves, in their empirical version, have many applications in various areas (see for example Hastie and Stuetzle (1989), Friedsam and Oren (1989) for applications in physics, Kégl and Krzyżak (2002), Reinhard and Niranjan (1999) in character and speech recognition, Brunsdon (2007), Stanford and Raftery (2000), Banfield and Raftery (1992), Einbeck et al. (2005a,b) in mapping and geology, De'ath (1999), Corkeron et al. (2004), Einbeck et al. (2005a) in natural sciences, Caffo et al. (2008) in pharmacology, and Wong and Chung (2008), Drier et al. (2013) in medicine, for the study of cardiovascular disease or cancer).

1.2 Description of our results

In this paper, we consider general distributions, assuming only that X has a second order moment, and search for a curve which is optimal for problem (1). We deal with open curves (with endpoints), as well as closed curves ($f(0) = f(1)$). Throughout, we will assume that the length-constraint is effective, that is the support of X is not the image of a curve with length less than or equal to

L . In this context, we prove that a minimizing curve cannot be self-consistent. We also show that, for an optimal curve, the set of points with several different projections of the curve, called ridge set in studies about the “average-distance problem” (see Section 1.3), or ambiguity points in the principal curves literature, is negligible for the distribution of X . Then, we establish that an optimal curve is right- and left-differentiable everywhere and has bounded curvature. Moreover, we obtain a first order Euler-Lagrange equation: we show that there exist $\lambda > 0$ and a random variable \hat{t} taking its values in $[0, 1]$ such that $\|X - f(\hat{t})\| = d(X, \text{Im} f)$ a.s. and

$$\mathbb{E} [X - f(\hat{t}) | \hat{t} = t] m_{\hat{t}}(dt) = -\lambda f''(dt), \quad (2)$$

where $m_{\hat{t}}$ stands for the distribution of \hat{t} . To obtain that $\lambda \neq 0$, we use the fact that an optimal curve is not self-consistent. Formula (2) allows us to propose in dimension $d = 2$ a proof of the injectivity of an open principal curve as well as of a closed principal curve restricted to $[0, 1]$.

1.3 Comparison with previous results

Our framework is related to the constrained problem:

$$\text{minimize } \int_{\mathbb{R}^d} d(x, \Sigma)^p d\mu(x) \text{ over compact connected sets } \Sigma \text{ such that } \mathcal{H}^1(\Sigma) \leq L. \quad (3)$$

Here, \mathcal{H}^ℓ denotes ℓ -dimensional Hausdorff measure. A connected question is the minimization of the penalized version of the criterion:

$$\int_{\mathbb{R}^d} d(x, \Sigma)^p d\mu(x) + \lambda \mathcal{H}^1(\Sigma). \quad (4)$$

This issue, called in the calculus of variations and shape optimization community “average-distance problem” or, for $p = 1$, “irrigation problem”, has been introduced by Buttazzo and Stepanov (2003); Buttazzo et al. (2002) (see also the survey Lemenant (2012), and the references therein). Considering a compactly supported distribution, the penalized form is studied for connected sets, with $p = 1$, in Lu and Slepčev (2013), and for curves, with $p \geq 1$, in Lu and Slepčev (2016). In the first article, the authors prove that a minimizer is a tree made of a finite union of curves with finite length, and they provide a bound on the total curvature of these curves. In the second one, they show existence of a curve minimizing the penalized criterion

$$\int_{\mathbb{R}^d} d(x, \text{Im} f)^p d\mu(x) + \lambda \mathcal{L}(f). \quad (5)$$

They give a bound on the curvature of the minimizer, and prove that, in two dimensions, if $p \geq 2$ or the distribution μ has a bounded density with respect to Lebesgue measure, a minimizing curve is injective.

For the penalized irrigation problem (4), under the assumption that the distribution μ , with compact support, does not charge the sets that have finite \mathcal{H}^{d-1} measure, which is true for instance if it has a density with respect to Lebesgue measure, an Euler-Lagrange equation is obtained for $p = 1$ in Buttazzo et al. (2009), whereas Lemenant (2011) uses arguments involving endpoints to derive one in the case of the constrained version (3), in \mathbb{R}^2 , under the same assumption on μ . This assumption implies that X is almost surely different from its projection on the curve, which is required for differentiability when $p = 1$, and, moreover, it is used to ensure negligibility of the ridge set.

For the constrained problem (3), if Σ^* denotes a minimizer and $\int_{\mathbb{R}^d} d(x, \Sigma)^p d\mu(x) > 0$, it is shown in Paolini and Stepanov (2004) that $\mathcal{H}^1(\Sigma^*) = L$. A similar result in our context is stated in Corollary 3.1 below.

Another related setting is the “lazy travelling salesman problem” of Polak and Wolansky (2007): in \mathbb{R}^2 , taking for μ an empirical distribution and considering closed curves, the authors study the penalized problem (5) for $p = 2$ (with $\lambda \mathcal{L}(f)$ replaced by $\lambda \mathcal{L}^2(f)$). They show that for λ large enough, the problem is reduced to a convex optimization.

Recall that we study in this manuscript the constrained problem (1), for open or closed curves. In our context, the distribution of X is not required to be compactly supported, and we do not need

to assume that μ does not charge the sets with finite \mathcal{H}^{d-1} measure to derive an Euler-Lagrange equation. Indeed, our proof does not rely on the fact that the ridge set is negligible. Besides, we prove that ambiguity points are actually negligible, which implies in particular that, for a given optimal curve, the Lagrange multiplier λ in equation (2) only depends on the curve f . We decided to focus on the case $p = 2$ for which we can state the more complete results. In particular, we are only able to show the default of self-consistency of an optimal curve when $p = 2$. As already mentioned, this is a key point to get the main result. Observe that it would be interesting to define a counterpart of the default of self-consistency when considering other values of p .

1.4 Organization of the paper

Our document is organized as follows. Section 2 introduces relevant notation and recalls some basic facts about length-constrained principal curves. In Section 3, negligibility of ambiguity points is given in Proposition 3.1, and the main result is stated in his complete form in Theorem 3.1.

Injectivity results are presented in 4. Finally, we give in Section 5.2 explicit examples of optimal curves.

2 Definitions and notation

For $d \geq 1$, the space \mathbb{R}^d is equipped with the standard Euclidean norm, denoted by $\|\cdot\|$. The associated inner product between two elements u and v is denoted by $\langle u, v \rangle$. Let \mathcal{H}^1 denotes the 1-dimensional Hausdorff measure in \mathbb{R}^d .

For $x \in \mathbb{R}^d$, $A \subset \mathbb{R}^d$, let $d(x, A) = \inf_{y \in A} \|x - y\|$ denote the distance from point x to set A . For $r > 0$, let $B(x, r)$ and $\bar{B}(x, r)$ denote, respectively, the open and the closed balls with center x and radius r . Also, let ∂A stand for the boundary of A , $\text{Card}(A)$ for its cardinality, and $\text{diam}(A) = \sup_{x, y \in A} \|x - y\|$ for its diameter.

For every $x \in \mathbb{R}^d$, let x^j be its j -th component, for $j = 1, \dots, d$, that is $x = (x^1, \dots, x^d)$. For every $x = (x^1, \dots, x^d) \in \mathbb{R}^d$, we set $\|x\|_\infty = \max_{1 \leq i \leq d} |x^i|$.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X a random vector on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in \mathbb{R}^d , such that $\mathbb{E}[\|X\|^2] < \infty$. We will consider curves, that are continuous functions

$$\begin{aligned} f : [0, 1] &\rightarrow \mathbb{R}^d \\ t &\mapsto (f^1(t), \dots, f^d(t)). \end{aligned}$$

For such a curve $f : [0, 1] \rightarrow \mathbb{R}^d$, let $\mathcal{L}(f) \in [0, \infty]$ denote its length, defined by

$$\mathcal{L}(f) = \sup \sum_{i=1}^n \|f(t_i) - f(t_{i-1})\|, \quad (6)$$

where the supremum is taken over all possible subdivisions $0 = t_0 \leq \dots \leq t_n = 1$, $n \geq 1$ (see, e.g., Alexandrov and Reshetnyak (1989)). Let $\text{Im} f$ denote the image of f .

Let

$$\Delta(f) = \mathbb{E} [d(X, \text{Im} f)^2],$$

and, for $L \geq 0$,

$$G(L) = \min\{\Delta(f), f \in \mathcal{C}_L\},$$

where, in the sequel, \mathcal{C}_L will denote either one of the following sets of curves:

$$\begin{aligned} &\{f \in [0, 1] \rightarrow \mathbb{R}^d, \mathcal{L}(f) \leq L\}, \\ &\{f \in [0, 1] \rightarrow \mathbb{R}^d, \mathcal{L}(f) \leq L, f(0) = f(1)\}. \end{aligned}$$

Curves belonging to the latter set are closed curves. Note that G is well-defined. Indeed, Kégl et al. (2000) have shown the existence of an open curve f with $\mathcal{L}(f) \leq L$ achieving the infimum of the criterion $\Delta(f)$, and the same proof applies for closed curves.

It will be useful to rewrite $G(L)$, for every $L \geq 0$, as the minimum of the quantity

$$\mathbb{E}[\|X - \hat{X}\|^2]$$

over all possible random vectors \hat{X} taking their values in the image $\text{Im} f$ of a curve $f \in \mathcal{C}_L$.

Remark 1. If $f : [0, 1] \rightarrow \mathbb{R}^d$ is Lipschitz with constant L , its length is at most L . This follows directly from the definition of the length (6). Conversely, if the curve $f : [0, 1] \rightarrow \mathbb{R}^d$ has length $\mathcal{L}(f) \leq L$, then there exists a curve with the same image which is Lipschitz with constant L . Indeed, a curve with finite length may be parameterized by arc-length (1-Lipschitz) (see, e.g., Alexandrov and Reshetnyak (1989, Theorem 2.1.4)).

Remark 2. Let $L \geq 0$. Suppose that \hat{X} satisfies $G(L) = \mathbb{E}[\|X - \hat{X}\|^2]$. Writing

$$\mathbb{E}[\|X - \hat{X}\|^2] = \mathbb{E}[\|X - \hat{X} - \mathbb{E}[X - \hat{X}]\|^2] + \|\mathbb{E}[X] - \mathbb{E}[\hat{X}]\|^2,$$

we see that, necessarily,

$$\mathbb{E}[X] = \mathbb{E}[\hat{X}], \tag{7}$$

since, otherwise, the criterion could be made strictly smaller by replacing \hat{X} by the translated variable $\hat{X} + \mathbb{E}[X] - \mathbb{E}[\hat{X}]$, which contradicts the optimality of \hat{X} .

Observe that (7) remains true in a more general setting, as soon as the constraint corresponds to a quantity invariant by translation.

3 Main results and proofs

3.1 Negligibility of the ridge set

Given a curve $f : [0, 1] \rightarrow \mathbb{R}^d$, consider the set

$$\mathcal{P}_f(x) = \{y \in \text{Im}f, \|x - y\| = d(x, \text{Im}f)\} = \bar{B}(x, d(x, \text{Im}f)) \cap \text{Im}f.$$

If $\mathcal{P}_f(x)$ has cardinality at least 2, x is called an ambiguity point in the principal curves literature (see Hastie and Stuetzle (1989)). Properties of the set of such points, named ridge set in the shape optimization community, have been studied for instance in Mantegazza and Menzucchi (2003). In particular, the ridge set is measurable. Using property (7), it may be shown that the ridge set of an optimal curve for X is negligible for the distribution of X . Section 3.3 below presents the proof of this result, as well as the proof of measurability, provided for the sake of completeness.

Proposition 3.1. *Let $f \in \mathcal{C}_L$ be an optimal curve for X ($\Delta(f) = G(L)$).*

1. *The set $\mathcal{A}_f = \{x \in \mathbb{R}^d, \text{Card}(\mathcal{P}_f(x)) \geq 2\}$ of ambiguity points is measurable.*
2. *The set \mathcal{A}_f is negligible for the distribution of X .*

Remark 3. The fact that the ridge set is negligible for the distribution of X may be extended to the context of computing optimal trees under \mathcal{H}^1 constraint. Indeed, the result relies on property (7), and \mathcal{H}^1 measure is translation invariant.

3.2 Main theorem and comments

Recall that a signed measure on (Ω, \mathcal{F}) is a function $m : \mathcal{F} \rightarrow \mathbb{R}$ such that $m(\emptyset) = 0$ and m is σ -additive, that is $m\left(\bigcup_{k \geq 1} A_k\right) = \sum_{k \geq 1} m(A_k)$ for any sequence $(A_k)_{k \geq 1}$ of pairwise disjoint sets. For an \mathbb{R}^d -valued signed measure m on $[0, 1]$, that is $m = (m^1, \dots, m^d)$, where each m^j is a signed measure, and for $g : [0, 1] \rightarrow \mathbb{R}^d$ a measurable function, we will use the following notation: $\int \langle g(t), m(dt) \rangle = \sum_{j=1}^d \int g^j(t) m^j(dt)$.

A probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ will be called an extension of $(\Omega, \mathcal{F}, \mathbb{P})$ if there exists a random vector \tilde{X} defined on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$, with the same distribution μ as X . For simplicity, we still denote this random vector by X throughout the paper.

Theorem 3.1. *Let $L > 0$ such that $G(L) > 0$ and let $f \in \mathcal{C}_L$ such that $\Delta(f) = G(L)$. Then, $\mathcal{L}(f) = L$. Assuming that f is L -Lipschitz, we obtain that*

- *f is right-differentiable on $[0, 1)$, $\|f'_r(t)\| = L$ for all $t \in [0, 1)$,*

- f is left-differentiable on $(0, 1]$, $\|f'_\ell(t)\| = L$ for all $t \in (0, 1]$,

and there exists a unique signed measure f'' on $[0, 1]$ (with values in \mathbb{R}^d) such that

- $f''((s, t]) = f'_r(t) - f'_r(s)$ for all $0 \leq s \leq t < 1$,
- $f''([0, 1]) = 0$.

In the case $\mathcal{C}_L = \{f : [0, 1] \rightarrow \mathbb{R}^d, \mathcal{L}(f) \leq L\}$, we also have

- $f''(\{0\}) = f'_r(0)$,
- $f''(\{1\}) = -f'_\ell(1)$.

Moreover, there exists a unique $\lambda > 0$ and, there exists a random variable \hat{t} with values in $[0, 1]$, defined on an extension $(\tilde{\Omega}, \tilde{\mathcal{F}}, \mathbb{P})$ of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, such that

- $\|X - f(\hat{t})\| = d(X, \text{Im}f)$ a.s.,
- for every bounded Borel function $g : [0, 1] \rightarrow \mathbb{R}^d$,

$$\mathbb{E} [\langle X - f(\hat{t}), g(\hat{t}) \rangle] = -\lambda \int_{[0,1]} \langle g(t), f''(dt) \rangle. \quad (8)$$

Remark 4. Let $m_{\hat{t}|X}$ denote the conditional distribution of \hat{t} given X . Then, equation (8) can be written in the following form:

$$\int_{\mathbb{R}^d} \int_{[0,1]} \langle x - f(t), g(t) \rangle m_{\hat{t}|X}(x, dt) d\mu(x) = -\lambda \int_{[0,1]} \langle g(t), f''(dt) \rangle.$$

Remark 5. Whenever the function g is absolutely continuous, an integration by parts (see for instance Hewitt and Stromberg (1965, Theorem 21.67 & Remarks 21.68)) shows that equation (8) may also be written

$$\mathbb{E} [\langle X - f(\hat{t}), g(\hat{t}) \rangle] = \lambda \int_0^1 \langle g'(t), f'_r(t) \rangle dt. \quad (9)$$

To see this, let us write

$$f''([0, 1])g(1) = f''(\{0\})g(0) + \int_{(0,1]} \langle g(t), f''(dt) \rangle + \int_{(0,1]} \langle g'(s), f''([0, s]) \rangle ds.$$

Since $f''([0, 1]) = 0$, we have

$$0 = \int_{[0,1]} \langle g(t), f''(dt) \rangle + \int_{(0,1]} \langle g'(s), f'_r(s) \rangle ds,$$

which, combined with (8), implies the announced formula (9).

Remark 6. If the curve f has an angle at t , which means that $f'_r(t) \neq f'_\ell(t)$, we see that

$$\mathbb{E}[(X - f(\hat{t}))\mathbf{1}_{\{\hat{t}=t\}}] = -\lambda f''(\{t\}) = \lambda(f'_\ell(t) - f'_r(t)) \neq 0.$$

So, at an angle, $\mathbb{P}(\hat{t} = t) > 0$.

Besides, when $\mathcal{C}_L = \{f : [0, 1] \rightarrow \mathbb{R}^d, \mathcal{L}(f) \leq L\}$, we have

$$\mathbb{E}[(X - f(\hat{t}))\mathbf{1}_{\{\hat{t}=0\}}] = -\lambda f''(\{0\}) = -\lambda f'_r(0),$$

which cannot be zero, since $f'_r(0)$ has norm $L > 0$. This implies that $\mathbb{P}(\hat{t} = 0) > 0$.

Remark 7. Regarding the random variable \hat{t} , let us mention that \hat{t} is unique almost surely whenever the curve is injective since $f(\hat{t})$ is unique almost surely (it is the case in dimension $d \leq 2$; see Section 4). In general, it is worth pointing out that Theorem 3.1 does not ensure that it is a function of X , as (X, \hat{t}) is, in fact, obtained as a limit in distribution of (X, \hat{t}_n) for some sequence $(\hat{t}_n)_{n \geq 1}$. Besides, note that we do not know whether λ depends on the curve f .

Remark 8 (Principal curves in dimension 1). Let $\mathcal{C}_L = \{f : [0, 1] \rightarrow \mathbb{R}^d, \mathcal{L}(f) \leq L\}$. It may be of interest to consider the simplest case of dimension 1, where the problem may be solved entirely and explicitly. Assume that X is a real-valued random variable, and that, for some length $L > 0$, $G(L) > 0$. Consider an optimal curve f with length $\mathcal{L}(f) \leq L$. Using Corollary 3.1 below, we have that, in fact, $\mathcal{L}(f) = L$, so that the image of f is given by an interval $[a, a + L]$. In this context, solving directly the length-constrained principal curve problem in dimension 1 leads to minimizing in a the quantity

$$\Delta(a) := \mathbb{E}[d(X, \text{Im}f)^2] = \mathbb{E}[(X - a)^2 \mathbf{1}_{\{X < a\}}] + \mathbb{E}[(X - a - L)^2 \mathbf{1}_{\{X > a + L\}}].$$

The function Δ is differentiable in a , with derivative given by

$$\Delta'(a) = 2\mathbb{E}[(a - X)\mathbf{1}_{\{X < a\}}] + 2\mathbb{E}[(a + L - X)\mathbf{1}_{\{X > a + L\}}].$$

Moreover, Δ' admits a right-derivative $\Delta_r''(a) = 2(\mathbb{P}(X < a) + \mathbb{P}(X > a + L))$, which is positive since $G(L) > 0$ implies that we do not have $X \in [a, a + L]$ almost surely. Hence, Δ is strictly convex, which shows that the minimizing a is unique, so that the image of the principal curve f is also uniquely defined.

Besides, observe that equation (8) from Theorem 3.1 takes the following form in dimension 1: for every bounded Borel function $g : [0, 1] \rightarrow \mathbb{R}^d$,

$$\mathbb{E}[(X - a)\mathbf{1}_{\{X < a\}}g(0)] + \mathbb{E}[(X - a - L)\mathbf{1}_{\{X > a + L\}}g(1)] = \lambda L(g(1) - g(0)).$$

In particular, we get

$$\begin{aligned} \mathbb{E}[(X - a)\mathbf{1}_{\{X < a\}}] &= -\lambda L, \\ \mathbb{E}[(X - a - L)\mathbf{1}_{\{X > a + L\}}] &= \lambda L, \end{aligned}$$

which characterizes λ . Let us stress that we directly see in this case that $\lambda > 0$, since, otherwise $X \in [a, a + L]$ almost surely, which contradicts the fact that $G(L) > 0$.

3.3 Proof of Proposition 3.1

1. Note that

$$\begin{aligned} \mathcal{A} &= \{x \in \mathbb{R}^d, \text{Card}(\bar{B}(x, d(x, \text{Im}f)) \cap \text{Im}f) \geq 2\} \\ &= \{x \in \mathbb{R}^d, \text{diam}(\bar{B}(x, d(x, \text{Im}f)) \cap \text{Im}f) > 0\} \\ &= \mathbb{R}^d \setminus \{x \in \mathbb{R}^d, \text{diam}(\bar{B}(x, d(x, \text{Im}f)) \cap \text{Im}f) = 0\}. \end{aligned}$$

For every $x \in \mathbb{R}^d$, we may write

$$\text{diam}(\bar{B}(x, d(x, \text{Im}f)) \cap \text{Im}f) = \lim_{n \rightarrow \infty} \text{diam}(B(x, d(x, \text{Im}f) + 1/n) \cap \text{Im}f).$$

Since f is continuous, $f([0, 1] \cap \mathbb{Q})$ is dense in $\text{Im}f$. For every $n \geq 1$, the countable set $B(x, d(x, \text{Im}f) + 1/n) \cap f([0, 1] \cap \mathbb{Q})$ is dense in $B(x, d(x, \text{Im}f) + 1/n) \cap \text{Im}f$, so that both sets have the same diameter. Yet, it can be easily checked that the diameter of a countable set is measurable, and finally, we obtain that the set \mathcal{A} of ambiguity points is measurable.

2. To begin with, we prove that, for every $j = 1, \dots, d$, it is possible to construct a random vector \hat{X} with values in $\text{Im}f$ such that $\|X - \hat{X}\| = d(X, \text{Im}f)$ a.s., and

$$\hat{X}^j = \max \pi_j(\bar{B}(X, d(X, \text{Im}f)) \cap \text{Im}f).$$

Here, π_j stands for the projection onto direction j , that is, for $x = (x^1, \dots, x^d) \in \mathbb{R}^d$, $\pi_j(x) = x^j$. Let $\{t_1, t_2, \dots\}$ be an enumeration of the countable set $[0, 1] \cap \mathbb{Q}$. Let $\varepsilon > 0$, $x \in \mathbb{R}^d$. First, note that the set $\{t \in [0, 1], \|f(t) - x\| < d(x, \text{Im}f) + \varepsilon\}$ is open. It is nonempty since the distance from x to the closed set $\text{Im}f$ is attained. We deduce from this

that $\text{Card}(\{t \in [0, 1] \cap \mathbb{Q}, \|f(t) - x\| \leq d(x, \text{Im}f) + \varepsilon\}) = \infty$. Let us define the sequence $(k_\varepsilon^n(x))_{n \in \mathbb{N}}$ by

$$\begin{aligned} k_\varepsilon^1(x) &= \min\{k : \|f(t_k) - x\| \leq d(x, \text{Im}f) + \varepsilon\} \\ k_\varepsilon^{m+1}(x) &= \min\{k > k_\varepsilon^m(x) : \|f(t_k) - x\| \leq d(x, \text{Im}f) + \varepsilon\}, \quad m \in \mathbb{N}. \end{aligned}$$

Let $j \in \{1, \dots, d\}$. We set

$$p^*(x) = \min\{p \geq 1, f^j(t_{k_\varepsilon^p(x)}) \geq \sup_{m \in \mathbb{N}} f^j(t_{k_\varepsilon^m(x)}) - \varepsilon\}.$$

We define $\hat{X}_\varepsilon(x) = f(t_{k_\varepsilon^{p^*(x)}}(x))$, which is a measurable choice. Notice that, since $\{f^j(t_{k_\varepsilon^m(x)}), m \in \mathbb{N}\} = \pi_j(\bar{B}(x, d(x, \text{Im}f) + \varepsilon) \cap f([0, 1] \cap \mathbb{Q}))$ is dense in $\pi_j(\bar{B}(x, d(x, \text{Im}f) + \varepsilon) \cap \text{Im}f)$, both sets have the same supremum.

Let

$$\Pi_\varepsilon(x) = \pi_j(\bar{B}(x, d(x, \text{Im}f) + \varepsilon) \cap \text{Im}f), \quad \Pi(x) = \pi_j(\bar{B}(x, d(x, \text{Im}f)) \cap \text{Im}f).$$

The limit of $\hat{X}_\varepsilon^j(x)$ is given by $\lim_{\varepsilon \rightarrow 0} \max \Pi_\varepsilon(x)$. Yet, note that, for every ε , $\Pi(x) \subset \Pi_\varepsilon(x)$ so that

$$\max \Pi(x) \leq \max \Pi_\varepsilon(x). \quad (10)$$

Moreover, if ε is small enough, then for all $y \in \Pi_\varepsilon(x)$, $d(y, \Pi(x)) \leq \eta(\varepsilon)$, where η tends to 0 with ε , and, thus,

$$\max \Pi_\varepsilon(x) \leq \max \Pi(x) + \eta(\varepsilon). \quad (11)$$

Combining inequalities (10) and (11), we obtain that $\lim_{\varepsilon \rightarrow 0} \max \Pi_\varepsilon(x) = \max \Pi(x)$.

Set $\varepsilon_n = 1/n$. Up to an extraction, we may assume that $(\hat{X}_{\varepsilon_n}(X), X)$ converges in distribution to (\hat{X}, X) as $n \rightarrow \infty$. The random vector \hat{X} satisfies $\|X - \hat{X}\| = d(X, \text{Im}f)$ and $\hat{X}^j = \max \Pi(X)$.

Similarly, as may be seen by replacing X by $-X$, there exists a random vector \hat{Y} with values in $\text{Im}f$ such that $\|X - \hat{Y}\| = d(X, \text{Im}f)$ a.s., and

$$\hat{Y}^j = \min \pi_j(\bar{B}(X, d(X, \text{Im}f)) \cap \text{Im}f).$$

Now, we use this result to show that \mathcal{A} is negligible for the distribution of X . Assume that $\mathbb{P}(\text{Card}(\mathcal{P}_f(X)) \geq 2) > 0$. There exists a first coordinate j such that $\mathbb{P}(\text{Card}(\pi_j(\mathcal{P}_f(X))) \geq 2) > 0$. Then, it is possible to construct \hat{X}^j and \hat{Y}^j such that $\mathbb{P}(\hat{X}^j \geq \hat{Y}^j) = 1$ and $\mathbb{P}(\hat{X}^j > \hat{Y}^j) > 0$. Yet, by property (7), $\mathbb{E}[\hat{X}] = \mathbb{E}[X] = \mathbb{E}[\hat{Y}]$, and, in particular, $\mathbb{E}[\hat{X}^j] = \mathbb{E}[\hat{Y}^j]$, which leads to a contradiction. Thus, $\mathbb{P}(\text{Card}(\mathcal{P}_f(X)) = 1) = 1$.

In the next sections, we present two lemmas, which are important both independently and for obtaining the main result Theorem 3.1.

3.4 Properties of the function G

The first lemma is about the monotonicity and continuity properties of the function G . Observe that G is nonincreasing, since $\{f : [0, 1] \rightarrow \mathbb{R}^d, \mathcal{L}(f) \leq L_1\} \subset \{f : [0, 1] \rightarrow \mathbb{R}^d, \mathcal{L}(f) \leq L_2\}$ when $L_1 < L_2$, so that $G(L_2) \leq G(L_1)$.

Lemma 3.1. 1. *The function G is continuous.*

2. *The function G is strictly decreasing over $[0, L_0)$, where $L_0 = \inf\{L \geq 0, G(L) = 0\} \in \mathbb{R}_+ \cup \{\infty\}$.*

In particular, Lemma 3.1 admits the next useful corollary.

Corollary 3.1. *For $L > 0$, if $G(L) > 0$ and $f \in \mathcal{C}_L$ is such that $\Delta(f) = G(L)$, then $\mathcal{L}(f) = L$.*

Proof. If $\mathcal{L}(f) < L$, then Lemma 3.1 would imply $G(\mathcal{L}(f)) > G(L) = \Delta(f)$, which contradicts the definition of G . \square

Proof of Lemma 3.1. 1. Set $L \geq 0$. Let us show that G is continuous at the point L . Let $(L_k)_{k \in \mathbb{N}}$ be a sequence in \mathbb{R}_+ converging to L , with $L_k \neq L$ for all $k \in \mathbb{N}$. Let $f \in \mathcal{C}_L$ be such that $\Delta(f) = G(L)$, and let \hat{X} stands for a random vector taking its values in $\text{Im} f$ such that $\|X - \hat{X}\| = d(X, \text{Im} f)$ a.s. For every $k \in \mathbb{N}$, let $f_k : [0, 1] \rightarrow \mathbb{R}^d$ be a curve such that $\mathcal{L}(f_k) \leq L_k$, $\Delta(f_k) = G(L_k)$ and $\|f_k(t) - f_k(t')\| \leq L_k|t - t'|$ for $t, t' \in [0, 1]$.

Observe that the sequence $(G(L_k))_{k \in \mathbb{N}}$ is bounded since $\mathbb{E}[\|X\|^2] < \infty$. Let us show that $G(L)$ is the unique limit point of this sequence. Let $\gamma : \mathbb{N} \rightarrow \mathbb{N}$ be any increasing function. Our purpose is to show that the sequence $(G(L_{\gamma(k)}))_{k \in \mathbb{N}}$ converges to $G(L)$.

Let us check that the f_k are equi-uniformly continuous and that the sequence $(f_k(0))$ is bounded. Since the sequence $(L_k)_{k \in \mathbb{N}}$ is bounded, say by L' , the f_k are Lipschitz with common Lipschitz constant L' , and, thus, they are equi-uniformly continuous. For every $k \in \mathbb{N}$, $t \in [0, 1]$, we have $\|f_k(t)\| \geq \|f_k(0)\| - L't \geq \|f_k(0)\| - L'$. Thus, if there exists an increasing function $\kappa : \mathbb{N} \rightarrow \mathbb{N}$ such that $\|f_{\kappa(k)}(0)\| \rightarrow \infty$, one has $G(L_{\kappa(k)}) \rightarrow \infty$, which is impossible since $G(L_k) \leq \mathbb{E}[\|X\|^2] < \infty$. So, the sequence $(f_k(0))_{k \in \mathbb{N}}$ is bounded.

Consequently, there exists an increasing function $\sigma : \mathbb{N} \rightarrow \mathbb{N}$ such that the subsequence $(f_{\sigma \circ \gamma(k)})_{k \in \mathbb{N}}$ converges uniformly to some function $\varphi : [0, 1] \rightarrow \mathbb{R}^d$. Note that the curve φ is L -Lipschitz, since for all t, t' ,

$$\begin{aligned} \|\varphi(t) - \varphi(t')\| &\leq \|\varphi(t) - f_{\sigma \circ \gamma(k)}(t)\| + \|f_{\sigma \circ \gamma(k)}(t) - f_{\sigma \circ \gamma(k)}(t')\| - \|f_{\sigma \circ \gamma(k)}(t') - \varphi(t')\| \\ &\leq \|\varphi(t) - f_{\sigma \circ \gamma(k)}(t)\| + L_{\sigma \circ \gamma(k)}|t - t'| - \|f_{\sigma \circ \gamma(k)}(t') - \varphi(t')\|, \end{aligned}$$

which implies, taking the limit as $k \rightarrow \infty$, $\|\varphi(t) - \varphi(t')\| \leq L|t - t'|$. We have $\mathcal{L}(\varphi) \leq \lim_{k \rightarrow \infty} L_k = L$. Now, observe that

$$\begin{aligned} &\min_t \|X - f_{\sigma \circ \gamma(k)}(t)\|^2 - \min_t \|X - \varphi(t)\|^2 \\ &= \left(\min_t \|X - f_{\sigma \circ \gamma(k)}(t)\| - \min_t \|X - \varphi(t)\| \right) \left(\min_t \|X - f_{\sigma \circ \gamma(k)}(t)\| + \min_t \|X - \varphi(t)\| \right) \\ &\leq \|\varphi(t^*) - f_{\sigma \circ \gamma(k)}(t^*)\| (\|X - f_{\sigma \circ \gamma(k)}(t^*)\| + \|X - \varphi(t^*)\|), \end{aligned}$$

where $\|X - \varphi(t^*)\| = \min_t \|X - \varphi(t)\|$. Since $\mathbb{E}[\|X\|^2] < \infty$ and $f_{\sigma \circ \gamma(k)}$ converges uniformly to φ , this shows that $\Delta(f_{\sigma \circ \gamma(k)})$ converges to $\Delta(\varphi)$.

Finally, let us check that $\Delta(\varphi) = G(L)$. If $L = 0$, then for every k , $L_k \geq L$, thus $\Delta(f_{\sigma \circ \gamma(k)}) = G(f_{\sigma \circ \gamma(k)}) \leq G(0)$ for every k . Consequently, $\Delta(\varphi) \leq G(0)$, which implies $\Delta(\varphi) = G(0)$ since φ has length 0. If $L > 0$, note that, for every k , $\frac{L_k}{L}\hat{X}$ is a random vector with values in $\frac{L_k}{L}\text{Im} f$ since \hat{X} is taking its values in $\text{Im} f$. Moreover, $\frac{L_k}{L}f$ has length at most L_k since f has length L . Thus, for every k ,

$$\mathbb{E} \left[\left\| X - \frac{L_{\sigma \circ \gamma(k)}}{L} \hat{X} \right\|^2 \right] \geq G(L_{\sigma \circ \gamma(k)}) = \Delta(f_{\sigma \circ \gamma(k)}).$$

taking the limit as $k \rightarrow \infty$, we obtain

$$\mathbb{E} [\|X - \hat{X}\|^2] \geq \Delta(\varphi),$$

which means that $\Delta(\varphi) = G(L)$ since $\mathcal{L}(\varphi) \leq L$.

2. We have to show that G is strictly decreasing as long as the length constraint is effective (that is $G(L) > 0$). Let us prove that for $0 \leq L_1 < L_2$, we have $G(L_2) < G(L_1)$ if $G(L_1) > 0$. Let $f : [0, 1] \rightarrow \mathbb{R}^d$ such that $\mathcal{L}(f) \leq L_1$ and $\Delta(f) = G(L_1)$. For $t_0 \in [0, 1]$ and $r > 0$, we define $\hat{Z}_{t_0, r}$ by

$$\begin{cases} \hat{Z}_{t_0, r}^J = f^J(t_0) + r \wedge (X^J - f^J(t_0)) \mathbf{1}_{\{X^J \geq f^J(t_0)\}} + (-r) \vee (X^J - f^J(t_0)) \mathbf{1}_{\{X^J < f^J(t_0)\}}, \\ \text{where } J = \min\{i : |X^i - f^i(t_0)| = \|X - f(t_0)\|_\infty\} \\ \hat{Z}_{t_0, r}^i = f^i(t_0) \text{ if } i \neq J, i = 1, \dots, d. \end{cases}$$

Observe that $\hat{Z}_{t_0,r}$ takes its values in

$$\mathcal{C}(t_0, r) = \bigcup_{j=1}^d \{x \in \mathbb{R}^d : x^i = f^i(t_0) \text{ for } i \neq j, |x^j - f^j(t_0)| \leq r\}.$$

Indeed, all coordinates of $\hat{Z}_{t_0,r}$ are equal to the corresponding coordinate of $f(t_0)$ apart from the J -th coordinate, that is the first coordinate for which the distance between X and $f(t_0)$ is the largest one. Let us check that $|\hat{Z}_{t_0,r}^J - f^J(t_0)| \leq r$.

If $X^J \geq f^J(t_0)$, either $\hat{Z}_{t_0,r}^J - f^J(t_0) = r$, or $\hat{Z}_{t_0,r}^J - f^J(t_0) = X^J - f^J(t_0) \leq r$.

If $X^J < f^J(t_0)$, either $f^J(t_0) - \hat{Z}_{t_0,r}^J = r$, or $f^J(t_0) - \hat{Z}_{t_0,r}^J = f^J(t_0) - X^J \leq r$.

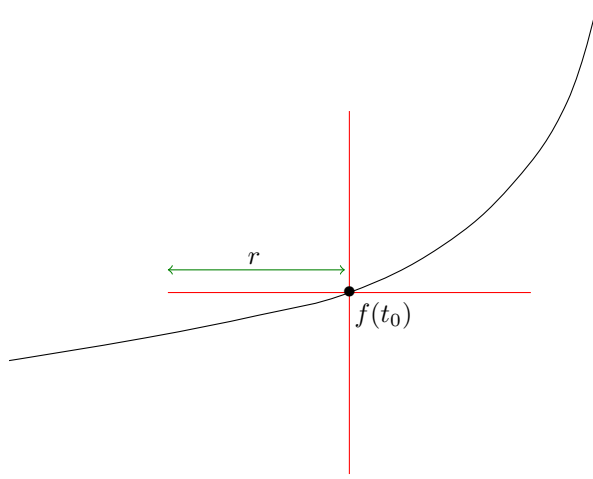


Figure 2: Example in \mathbb{R}^2 , illustrating the support of $\hat{Z}_{t_0,r}$.

Then, letting again \hat{X} be a random vector with values in $\text{Im} f$ such that $\|X - \hat{X}\| = d(X, \text{Im} f)$ a.s., we set

$$\hat{X}_{t_0,r} = \hat{X} \mathbf{1}_{\{\|X - \hat{X}\| \leq \|X - \hat{Z}_{t_0,r}\|\}} + \hat{Z}_{t_0,r} \mathbf{1}_{\{\|X - \hat{X}\| > \|X - \hat{Z}_{t_0,r}\|\}}.$$

Since $\|X - \hat{Z}_{t_0,r}\|^2 = \|X - f(t_0)\|^2 - \|X - f(t_0)\|_\infty^2 + (\|X - f(t_0)\|_\infty - r)_+^2$,

$$\begin{aligned} & \|X - \hat{X}\|^2 - \|X - \hat{X}_{t_0,r}\|^2 \\ &= \left[\|X - \hat{X}\|^2 - \|X - \hat{Z}_{t_0,r}\|^2 \right]_+ \\ &= \left[\|X - \hat{X}\|^2 - \|X - f(t_0)\|^2 + \|X - f(t_0)\|_\infty^2 - (\|X - f(t_0)\|_\infty - r)_+^2 \right]_+ \\ &\geq \left[\|X - \hat{X}\|^2 - \|X - f(t_0)\|^2 + \|X - f(t_0)\|_\infty^2 - (\|X - f(t_0)\|_\infty - r)^2 \right]_+ \\ &= \left[\|X - \hat{X}\|^2 - \|X - f(t_0)\|^2 + 2r\|X - f(t_0)\|_\infty - r^2 \right]_+ \\ &= \left[\|f(t_0) - \hat{X}\|^2 + 2\langle X - f(t_0), f(t_0) - \hat{X} \rangle + 2r\|X - f(t_0)\|_\infty - r^2 \right]_+ \\ &= \left[-\|f(t_0) - \hat{X}\|^2 + 2\langle X - \hat{X}, f(t_0) - \hat{X} \rangle + 2r\|X - f(t_0)\|_\infty - r^2 \right]_+ \\ &\geq \left[-\|f(t_0) - \hat{X}\|^2 + 2\langle X - \hat{X}, f(t_0) - \hat{X} \rangle + \frac{2r}{\sqrt{d}}\|X - f(t_0)\| - r^2 \right]_+ \\ &\quad \text{since for every } x \in \mathbb{R}^d, \|x\| \leq \sqrt{d}\|x\|_\infty \\ &\geq \left[-\|f(t_0) - \hat{X}\|^2 + 2\langle X - \hat{X}, f(t_0) - \hat{X} \rangle + \frac{2r}{\sqrt{d}}\|X - \hat{X}\| - r^2 \right]_+ \\ &\quad \text{since } \|X - \hat{X}\| \leq \|X - f(t_0)\|. \end{aligned}$$

Besides, $\hat{X}_{t_0, r}$ takes its values in $\text{Im}f \cup \mathcal{C}(t_0, r)$, which is the image of a curve with length at most $L_1 + 4dr$, so that $\mathbb{E}[\|X - \hat{X}_{t_0, r}\|^2] \geq G(L_1 + 4dr)$.

Thus,

$$G(L_1) \geq G(L_1 + 4dr) + \mathbb{E} \left[\left[-\|f(t_0) - \hat{X}\|^2 + 2\langle X - \hat{X}, f(t_0) - \hat{X} \rangle + \frac{2r}{\sqrt{d}}\|X - \hat{X}\| - r^2 \right]_+ \right]. \quad (12)$$

Since $G(L_1) > 0$, $\mathbb{P}(\|X - \hat{X}\| > 0) > 0$, thus there exist $\delta > 0$ and $K < \infty$ such that $\eta := P(K \geq \|X - \hat{X}\| \geq \delta) > 0$.

Recall that, for all (t, t') , we have $\|f(t) - f(t')\| \leq L_1|t - t'|$. Then, for every $p \geq 1$, there exists k , $1 \leq k \leq p$, such that $\|\hat{X} - f(\frac{k}{p})\| \leq \frac{L_1}{p}$ and so, we have

$$\sum_{k=1}^p \mathbf{1}_{\{\|\hat{X} - f(\frac{k}{p})\| \leq \frac{L_1}{p}\}} \geq 1.$$

Thus,

$$\sum_{k=1}^p \mathbb{P} \left(K \geq \|X - \hat{X}\| \geq \delta, \left\| \hat{X} - f\left(\frac{k}{p}\right) \right\| \leq \frac{L_1}{p} \right) \geq \eta.$$

Consequently, for every $p \geq 1$, there exists $t_p \in [0, 1]$ such that

$$\mathbb{P} \left(K \geq \|X - \hat{X}\| \geq \delta, \|\hat{X} - f(t_p)\| \leq \frac{L_1}{p} \right) \geq \frac{\eta}{p} > 0.$$

According to (12), we obtain

$$\begin{aligned} G(L_1) &\geq G(L_1 + 4dr) + \mathbb{E} \left[-\|f(t_p) - \hat{X}\|^2 + 2\langle X - \hat{X}, f(t_p) - \hat{X} \rangle + \frac{2r}{\sqrt{d}}\|X - \hat{X}\| - r^2 \right]_+ \\ &\geq G(L_1 + 4dr) + \mathbb{E} \left[\mathbf{1}_{\{K \geq \|X - \hat{X}\| \geq \delta, \|\hat{X} - f(t_p)\| \leq \frac{L_1}{p}\}} \left(-\frac{L_1^2}{p^2} - \frac{2KL_1}{p} + \frac{2r\delta}{\sqrt{d}} - r^2 \right) \right] \\ &\geq G(L_1 + 4dr) + \frac{\eta}{p} \left(-\frac{L_1^2}{p^2} - \frac{2KL_1}{p} + \frac{2r\delta}{\sqrt{d}} - r^2 \right). \end{aligned}$$

Now, choosing $r > 0$ such that $\frac{2r\delta}{\sqrt{d}} - r^2 > 0$ and $L_1 + 4dr \leq L_2$, we finally obtain, taking p large enough,

$$G(L_1) > G(L_1 + 4dr) \geq G(L_2).$$

□

3.5 Default of self-consistency

The next lemma states that a principal curve with length $\leq L$ does not satisfy the so-called self-consistency property, provided that the constraint is effective, that is $G(L) > 0$.

Lemma 3.2. *Let $L > 0$ such that $G(L) > 0$, and let $f \in \mathcal{C}_L$ be such that $\Delta(f) = G(L)$. If \hat{X} is a random vector with values in $\text{Im}f$ such that $\|X - \hat{X}\| = d(X, \text{Im}f)$ a.s., then $\mathbb{P}(\mathbb{E}[X|\hat{X}] \neq \hat{X}) > 0$.*

Proof. First of all, observe that $\mathcal{L}(f) = L$ since $G(L) > 0$, according to Corollary 3.1. Assume that $\mathbb{E}[X|\hat{X}] = \hat{X}$ a.s..

For $\varepsilon \in [0, 1]$, we set $\hat{X}_\varepsilon = (1 - \varepsilon)\hat{X}$. Then,

$$\|X - \hat{X}_\varepsilon\|^2 = \|X - \hat{X} + \varepsilon\hat{X}\|^2 = \|X - \hat{X}\|^2 + \varepsilon^2\|\hat{X}\|^2 + 2\varepsilon\langle X - \hat{X}, \hat{X} \rangle.$$

Since $\mathbb{E}[X|\hat{X}] = \hat{X}$ a.s., $\mathbb{E}[X - \hat{X}|\hat{X}] = 0$ a.s., and thus, $\mathbb{E}[\langle X - \hat{X}, \hat{X} \rangle] = \mathbb{E}[\langle \mathbb{E}[X - \hat{X}|\hat{X}], \hat{X} \rangle] = 0$, so that

$$\mathbb{E}[\|X - \hat{X}_\varepsilon\|^2] = \mathbb{E}[\|X - \hat{X}\|^2] + \varepsilon^2\mathbb{E}[\|\hat{X}\|^2]. \quad (13)$$

The random vector \hat{X}_ε is taking its values in the image of $(1 - \varepsilon)f$, which has length $(1 - \varepsilon)L$. Observe that

$$\mathbb{E}[\|\hat{X}\|^2] < \infty, \quad (14)$$

since $\mathbb{E}[\|X\|^2] < \infty$ and

$$\begin{aligned} \mathbb{E}[\|\hat{X}\|^2] &\leq 2\mathbb{E}[\|X - \hat{X}\|^2] + 2\mathbb{E}[\|X\|^2] \\ &\leq 2\mathbb{E}[\|X - f(0)\|^2] + 2\mathbb{E}[\|X\|^2] \\ &\leq 6\mathbb{E}[\|X\|^2] + 4\|f(0)\|^2. \end{aligned}$$

We will show that, adding to $(1 - \varepsilon)f$ a curve with length εL , it is possible to build \hat{Y}_ε with $\mathbb{E}[\|X - \hat{Y}_\varepsilon\|^2] < \mathbb{E}[\|X - \hat{X}\|^2]$, which contradicts the optimality of f .

For $\varepsilon \in [0, 1]$, let $f_\varepsilon = (1 - \varepsilon)f$. We then define $\hat{X}_{\varepsilon, t_0, r}$ as the variable $\hat{X}_{t_0, r}$ corresponding to f_ε . More precisely, similarly to the proof of Lemma 3.1, we define, for $t_0 \in [0, 1]$ and $r > 0$, the random vector $\hat{Z}_{\varepsilon, t_0, r}$, with values in

$$\mathcal{C}(t_0, r) = \bigcup_{j=1}^d \{x \in \mathbb{R}^d : x^i = f_\varepsilon^i(t_0) \text{ for } i \neq j, |x^j - f_\varepsilon^j(t_0)| \leq r\},$$

by

$$\begin{cases} \hat{Z}_{\varepsilon, t_0, r}^J = f_\varepsilon^J(t_0) + r \wedge (X^J - f_\varepsilon^J(t_0)) \mathbf{1}_{\{X^J \geq f_\varepsilon^J(t_0)\}} + (-r) \vee (X^J - f_\varepsilon^J(t_0)) \mathbf{1}_{\{X^J < f_\varepsilon^J(t_0)\}}, \\ \text{where } J = \min\{i : |X^i - f_\varepsilon^i(t_0)| = \|X - f_\varepsilon(t_0)\|_\infty\} \\ \hat{Z}_{\varepsilon, t_0, r}^i = f_\varepsilon^i(t_0) \text{ if } i \neq J, i = 1, \dots, d. \end{cases}$$

We set

$$\hat{X}_{\varepsilon, t_0, r} = \hat{X} \mathbf{1}_{\{\|X - \hat{X}_\varepsilon\| \leq \|X - \hat{Z}_{\varepsilon, t_0, r}\|\}} + \hat{Z}_{\varepsilon, t_0, r} \mathbf{1}_{\{\|X - \hat{X}_\varepsilon\| > \|X - \hat{Z}_{\varepsilon, t_0, r}\|\}}.$$

By the same calculation as in the proof of Lemma 3.1, we obtain

$$\|X - \hat{X}_\varepsilon\|^2 - \|X - \hat{X}_{\varepsilon, t_0, r}\|^2 \geq \left[-\|f_\varepsilon(t_0) - \hat{X}_\varepsilon\|^2 + 2\langle X - \hat{X}_\varepsilon, f_\varepsilon(t_0) - \hat{X}_\varepsilon \rangle + \frac{2r}{\sqrt{d}} \|X - f_\varepsilon(t_0)\| - r^2 \right]_+.$$

Since $\|X - f_\varepsilon(t_0)\| \geq \|X - f(t_0)\| - \varepsilon\|f(t_0)\| \geq \|X - \hat{X}\| - \varepsilon\|f(t_0)\|$, we get

$$\begin{aligned} \|X - \hat{X}_\varepsilon\|^2 - \|X - \hat{X}_{\varepsilon, t_0, r}\|^2 &\geq \left[-(1 - \varepsilon)^2 \|f(t_0) - \hat{X}\|^2 + 2(1 - \varepsilon) \langle X - \hat{X}_\varepsilon, f(t_0) - \hat{X} \rangle \right. \\ &\quad \left. + \frac{2r}{\sqrt{d}} \|X - \hat{X}\| - \frac{2r}{\sqrt{d}} \varepsilon \|f(t_0)\| - r^2 \right]_+. \end{aligned}$$

Thus,

$$\begin{aligned} &\mathbb{E} \left[\|X - \hat{X}_\varepsilon\|^2 - \|X - \hat{X}_{\varepsilon, t_0, r}\|^2 \middle| \hat{X} \right] \\ &\geq \left[-\|f(t_0) - \hat{X}\|^2 + 2(1 - \varepsilon) \langle \mathbb{E}[X | \hat{X}] - \hat{X}_\varepsilon, f(t_0) - \hat{X} \rangle + \frac{2r}{\sqrt{d}} \mathbb{E} \left[\|X - \hat{X}\| \middle| \hat{X} \right] - \frac{2r}{\sqrt{d}} \varepsilon \|f(t_0)\| - r^2 \right]_+ \\ &= \left[-\|f(t_0) - \hat{X}\|^2 + 2(1 - \varepsilon) \langle \varepsilon \hat{X}, f(t_0) - \hat{X} \rangle + \frac{2r}{\sqrt{d}} \mathbb{E} \left[\|X - \hat{X}\| \middle| \hat{X} \right] - \frac{2r}{\sqrt{d}} \varepsilon \|f(t_0)\| - r^2 \right]_+ \\ &\geq \left[-\|f(t_0) - \hat{X}\|^2 - 2\varepsilon \|\hat{X}\| \|f(t_0) - \hat{X}\| + \frac{2r}{\sqrt{d}} \mathbb{E} \left[\|X - \hat{X}\| \middle| \hat{X} \right] - \frac{2r}{\sqrt{d}} \varepsilon \|f(t_0)\| - r^2 \right]_+. \end{aligned} \quad (15)$$

Besides, since $G(L) > 0$, there exist $\delta > 0$, $K < \infty$, such that

$$\eta = \mathbb{P} \left(\|\hat{X}\| \leq K, \mathbb{E} \left[\|X - \hat{X}\| \middle| \hat{X} \right] \geq \delta \right) > 0.$$

Moreover, for every $p \geq 1$, $\sum_{k=1}^p \mathbf{1}_{\{\|\hat{X} - f(\frac{k}{p})\| \leq \frac{L}{p}\}} \geq 1$ since f is L -Lipschitz. Consequently,

$$\sum_{k=1}^p \mathbb{P} \left(\|\hat{X}\| \leq K, \mathbb{E} \left[\|X - \hat{X}\| \middle| \hat{X} \right] \geq \delta, \left\| \hat{X} - f \left(\frac{k}{p} \right) \right\| \leq \frac{L}{p} \right) \geq \eta.$$

Hence, setting

$$A_p = \left\{ \|\hat{X}\| \leq K, \mathbb{E} \left[\|X - \hat{X}\| \middle| \hat{X} \right] \geq \delta, \left\| \hat{X} - f\left(\frac{k}{p}\right) \right\| \leq \frac{L}{p} \right\},$$

we see that there exists $t_p \in [0, 1]$ such that $\mathbb{P}(A_p) \geq \frac{\eta}{p}$. From (15), we get

$$\begin{aligned} & \mathbb{E} \left[\|X - \hat{X}_\varepsilon\|^2 - \|X - \hat{X}_{\varepsilon, t_p, r}\|^2 \right] \\ & \geq \mathbb{E} \left[\mathbf{1}_{A_p} \left[-\|f(t_p) - \hat{X}\|^2 - 2\varepsilon \|\hat{X}\| \|f(t_p) - \hat{X}\| + \frac{2r}{\sqrt{d}} \mathbb{E} \left[\|X - \hat{X}\| \middle| \hat{X} \right] - \frac{2r}{\sqrt{d}} \varepsilon \|f(t_p)\| - r^2 \right]_+ \right] \\ & \geq \mathbb{P}(A_p) \left[-\frac{L^2}{p^2} - \frac{2\varepsilon KL}{p} + \frac{2r\delta}{\sqrt{d}} - \frac{2r\varepsilon M}{\sqrt{d}} - r^2 \right], \end{aligned}$$

where $M = \sup_{t \in [0, 1]} \|f(t)\|$. Since $\hat{X}_{\varepsilon, t_p, r}$ takes its values in $f_\varepsilon([0, 1]) \cup \mathcal{C}(\varepsilon, t_p, r)$, which is the image of a curve with length at most $(1 - \varepsilon)L + 4dr$, then choosing r such that $4dr = \varepsilon L$, we have

$$\begin{aligned} \mathbb{E} \left[\left\| X - \hat{X}_{\varepsilon, t_p, \frac{\varepsilon L}{4d}} \right\|^2 \right] & \leq \mathbb{E} \left[\|X - \hat{X}_\varepsilon\|^2 \right] - \frac{\eta}{p} \left(-\frac{L^2}{p^2} - \frac{2KL\varepsilon}{p} + \frac{L\delta\varepsilon}{2d^{3/2}} - \frac{ML\varepsilon^2}{2d^{3/2}} - \frac{L^2\varepsilon^2}{16d^2} \right) \\ & = \mathbb{E}[\|X - \hat{X}\|^2] + \varepsilon^2 \mathbb{E}[\|\hat{X}\|^2] + \frac{\eta L^2}{p^3} + \frac{2\eta KL\varepsilon}{p^2} - \frac{\eta L\delta\varepsilon}{2d^{3/2}p} + \frac{\eta ML\varepsilon^2}{2d^{3/2}p} - \frac{\eta L^2\varepsilon^2}{16d^2p^3}, \end{aligned}$$

using (13). Then, taking $\varepsilon = \frac{\rho}{p}$, we get

$$\mathbb{E} \left[\left\| X - \hat{X}_{\frac{\rho}{p}, t_p, \frac{\rho L}{4dp}} \right\|^2 \right] \leq \mathbb{E}[\|X - \hat{X}\|^2] + \frac{\rho^2}{p^2} \mathbb{E}[\|\hat{X}\|^2] + \frac{\eta L^2}{p^3} + \frac{2\eta KL\rho}{p^3} - \frac{\eta L\delta\rho}{2d^{3/2}p^2} + \frac{\eta ML\rho^2}{2d^{3/2}p^3} - \frac{\eta L^2\rho^2}{16d^2p^3}.$$

If ρ is small enough, then $\rho^2 \mathbb{E}[\|\hat{X}\|^2] - \frac{\eta L\delta\rho}{2d^{3/2}} < 0$. Then, taking p large enough, this leads to a random vector \hat{Y} , with values in the image of a curve with length at most L , such that $\mathbb{E}[\|X - \hat{Y}\|^2] < \mathbb{E}[\|X - \hat{X}\|^2]$. □

Equipped with lemmas 3.1 and 3.2, we can present the proof of the main result.

3.6 Proof of Theorem 3.1

To obtain a length-constrained principal curve, we have to minimize a function which may not be differentiable. We propose to build a discrete approximation of the principal curve f , using a chain of points v_1^n, \dots, v_n^n , $n \geq 1$, in \mathbb{R}^d . For every $n \geq 1$, linking the points yields a polygonal curve f_n . The properties of the principal curve f will be shown by passing to the limit. The chain of points is obtained by minimizing a k -means-like criterion, which is differentiable, under a length-constraint. This criterion is based on the distances from the random vector X to the n points and not to the corresponding segments of the polygonal line f_n , which allows to simplify the computation of the gradients.

We have chosen to present the proof for open curves, that is in the case $\mathcal{C}_L = \{\varphi : [0, 1] \rightarrow \mathbb{R}^d, \mathcal{L}(\varphi) \leq L\}$. It adapts straightforwardly to the case of closed curves, which turns out to be even simpler since there are no endpoints and so all points of the curve play the same role. Note that the normalization factor “ $n - 1$ ” below becomes “ n ” in the closed curve context.

First insight into the proof

To facilitate understanding, we sketch the proof in a simpler case. Assume that X has a density with respect to Lebesgue measure, and consider a polygonal line f_n with vertices v_1^n, \dots, v_n^n obtained by minimizing under length constraint the criterion

$$F_n^0(x_1, \dots, x_n) = \mathbb{E} \left[\min_{1 \leq i \leq n} \|X - x_i\|^2 \right]. \quad (16)$$

For $h = (h_1, \dots, h_n) \in (\mathbb{R}^d)^n$, $\nabla F_n^0 \cdot h = \sum_{i=1}^n \mathbb{E} \left[-2 \langle X - \hat{X}_n, h_i \rangle \mathbf{1}_{\{\hat{X} = v_i^n\}} \right]$, where \hat{X} is such that $\|X - \hat{X}\| = \min_{1 \leq j \leq n} \|X - v_j^n\|$. For differentiability, it is convenient to write the length constraint as follows:

$$(n-1) \sum_{i=2}^n \|x_i - x_{i-1}\|^2 \leq L^2.$$

Let \hat{t}_n be defined by $\hat{t}_n = \frac{i-1}{n-1}$ on the event $\{\hat{X} = v_i^n\}$. For a test function g , set $h_i = g(\frac{i-1}{n-1})$ for $i = 1, \dots, n$. Then, we obtain the Euler-Lagrange equation

$$\mathbb{E} [\langle X - f_n(\hat{t}_n), g(\hat{t}_n) \rangle] = -\lambda_n \int_{[0,1]} \langle g(t), f_n''(dt) \rangle. \quad (17)$$

Up to an extraction, f_n converges uniformly to an optimal curve and \hat{t}_n converges in distribution. Using the default of self-consistency (3.2), it may be shown that every limit point of the sequence $(\lambda_n)_{n \geq 1}$ is positive. Together with the discrete Euler-Lagrange equation (17), this allows to prove that f_n'' converges weakly to a signed measure f'' . Finally, the desired Euler-Lagrange equation is obtained as the limit of (17).

Complete proof

Let us now start the complete proof of the theorem.

First, some notation is in order. Let Z be a standard d -dimensional Gaussian vector, independent of X . Let (ζ_n) , (η_n) and (ε_n) be sequences of positive real numbers such that

$$\zeta_n = \mathcal{O}(1/n), \quad \eta_n = \mathcal{O}(1/n), \quad n\varepsilon_n \rightarrow \infty, \quad \varepsilon_n \rightarrow 0.$$

We also introduce i.i.d. random vectors ξ_1^n, \dots, ξ_n^n , independent of X and Z , with same distribution as a centered random vector ξ with compactly supported density, such that $\|\xi\| \leq \eta_n$.

We will construct a sequence of polygonal lines converging to the optimal curve f by linking points v_1^n, \dots, v_n^n obtained by minimization of a criterion generalizing 16. Proving differentiability in this case is a little more involved in this case.

To begin with, since the random vector X is not assumed to have a density with respect to Lebesgue measure, we convolve it with a Gaussian random vector: we define, for $n \geq 1$, $X_n = X + \zeta_n Z$. So, X is approximated by a sequence $(X_n)_{n \geq 1}$ of continuous random variables.

For $1 \leq i \leq n$, let

$$t_i^n := \frac{i-1}{n-1}.$$

In order to be able to prove results which are true for any optimal curve f , we have to ensure that the points v_1^n, \dots, v_n^n , $n \geq 1$ are located on this curve f . To this aim, we add to the criterion 16 a penalty proportional to

$$\sum_{i=1}^n \|x_i - f(t_i^n)\|^2.$$

With this penalty, we cannot affirm any more that the x_i 's are pairwise distinct. To overcome this difficulty, a random vector ξ_i^n is added to each x_i : the points $x_i + \xi_i^n$, approximating the x_i 's, are almost surely pairwise distinct.

The desired chain of points is then defined, for $n \geq 1$, by minimizing in $x = (x_1, \dots, x_n) \in (\mathbb{R}^d)^n$ the criterion

$$F_n(x_1, \dots, x_n) = \mathbb{E} \left[\min_{1 \leq i \leq n} \|X_n - x_i - \xi_i^n\|^2 \right] + \varepsilon_n \sum_{i=1}^n \|x_i - f(t_i^n)\|^2, \quad (18)$$

under the constraint

$$(n-1) \sum_{i=2}^n \|x_i - x_{i-1}\|^2 \leq L^2. \quad (19)$$

Lemma 3.3. *There exists $(v_1^n, \dots, v_n^n) \in (\mathbb{R}^d)^n$, satisfying*

$$(n-1) \sum_{i=2}^n \|v_i^n - v_{i-1}^n\|^2 \leq L^2,$$

such that

$$F_n(v_1^n, \dots, v_n^n) = \min \left\{ F_n(x_1, \dots, x_n); (n-1) \sum_{i=2}^n \|x_i - x_{i-1}\|^2 \leq L^2 \right\}.$$

Let \hat{X}_n^x be such that $\hat{X}_n^x \in \{x_1 + \xi_1^n, \dots, x_n + \xi_n^n\}$ and

$$\|X_n - \hat{X}_n^x\| = \min_{1 \leq i \leq n} \|X_n - x_i - \xi_i^n\| \quad (20)$$

almost surely. In the sequel, \hat{X}_n will stand for $\hat{X}_n^{(v_1^n, \dots, v_n^n)}$.

Lemma 3.4.

$$\sup_{n \geq 1} F_n(v_1^n, \dots, v_n^n) < \infty.$$

We define the sequence $(f_n)_{n \geq 1}$ of polygonal lines approximating f , where each $f_n : [0, 1] \rightarrow \mathbb{R}^d$, $n \geq 1$, is given by

$$f_n(t) = v_i^n + (n-1)(t - t_i^n)(v_{i+1}^n - v_i^n), \quad t_i^n \leq t \leq t_{i+1}^n, \quad 1 \leq i \leq n-1.$$

This function f_n is absolutely continuous and we have $f_n'(t) = (n-1)(v_{i+1}^n - v_i^n)$ for $t \in (t_i^n, t_{i+1}^n)$. Using the definition of f_n' , we obtain the following regularity properties of f_n .

Lemma 3.5. *For $n \geq 1$, the curve f_n satisfies:*

1. $\mathcal{L}(f_n) \leq L$.
2. For all $t, t' \in [0, 1]$, $\|f_n(t) - f_n(t')\| \leq L\sqrt{|t - t'|}$.

Asymptotically, the penalty term ensuring that the points v_1^n, \dots, v_n^n , $n \geq 1$ belong to the curve f can be neglected.

Lemma 3.6. *There exists $c \geq 0$ such that, for all $n \geq 1$,*

$$\varepsilon_n \sum_{i=1}^n \|v_i^n - f(t_i^n)\|^2 \leq \frac{c}{n}.$$

Lemma 3.7. *The sequence $(f_n)_{n \geq 1}$ converges uniformly to the curve f .*

Let $\hat{t}_n = t_i^n$ on the event $\{\hat{X}_n = v_i^n + \xi_i^n\}$, $1 \leq i \leq n$. Note that the sequence $(\hat{t}_n)_{n \geq 1}$ is bounded. Thus, up to extending the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and extracting a subsequence, we may assume that (X_n, \hat{t}_n) converge in distribution to a tuple (X, \hat{t}) . This implies the next result.

Lemma 3.8. *There exists a random variable \hat{t} with values in $[0, 1]$, defined on an extension of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, such that*

$$\|X - f(\hat{t})\| = d(X, \text{Im}f) \quad \text{a.s.}$$

In order to be able to state a first order Euler-Lagrange equation for the criterion (18), we show that the quantity $\mathbb{E}[\min_{1 \leq i \leq n} \|X_n - x_i - \xi_i^n\|^2]$ is differentiable at x_i . Recall the definition (20) of \hat{X}_n^x .

Lemma 3.9. *The function $(x_1, \dots, x_n) \mapsto \mathbb{E}[\min_{1 \leq i \leq n} \|X_n - x_i - \xi_i^n\|^2]$ is differentiable, and, for $1 \leq i \leq n$, the gradient with respect to x_i is given by*

$$\frac{\partial}{\partial x_i} \mathbb{E} \left[\min_{1 \leq j \leq n} \|X_n - x_j - \xi_j^n\|^2 \right] = -2\mathbb{E} \left[(X_n - \hat{X}_n^x) \mathbf{1}_{\{\hat{X}_n^x = x_i + \xi_i\}} \right].$$

The Lagrange multiplier method then leads to the next system of equations satisfied by v_1^n, \dots, v_n^n .

Lemma 3.10. *For $n \geq 1$, there exists a Lagrange multiplier $\lambda_n \geq 0$ such that*

$$\begin{cases} -\mathbb{E} \left[(X_n - \hat{X}_n) \mathbf{1}_{\{\hat{X}_n = v_i^n + \xi_i^n\}} \right] + \varepsilon_n (v_i^n - f(t_i^n)) + \lambda_n (n-1) (v_i^n - v_{i-1}^n - (v_{i+1}^n - v_i^n)) = 0, \\ \hspace{25em} 2 \leq i \leq n-1, \\ -\mathbb{E} \left[(X_n - \hat{X}_n) \mathbf{1}_{\{\hat{X}_n = v_1^n + \xi_1^n\}} \right] + \varepsilon_n (v_1^n - f(0)) - \lambda_n (n-1) (v_2^n - v_1^n) = 0, \\ -\mathbb{E} \left[(X_n - \hat{X}_n) \mathbf{1}_{\{\hat{X}_n = v_n^n + \xi_n^n\}} \right] + \varepsilon_n (v_n^n - f(1)) + \lambda_n (n-1) (v_n^n - v_{n-1}^n) = 0. \end{cases}$$

Lemma 3.11. *If λ is a limit point of the sequence $(\lambda_n)_{n \geq 1}$, then $\lambda \in (0, \infty]$.*

Hence, up to an extraction, we may assume that the sequence $(\lambda_n)_{n \geq 1}$ converges to a limit $\lambda \in (0, \infty]$.

Let δ_ℓ denote the Dirac mass at ℓ . For every $n \geq 2$, we define f_n'' on $[0, 1]$ by

$$f_n'' = (n-1) \left[\sum_{i=2}^{n-1} (v_{i+1}^n - v_i^n - (v_i^n - v_{i-1}^n)) \delta_{t_i^n} + (v_2^n - v_1^n) \delta_0 - (v_n^n - v_{n-1}^n) \delta_1 \right], \quad (21)$$

which is a vector-valued signed measure.

Lemma 3.12. *The sequence $(f_n'')_{n \geq 1}$ converges weakly to a signed measure f'' on $[0, 1]$, with values in \mathbb{R}^d , which is the second derivative of f . The following regularity properties hold:*

- f is right-differentiable on $[0, 1)$, $\|f_r'(t)\| = L$ for all $t \in [0, 1)$,
- f is left-differentiable on $(0, 1]$, $\|f_\ell'(t)\| = L$ for all $t \in (0, 1]$,
- $f''((s, t]) = f_r'(t) - f_r'(s)$ for all $0 \leq s \leq t < 1$,
- $f''([0, 1]) = 0$.
- $f''(\{0\}) = f_r'(0)$,
- $f''(\{1\}) = -f_\ell'(1)$.

These properties imply in particular that λ is finite.

Lemma 3.13. *We have $\lambda < \infty$.*

Finally, collecting all the results allows to derive the Euler-Lagrange equation.

Lemma 3.14. *For every bounded Borel function $g : [0, 1] \rightarrow \mathbb{R}^d$,*

$$\mathbb{E} [\langle X - f(\hat{t}), g(\hat{t}) \rangle] = -\lambda \int_{[0, 1]} \langle g(t), f''(dt) \rangle.$$

Moreover, λ depends only on the curve f .

Proof of Lemma 3.3. Since $\mathbb{E}[\|X_n\|^2] \leq 2\mathbb{E}[\|X\|^2] + 2d\zeta_n^2 < \infty$ and $\mathbb{E}[\|\xi\|^2] \leq \eta_n^2 < \infty$, F_n takes its values in $[0, \infty)$ and is continuous. The constraint (19) defines a nonempty closed set D_n . Since

$$\lim_{\substack{\|x_1\| + \dots + \|x_n\| \rightarrow \infty \\ (x_1, \dots, x_n) \in D_n}} F_n(x_1, \dots, x_n) = \infty,$$

the optimization problem reduces thus to minimizing a continuous function on a compact set. \square

Proof of Lemma 3.4. Recall that, for all $t, t' \in [0, 1]$, $\|f(t) - f(t')\| \leq L|t - t'|$. Hence, we have

$$(n-1) \sum_{i=2}^n \|f(t_i^n) - f(t_{i-1}^n)\|^2 \leq L^2,$$

and consequently, we may consider $(x_1, \dots, x_n) = (f(t_1^n), \dots, f(t_n^n))$. We see that

$$\begin{aligned} F_n(v_1^n, \dots, v_n^n) &\leq \mathbb{E} [\|X_n - f(0) - \xi_1^n\|^2] \\ &\leq 2\mathbb{E} [\|X_n - \xi_1^n\|^2] + 2\|f(0)\|^2 \\ &\leq 2\mathbb{E} [\|X\|^2] + 2d\zeta_n^2 + 2\eta_n^2 + 2\|f(0)\|^2. \end{aligned}$$

□

Proof of Lemma 3.5. By definition of f'_n , and using that v_1^n, \dots, v_n^n satisfy constraint (19), we have

$$\int_0^1 \|f'_n(t)\|^2 dt = \sum_{i=1}^{n-1} (n-1)^2 \|v_{i+1}^n - v_i^n\|^2 \times \frac{1}{n-1} = (n-1) \sum_{i=1}^{n-1} \|v_{i+1}^n - v_i^n\|^2 \leq L^2.$$

Hence,

$$\mathcal{L}(f_n) \leq \left(\int_0^1 \|f'_n(t)\|^2 dt \right)^{1/2} \leq L,$$

and for all $t, t' \in [0, 1]$,

$$\|f_n(t) - f_n(t')\| = \left\| \int_0^1 \mathbf{1}_{[t \wedge t', t \vee t']} f'_n(u) du \right\| \leq L\sqrt{|t - t'|}.$$

□

Proof of Lemma 3.6. The aim is to show that there exists $c \geq 0$ such that, for all $n \geq 1$,

$$\varepsilon_n \sum_{i=1}^n \|v_i^n - f(t_i^n)\|^2 \leq \frac{c}{n}.$$

The following upper bound will be useful:

$$\left| \min_{1 \leq i \leq n} \|X_n - f(t_i^n) - \xi_i^n\| - \min_{1 \leq i \leq n} \|X - f(t_i^n)\| \right| \leq \zeta_n \|Z\| + \eta_n.$$

By definition of (v_1^n, \dots, v_n^n) , since $(f(t_1^n), \dots, f(t_n^n))$ satisfies constraint 19 as already mentioned in the proof of Lemma 3.4, we may write

$$F_n(v_1^n, \dots, v_n^n) \leq \mathbb{E} \left[\min_{1 \leq i \leq n} \|X_n - f(t_i^n) - \xi_i^n\|^2 \right].$$

Observe that

$$\begin{aligned} &\left| \min_{1 \leq i \leq n} \|X_n - f(t_i^n) - \xi_i^n\| - \min_{t \in [0, 1]} \|X - f(t)\| \right| \\ &\leq \left| \min_{1 \leq i \leq n} \|X_n - f(t_i^n) - \xi_i^n\| - \min_{1 \leq i \leq n} \|X - f(t_i^n)\| \right| + \left| \min_{1 \leq i \leq n} \|X - f(t_i^n)\| - \min_{t \in [0, 1]} \|X - f(t)\| \right| \\ &\leq \zeta_n \|Z\| + \eta_n + \frac{L}{n-1}, \end{aligned}$$

so that

$$\begin{aligned} \min_{1 \leq i \leq n} \|X_n - f(t_i^n) - \xi_i^n\|^2 &\leq \min_{t \in [0, 1]} \|X - f(t)\|^2 + \left(\eta_n + \zeta_n \|Z\| + \frac{L}{n-1} \right)^2 \\ &\quad + 2 \left(\eta_n + \zeta_n \|Z\| + \frac{L}{n-1} \right) \min_{t \in [0, 1]} \|X - f(t)\|. \end{aligned}$$

Consequently, there exists $c_1 \geq 0$, such that

$$F_n(v_1^n, \dots, v_n^n) \leq G(L) + \frac{c_1}{n}.$$

Besides,

$$F_n(v_1^n, \dots, v_n^n) = \mathbb{E} \left[\min_{1 \leq i \leq n} \|X_n - f_n(t_i^n) - \xi_i^n\|^2 \right] + \varepsilon_n \sum_{i=1}^n \|f_n(t_i^n) - f(t_i^n)\|^2,$$

and, writing

$$\begin{aligned} & \left| \min_{1 \leq i \leq n} \|X_n - f_n(t_i^n) - \xi_i^n\|^2 - \min_{1 \leq i \leq n} \|X - f_n(t_i^n)\|^2 \right| \\ & \leq \left| \min_{1 \leq i \leq n} \|X_n - f_n(t_i^n) - \xi_i^n\| - \min_{1 \leq i \leq n} \|X - f_n(t_i^n)\| \right| \times \left(\min_{1 \leq i \leq n} \|X_n - f_n(t_i^n) - \xi_i^n\| + \min_{1 \leq i \leq n} \|X - f_n(t_i^n)\| \right) \\ & \leq \left(\zeta_n \|Z\| + \eta_n \right) \left(\zeta_n \|Z\| + \eta_n + 2 \min_{1 \leq i \leq n} \|X - f_n(t_i^n)\| \right) \\ & = \left(\zeta_n \|Z\| + \eta_n \right)^2 + 2 \left(\zeta_n \|Z\| + \eta_n \right) \min_{1 \leq i \leq n} \|X - f_n(t_i^n)\|, \end{aligned}$$

we obtain

$$\begin{aligned} F_n(v_1^n, \dots, v_n^n) & \geq \mathbb{E} \left[\min_{1 \leq i \leq n} \|X - f_n(t_i^n)\|^2 \right] - \mathbb{E} [(\zeta_n \|Z\| + \eta_n)^2] - 2(\zeta_n \mathbb{E}[\|Z\|] + \eta_n) \mathbb{E} \left[\min_{1 \leq i \leq n} \|X - f_n(t_i^n)\| \right] \\ & \quad + \varepsilon_n \sum_{i=1}^n \|f_n(t_i^n) - f(t_i^n)\|^2 \\ & \geq \mathbb{E} \left[\min_{t \in [0,1]} \|X - f_n(t)\|^2 \right] - \zeta_n^2 \mathbb{E}[\|Z\|^2] - \eta_n^2 - 2\eta_n \zeta_n \mathbb{E}[\|Z\|] \\ & \quad - 2(\zeta_n \mathbb{E}[\|Z\|] + \eta_n) \mathbb{E} \left[\min_{1 \leq i \leq n} \|X - f_n(t_i^n)\| \right] + \varepsilon_n \sum_{i=1}^n \|f_n(t_i^n) - f(t_i^n)\|^2 \\ & \geq G(L) - \frac{c_2}{n} + \varepsilon_n \sum_{i=1}^n \|f_n(t_i^n) - f(t_i^n)\|^2, \end{aligned}$$

for some constant $c_2 \geq 0$. Indeed, $\mathcal{L}(f_n) \leq L$ according to point 1 in Lemma 3.5, which allows to lower bound $\mathbb{E} \left[\min_{t \in [0,1]} \|X - f_n(t)\|^2 \right]$ by $G(L)$, and moreover, $\mathbb{E} [\min_{1 \leq i \leq n} \|X - f_n(t_i^n)\|]$ is bounded since $(f_n)_{n \geq 1}$ is uniformly bounded and $\mathbb{E}[\|X\|^2] < \infty$. Thus, there exists a constant c_3 such that $G(L) - \frac{c_3}{n} + \varepsilon_n \sum_{i=1}^n \|f_n(t_i^n) - f(t_i^n)\|^2 \leq G(L) + \frac{c_3}{n}$, which shows that $\varepsilon_n \sum_{i=1}^n \|f_n(t_i^n) - f(t_i^n)\|^2 \leq \frac{2c_3}{n}$. \square

Proof of Lemma 3.7. Point 2 in Lemma 3.5 and Lemma 3.6, together with the assumption $n\varepsilon_n \rightarrow \infty$, imply that the sequence $(f_n)_{n \geq 1}$ converges uniformly to the curve f . \square

Proof of Lemma 3.8. For every $n \geq 1$,

$$\begin{aligned} & \left| \|X_n - f_n(\hat{t}_n)\| - \min_{1 \leq i \leq n} \|X_n - f_n(t_i^n)\| \right| \\ & \leq \left| \|X_n - f_n(\hat{t}_n)\| - \min_{1 \leq i \leq n} \|X_n - f_n(t_i^n) - \xi_i^n\| \right| + \left| \min_{1 \leq i \leq n} \|X_n - f_n(t_i^n) - \xi_i^n\| - \min_{1 \leq i \leq n} \|X_n - f_n(t_i^n)\| \right| \\ & \leq \left| \|X_n - f_n(\hat{t}_n)\| - \sum_{i=1}^n \|X_n - f_n(\hat{t}_n) - \xi_i^n\| \mathbf{1}_{\{\hat{X}_n = f_n(t_i^n) + \xi_i^n\}} \right| + \eta_n \\ & \leq \sum_{i=1}^n \left| \|X_n - f_n(\hat{t}_n)\| - \|X_n - f_n(\hat{t}_n) - \xi_i^n\| \mathbf{1}_{\{\hat{X}_n = f_n(t_i^n) + \xi_i^n\}} \right| + \eta_n \\ & \leq 2\eta_n, \end{aligned}$$

Hence, we obtain

$$\|X - f(\hat{t})\| = \min_{t \in [0,1]} \|X - f(t)\| \quad a.s.$$

□

Proof of Lemma 3.9. For $x = (x_1, \dots, x_n) \in (\mathbb{R}^d)^n$ and $\omega \in \Omega$, we set

$$G_n(x, \omega) := \min_{1 \leq i \leq n} \|X_n(\omega) - x_i - \xi_i^n(\omega)\|^2.$$

For every x , since the distribution of X_n gives zero measure to affine hyperplanes of \mathbb{R}^d and the vectors $x_i + \xi_i^n$, $1 \leq i \leq n$, are mutually distinct $\mathbb{P}(d\omega)$ almost surely, we have $\mathbb{P}(d\omega)$ almost surely,

$$G_n(x, \omega) = \sum_{i=1}^n \|X_n(\omega) - x_i - \xi_i^n(\omega)\|^2 \mathbf{1}_{\{\|X_n(\omega) - x_i - \xi_i^n(\omega)\| < \min_{j \neq i} \|X_n(\omega) - x_j - \xi_j^n(\omega)\|\}}.$$

For every $x \in (\mathbb{R}^d)^n$, $\mathbb{P}(d\omega)$ almost surely, $y \mapsto G_n(y, \omega)$ is differentiable at x and for $1 \leq i \leq n$,

$$\begin{aligned} \frac{\partial}{\partial x_i} G_n(x, \omega) &= -2(X_n(\omega) - x_i - \xi_i^n(\omega)) \mathbf{1}_{\{\|X_n(\omega) - x_i - \xi_i^n(\omega)\| < \min_{j \neq i} \|X_n(\omega) - x_j - \xi_j^n(\omega)\|\}} \\ &= -2(X_n(\omega) - \hat{X}_n^x(\omega)) \mathbf{1}_{\{\hat{X}_n^x(\omega) = x_i + \xi_i^n(\omega)\}}. \end{aligned}$$

For every $u = (u_1, \dots, u_n) \in (\mathbb{R}^d)^n$, we set $\|u\| = (\sum_{i=1}^n \|u_i\|^2)^{1/2}$. Let $x^{(k)} = (x_1^{(k)}, \dots, x_n^{(k)})$ be a sequence tending to $x = (x_1, \dots, x_n) \in (\mathbb{R}^d)^n$ as k tends to infinity. Then,

$$\left[G_n(x^{(k)}, \omega) - G_n(x, \omega) - \sum_{i=1}^n \left\langle \frac{\partial}{\partial x_i} G_n(x, \omega), x_i^{(k)} - x_i \right\rangle \right] \times \frac{1}{\|x^{(k)} - x\|}$$

converges $\mathbb{P}(d\omega)$ almost surely to 0 as k tends to infinity. Moreover,

$$\begin{aligned} & \left| G_n(x, \cdot) - G_n(x^{(k)}, \cdot) \right| \\ &= \left(\min_{1 \leq i \leq n} \|X_n - x_i - \xi_i^n\| + \min_{1 \leq i \leq n} \|X_n - x_i^{(k)} - \xi_i^n\| \right) \left| \min_{1 \leq i \leq n} \|X_n - x_i - \xi_i^n\| - \min_{1 \leq i \leq n} \|X_n - x_i^{(k)} - \xi_i^n\| \right| \\ &\leq 2 \left(\|X_n\| + \eta_n + \|x_1\| + \|x_1^{(k)}\| \right) \max_{1 \leq i \leq n} \|x_i - x_i^{(k)}\|, \end{aligned}$$

so that

$$\frac{|G_n(x, \cdot) - G_n(x^{(k)}, \cdot)|}{\|x - x^{(k)}\|} \leq C(\|X_n\| + 1),$$

where C is a constant which does not depend on k . Similarly, we have, for $1 \leq i \leq n$,

$$\left\| \frac{\partial}{\partial x_i} G_n(x, \cdot) \right\| \leq C'(\|X_n\| + 1),$$

where C' does not depend on k , and, thus,

$$\begin{aligned} \frac{1}{\|x - x^{(k)}\|} \left| \sum_{i=1}^n \left\langle \frac{\partial}{\partial x_i} G_n(x, \cdot), x_i^{(k)} - x_i \right\rangle \right| &\leq C'(\|X_n\| + 1) \frac{\sum_{i=1}^n \|x_i^{(k)} - x_i\|}{\|x - x^{(k)}\|} \\ &\leq C' \sqrt{n}(\|X_n\| + 1). \end{aligned}$$

Since $\mathbb{E}[\|X_n\|] < \infty$, the result follows from Lebesgue's dominated convergence theorem. □

Proof of Lemma 3.10. By Lemma 3.9, we obtain that F_n is differentiable, and for $1 \leq i \leq n$, the gradient with respect to x_i is given by

$$\frac{\partial}{\partial x_i} F_n(x_1, \dots, x_n) = -2\mathbb{E} \left[(X_n - \hat{X}_n^x) \mathbf{1}_{\{\hat{X}_n^x = x_i + \xi_i^n\}} \right] + 2\varepsilon_n(x_i - f(t_i^n)), \quad 1 \leq i \leq n.$$

Consequently, considering the minimization of F_n under the length constraint (19), there exists a Lagrange multiplier $\lambda_n \geq 0$ such that

$$\begin{cases} -2\mathbb{E} \left[(X_n - \hat{X}_n) \mathbf{1}_{\{\hat{X}_n = v_i^n + \xi_i^n\}} \right] + 2\varepsilon_n (v_i^n - f(t_i^n)) + 2\lambda_n (n-1) (v_i^n - v_{i-1}^n - (v_{i+1}^n - v_i^n)) = 0, \\ \hspace{25em} 2 \leq i \leq n-1, \\ -2\mathbb{E} \left[(X_n - \hat{X}_n) \mathbf{1}_{\{\hat{X}_n = v_1^n + \xi_1^n\}} \right] + 2\varepsilon_n (v_1^n - f(0)) - 2\lambda_n (n-1) (v_2^n - v_1^n) = 0, \\ -2\mathbb{E} \left[(X_n - \hat{X}_n) \mathbf{1}_{\{\hat{X}_n = v_n^n + \xi_n^n\}} \right] + 2\varepsilon_n (v_n^n - f(1)) + 2\lambda_n (n-1) (v_n^n - v_{n-1}^n) = 0, \end{cases}$$

that is,

$$\begin{cases} -\mathbb{E} \left[(X_n - \hat{X}_n) \mathbf{1}_{\{\hat{X}_n = v_i^n + \xi_i^n\}} \right] + \varepsilon_n (v_i^n - f(t_i^n)) + \lambda_n (n-1) (v_i^n - v_{i-1}^n - (v_{i+1}^n - v_i^n)) = 0, \\ \hspace{25em} 2 \leq i \leq n-1, \\ -\mathbb{E} \left[(X_n - \hat{X}_n) \mathbf{1}_{\{\hat{X}_n = v_1^n + \xi_1^n\}} \right] + \varepsilon_n (v_1^n - f(0)) - \lambda_n (n-1) (v_2^n - v_1^n) = 0, \\ -\mathbb{E} \left[(X_n - \hat{X}_n) \mathbf{1}_{\{\hat{X}_n = v_n^n + \xi_n^n\}} \right] + \varepsilon_n (v_n^n - f(1)) + \lambda_n (n-1) (v_n^n - v_{n-1}^n) = 0. \end{cases}$$

□

Proof of Lemma 3.11. Let $g : [0, 1] \rightarrow \mathbb{R}^d$ be an absolutely continuous function such that $\int_0^1 \|g'(t)\|^2 dt < \infty$. For $n \geq 1$, we may write

$$\begin{aligned} & \mathbb{E}[\langle X_n - f_n(\hat{t}_n), g(\hat{t}_n) \rangle] \\ &= \sum_{i=1}^n \left\langle \mathbb{E} \left[(X_n - \hat{X}_n + \xi_i^n) \mathbf{1}_{\{\hat{X}_n = v_i^n + \xi_i^n\}} \right], g(t_i^n) \right\rangle \\ &= \sum_{i=1}^n \left\langle \mathbb{E} \left[(X_n - \hat{X}_n) \mathbf{1}_{\{\hat{X}_n = v_i^n + \xi_i^n\}} \right], g(t_i^n) \right\rangle + \sum_{i=1}^n \left\langle \mathbb{E} \left[\xi_i^n \mathbf{1}_{\{\hat{X}_n = v_i^n + \xi_i^n\}} \right], g(t_i^n) \right\rangle \\ &= \sum_{i=1}^n \left\langle \mathbb{E} \left[\xi_i^n \mathbf{1}_{\{\hat{X}_n = v_i^n + \xi_i^n\}} \right], g(t_i^n) \right\rangle + \varepsilon_n \sum_{i=1}^n \langle v_i^n - f(t_i^n), g(t_i^n) \rangle \\ &\quad + \lambda_n (n-1) \left[-\langle v_2^n - v_1^n, g(0) \rangle + \sum_{i=2}^{n-1} \langle v_i^n - v_{i-1}^n - (v_{i+1}^n - v_i^n), g(t_i^n) \rangle + \langle v_n^n - v_{n-1}^n, g(1) \rangle \right] \\ &= \sum_{i=1}^n \left\langle \mathbb{E} \left[\xi_i^n \mathbf{1}_{\{\hat{X}_n = v_i^n + \xi_i^n\}} \right], g(t_i^n) \right\rangle + \varepsilon_n \sum_{i=1}^n \langle v_i^n - f(t_i^n), g(t_i^n) \rangle \\ &\quad + \lambda_n (n-1) \left[\sum_{i=1}^{n-2} \langle v_{i+1}^n - v_i^n, g(t_{i+1}^n) \rangle - \sum_{i=2}^{n-1} \langle v_{i+1}^n - v_i^n, g(t_i^n) \rangle - \langle v_2^n - v_1^n, g(0) \rangle \right. \\ &\quad \left. + \langle v_n^n - v_{n-1}^n, g(1) \rangle \right] \\ &= \sum_{i=1}^n \left\langle \mathbb{E} \left[\xi_i^n \mathbf{1}_{\{\hat{X}_n = v_i^n + \xi_i^n\}} \right], g(t_i^n) \right\rangle \\ &\quad + \varepsilon_n \sum_{i=1}^n \langle v_i^n - f(t_i^n), g(t_i^n) \rangle + \lambda_n (n-1) \sum_{i=1}^{n-1} \langle v_{i+1}^n - v_i^n, g(t_{i+1}^n) - g(t_i^n) \rangle. \end{aligned} \tag{22}$$

Note first that

$$\begin{aligned} \left| \sum_{i=1}^n \left\langle \mathbb{E} \left[\xi_i^n \mathbf{1}_{\{\hat{X}_n = v_i^n + \xi_i^n\}} \right], g(t_i^n) \right\rangle \right| &\leq \eta_n \|g\|_\infty \sum_{i=1}^n \mathbb{E} \left[\mathbf{1}_{\{\hat{X}_n = v_i^n + \xi_i^n\}} \right] \\ &= \eta_n \|g\|_\infty. \end{aligned} \tag{23}$$

Then,

$$\begin{aligned}
\left| \varepsilon_n \sum_{i=1}^n \langle v_i^n - f(t_i^n), g(t_i^n) \rangle \right| &\leq \varepsilon_n \sum_{i=1}^n \|v_i^n - f(t_i^n)\| \|g\|_\infty \\
&\leq \varepsilon_n \left(\sum_{i=1}^n \|v_i^n - f(t_i^n)\|^2 \right)^{1/2} \sqrt{n} \|g\|_\infty \\
&\leq \sqrt{c\varepsilon_n} \|g\|_\infty,
\end{aligned} \tag{24}$$

according to Lemma 3.6. Regarding the last term, we may write

$$\begin{aligned}
\left| (n-1) \sum_{i=1}^{n-1} \langle v_{i+1}^n - v_i^n, g(t_{i+1}^n) - g(t_i^n) \rangle \right| &\leq (n-1) \left[\sum_{i=1}^{n-1} \|v_{i+1}^n - v_i^n\|^2 \sum_{i=1}^{n-1} \|g(t_{i+1}^n) - g(t_i^n)\|^2 \right]^{1/2} \\
&\leq L\sqrt{n-1} \left[\sum_{i=1}^{n-1} \left\| \int_{t_i^n}^{t_{i+1}^n} g'(t) dt \right\|^2 \right]^{1/2} \\
&\leq L \left[\int_0^1 \|g'(t)\|^2 dt \right]^{1/2}.
\end{aligned}$$

Thus, if $h : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is continuously differentiable, we have

$$\begin{aligned}
|\mathbb{E}[\langle X_n - f_n(\hat{t}_n), h(f_n(\hat{t}_n)) \rangle]| &\leq \sqrt{c\varepsilon_n} \|h\|_\infty + \lambda_n L \left[\int_0^1 \|\nabla h(f_n(t)), f_n'(t)\|^2 dt \right]^{1/2} \\
&\leq \sqrt{c\varepsilon_n} \|h\|_\infty + \lambda_n L \sup_{t \in [0,1]} \|\nabla h(f_n(t))\| \left[\int_0^1 \|f_n'(t)\|^2 dt \right]^{1/2} \\
&\leq \sqrt{c\varepsilon_n} \|h\|_\infty + \lambda_n L^2 \sup_{t \in [0,1]} \|\nabla h(f_n(t))\|.
\end{aligned}$$

Recall that (X_n, \hat{t}_n) is assumed to converge to (X, \hat{t}) . Since $\varepsilon_n \rightarrow 0$ and $(f_n)_{n \geq 1}$ is uniformly bounded, we see that $\lambda = 0$ would imply that

$$\mathbb{E}[\langle X - f(\hat{t}), h(f(\hat{t})) \rangle] = 0,$$

so that $\mathbb{E}[X - f(\hat{t})|f(\hat{t})] = 0$ a.s. by density of continuously differentiable functions since h is an arbitrary such function. This contradicts Lemma 3.2. \square

Proof of Lemma 3.12. For an \mathbb{R}^d -valued signed measure $m = (m^1, \dots, m^d)$ on $[0, 1]$, we set

$$\|m\| = \left(\sum_{j=1}^d \|m^j\|_{TV}^2 \right)^{1/2} \tag{25}$$

where $\|m^j\|_{TV}$ denotes the total variation norm of m^j . Recall that

$$f_n'' = (n-1) \left[\sum_{i=2}^{n-1} (v_{i+1}^n - v_i^n - (v_i^n - v_{i-1}^n)) \delta_{t_i^n} + (v_2^n - v_1^n) \delta_0 - (v_n^n - v_{n-1}^n) \delta_1 \right].$$

Thanks to the Euler-Lagrange system of equations obtained in Lemma 3.10, we may write

$$\begin{aligned}
\lambda_n \times \|f_n''\| &\leq \lambda_n \sum_{i=1}^n \|f_n''(\{t_i^n\})\| \\
&\leq \sum_{i=1}^n \left\| \mathbb{E} \left[(X_n - \hat{X}_n) \mathbf{1}_{\{\hat{X}_n = v_i^n + \xi_i^n\}} \right] \right\| + \varepsilon_n \sum_{i=1}^n \|v_i^n - f(t_i^n)\| \\
&\leq \mathbb{E}[\|X_n - \hat{X}_n\|] + \varepsilon_n \sqrt{n} \left(\sum_{i=1}^n \|v_i^n - f(t_i^n)\|^2 \right)^{1/2} \\
&\leq F_n(v_1^n, \dots, v_n^n)^{1/2} + \varepsilon_n \sqrt{n} \left(\sum_{i=1}^n \|v_i^n - f(t_i^n)\|^2 \right)^{1/2}.
\end{aligned}$$

Consequently, using Lemma 3.4 and Lemma 3.6, $\varepsilon_n \rightarrow 0$ and $\lim_{n \rightarrow \infty} \lambda_n = \lambda \in (0, \infty]$, we obtain that $\sup_{n \geq 1} \|f_n''\| < \infty$, that is, the sequence of signed measures $(f_n'')_{n \geq 1}$ is uniformly bounded in total variation norm. Hence, it is relatively compact for the topology induced by the bounded Lipschitz norm defined for every signed measure m by

$$\|m\|_{\text{BL}} = \sup \left\{ \left\| \int g(x) m(dx) \right\|, \|g\|_{\infty} \leq 1, \sup_{t \neq x} \frac{|g(x) - g(t)|}{|x - t|} \leq 1 \right\}.$$

Let us show that the sequence $(f_n'')_{n \geq 1}$ converges weakly to some signed measure. Let ν be a limit point of $(f_n'')_{n \geq 1}$. There exists an increasing function $\sigma : \mathbb{N} \rightarrow \mathbb{N}$, such that, for every (s, t) such that $\nu(\{s\}) = \nu(\{t\}) = 0$,

$$f_{\sigma(n)}''((s, t]) \rightarrow \nu((s, t]), \quad (26)$$

$$f_{\sigma(n)}''([0, t]) \rightarrow \nu([0, t]), \quad f_{\sigma(n)}''([0, t)) \rightarrow \nu([0, t)). \quad (27)$$

Since, for $0 \leq s \leq t < 1$, $f_n''((s, t]) = f_{n,r}'(t) - f_{n,r}'(s)$, we have, for $0 \leq t < 1$,

$$f_n(t) = f_n(0) + t f_{n,r}'(0) + \int_0^t f_n''((0, u]) du.$$

Note that $f_{n,r}'(0) = f_n''(\{0\})$, so that the fact that $\sup_{n \geq 1} \|f_n''\| < \infty$ implies in particular that $(f_{n,r}'(0))_{n \geq 1}$ is bounded. Thus, up to an extraction, by (26), all terms converge: there exists a vector $v \in \mathbb{R}^d$, such that, for $0 \leq t < 1$,

$$f(t) = f(0) + tv + \int_0^t \nu((0, u]) du.$$

Consequently, $v = f_r'(0)$, and, for $0 \leq s \leq t < 1$,

$$\nu((s, t]) = f_r'(t) - f_r'(s).$$

In other words, the signed measure ν is the second derivative of f in the distribution sense, called hereafter f'' , and $(f_n'')_{n \geq 1}$ converges weakly to f'' .

Observe, on the definition (21), that $f_n''([0, 1]) = 0$, so that $f''([0, 1]) = 0$.

We have, for t such that $f''(\{t\}) = 0$,

$$f_n''([0, t]) \rightarrow f''([0, t]), \quad f_n''([0, t)) \rightarrow f''([0, t)).$$

Hence, since, for $t \in [0, 1]$, $f_n''([0, t]) = f_{n,r}'(t)$, and $t \mapsto f''([0, t])$ is right-continuous, $f_r'(t) = f''([0, t])$ for $t \in [0, 1]$. Similarly, for $t \in (0, 1]$, $f_n''([0, t)) = f_{n,\ell}'(t)$, and $t \mapsto f''([0, t))$ is left-continuous, so that $f_{\ell}'(t) = f''([0, t))$ for $t \in (0, 1]$.

Recall that f is L -Lipschitz. Moreover, according to Corollary 3.1, $\mathcal{L}(f) = L$ since $G(L) > 0$. Thus, we have $\|f_r'(t)\| = L$ dt-a.e., and, since f_r' is right-continuous, this implies that $\|f_r'(t)\| = L$ for all $t \in [0, 1]$. Similarly, we obtain that $\|f_{\ell}'(t)\| = L$ for all $t \in (0, 1]$. \square

Proof of Lemma 3.13. Observe that $f'' \neq 0$. Indeed, we have, for example, $f''(\{0\}) = f'_r(0)$, with $\|f'_r(0)\| = L > 0$. Yet, $\lambda = \infty$ would imply $f'' = 0$ since $\sup_{n \geq 1} (\lambda_n \times \|f''_n\|) < \infty$. \square

Proof of Lemma 3.14. Clearly, it suffices to consider the case where the test function g is continuous. Using equation (22) and the upper bounds (23) and (24) in the proof of Lemma 3.11, we obtain, for $n \geq 2$,

$$\left| \mathbb{E}[\langle X_n - f_n(\hat{t}_n), g(\hat{t}_n) \rangle] - \lambda_n(n-1) \sum_{i=1}^{n-1} \langle v_{i+1}^n - v_i^n, g(t_{i+1}^n) - g(t_i^n) \rangle \right| \leq (\eta_n + c\sqrt{\varepsilon_n}) \|g\|_\infty,$$

and besides

$$\lambda_n(n-1) \sum_{i=1}^{n-1} \langle v_{i+1}^n - v_i^n, g(t_{i+1}^n) - g(t_i^n) \rangle = -\lambda_n \int_{[0,1]} \langle g(t), f''_n(dt) \rangle.$$

Thus, passing to the limit, we see that f satisfies equation (8).

Finally, the uniqueness of λ follows from the uniqueness of \hat{X} (Proposition 3.1), and the fact that

$$\mathbb{E}[\langle X - \hat{X}, \hat{X} \rangle] = \lambda \int_0^1 \|f'_r(s)\|^2 ds = \lambda L^2$$

obtained thanks to equation (9) in Remark 5. \square

4 An application: injectivity of f

In this section, we present an application of the formula (8) of Theorem 3.1. We will use this first order condition to show in dimension $d = 2$ that an open optimal curve is injective, and a closed optimal curve restricted to $[0, 1]$ is injective, except in the case where its image is a segment. To obtain the result, we follow arguments exposed in Lu and Slepčev (2016) in the frame of the penalized problem, for open curves. The main difference is the fact that we have at hand the Euler-Lagrange equation, which allows to simplify the proof.

Again, we consider $L > 0$ such that $G(L) > 0$ and a curve $f \in \mathcal{C}_L$ such that $\Delta(f) = G(L)$, which is L -Lipschitz. We let \hat{t} be defined as in Theorem 3.1. The random vector $f(\hat{t})$ will sometimes be denoted by \hat{X} . Recall that $\|X - \hat{X}\| = d(X, \text{Im} f)$ a.s. by Theorem 3.1.

To prove the injectivity of f , we will need several preliminary lemmas. Let us point out that Lemma 4.1 to Lemma 4.5 below are valid for every $d \geq 1$.

First of all, we state the next lemma, which will be useful in the sequel, providing a lower bound on the curvature of any closed arc of f . Recall that the total variation of a signed measure ν is defined by

$$\|\nu\| = \left(\sum_{j=1}^d \|\nu^j\|_{TV}^2 \right)^{1/2},$$

where $\|\nu^j\|_{TV}$ denotes the total variation norm of ν^j . For a Borel set $A \subset [0, 1]$, f''_A denotes the vector-valued signed measure defined by $f''_A(B) = f''(A \cap B)$ for all Borel set $B \subset [0, 1]$.

Lemma 4.1. *If $0 \leq a < b \leq 1$ and $f(a) = f(b)$, then $\|f''_{(a,b]}\| \geq L$.*

Proof of Lemma 4.1. Let us write

$$\begin{aligned} 0 &= f(b) - f(a) = \int_a^b f'_r(t) dt \\ &= \int_a^b \left[f'_r(0) + \int_{(0,t]} f''(ds) \right] dt = (b-a)f'_r(0) + \int_{(0,b]} (b - (s \vee a)) f''(ds) \\ &= (b-a)f'_r(0) + (b-a)f''((0, a]) + \int_{(a,b]} (b-s) f''(ds) \\ &= (b-a)f'_r(a) + \int_{(a,b]} (b-s) f''(ds). \end{aligned}$$

Thus, $\int_{(a,b)} \frac{b-s}{b-a} f''(ds) = -f'_r(a)$, which implies $\|f''_{(a,b)}\| \geq \|f'_r(a)\| = L$. \square

As a first step toward injectivity, we now show that, if a point is multiple, it is only visited finitely many times.

Lemma 4.2. *For every $t \in [0, 1]$, the set $f^{-1}(\{f(t)\})$ is finite.*

Proof. Let $t \in [0, 1]$. Suppose that $f^{-1}(\{f(t)\})$ is infinite. Then, for all $k \geq 1$, there exist $t_0, t_1, \dots, t_k \in f^{-1}(\{f(t)\})$ such that $0 \leq t_0 < t_1 < \dots < t_k \leq 1$. So, by Lemma 4.1, $\|f''\| \geq \sum_{i=1}^k \|f''_{(t_{i-1}, t_i)}\| \geq kL$, which contradicts the fact that f has finite curvature. \square

In the case $\mathcal{C}_L = \{\varphi : [0, 1] \rightarrow \mathbb{R}^d, \mathcal{L}(\varphi) \leq L\}$, the endpoints of the curve f cannot be multiple points.

Lemma 4.3. *Let $\mathcal{C}_L = \{\varphi : [0, 1] \rightarrow \mathbb{R}^d, \mathcal{L}(\varphi) \leq L\}$. We have $f^{-1}(\{f(0)\}) = \{0\}$ and $f^{-1}(\{f(1)\}) = \{1\}$.*

Proof. Observe that, by symmetry, we only need to prove the first statement since the second one follows then by considering the curve $t \mapsto f(1-t)$. Assume that the set $f^{-1}(\{f(0)\})$ has cardinality at least 2. Thanks to Lemma 4.2, we may consider $t_0 = \min\{t > 0 : f(t) = f(0)\}$. For $x \in \text{Im}f$, we set $\hat{t}(x) = \inf\{t \in [0, 1], f(t) = x\}$. For every $\varepsilon \in (0, t_0)$, we let

$$\hat{X}_\varepsilon = f(\hat{t} \vee \varepsilon) \mathbf{1}_{\{\hat{t} > 0\}} + f(0) \mathbf{1}_{\{\hat{t} = 0\}}.$$

With this definition, the random vector \hat{X}_ε takes its values in $f([\varepsilon, 1]) \cup \{f(0)\}$, that is in $f([\varepsilon, 1])$ since $f(t_0) = f(0)$ and $\varepsilon < t_0$. Thus, $\frac{\hat{X}_\varepsilon}{1-\varepsilon}$ takes its values in $\frac{f([\varepsilon, 1])}{1-\varepsilon}$, which is the image of a curve with length at most L . Consequently, by optimality of f , we have

$$\mathbb{E} \left[\left\| X - \frac{\hat{X}_\varepsilon}{1-\varepsilon} \right\|^2 \right] \geq \mathbb{E}[\|X - \hat{X}\|^2].$$

Besides, we may write

$$\begin{aligned} \left\| X - \frac{\hat{X}_\varepsilon}{1-\varepsilon} \right\|^2 &= \left\| X - \hat{X} + \hat{X} - \frac{\hat{X}_\varepsilon}{1-\varepsilon} \right\|^2 \\ &= \|X - \hat{X}\|^2 + \left\| \hat{X} - \frac{\hat{X}_\varepsilon}{1-\varepsilon} \right\|^2 + 2 \left\langle X - \hat{X}, \hat{X} - \frac{\hat{X}_\varepsilon}{1-\varepsilon} \right\rangle \\ &= \|X - \hat{X}\|^2 + \frac{1}{(1-\varepsilon)^2} \|\hat{X} - \hat{X}_\varepsilon - \varepsilon \hat{X}\|^2 + \frac{2}{1-\varepsilon} \left(\langle X - \hat{X}, \hat{X} - \hat{X}_\varepsilon \rangle - \varepsilon \langle X - \hat{X}, \hat{X} \rangle \right). \end{aligned}$$

As $\|\hat{X} - \hat{X}_\varepsilon\| \leq L\varepsilon$ since f is L -Lipschitz, we get

$$\mathbb{E}[\|\hat{X} - \hat{X}_\varepsilon - \varepsilon \hat{X}\|^2] \leq 2L^2\varepsilon^2 + 2\varepsilon^2\mathbb{E}[\|\hat{X}\|^2] = 2(L^2 + \mathbb{E}[\|\hat{X}\|^2])\varepsilon^2.$$

Note that $\mathbb{E}[\|\hat{X}\|^2] < \infty$ by the same argument as in (14). Moreover, thanks to equation (9) in Remark 5, we have

$$\mathbb{E}[\langle X - \hat{X}, \hat{X} \rangle] = \lambda \int_0^1 \|f'_r(s)\|^2 ds = \lambda L^2. \quad (28)$$

Furthermore, $\hat{X} - \hat{X}_\varepsilon = (f(\hat{t}) - f(\varepsilon)) \mathbf{1}_{\{0 < \hat{t} \leq \varepsilon\}}$, so that equation (8) implies

$$\mathbb{E}[\langle X - \hat{X}, \hat{X} - \hat{X}_\varepsilon \rangle] = -\lambda \int_{[0,1]} \langle (f(t) - f(\varepsilon)) \mathbf{1}_{\{0 < t \leq \varepsilon\}}, f''(dt) \rangle.$$

Hence,

$$\begin{aligned} |\mathbb{E}[\langle X - \hat{X}, \hat{X} - \hat{X}_\varepsilon \rangle]| &\leq \lambda \sum_{j=1}^d \int_{(0,\varepsilon]} |f^j(t) - f^j(\varepsilon)| |(f'')^j|(dt) \\ &\leq \lambda L \varepsilon \sum_{j=1}^d |(f'')^j|((0, \varepsilon]), \end{aligned}$$

where $|(f'')^j|$ stands for the total variation of the signed measure $(f'')^j$. Finally, we obtain

$$\mathbb{E} \left[\left\| X - \frac{\hat{X}_\varepsilon}{1-\varepsilon} \right\|^2 \right] \leq \mathbb{E} [\|X - \hat{X}\|^2] + 2(L^2 + \mathbb{E}[\|\hat{X}\|^2])\varepsilon^2 + \lambda L \varepsilon \rho(\varepsilon) - \frac{2\varepsilon}{1-\varepsilon} \lambda L^2,$$

where $\rho(\varepsilon)$ tends to 0 as $\varepsilon \rightarrow 0$. This inequality shows that, for ε small enough, $\mathbb{E} \left[\left\| X - \frac{\hat{X}_\varepsilon}{1-\varepsilon} \right\|^2 \right] < \mathbb{E}[\|X - \hat{X}\|^2]$, which contradicts the optimality of f . \square

For an open curve, there exists a multiple point which is the last multiple point.

Lemma 4.4. *Let $\mathcal{C}_L = \{\varphi : [0, 1] \rightarrow \mathbb{R}^d, \mathcal{L}(\varphi) \leq L\}$. There exists $\delta > 0$ such that for every $t \in [1 - \delta, 1]$, $f^{-1}(\{f(t)\}) = \{t\}$.*

Proof. Otherwise, we can build sequences $(t_k)_{k \geq 1}$ and $(s_k)_{k \geq 1}$ such that $t_k \rightarrow 1$ and $f(t_k) = f(s_k)$, with $s_k \neq t_k$ for all $k \geq 1$. Up to extraction of a subsequence, we may assume that (s_k) converges to a limit $s \in [0, 1]$. Hence, we have $f(s) = f(1)$, which implies $s = 1$ by Lemma 4.3. Up to another extraction, we may consider that the intervals $[s_k \wedge t_k, s_k \vee t_k]$, $k \geq 1$, are mutually disjoint. Finally, using Lemma 4.1, we obtain

$$\|f''\| \geq \sum_{k \geq 1} \|f''_{(s_k \wedge t_k, s_k \vee t_k)}\| = \infty,$$

which yields a contradiction since we have shown that an optimal curve has finite curvature. \square

Now, we show that the two branches of the curve are necessarily tangent at a multiple point.

Lemma 4.5. (i) *If there exist $0 < t_0 < t_1 < 1$ such that $f(t_0) = f(t_1)$, then $f'_\ell(t_0) = f'_r(t_0) = -f'_r(t_1) = -f'_\ell(t_1)$.*

(ii) *In the case $\mathcal{C}_L = \{\varphi : [0, 1] \rightarrow \mathbb{R}^d, \mathcal{L}(\varphi) \leq L, \varphi(0) = \varphi(1)\}$, if there exists $0 < t < 1$ such that $f(t) = f(0)$, then $f'_\ell(t) = f'_r(t) = -f'_r(0) = -f'_\ell(1)$.*

Proof. First, we show that point (ii) follows from point (i). Let $t \in (0, 1)$ such that $f(t) = f(0)$. Define the curve g by $g(s) = f(s + t/2)$ for $s \in [0, 1 - t/2]$ and $g(s) = f(s + t/2 - 1)$ for $s \in [1 - t/2, 1]$. Clearly, g is a closed curve, $\Delta(g) = \Delta(f)$ and g is L -Lipschitz. Moreover, one has: $g(t/2) = g(1 - t/2)$, $g'_r(t/2) = f'_r(t)$, $g'_\ell(t/2) = f'_\ell(t)$, $g'_r(1 - t/2) = f'_r(0)$ and $g'_\ell(1 - t/2) = f'_\ell(1)$. Consequently, if (i) holds true for g , one deduces (ii).

It remains to show point (i). Suppose that $f'_\ell(t_0) \neq f'_r(t_0)$. Let $\gamma \in (0, 1)$ and $\varepsilon > 0$. We introduce the random vectors $\hat{X}_{0,\gamma} = (1 + \gamma)\hat{X}$ and

$$\hat{X}_{\varepsilon,\gamma} = (1 + \gamma) \left[\hat{X} \mathbf{1}_{\hat{t} \in [0, t_0 - \varepsilon] \cup (t_0 + \varepsilon, 1] \cup \{t_0\}} + h_\varepsilon(\hat{t}) \mathbf{1}_{\hat{t} \in [t_0 - \varepsilon, t_0 + \varepsilon] \setminus \{t_0\}} \right],$$

where $h_\varepsilon(t) = \left(\frac{f(t_0 + \varepsilon) - f(t_0 - \varepsilon)}{2\varepsilon} (t - (t_0 - \varepsilon)) + f(t_0 - \varepsilon) \right)$.

Let us write

$$\begin{aligned} \mathbb{E}[\|X - \hat{X}_{0,\gamma}\|^2] &= \mathbb{E}[\|X - \hat{X}\|^2] + \mathbb{E}[\|\hat{X} - \hat{X}_{0,\gamma}\|^2] + 2\mathbb{E}[\langle X - \hat{X}, \hat{X} - \hat{X}_{0,\gamma} \rangle] \\ &= \mathbb{E}[\|X - \hat{X}\|^2] + \gamma^2 \mathbb{E}[\|\hat{X}\|^2] - 2\mathbb{E}[\langle X - \hat{X}, \gamma \hat{X} \rangle] \\ &= \mathbb{E}[\|X - \hat{X}\|^2] + \gamma^2 \mathbb{E}[\|\hat{X}\|^2] - 2\gamma \lambda L^2. \end{aligned} \quad (29)$$

For the last equality, we used equation (28).

Note that $\hat{X}_{\varepsilon,\gamma} = \hat{X}_{0,\gamma} + (1 + \gamma)(h_\varepsilon(\hat{t}) - f(\hat{t})) \mathbf{1}_{\hat{t} \in [t_0 - \varepsilon, t_0 + \varepsilon] \setminus \{t_0\}}$ and that $\|h_\varepsilon(\hat{t}) - f(\hat{t})\| \leq 4\varepsilon L$. So, we have

$$\begin{aligned} \mathbb{E}[\|X - \hat{X}_{\varepsilon,\gamma}\|^2] &= \mathbb{E}[\|X - \hat{X}_{0,\gamma}\|^2] + (1 + \gamma)^2 \mathbb{E}[\|h_\varepsilon(\hat{t}) - f(\hat{t})\|^2 \mathbf{1}_{\hat{t} \in [t_0 - \varepsilon, t_0 + \varepsilon] \setminus \{t_0\}}] \\ &\quad + 2(1 + \gamma) \mathbb{E}[\langle X - \hat{X}_{0,\gamma}, (h_\varepsilon(\hat{t}) - f(\hat{t})) \mathbf{1}_{\hat{t} \in [t_0 - \varepsilon, t_0 + \varepsilon] \setminus \{t_0\}} \rangle] \\ &= \mathbb{E}[\|X - \hat{X}_{0,\gamma}\|^2] + \mathcal{O}(\varepsilon^2) + o(\varepsilon). \end{aligned} \quad (30)$$

Indeed, $\mathbb{P}([t_0 - \varepsilon, t_0 + \varepsilon] \setminus \{t_0\})$ tends to 0 as ε tends to 0. Besides, the random vector $\hat{X}_{\varepsilon, \gamma}$ is taking its values in the image of a curve of length

$$L_{\varepsilon, \gamma} := (1 + \gamma)(L(1 - 2\varepsilon) + \|f(t_0 + \varepsilon) - f(t_0 - \varepsilon)\|).$$

Yet, since $f'_\ell(t_0) \neq f'_r(t_0)$, if ε is small enough, there exists $\alpha \in [0, 1)$ such that

$$\begin{aligned} \|f(t_0 + \varepsilon) - f(t_0 - \varepsilon)\|^2 &= \|f(t_0 + \varepsilon) - f(t_0) + f(t_0) - f(t_0 - \varepsilon)\|^2 \\ &= \varepsilon^2 \left[\left\| \frac{f(t_0 + \varepsilon) - f(t_0)}{\varepsilon} \right\|^2 + \left\| \frac{f(t_0) - f(t_0 - \varepsilon)}{\varepsilon} \right\|^2 \right. \\ &\quad \left. + 2 \left\langle \frac{f(t_0 + \varepsilon) - f(t_0)}{\varepsilon}, \frac{f(t_0) - f(t_0 - \varepsilon)}{\varepsilon} \right\rangle \right] \\ &\leq \varepsilon^2 (2L^2 + 2L^2\alpha). \end{aligned}$$

Hence, $\|f(t_0 + \varepsilon) - f(t_0 - \varepsilon)\| < \varepsilon L \sqrt{2(1 + \alpha)}$, and, thus,

$$L_{\varepsilon, \gamma} \leq (1 + \gamma)(L - 2\varepsilon L + \varepsilon L \sqrt{2(1 + \alpha)}) = (1 + \gamma)(L - \eta\varepsilon),$$

where $\eta > 0$. Let $\gamma = \frac{\eta\varepsilon}{L}$. Then, for ε small enough, we get $L_{\varepsilon, \gamma} \leq L - \frac{(\eta\varepsilon)^2}{L} < L$ and, using equations (29) and (30), we have $\mathbb{E}[\|X - \hat{X}_{\varepsilon, \gamma}\|^2] < \mathbb{E}[\|X - \hat{X}\|^2]$. This contradicts the optimality of f . So, $f'_\ell(t_0) = f'_r(t_0)$. Similarly, we obtain that $f'_\ell(t_1) = f'_r(t_1)$. Finally, consider the curve g , defined by

$$g(t) = \begin{cases} f(t) & \text{if } t \in [0, t_0] \cup [t_1, 1] \\ f(t_0 + t_1 - t) & \text{if } t \in (t_0, t_1). \end{cases}$$

This definition means that g has the same image as f but the arc between t_0 and t_1 is traveled along in the reverse direction. Since g , having the same image and length as f , is an optimal curve, which satisfies $g(t_0) = g(t_1)$, we have $g'_\ell(t_0) = g'_r(t_0)$ and $g'_\ell(t_1) = g'_r(t_1)$. On the other hand, by the definition of g , we know that $f'(t_0) = g'_\ell(t_0) = -g'_\ell(t_1)$ and $f'(t_1) = g'_r(t_1) = -g'_r(t_0)$. Hence, $f'(t_0) = -f'(t_1)$. \square

We introduce the set

$$D = \left\{ t \in [0, 1] \mid \text{Card}(f^{-1}(\{f(t)\}) \cap [0, 1]) \geq 2 \right\}.$$

Lemma 4.6. *If $f(t)$, $t \in (0, 1)$, is a multiple point of $f : [0, 1] \rightarrow \mathbb{R}^2$, then t cannot be right- or left-isolated:*

for all $t \in D \cap (0, 1)$, for all $\varepsilon > 0$, $(t, t + \varepsilon) \cap D \neq \emptyset$ and $(t - \varepsilon, t) \cap D \neq \emptyset$.

Proof. Let $t_0 \in D \cap (0, 1)$. Assume that there exists $\varepsilon > 0$ such that $(t_0, t_0 + \varepsilon) \cap D = \emptyset$ or $(t_0 - \varepsilon, t_0) \cap D = \emptyset$. We will show that this leads to a contradiction. Without loss of generality, up to considering $t \mapsto f(1 - t)$, we assume that $(t_0 - \varepsilon, t_0) \cap D = \emptyset$. Let $t_1 \in [0, 1)$ such that $t_0 \neq t_1$ and $f(t_0) = f(t_1)$. By Lemma 4.5, one has $f'_\ell(t_0) = -f'_r(t_1)$.

Let

$$y = \frac{f'_r(t_1)}{L}$$

and define the functions α and β by

$$\begin{aligned} \alpha(t) &= \langle f(t) - f(t_1), y \rangle \text{ for } t \in [t_1, t_1 + \varepsilon) \\ \beta(t) &= \langle f(t) - f(t_0), y \rangle \text{ for } t \in (t_0 - \varepsilon, t_0]. \end{aligned}$$

Notice, since $f(t_0) = f(t_1)$, that α and β are restrictions, to $[t_1, t_1 + \varepsilon)$ and $(t_0 - \varepsilon, t_0]$ respectively, of the same function. Nevertheless, this notation α, β were chosen for readability.

The functions α and β satisfy the following properties:

- α is right-differentiable and $\alpha'_r(t) = \langle f'_r(t), y \rangle$ for every $t \in [t_1, t_1 + \varepsilon)$. Since $\alpha'_r(t_1) = L > 0$ and α'_r is right-continuous, there exists $\delta \in (0, \varepsilon)$, such that $\alpha'_r(t) \geq \delta L$ for every $t \in [t_1, t_1 + \delta]$.

- β is left-differentiable and $\beta'_\ell(t) = \langle f'_\ell(t), y \rangle$ for every $t \in (t_0 - \varepsilon, t_0]$. Since $\beta'_\ell(t_0) = -L < 0$ and β'_ℓ is left-continuous, there exists $\delta' \in (0, \varepsilon)$ such that $\beta'_\ell(t) \leq -\delta' L$ for every $t \in [t_0 - \delta', t_0]$.

Without loss of generality, we may assume that $\delta' = \delta$, since it suffices to pick the smallest of both values to have the properties on α'_r and β'_ℓ . In particular, we see that

- α is a bijection from $[t_1, t_1 + \delta]$ onto its image $\alpha([t_1, t_1 + \delta]) = [0, a]$, where $a := \alpha(t_1 + \delta) > 0$,
- β is a bijection from $[t_0 - \delta, t_0]$ onto its image $\beta([t_0 - \delta, t_0]) = [0, b]$, where $b := \beta(t_0 - \delta) > 0$.

We denote by α^{-1} and β^{-1} their inverse functions.

Let $z \in \mathbb{R}^2$ be such that $\|z\| = 1$ and $\langle z, y \rangle = 0$. For every $t \in (t_1, \alpha^{-1}(b)]$, we have $\langle f(t) - f(\beta^{-1}(\alpha(t))), y \rangle = 0$. Then, we may write $f(t) - f(\beta^{-1}(\alpha(t))) = \langle f(t) - f(\beta^{-1}(\alpha(t))), z \rangle z$. Moreover, for $t \in (t_1, \alpha^{-1}(b)]$, since there are no further multiple point before t_0 , $f(t) - f(\beta^{-1}(\alpha(t))) \neq 0$. Thus, there exists $\sigma \in \{-1, 1\}$ such that

$$\frac{f(t) - f(\beta^{-1}(\alpha(t)))}{\|f(t) - f(\beta^{-1}(\alpha(t)))\|} = \sigma z.$$

We suppose, without loss of generality, that the vector z was chosen such that $\sigma = 1$. Now, let us show that, for $t \in (t_1, \alpha^{-1}(b)]$,

$$\langle z, f'_r(t) \rangle \leq \frac{1}{2\lambda} \sup_{t_1 \leq s \leq t} \|f(s) - f(\beta^{-1}(\alpha(s)))\|.$$

Since $\langle z, f'_r(t_1) \rangle = 0$, we have, according to Theorem 3.1,

$$\begin{aligned} \langle z, f'_r(t) \rangle &= \langle z, f'_r(t) - f'_r(t_1) \rangle \\ &= \int_{(t_1, t]} \langle z, f''(ds) \rangle \\ &= -\frac{1}{\lambda} \mathbb{E} \left[\langle X - f(\hat{t}), z \rangle \mathbf{1}_{\{t_1 < \hat{t} \leq t\}} \right] \\ &= -\frac{1}{\lambda} \mathbb{E} \left[\left\langle X - f(\hat{t}), \frac{f(\hat{t}) - f(\beta^{-1}(\alpha(\hat{t})))}{\|f(\hat{t}) - f(\beta^{-1}(\alpha(\hat{t})))\|} \right\rangle \mathbf{1}_{\{t_1 < \hat{t} \leq t\}} \right] \end{aligned}$$

Besides, for $t \in [0, 1]$, starting from

$$\|X - f(t)\|^2 = \|X - f(\hat{t})\|^2 + \|f(\hat{t}) - f(t)\|^2 + 2\langle X - f(\hat{t}), f(\hat{t}) - f(t) \rangle,$$

we deduce, by optimality of \hat{t} , the inequality

$$-\langle X - f(\hat{t}), f(\hat{t}) - f(t) \rangle \leq \frac{1}{2} \|f(\hat{t}) - f(t)\|^2 \quad a.s.$$

Hence, we obtain

$$\begin{aligned} \langle z, f'_r(t) \rangle &\leq \frac{1}{2\lambda} \mathbb{E} \left[\|f(\hat{t}) - f(\beta^{-1}(\alpha(\hat{t})))\| \mathbf{1}_{\{t_1 < \hat{t} \leq t\}} \right] \\ &\leq \frac{1}{2\lambda} \sup_{t_1 < s \leq t} \|f(s) - f(\beta^{-1}(\alpha(s)))\|. \end{aligned} \tag{31}$$

Similarly, we get, for every $t \in [\beta^{-1}(a), t_0]$,

$$\langle z, f'_\ell(t) \rangle \leq \frac{1}{2\lambda} \sup_{t \leq s < t_0} \|f(s) - f(\alpha^{-1}(\beta(s)))\|. \tag{32}$$

This may be seen for instance by considering the optimal curve parameterized in the reverse direction $t \mapsto f(1 - t)$. For $x \in [0, a \wedge b]$, let $D(x) = f(\alpha^{-1}(x)) - f(\beta^{-1}(x))$. This function D is right-differentiable and

$$D'_r(x) = \frac{f'_r(\alpha^{-1}(x))}{\alpha'_r(\alpha^{-1}(x))} - \frac{f'_\ell(\beta^{-1}(x))}{\beta'_\ell(\beta^{-1}(x))}.$$

Moreover, $\alpha'_r(\alpha^{-1}(x)) \geq \delta L$ and $-\beta'_\ell(\beta^{-1}(x)) \geq \delta L$, so that

$$\begin{aligned} \langle D'_r(x), z \rangle &\leq \frac{1}{\delta L} (\langle z, f'_r(\alpha^{-1}(x)) \rangle + \langle z, f'_\ell(\beta^{-1}(x)) \rangle) \\ &\leq \frac{1}{\delta L \lambda} \sup_{u \leq x} \|D(u)\|. \end{aligned}$$

For the last inequality, we used the upper bounds (31) and (32) together with the monotony of α and β . Observe, since $z = \frac{D(x)}{\|D(x)\|}$, that $\langle D'_r(x), z \rangle$ is the right-derivative of $\|D(x)\|$. As $D(0) = 0$, the Gronwall Lemma implies that $D(x) = 0$ for all $x \in [0, a \wedge b)$, which yields a contradiction, since the considered multiple point is supposed to be left-isolated. \square

We may now state the injectivity result in dimension 2, for open and closed curves.

Proposition 4.1. (i) *If $\mathcal{C}_L = \{\varphi \in [0, 1] \rightarrow \mathbb{R}^2, \mathcal{L}(\varphi) \leq L\}$, then f is injective.*

(ii) *If $\mathcal{C}_L = \{\varphi \in [0, 1] \rightarrow \mathbb{R}^2, \mathcal{L}(\varphi) \leq L, \varphi(0) = \varphi(1)\}$, then either f restricted to $[0, 1)$ is injective or $\text{Im}f$ is a segment.*

Proof. (i) $\mathcal{C}_L = \{\varphi \in [0, 1] \rightarrow \mathbb{R}^2, \mathcal{L}(\varphi) \leq L\}$.

Thanks to Lemma 4.4, if f has multiple points, there exists a last multiple point. As such, this multiple point is right-isolated. However, by Lemma 4.6, this cannot happen. So, f is injective.

(ii) $\mathcal{C}_L = \{\varphi \in [0, 1] \rightarrow \mathbb{R}^2, \mathcal{L}(\varphi) \leq L, \varphi(0) = \varphi(1)\}$.

We assume that f restricted to $[0, 1)$ is not injective. So, our aim is to prove that $\text{Im}f$ is a segment. As f is supposed not to be injective, the set $D = \{t \in [0, 1) \mid \text{Card}([0, 1) \cap f^{-1}(\{f(t)\})) \geq 2\}$ is non-empty. Without loss of generality, we can assume that $D \cap (0, 1) \neq \emptyset$. Indeed, if $D = \{0\}$, we can replace f by the curve $t \mapsto f((t + 1/2) \bmod 1)$ for which $D = \{1/2\}$.

Let us show that D is dense in $(0, 1)$. Proceeding by contradiction, we assume that there exists a non-empty open interval $(a, b) \subset (0, 1)$ such that $D \cap (a, b) = \emptyset$. Since $D \cap (0, 1) \neq \emptyset$, one has $D \cap (0, a] \neq \emptyset$ or $D \cap [b, 1) \neq \emptyset$. Consider the case where $D \cap [b, 1) \neq \emptyset$. Define $\beta = \inf(D \cap [b, 1))$. There exist two sequences $(t_k)_{k \geq 1} \subset D$ and $(s_k)_{k \geq 1} \subset D$ such that $t_k \downarrow \beta$, $f(t_k) = f(s_k)$ and $s_k \neq t_k$ for all $k \geq 1$. Up to an extraction, s_k converges to a limit $s \in [0, 1]$. If $\beta \neq s$ then $\beta \in D$ is left-isolated which is impossible by Lemma 4.6. Thus $s = \beta$ and consequently $s_k \geq \beta$ for k large enough. This yields $f'_r(s_k) \rightarrow f'_r(\beta)$. Besides, for all k , $f'_r(t_k) \rightarrow f'_r(\beta)$ and, by Lemma 4.5, $f'_r(t_k) = -f'_r(s_k)$, which contradicts the fact that f has speed L . The case where $D \cap (0, a] \neq \emptyset$ is similar.

The next step is to prove that the set $[0, 1) \setminus D$ is finite. Let $t \in (0, 1) \setminus D$. Since D is dense, there exists a sequence $(t_k)_{k \geq 1} \in D$ such that $t_k \downarrow t$. For every $k \geq 1$, there exists $s_k \neq t_k$ such that $f(t_k) = f(s_k)$. If $s \in [0, 1]$ is a limit point of (s_k) , then $f(t) = f(s)$ which implies $t = s$ since $t \notin D$ and $t \neq 0$. Therefore $\lim_{k \rightarrow \infty} s_k = t$. Up to an extraction, we may assume that (s_k) converges increasingly or decreasingly to t . By Lemma 4.5, one has $f'(t_k) = -f'(s_k)$ for k large enough. If $s_k \downarrow t$, one obtains a contradiction: $f'_r(t) = \lim_k f'_r(t_k) = -\lim_k f'_r(s_k) = -f'_r(t)$. Thus $s_k \uparrow t$ and one gets $f'_r(t) = -f'_\ell(t)$. This means that $f(t)$ is a cusp. Since $\|f''\|([0, 1]) < \infty$, there are only a finite number of such points.

Observe that, as a consequence of Lemma 4.5, for every $t \in [0, 1)$, $\text{Card}([0, 1) \cap f^{-1}(\{f(t)\})) < 3$. Indeed, if a point has multiplicity at least 3, that is there exist $0 \leq t_1 < t_2 < t_3 < 1$ such that $f(t_1) = f(t_2) = f(t_3)$, then, on the one hand, $f'_r(t_1) = -f'(t_2) = -f'(t_3)$, and on the other hand, $f'(t_2) = -f'(t_3)$. Thus, one obtains again a contradiction: $f'_r(t_1) = f'(t_2) = f'(t_3) = 0$. In other words, $D = \{t \in [0, 1) \mid \text{Card}([0, 1) \cap f^{-1}(\{f(t)\})) = 2\}$.

We introduce the function $\varphi : [0, 1) \rightarrow [0, 1)$, defined as follows: for $t \in [0, 1) \setminus D$, set $\varphi(t) = t$ and for $t \in D$, set $\varphi(t) = t'$ where $t' \in f^{-1}(\{f(t)\})$ and $t' \neq t$. Note that φ is an involution.

Let us show that the function φ is continuous on $(0, 1) \setminus \{\varphi(0)\}$. First, observe that f is derivable on $D \cap (0, 1)$ by Lemma 4.5, and that f' is continuous on $D \cap (0, 1)$ since f'_r is right-continuous and f'_ℓ is left-continuous. Let $t \in (0, 1)$ such that $t \neq \varphi(0)$ and let $(t_k)_{k \geq 1}$ be a sequence converging to t . Let $s \in [0, 1]$ be a limit point of $(\varphi(t_k))$. Since $f(t_k) = f(\varphi(t_k))$, for all $k \geq 1$, one has $f(s) = f(t)$. Necessarily, $s \in (0, 1)$ since $t \neq \varphi(0)$. If $t \notin D$, one has $s = t = \varphi(t)$. If $t \in D$, then $s \in \{t, \varphi(t)\}$. Since $D \cap (0, 1)$ is open, $t_k \in D$ for k large enough, hence $f'(\varphi(t_k)) = -f'(t_k)$ for k large enough. Thus $f'(s) = -f'(t)$ and consequently $s = \varphi(t)$.

Let us show that φ is derivable on $D \cap (0, 1) \setminus \{\varphi(0)\}$ and $\varphi'(t) = -1$ for all $t \in D \cap (0, 1) \setminus \{\varphi(0)\}$. Let $t \in D \cap (0, 1)$, $t \neq \varphi(0)$. For all $h \in \mathbb{R}$ such that $|h| < t \wedge (1 - t)$, we have

$$\begin{aligned} f(t+h) - f(t) &= f(\varphi(t+h)) - f(\varphi(t)) \\ &= \int_{\varphi(t)}^{\varphi(t+h)} f'(s) ds \\ &= (\varphi(t+h) - \varphi(t)) \int_0^1 f'(\varphi(t) + u(\varphi(t+h) - \varphi(t))) du. \end{aligned}$$

Besides, since f' is continuous at the point $\varphi(t) \in D \cap (0, 1)$ and φ is continuous at the point t , one has $\lim_{h \rightarrow 0} \int_0^1 f'(\varphi(t) + u(\varphi(t+h) - \varphi(t))) du = f'(\varphi(t)) = -f'(t)$. One deduces that $\lim_{h \rightarrow 0} (\varphi(t+h) - \varphi(t))/h = -1$.

Let us prove that $\varphi(\varphi(0)/2 + t) = \varphi(0)/2 + 1 - t \pmod{1}$ for all $t \in [-\varphi(0)/2, 1 - \varphi(0)/2]$. From the two previous steps, one deduces that if $\varphi(0) = 0$, $\varphi(t) = 1 - t$ for all $t \in (0, 1)$, as desired, while, if $\varphi(0) \in (0, 1)$, there exist two constants c_1 and c_2 such that

$$\varphi(t) = c_1 - t \quad \forall t \in (0, \varphi(0)), \quad \varphi(t) = c_2 - t \quad \forall t \in (\varphi(0), 1).$$

It remains to prove that $c_1 = \varphi(0)$ and $c_2 = 1 + \varphi(0)$. As φ takes its values in $[0, 1]$, one has $\varphi(0) \leq c_1 \leq 1$ and $1 \leq c_2 \leq 1 + \varphi(0)$. Moreover, since φ is a bijection, $c_2 - t \geq c_1$ for $t \geq \varphi(0)$ or $c_2 - t \leq c_1 - \varphi(0)$ for $t \geq \varphi(0)$, that is $c_2 - 1 \geq c_1$ or $c_2 \leq c_1$. In the first case, one gets $c_1 = \varphi(0)$ and $c_2 = 1 + \varphi(0)$. In the second case, one gets $c_1 = c_2 = 1$, which is not possible: necessarily, $\varphi(0) = 1/2$, since otherwise $\varphi(1 - \varphi(0)) = \varphi(0)$ which yields $1 - \varphi(0) = 0$, and we see that the restriction of f to $[0, 1/2]$ is a closed curve with the same image as f , hence f is not optimal.

Finally, define the curve \tilde{f} by

$$\tilde{f}(t) = f((\varphi(0)/2 + t) \pmod{1}).$$

This curve \tilde{f} has the same image as f and, from the last step, $\tilde{f}(t) = \tilde{f}(1 - t)$ for all $t \in [0, 1]$. Let us show that $\text{Im} f$ is a segment. Otherwise, the curve g defined by

$$g(t) = \tilde{f}(t) \quad \text{if } t \in [0, 1/2], \quad g(t) = \tilde{f}(1/2) + 2(t - 1/2)(\tilde{f}(1) - \tilde{f}(1/2)) \quad \text{if } t \in [1/2, 1]$$

satisfies $\mathcal{L}(g) < \mathcal{L}(f)$ and $\Delta(g) \leq \Delta(f)$, since $\text{Im} f = \tilde{f}([0, 1/2])$, thus f cannot be optimal. \square

5 Examples of principal curves

5.1 Uniform distribution on an enlargement of a curve

The purpose of this section is to study the principal curve problem for the uniform distribution on an enlargement of some generative curve. For $A \subset \mathbb{R}^d$ and $r \geq 0$, we denote by

$$A \oplus r = \{x \in \mathbb{R}^d \mid d(x, A) \leq r\}$$

the r -enlargement of A . Under some conditions on the generative curve $f : [0, 1] \rightarrow \mathbb{R}^d$, for r small enough, it turns out that the image of an optimal curve with length $\mathcal{L}(f)$ for the uniform distribution on an r -enlargement of $\text{Im} f$ is necessarily $\text{Im} f$. More specifically, the radius r must not exceed the reach of $\text{Im} f$.

The reach of a set $A \subset \mathbb{R}^d$ is the supremum of the radii ρ such that every point at distance at most ρ of A has a unique projection on A . More formally, following Federer (1959), we define for $A \subset \mathbb{R}^d$

$$\text{reach}(A) = \sup \{ \rho \geq 0 \mid \forall x \in \mathbb{R}^d \quad d(x, A) \leq \rho \Rightarrow \exists! a \in A \quad d(x, a) = d(x, A) \} \in [0, +\infty].$$

The question of the optimality of the generative curve when considering the uniform distribution on an enlargement has been first addressed in dimension $d = 2$ in Mosconi and Tilli (2005). Observe that related ideas can be found in Genovese et al. (2012). Our proof in arbitrary dimension $d \geq 1$ relies on arguments in Federer (1959), which moreover allow to show uniqueness.

Theorem 5.1. *Let $f : [0, 1] \rightarrow \mathbb{R}^d$ be a curve. Suppose that f is injective, differentiable, f' is Lipschitz, and there exists $c > 0$ such that $\|f'(t)\| \geq c$ for all $t \in [0, 1]$. Then, the reach of $\text{Im}f$ is positive. Let $r \in (0, \text{reach}(\text{Im}f))$ and let X be a random vector uniformly distributed on $\text{Im}f \oplus r$. Consider a function $V : [0, \infty) \rightarrow [0, \infty)$ continuous, increasing and such that $V(0) = 0$. Then, for every curve $g : [0, 1] \rightarrow \mathbb{R}^d$ such that $\mathcal{L}(g) \leq \mathcal{L}(f)$ one has*

$$\mathbb{E}[V(d(X, \text{Im}f))] \leq \mathbb{E}[V(d(X, \text{Im}g))].$$

with equality if and only if $\text{Im}g = \text{Im}f$.

The proof of the theorem is based on two lemmas. For $k \geq 1$, λ_k denotes the Lebesgue measure on \mathbb{R}^k and α_k the volume of the unit ball in \mathbb{R}^k . From (Mosconi and Tilli, 2005, Lemma 42), we have the next result.

Lemma 5.1. *Let A be a compact connected subset of \mathbb{R}^d with $\mathcal{H}^1(A) < \infty$. Then for all $r \geq 0$ one has*

$$\lambda_d(A \oplus r) \leq \mathcal{H}^1(A)\alpha_{d-1}r^{d-1} + \alpha_d r^d.$$

Lemma 5.2. *Let $f : [0, 1] \rightarrow \mathbb{R}^d$ be a curve. Suppose that f is injective, f is differentiable, f' is Lipschitz, and there exists $c > 0$ such that $\|f'(t)\| \geq c$ for all $t \in [0, 1]$. Then, the reach of $A = \text{Im}f$ is positive and for all $r \leq \text{reach}(A)$ one has*

$$\lambda_d(A \oplus r) = \mathcal{L}(f)\alpha_{d-1}r^{d-1} + \alpha_d r^d \quad (33)$$

Moreover, one has

$$\{x \in A \oplus r \mid d(x, \partial(A \oplus r)) \geq r\} \subset A. \quad (34)$$

Proof. The assumptions on f imply that there exists $\varepsilon > 0$, a set $B \subset \mathbb{R}^d$ and a function $\varphi : (-\varepsilon, 1 + \varepsilon) \rightarrow B$ such that φ is bijective, $\varphi = f$ on $[0, 1]$, φ is differentiable, φ' is Lipschitz and φ^{-1} is Lipschitz. From (Federer, 1959, Theorem 4.19), we deduce that $\text{reach}(A) > 0$. For $r \in (0, \text{reach}(A))$, equality (33) follows from (Federer, 1959, Theorem 5.6 & Remark 6.14). For $r = \text{reach}(A)$, we can write

$$\mathcal{L}(f)\alpha_{d-1}r^{d-1} + \alpha_d r^d = \lim_{n \rightarrow \infty} \lambda_d(A \oplus (r - 1/n)) = \lambda_d(\{x \in \mathbb{R}^d \mid d(x, A) < r\})$$

and $\lambda_d(A \oplus r) \leq \mathcal{L}(f)\alpha_{d-1}r^{d-1} + \alpha_d r^d$ by Lemma 5.1. Thus, equality (33) holds.

Now, we prove (34). Let $x \in A \oplus r$ such that $d(x, \partial(A \oplus r)) \geq r$. According to (Federer, 1959, Corollary 4.9), if $0 < s < \text{reach}(A)$ and $A'_s = \{y \in \mathbb{R}^d \mid d(y, A) \geq s\}$ then

$$d(y, A'_s) = s - d(y, A) \quad \text{whenever } 0 < d(y, A) \leq s.$$

Suppose that $d(x, A) > 0$, then for all $s \in [d(x, A), r)$ one has $d(x, A) = s - d(x, A'_s)$. Since $\lim_{s \rightarrow r} d(x, A'_s) = d(x, A'_r) = d(x, \partial(A \oplus r))$, one gets $d(x, A) \leq 0$. This proves that $x \in A$. \square

Proof of Theorem 5.1. We set $A = \text{Im}f$ and $B = \text{Im}g$. On the one hand, denoting by V^{-1} the inverse of $V : [0, \infty) \rightarrow [0, V(\infty))$,

$$\begin{aligned} \mathbb{E}[V(d(X, B))] &= \int_0^\infty \mathbb{P}[V(d(X, B)) > t] dt \\ &= \int_0^{V(\infty)} \left(1 - \mathbb{P}(d(X, B) \leq V^{-1}(t))\right) dt \\ &= \int_0^{V(\infty)} \left(1 - \frac{\lambda_d((B \oplus V^{-1}(t)) \cap (A \oplus r))}{\lambda_d(A \oplus r)}\right) dt. \end{aligned}$$

On the other hand, for $t \leq V(r)$, by Lemma 5.1 and equation (33), one gets

$$\begin{aligned} \lambda_d((B \oplus V^{-1}(t)) \cap (A \oplus r)) &\leq \lambda_d(B \oplus V^{-1}(t)) \leq \mathcal{H}^1(\text{Im}g)\alpha_{d-1}V^{-1}(t)^{d-1} + \alpha_d V^{-1}(t)^d \\ &\leq \mathcal{L}(f)\alpha_{d-1}V^{-1}(t)^{d-1} + \alpha_d V^{-1}(t)^d = \lambda_d(A \oplus V^{-1}(t)). \end{aligned} \quad (35)$$

Therefore, for all $t \in [0, V(\infty))$,

$$1 - \frac{\lambda_d((B \oplus V^{-1}(t)) \cap (A \oplus r))}{\lambda_d(A \oplus r)} \geq \left[1 - \frac{\lambda_d(A \oplus V^{-1}(t))}{\lambda_d(A \oplus r)} \right]_+.$$

Consequently,

$$\mathbb{E}[V(d(X, B))] \geq \int_0^{V(r)} \left(1 - \frac{\lambda_d(A \oplus V^{-1}(t))}{\lambda_d(A \oplus r)} \right) dt = \mathbb{E}[V(d(X, A))].$$

Suppose that $\mathbb{E}[V(d(X, B))] = \mathbb{E}[V(d(X, A))]$. Then, we have

$$\lambda_d(A \oplus r) = \lambda_d((B \oplus V^{-1}(t)) \cap (A \oplus r)) \, dt\text{-a.e. on } [V(r), +\infty).$$

By right continuity with respect to t , we obtain that $\lambda_d(A \oplus r) = \lambda_d((B \oplus r) \cap (A \oplus r))$. From the chain of inequalities (35) with $t = V(r)$, we deduce that $\mathcal{L}(f) = \mathcal{H}^1(\text{Img})$ and

$$\lambda_d((A \oplus r) \cap (B \oplus r)^c) = \lambda_d((A \oplus r)^c \cap (B \oplus r)) = 0.$$

Let us show that $A \oplus r = B \oplus r$. Suppose that $(A \oplus r) \cap (B \oplus r)^c \neq \emptyset$. Then one can find $x \in \mathbb{R}^d$ and $a \in A$ such that $d(x, a) \leq r$ and $d(x, B) > r$. Set $y = x - \varepsilon(x - a)$ where $0 < \varepsilon \leq 1$ and $\varepsilon < d(x, B)/r - 1$. One has $d(y, a) \leq r - \varepsilon r < r$ and $d(y, B) > d(x, B) - \varepsilon r > r$. Thus y belongs to the interior of $(A \oplus r) \cap (B \oplus r)^c$ which implies $\lambda_d((A \oplus r) \cap (B \oplus r)^c) > 0$. Therefore $(A \oplus r) \cap (B \oplus r)^c = \emptyset$. Similarly one can prove that $(A \oplus r)^c \cap (B \oplus r) = \emptyset$.

Finally, from (34), we deduce that

$$B \subset \{x \in B \oplus r \mid d(x, \partial(B \oplus r)) \geq r\} = \{x \in A \oplus r \mid d(x, \partial(A \oplus r)) \geq r\} \subset A.$$

Since $\mathcal{L}(f) = \mathcal{H}^1(\text{Img})$, this implies that $A = B$. \square

5.2 Uniform distribution on a circle

In this section, we investigate the principal curve problem for a particular distribution, the uniform distribution on a circle.

Proposition 5.1. *Consider the unit circle centered at the origin with parameterization given by*

$$g(t) = (\cos(2\pi t), \sin(2\pi t))$$

for $t \in [0, 1]$. Let U be a uniform random variable on $[0, 1]$ and let $X = g(U)$. Then, for every $L < 2\pi$, the circle centered at the origin with radius $\frac{L}{2\pi}$ is the unique closed principal curve with length L for X .

Proof. Let $f : [0, 1] \rightarrow \mathbb{R}^2$ be an optimal closed curve with length L . We denote by K the convex hull of $\text{Im}f$. Since $\text{Im}f$ is compact, K is a compact convex set (consequence of Caratheodory's theorem; see, e.g., Hiriart-Urruty and Lemaréchal (2012)). Notice that $\text{Im}f$ is included in the unit disk: indeed, if not, since f is a closed curve, with $\mathcal{L}(f) < 2\pi$, there exist u_1 and u_2 , such that $f(u_1)$ and $f(u_2)$ belong to the unit circle and the arc $t \in (u_1, u_2) \mapsto f(t)$ is outside the disk, which is not optimal since replacing this arc by the corresponding unit circle arc yields a better and shorter curve. In turn, the convex hull K is also included in the unit disk, by convexity of the latter. Let $\pi_K : \mathbb{R}^2 \rightarrow K$ denote the projection onto K and define the curve h by $h(t) = \pi_K(g(t))$ for $t \in [0, 1]$. By this definition of h as projection of the unit circle on a set included in the unit disk containing $\text{Im}f$, we have

$$\Delta(h) \leq \Delta(f).$$

- Let us prove that h has length at most L . First, note that h has finite length, since π_K is Lipschitz. By properties of the projection on a closed convex set, we know that the set of points of \mathbb{R}^2 projecting onto a given element of the boundary ∂K of K is a cone. This ensures that $h : [0, 1] \rightarrow \partial K$ is onto, because a cone with vertex in the unit disk intersects the unit

circle $\text{Im}g$ at least once. More specifically, if the cone reduces to a half-line (degenerated case), then it intersects $\text{Im}g$ exactly once. Otherwise, the cone is the region delimited by two distinct half-lines with common origin in the disk, and, thus, contains an infinity of such distinct half-lines, each of them intersecting $\text{Im}g$ once. Hence, for every $v \in \text{Im}h$, there is either one t such that $v = h(t)$, or an infinity.

We will use Cauchy-Crofton's formula on the length of a curve (for a proof, see, e.g., Ayari and Dubuc (1997)). Let $d_{r,\theta}$ denote the line with equation $x \cos \theta + y \sin \theta = r$. For every curve $\varphi = (\varphi^1, \varphi^2)$, if

$$N_\varphi(r, \theta) = \text{Card}(\{t \in [0, 1], \varphi(t) \in d_{r,\theta}\}) = \text{Card}(\{t \in [0, 1], \varphi^1(t) \cos \theta + \varphi^2(t) \sin \theta = r\}),$$

then the length of φ is given by

$$\frac{1}{4} \int_0^{2\pi} \int_{-\infty}^{\infty} N_\varphi(r, \theta) dr d\theta.$$

Let us compare $N_h(r, \theta)$ and $N_f(r, \theta)$ for $(r, \theta) \in \mathbb{R} \times [0, 2\pi]$. To begin with, note that $N_h(r, \theta)$ is finite almost everywhere since h has finite length. So, we need only consider the cases where $N_h(r, \theta)$ is finite. This allows to exclude the points $v \in \text{Im}h$ such that $h^{-1}(\{v\})$ is infinite, as well as the cases where a line $d_{r,\theta}$ and $\text{Im}h$ have a whole segment in common. Observing that, if the line $d_{r,\theta}$ does not intersect $\text{Im}h$, then it does not intersect $\text{Im}f$ either, since $\text{Im}h$ is the boundary of the convex hull of $\text{Im}f$, it remains to look at the two following cases for comparing $N_h(r, \theta)$ and $N_f(r, \theta)$.

- If the line $d_{r,\theta}$ intersects $\text{Im}h$ at a single point, then this point belongs to $\text{Im}f$.
- If the line $d_{r,\theta}$ intersects $\text{Im}h$ at exactly two points, then $\text{Im}f$ crosses the line. If $\text{Im}f$ were located on one side of the line, K were not the convex hull. Since f is a closed curve, $\text{Im}f$ crosses the line at least twice.

So, $N_h(r, \theta) \leq N_f(r, \theta)$ almost everywhere, that is $\mathcal{L}(h) \leq \mathcal{L}(f) = L$.

- Now, observe that $\text{Im}h \subset \text{Im}f$. Indeed, otherwise, there exists $t \in [0, 1]$ such that $h(t) \notin \text{Im}f$, which means that $d(g(t), \text{Im}f) > d(g(t), K)$. By continuity this implies that $d(g(s), \text{Im}f) > d(g(s), K)$ for all s in a non-empty open set and one obtains that $\Delta(h) < \Delta(f)$. By optimality of f , this is not possible since $\mathcal{L}(h) \leq L$.
- Since $\text{Im}h \subset \text{Im}f$ and $\mathcal{L}(f) = L$, to obtain that $\text{Im}f$ is the circle with center $(0, 0)$ and radius $L/2\pi$, it remains to show that $\text{Im}h$ is the circle with center $(0, 0)$ and radius $L/2\pi$. Let $\theta \in [0, 1]$ and let $A_\theta : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ denote the rotation with center $(0, 0)$ and angle $2\pi\theta$. We set $h_\theta(t) = \pi_{A_\theta(K)}(g(t))$, for every $t \in [0, 1]$. Since $h_\theta(t) = A_\theta \circ \pi_K(A_\theta^{-1}(g(t))) = A_\theta \circ \pi_K(g(t - \theta))$, h_θ is a curve with same length as h . Moreover, $A_\theta(X)$ has the same distribution as X , so that

$$\begin{aligned} \mathbb{E} [\|X - \pi_{A_\theta(K)}(X)\|^2] &= \mathbb{E} [\|A_\theta(X) - \pi_{A_\theta(K)}(A_\theta(X))\|^2] \\ &= \mathbb{E} [\|A_\theta(X) - A_\theta(\pi_K(X))\|^2] \\ &= \mathbb{E} [\|X - \pi_K(X)\|^2]. \end{aligned}$$

By strict convexity, we deduce from this equality that, if $\mathbb{P}(\pi_{A_\theta(K)}(X) \neq \pi_K(X)) > 0$, then

$$\mathbb{E} [\|X - (\pi_K(X) + \pi_{A_\theta(K)}(X))/2\|^2] < \Delta(h).$$

Since the random variable $(\pi_K(X) + \pi_{A_\theta(K)}(X))/2$ takes its values in the image of the curve $(h + h_\theta)/2$ with length smaller than $\mathcal{L}(h) \leq L$, that is not possible. Consequently, $\pi_{A_\theta(K)}(X) = \pi_K(X)$ almost surely. In other words, $\pi_{A_\theta(K)}(g(t)) = h(t)$ for almost every $t \in [0, 1]$, and, thus, by continuity, $h_\theta(t) = h(t)$ for every $t \in [0, 1]$. For $t \in [0, 1]$, let $\theta = t$. We have $h(t) = h_t(t) = A_t \circ \pi_K(g(0)) = A_t(h(0))$. Since $\Delta(f) = \Delta(h)$ and $\mathcal{L}(h) \leq L$, $\mathcal{L}(h) = L$. Hence, $\text{Im}h$ is the circle with center $(0, 0)$ and radius $L/2\pi$.

□

Remark 9. Observe that radial symmetry of a distribution is not sufficient to guarantee that a given circle will be a constrained principal curve for this distribution. Let us exhibit two counterexamples.

- Let $p > 0$ and let \mathcal{U} denote the uniform distribution on the unit circle. Consider a random variable X taking its values in \mathbb{R}^2 , distributed according to the mixture distribution

$$p\delta_{(0,0)} + (1-p)\mathcal{U},$$

where $\delta_{(0,0)}$ stands for the Dirac mass at the origin $(0,0)$. Then, for every circle with center $(0,0)$ and radius $r \in (0,1]$, because of the atom at the origin, the projection of X on the circle is not unique almost surely, which implies, thanks to Proposition 3.1, that none of these circles may be a constrained principal curve for X .

- We consider the case where X is a standard Gaussian random vector in \mathbb{R}^2 . Lemma 3.2 ensures that the circle with center $(0,0)$ and radius $\mathbb{E}[\|X\|] = \sqrt{\pi/2}$ cannot be a constrained principal curve for X because it is self-consistent.

References

- 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.
- J. D. Banfield and A. E. Raftery. Ice floe identification in satellite images using mathematical morphology and clustering about principal curves. *Journal of the American Statistical Association*, 87:7–16, 1992.
- G. Biau and A. Fischer. Parameter selection for principal curves. *IEEE Transactions on Information Theory*, 58:1924–1939, 2012.
- C. Brunson. Path estimation from GPS tracks. In *Proceedings of the 9th International Conference on GeoComputation, National Centre for Geocomputation, National University of Ireland, Maynooth, Eire*, 2007.
- 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.
- G. Buttazzo, E. Mainini, and E. Stepanov. Stationary configurations for the average distance functional and related problems. *Control and Cybernetics*, 38:1107–1130, 2009.
- B. S. Caffo, C. M. Crainiceanu, L. Deng, and C. W. Hendrix. A case study in pharmacologic colon imaging using principal curves in single photon emission computed tomography. *Journal of the American Statistical Association*, 103:1470–1480, 2008.
- P. J. Corkeron, P. Anthony, and R. Martin. Ranging and diving behaviour of two ‘offshore’ bottlenose dolphins, *Tursiops* sp., off eastern Australia. *Journal of the Marine Biological Association of the United Kingdom*, 84:465–468, 2004.
- G. De’ath. Principal curves: a new technique for indirect and direct gradient analysis. *Ecology*, 80:2237–2253, 1999.
- P. Delicado. Another look at principal curves and surfaces. *Journal of Multivariate Analysis*, 77: 84–116, 2001.

- Y. Drier, M. Sheffer, and E. Domany. Pathway-based personalized analysis of cancer. *PNAS*, 110(1):6388–6393, 2013.
- 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. Einbeck, G. Tutz, and L. Evers. Local principal curves. *Statistics and Computing*, 15:301–313, 2005a.
- J. Einbeck, G. Tutz, and L. Evers. Exploring multivariate data structures with local principal curves. In C. Weihs and W. Gaul, editors, *Classification – The Ubiquitous Challenge, Proceedings of the 28th Annual Conference of the Gesellschaft für Klassifikation, University of Dortmund*, Studies in Classification, Data Analysis, and Knowledge Organization, pages 256–263. Springer, Berlin, Heidelberg, 2005b.
- H. Federer. Curvature measures. *Transactions of the American Mathematical Society*, 93:418–491, 1959.
- A. Fischer. Selecting the length of a principal curve within a Gaussian model. *Electronic Journal of Statistics*, 7:342–363, 2013.
- H. Friedsam and W. A. Oren. The application of the principal curve analysis technique to smooth beamlines. In *Proceedings of the 1st International Workshop on Accelerator Alignment*, 1989.
- 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, 2012.
- S. Gerber and R. Whitaker. Regularization-free principal curve estimation. *Journal of Machine Learning Research*, 14:1285–1302, 2013.
- T. Hastie and W. Stuetzle. Principal curves. *Journal of the American Statistical Association*, 84:502–516, 1989.
- E. Hewitt and K. Stromberg. *Real and Abstract Analysis: A modern treatment of the theory of functions of a real variable*. Springer-Verlag, 1965.
- J. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of convex analysis*. Springer Science & Business Media, 2012.
- B. Kégl and A. Krzyżak. Piecewise linear skeletonization using principal curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:59–74, 2002.
- B. Kégl, A. Krzyżak, T. Linder, and K. Zeger. Learning and design of principal curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:281–297, 2000.
- A. Lemenant. About the regularity of average distance minimizers in \mathbb{R}^2 . *J. Convex Anal.*, 18:949–981, 2011.
- A. Lemenant. A presentation of the average distance minimizing problem. *Journal of Mathematical Sciences*, 181:820–836, 2012.
- X. Y. Lu and D. Slepčev. Properties of minimizers of average-distance problem via discrete approximation of measures. *SIAM J. Math. Anal.*, 45:3114–3131, 2013.
- X. Y. Lu and D. Slepčev. Average-distance problem for parameterized curves. *ESAIM: Control, Optimisation and Calculus of Variations*, 22:404–416, 2016.

- C. Mantegazza and A. Mennucci. Hamilton-Jacobi equations and distance functions in Riemannian manifolds. *Appl. Math. Optim.*, 47:1–25, 2003.
- S. J. N. Mosconi and P. Tilli. Γ -convergence for the irrigation problem. *Journal of Convex Analysis*, 12:145–158, 2005.
- U. Ozertem and D. Erdogmus. Locally defined principal curves and surfaces. *Journal of Machine Learning Research*, 12:1249–1286, 2011.
- E. Paolini and E. Stepanov. Qualitative properties of maximum and average distance minimizers in \mathbb{R}^n . *J. of Math. Sci.*, 122:3290–3309, 2004.
- P. Polak and G. Wolansky. The lazy travelling salesman problem in \mathbb{R}^2 . *ESAIM: Control, Optimisation and Calculus of Variations*, 13:538–552, 2007.
- K. Reinhard and M. Niranjan. Parametric subspace modeling of speech transitions. *Speech Communication*, 27:19–42, 1999.
- S. Sandilya and S. R. Kulkarni. Principal curves with bounded turn. *IEEE Transactions on Information Theory*, 48:2789–2793, 2002.
- D. C. Stanford and A. E. Raftery. Finding curvilinear features in spatial point patterns: principal curve clustering with noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:2237–2253, 2000.
- T. Tarpey and B. Flury. Self-consistency: a fundamental concept in statistics. *Statistical Science*, 11:229–243, 1996.
- R. Tibshirani. Principal curves revisited. *Statistics and Computing*, 2:183–190, 1992.
- J. J. Verbeek, N. Vlassis, and B. Kröse. A soft k-segments algorithm for principal curves. In *Proceedings of International Conference on Artificial Neural Networks 2001*, pages 450–456, 2001.
- W. C. K. Wong and A. C. S. Chung. Principal curves to extract vessels in 3D angiograms. In *Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW’08)*, pages 1–8, 2008.

Sylvain Delattre & Aurélie Fischer
 Laboratoire de Probabilités et Modèles Aléatoires
 Université Paris Diderot
 Bâtiment Sophie Germain
 Case courrier 7012
 75205 Paris Cedex 13, France