



**HAL**  
open science

## **Fermat Distance-to-Measure: a robust Fermat-like metric**

Jérôme Taupin, Frédéric Chazal

► **To cite this version:**

Jérôme Taupin, Frédéric Chazal. Fermat Distance-to-Measure: a robust Fermat-like metric. 2025. ⟨hal-05015471v2⟩

**HAL Id: hal-05015471**

**<https://hal.science/hal-05015471v2>**

Preprint submitted on 21 Jan 2026

**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

# FERMAT DISTANCE-TO-MEASURE: A ROBUST MEASURE-BASED METRIC

Jérôme Taupin\*      Frédéric Chazal\*

January 21, 2026

## Abstract

Given a probability measure with density, Fermat distances and density-driven metrics are conformal transformations of the Euclidean metric that shrink distances in high density areas and enlarge distances in low density areas. Although they have been widely studied and have shown to be useful in various machine learning tasks, they are limited to measures with density (with respect to Lebesgue measure, or volume form on manifold). In this paper, by replacing the density with the *distance-to-measure* (DTM), we introduce a new metric, the *Fermat distance-to-measure* (FDTM), defined for any probability measure in  $\mathbb{R}^d$ . We derive strong stability properties for the FDTM with respect to the measure and propose an estimator from random sampling of the same measure, featuring an explicit bound on its convergence rate.

## 1 Introduction

Fermat distances belong to the class of density-driven metrics that have been established as a useful tool for machine learning tasks, such as semi-supervised learning or clustering, when the Euclidean metric does not sufficiently exploit the geometry of the distribution underlying the data. Given a measure  $\mu$  with density  $f$  with respect to the Lebesgue measure in a Euclidean space or the volume form in a more general Riemannian manifold, the Fermat distance between two points  $x$  and  $y$  is defined as

$$\inf_{\gamma} \int_{\gamma} f^{-\beta} \tag{1}$$

where the infimum is taken over all rectifiable paths  $\gamma$  from  $x$  to  $y$  and  $\beta \geq 0$  is a parameter. The Fermat distance has long been studied, with applications ranging from unsupervised and semi-supervised machine learning to persistent homology computation in topological data analysis and signal processing—see, e.g., the recent paper [10] and the references therein. It also possesses the remarkable property that it can be estimated from the metric induced by a simple weighted graph built on top of random samples of  $\mu$ , called the sample Fermat distance. Although this estimator comes with convergence guarantees as the sample size increases [11], quantitative bounds come on the other hand at the price of strict smoothness conditions on the density  $f$  [8, 10]. Moreover, the existence of the continuous Fermat distance requires the measure  $\mu$  to have a well-defined density  $f$ , restricting the convergence guarantees of the sample Fermat distance to such measures.

The aim of this paper is to introduce a new family of measure-based metrics—called FDTM—that overcomes the above limitations and holds strong stability properties with respect to perturbations of the measure. We rely on the notion of DTM functions introduced in [4]. Given any probability measure on  $\mathbb{R}^d$ , the DTM is a distance-like function that behaves similarly to the inverse of a density. Intuitively, the DTM function associated to a measure  $\mu$  at a point  $x \in \mathbb{R}^d$  represents how far  $x$  is from a given fraction  $m$  of the total mass of  $\mu$ . It has been proven [1] that the DTM can be viewed as an estimator of the density of  $\mu$  (when it is absolutely continuous with respect to the Lebesgue measure) and our approach roughly consists in replacing  $f^{-\beta}$  in Eq. (1) with the DTM. Thanks to regularity and stability properties of the DTM, the FDTM

---

\*Inria Saclay & Institut de Mathématiques d’Orsay

presents several benefits compared to the classical Fermat distance. First, it is well-defined for any measure  $\mu$ , regardless of the existence of a density. Second, we provide quantitative stability results for the FDTM with respect to (w.r.t.) perturbations of the measure in a Wasserstein metric and to perturbations of its support in Hausdorff distance. Last, we propose a natural estimator of the FDTM from random samples of the measure and, building on the stability properties of the FDTM, provide explicit convergence rates for this estimator that do not depend on the ambient dimension but on some mild regularity properties of the sampled measure.

The formal definition of FDTM and the statement of the above results are provided in Section 2. Section 3 is devoted to the study of geodesics and stability properties of the FDTM. The estimation of the FDTM from random samples is studied in Section 4 where an effective estimator is provided together with an upper bound on its convergence rate. A similar lower bound is also provided on the convergence rate of any estimator in the worst case w.r.t. the measure. Finally, we give some numerical illustrations of our results in Section 5. Although most of the proofs of the paper rely on simple ideas, they often require technical arguments and intermediate results that are detailed in the appendices.

## 2 Definitions and Overview of the Results

The setting of this paper is the space  $\mathbb{R}^d$  endowed with the Euclidean norm  $\|\cdot\|$ . A list of notations used throughout this paper is given in Table 1. In this section, we provide the definition of the DTM and FDTM then give an overview of the results we obtain.

### 2.1 The Fermat Distance-to-Measure Metric

Given a probability measure  $\mu$  over  $\mathbb{R}^d$  and two real parameters  $m \in (0, 1]$  and  $p \geq 1$ , the DTM [4] is defined for all  $x \in \mathbb{R}^d$  as

$$d_\mu(x) \stackrel{\text{def}}{=} \left( \frac{1}{m} \int_0^m \delta_{\mu,u}(x)^p \, du \right)^{1/p}$$

where

$$\delta_{\mu,u}(x) \stackrel{\text{def}}{=} \inf \{ r > 0 : \mu(\mathcal{B}(x, r)) > u \}$$

is the pseudo-DTM and represents the minimal radius of a ball centered at  $x$  that contains a mass  $u$  w.r.t.  $\mu$ . In this paper, the parameters  $m$  and  $p$  are fixed constants hence we omit the notational dependency of the DTM w.r.t. these parameters to avoid heavy notations. Notice that  $\delta_{\mu,u}$  is a finite function everywhere for all  $u < 1$ . Therefore, when  $m < 1$ ,  $d_\mu$  is well-defined and finite over the whole space  $\mathbb{R}^d$ . Moreover, it was proven in [4] that  $d_\mu$  is 1-Lipschitz, i.e.,

$$|d_\mu(x) - d_\mu(y)| \leq \|x - y\| \text{ for any } x, y \in \mathbb{R}^d, \quad (2)$$

and stable with respect to  $\mu$ , i.e.,

$$\|d_\mu - d_\nu\|_\infty \leq \frac{W_p(\mu, \nu)}{m^{1/p}} \quad (3)$$

where  $W_p$  denotes the  $p$ -Wasserstein distance defined by

$$W_p(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \left( \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^p \pi(dx, dy) \right)^{\frac{1}{p}}$$

with  $\Pi(\mu, \nu)$  the set of probability measures on  $\mathbb{R}^d \times \mathbb{R}^d$  that have marginals  $\mu$  and  $\nu$ .

The function  $\delta_{\mu,u}$  generalizes the distance function to the support  $\mathcal{S}(\mu)$  of  $\mu$ —the latter being exactly  $\delta_{\mu,0}$ —and reflects how far  $x$  is from a mass  $u$  of the measure  $\mu$ . The DTM, obtained by taking the  $p$ -Hölder mean of  $\delta_{\mu,u}$  over  $u \in [0, m]$ , intuitively behaves like the inverse of a density:  $d_\mu(x)$  is large when  $x$  belongs to a low-density area or when  $x$  is far away from the support of  $\mu$ , and  $d_\mu(x)$  is small when  $x$  is in a high-density

area. Precisely, in the case where  $\mu$  has a continuous density  $f$  w.r.t. the Lebesgue measure over  $\mathbb{R}^d$ , it has been shown [1] that  $md_\mu^{-d}$  converges to the density  $f$  as  $m$  goes to 0.

These properties motivate the introduction of the FDTM between two points as the integral of the DTM—possibly elevated at some power greater than 1—minimized over paths between  $x$  and  $y$ .

**Definition 1** (Fermat distance-to-measure (FDTM)). Let  $\mu$  be a probability measure over  $\mathbb{R}^d$ ,  $F$  be a closed subset of  $\mathbb{R}^d$  and  $\beta \geq 1$  be a power parameter. Given a rectifiable path  $\gamma : [0, 1] \rightarrow \mathbb{R}^d$ , we call FDTM length of the path  $\gamma$  the quantity

$$D_\mu(\gamma) \stackrel{\text{def}}{=} \int_\gamma d_\mu^\beta = \int_0^1 d_\mu(\gamma(t))^\beta \|\dot{\gamma}(t)\| dt .$$

Given  $x, y \in \mathbb{R}^d$ , we call FDTM between  $x$  and  $y$  with measure  $\mu$  and domain  $F$  the quantity

$$D_{\mu, F}(x, y) \stackrel{\text{def}}{=} \inf_{\gamma \in \Gamma_F(x, y)} D_\mu(\gamma) ,$$

where  $\Gamma_F(x, y)$  is the set of rectifiable paths  $\gamma : [0, 1] \rightarrow \mathbb{R}^d$  such that  $\gamma(0) = x$ ,  $\gamma(1) = y$  and with the additional constraint that  $\gamma$  may leave the domain  $F$  only by drawing straight lines: Formally, for all open interval  $I \subset [0, 1]$  such that  $\gamma(I) \subset \mathbb{R}^d \setminus F$ , there exists a unit vector  $u \in \mathbb{R}^d$  such that for all  $t \in I$ ,

$$\dot{\gamma}(t) = \|\dot{\gamma}(t)\| \cdot u .$$

As for  $m$  and  $p$ ,  $\beta$  is a fixed parameter and we omit the notational dependency w.r.t.  $\beta$ . In the following, we shall focus on the case where the domain  $F$  is chosen as the support of the measure  $\mu$ , see below for more explanations. The FDTM can be interpreted as a generalization of the Fermat distance. Indeed, the behavior of the latter is expected to be retrieved when choosing  $m$  close to 0 due to the DTM approaching the density.

The rectifiable paths  $\gamma$  belonging to  $\Gamma_F(x, y)$  are called *admissible paths*. Recall that a path  $\gamma$  is rectifiable if  $\|\dot{\gamma}\|$  is defined almost everywhere on  $[0, 1]$  and integrable. Its Euclidean length is then denoted by

$$|\gamma| \stackrel{\text{def}}{=} \int_0^1 \|\dot{\gamma}(t)\| dt < +\infty .$$

The notion of integration along a rectifiable path is well-defined. Moreover, such path may always be reparameterized as a Lipschitz path. In the sequel, paths  $\gamma$  will often be implicitly assumed to be parameterized with constant Euclidean velocity, i.e., such that  $t \mapsto \|\dot{\gamma}(t)\|$  is constant where it is defined. Some proofs shall instead consider an arc length parameterization, where  $\gamma$  is defined over  $[0, |\gamma|]$  and such that  $\|\dot{\gamma}(t)\| = 1$  almost everywhere. The most basic path in  $\Gamma_F(x, y)$  is the segment  $t \mapsto (1-t)x + ty$ , which will be written as  $[x, y]$  for short. Its FDTM length is given by

$$D_\mu([x, y]) = \|x - y\| \int_0^1 d_\mu((1-t)x + ty)^\beta dt .$$

Note that the straight path is always admissible regardless of  $F$  since its derivative is constant. More generally, we shall refer as  $[x_1, x_2, \dots, x_k]$  to the polygonal path going through  $x_1, x_2, \dots, x_k$  in this order, which is admissible if all intermediate points belong to the domain.

We now discuss a few remarks before stating our contributions.

*Restriction to a compact subset  $K$ .* Throughout the paper,  $K \subset \mathbb{R}^d$  denotes a compact convex set with diameter  $\text{diam}(K)$ . Without loss of generality and to achieve clearer results, we choose to restrict all objects to be contained within  $K$ , that is, support of measures, domains and endpoints. The main benefit of this assumption is to uniformly upper bound the DTM over  $K$  by  $\text{diam}(K)$ , which in particular also holds over all admissible paths. Indeed, if the support is in  $K$  then all points in  $K$  are at distance at most  $\text{diam}(K)$  from any point of the support, thus the DTM is upper bounded by  $\text{diam}(K)$ . Denote  $\mathcal{M}_K$  the set of measures with support included in  $K$ . Then for any such measure  $\mu \in \mathcal{M}_K$ , domain  $F \subset K$  and endpoints  $x, y \in K$ ,

$$\|d_\mu\|_{\infty, K} \leq \text{diam}(K) \text{ and } \|d_\mu\|_{\infty, \gamma} \leq \text{diam}(K) \text{ for any } \gamma \in \Gamma_F(x, y) . \quad (4)$$

Furthermore, intermediate results and proofs will often assume that  $\text{diam}(K) = 1$  for clarity purposes. The fact that these assumptions are without loss of generality is discussed in Section A.1. In particular, the implication that  $\mu$  has compact support in  $\mathbb{R}^d$  is not very restrictive, as we show in Section A.1 that given any measure  $\mu$  and any compact area of study  $A \subset \mathbb{R}^d$ , the restriction of the FDTM of  $\mu$  over  $A$  coincides with the FDTM of a “cropped” modification of  $\mu$  with support included in a larger compact subset  $K$ .

NOTATION	DEFINITION
$\ \cdot\ $	Euclidean norm over $\mathbb{R}^d$
$\mathcal{B}(x, r)$	open ball centered at $x \in \mathbb{R}^d$ with radius $r$
$\overline{\mathcal{B}}(x, r)$	closed ball centered at $x \in \mathbb{R}^d$ with radius $r$
$\mu, \nu$	measures over $\mathbb{R}^d$
$\mathcal{S}(\mu)$	support of the measure $\mu$
$F, G$	domains for paths
$m$	fraction of mass considered for the DTM
$p$	exponent of the Hölder mean considered for the DTM
$\beta$	exponent of the integrated DTM considered for the FDTM
$K$	compact convex subset of $\mathbb{R}^d$ (often with diameter 1)
$a, b$	parameters used of the standard assumption on measures
$\sigma$	lower bound on the DTM $d_\mu$
$\mathcal{M}_K$	set of measures with support included in $K$
$\mathcal{M}_{K,a,b,\sigma}$	set of measures satisfying Assumption (A)
$d_\mu$	DTM associated with $\mu$
$\gamma$	rectifiable path in $\mathbb{R}^d$
$ \gamma $	Euclidean length of path $\gamma$
$\ d_\mu\ _{\infty, \gamma}$	maximal value of the DTM reached on the path $\gamma$
$L_{\mu, \delta}$	sublevel set of $d_\mu$ below the value $\delta$
$D_\mu(\gamma)$	FDTM length of the path $\gamma$
$D_{\mu, F}(x, y)$	FDTM associated with $\mu$ and $F$ from $x$ to $y$
$D_\mu(x, y)$	Special case when $F = \mathcal{S}(\mu)$
$\Gamma_F(x, y)$	Set of rectifiable paths with domain $F$ from $x$ to $y$
$\Gamma_{\mu, F}^*(x, y)$	Set of geodesics paths for $\mu$ with domain $F$ from $x$ to $y$
$[x_1, x_2, \dots, x_k]$	polygonal path going through $x_1, x_2, \dots, x_k$
$W_p$	Wasserstein distance of order $p$
$d_{\text{TV}}$	Total-Variation distance
$d_{\text{H}}$	Hausdorff distance between closed subsets of $\mathbb{R}^d$
$n$	Number of sample points
$\mathbb{X}_n$	Family of $n$ random samples from a measure $\mu$
$\hat{\mu}_n$	Empirical measure associated with $\mathbb{X}_n$

Table 1: Notations used throughout the paper

*Lower bounding the DTM.* For  $d_\mu$  to vanish at a point  $x$ ,  $\mu$  needs to have an atom of mass at least  $m$  at  $x$ . Assuming that this does not happen anywhere,  $d_\mu$  is positive and, as a continuous and proper function, achieves a positive minimum over  $\mathbb{R}^d$ :

**Lemma 2.** Let  $\mu$  be a probability measure over  $\mathbb{R}^d$  such that  $\mu(\{x\}) < m$  for all  $x \in \mathbb{R}^d$ . Then  $\min d_\mu > 0$ .

*About the role of  $F$ .* The domain  $F \subset \mathbb{R}^d$  is introduced to achieve both general stability results and interesting convergence rates for the estimation of the FDTM in a way that allows to evaluate all admissible paths with reasonable computational complexity independent of the ambient dimension  $d$ —see Section 4. In this case  $F$  will be chosen to be the support of  $\mu$ , so that sampling points according to  $\mu$  creates a point cloud that fills the domain and allows to recover all admissible paths. On the other hand, straight edges outside the domain are allowed in order to improve the stability of the FDTM w.r.t. the domain, and do not affect the estimation properties by a point cloud as the straight edges can still be approximated.

In general, if  $F \subset G \subset \mathbb{R}^d$  then  $\Gamma_F(x, y) \subset \Gamma_G(x, y)$  for all endpoints  $x$  and  $y$ , and it follows immediately that  $D_{\mu, G}(x, y) \leq D_{\mu, F}(x, y)$ . Moreover,  $(F, D_{\mu, F})$  is a metric space (see Section A.2). In particular, the triangular inequality applies when considering endpoints within  $F$ . However, this is not true when considering general endpoints: If  $x, y \in F$  and  $z \notin F$ , it may happen that

$$D_\mu(x, z) + D_\mu(z, y) \leq D_\mu([x, z, y]) < D_\mu(x, y)$$

if the polygonal path  $[x, z, y]$  does not belong to  $\Gamma_F(x, y)$ .

## 2.2 Overview of the Results

We now summarize our contributions.

*Existence of geodesics and upper bound on their Euclidean length.* Without domain constraints, the FDTM metric can be seen as a continuous conformal transformation of the Euclidean metric via the DTM function, and the existence of minimizing geodesics follows from classical arguments (see for instance Proposition 2.5.19 in [3]). The proof is adapted to account for domain constraints in Section 3.1.

**Theorem 3** (Existence of geodesics). Let  $\mu$  be a probability measure over  $\mathbb{R}^d$  and  $F \subset \mathbb{R}^d$  a closed domain. Then for all  $x, y \in \mathbb{R}^d$ , there exists a path  $\gamma \in \Gamma_F(x, y)$  such that

$$D_{\mu, F}(x, y) = D_\mu(\gamma) .$$

We denote  $\Gamma_{\mu, F}^*(x, y)$  the set of such paths, called (minimizing) geodesics.

Notice that Theorem 3 holds regardless of the measure  $\mu$  or the domain  $F$ . In particular, the existence of geodesics holds for all measures in  $\mathbb{R}^d$ , including those without compact support. This is tied with the fact that geodesics may not escape a large ball around the endpoints due to the regularity properties of the DTM—see Corollary 6—regardless of the measure. This observation contrasts with the Fermat distances, for which one may construct a measure whose support is made of a family of disjoint strips from  $x$  to  $y$  that make increasingly longer detours while being thinner but with higher density. More precisely, consider a family of strips  $(s_k)_k$  with respective lengths  $(L_k)_k$ , thickness  $(h_k)_k$  and density  $(a_k)_k$  such that

- $L_k \rightarrow +\infty$ ,  $h_k \rightarrow 0$ ,  $a_k \rightarrow +\infty$ .
- The series with general term  $L_k h_k a_k$  converges to 1, ensuring that the underlying measure is a probability.
- $L_k a_k^{-\beta}$  is decreasing, ensuring that each strip is shorter in the sense of Eq. (1) despite being longer Euclidean-wise.

In this example each strip  $s_k$  yields a lower Fermat distance between  $x$  and  $y$  than the previous one, so that there are no geodesics for the Fermat metric.

In addition to the existence of geodesics, the 1-Lipschitz property of the DTM (Eq. 2) and the fact that it cannot take small values on a subset that is too large—see Lemma 13—allow to upper bound their Euclidean length, regardless of the underlying measure and domain. This upper bound then plays a fundamental role in establishing stability properties of the FDTM.

**Theorem 4** (Upper bound on the Euclidean length of geodesics). Let  $\mu$  be a probability measure over  $\mathbb{R}^d$  and  $F \subset \mathbb{R}^d$  a closed domain. Then for all geodesic  $\gamma \in \Gamma_{\mu, F}^*(x, y)$ ,

$$|\gamma| \lesssim \|d_\mu\|_{\infty, \gamma} \tag{5}$$

where  $\lesssim$  hides a multiplicative constant—made explicit in the proofs—only depending on  $m$  and  $\beta$ .

**Corollary 5.** Assume that  $\mathcal{S}(\mu) \subset K$ ,  $F \subset K$ ,  $x, y \in K$  and let  $\gamma \in \Gamma_{\mu, F}^*(x, y)$ . Then

$$|\gamma| \lesssim \text{diam}(K)$$

where  $\lesssim$  hides the same multiplicative constant as in Eq. (5).

**Corollary 6.** Let  $\gamma \in \Gamma_{\mu, F}^*(x, y)$ . Then

$$|\gamma| \lesssim \|x - y\|$$

where  $\lesssim$  hides a multiplicative constant depending on  $\beta$  and the constant from Eq. (5).

The reasoning to derive Theorem 4 is described in Section 3.2. Corollary 5 is a direct consequence of Eqs. (4) and (5). As for Corollary 6, it is also a direct consequence of Eq. (5) if  $\|d_\mu\|_{\infty, \gamma}$  is smaller than  $\|x - y\|$ . Else, it is still possible to obtain the domination by the Euclidean distance by simpler means, which is detailed in Section B.3.

Recall that under mild assumptions on the measure  $\mu$ , the DTM is lower bounded by a positive quantity (see Lemma 2). In this special case the Euclidean length of geodesics can be upper bounded in a straightforward manner, albeit with a dependency in the DTM lower bound. Proposition 7 is also proven in Section 3.2.

**Proposition 7.** Assume that  $\min d_\mu > 0$  and let  $\gamma \in \Gamma_{\mu, F}^*(x, y)$ . Then

$$|\gamma| \leq \left( \frac{\max_{[x, y]} d_\mu}{\min d_\mu} \right)^\beta \|x - y\| .$$

Under the same assumptions as in Corollary 5, this further implies that

$$|\gamma| \leq \left( \frac{\text{diam}(K)}{\min d_\mu} \right)^\beta \text{diam}(K) .$$

*Stability of the FDTM.* Beyond the fact that the FDTM metric is defined for any probability measure on  $\mathbb{R}^d$ , one of the main motivations to introduce it as a variant of the Fermat distance is that it turns out to be more stable w.r.t. the measure. This property follows from Theorem 4 and the Wasserstein stability of the DTM (Eq. 3) and is proven in Section 3.3.1.

**Theorem 8** (FDTM stability w.r.t. the measure). Let  $\mu, \nu \in \mathcal{M}_K$  and  $F \subset K$  a closed domain. Then

$$\|D_{\mu, F} - D_{\nu, F}\|_{\infty, K} \lesssim \text{diam}(K)^\beta W_p(\mu, \nu)$$

where  $\lesssim$  hides a multiplicative constant—made explicit in the proofs—only depending on  $m$  and  $\beta$ .

Likewise, the FDTM is stable w.r.t. the domain. This property is obtained by approximating a geodesic for the first domain using a path that is admissible for the second domain—see Section 3.3.2.

**Theorem 9** (FDTM stability w.r.t. the domain). Let  $\mu \in \mathcal{M}_K$  and  $F, G \subset K$  two closed domains. Then

$$\|D_{\mu, F} - D_{\mu, G}\|_{\infty, K} \lesssim \text{diam}(K)^{\beta + \frac{1}{2}} \sqrt{d_H(F, G)}$$

where  $\lesssim$  hides a multiplicative constant—made explicit in the proofs—only depending on  $m$  and  $\beta$  and  $d_H(F, G)$  denotes the Hausdorff distance between  $F$  and  $G$  defined as

$$d_H(F, G) \stackrel{\text{def}}{=} \max(\|d(\cdot, G)\|_{\infty, F}, \|d(\cdot, F)\|_{\infty, G}) .$$

Both Theorems 8 and 9 require no assumption on the regularity of the measures and domains. This is possible due to the definition of admissible paths allowing “shortcuts” to override irregularities in the domain.

*Empirical FDTM.* The previous stability results provide efficient tools to study the estimation of the FDTM from random samples. Given  $n \geq 1$  sample points  $(X_1, \dots, X_n)$  drawn independently and identically distributed (i.i.d.) from  $\mu$ , we denote  $\mathbb{X}_n = \{X_1, \dots, X_n\}$  the corresponding point cloud and  $\hat{\mu}_n = \frac{1}{n} \sum_{k=1}^n \delta_{X_k}$  the empirical measure. Estimation results are restricted to the special case where the domain  $F$  is chosen as the support  $\mathcal{S}(\mu)$  of the measure, and we denote  $D_\mu = D_{\mu, \mathcal{S}(\mu)}$  for simplicity. Indeed, to avoid a strong computation cost depending on the ambient dimension  $d$ , it is easier to compute paths made up of straight lines between sample points, which are exactly the paths admissible for the domain  $\mathbb{X}_n$ . Since  $d_{\hat{\mu}_n}$  is expected

to converge towards  $d_\mu$  and  $\mathbb{X}_n = \mathcal{S}(\hat{\mu}_n)$  is expected to fill up the whole support  $\mathcal{S}(\mu)$ , the FDTM  $D_\mu$  can be estimated by the empirical FDTM  $D_{\hat{\mu}_n} = D_{\hat{\mu}_n, \mathbb{X}_n}$ , which is the metric of a finite weighted graph. The weights are the integrals of the DTM over edges, hence cannot be reasonably computed in general. In practice, as the DTM is 1-Lipschitz, each integral can be approximated by a discrete sum over a regular subdivision of the underlying edge, using finer subdivisions as  $n$  grows to retain convergence properties. For simplicity we keep the original definition of the weights in our theoretical results. Practical considerations are discussed in Section 4.3.

In order to control the convergence rate of the empirical FDTM over  $K$ , the offset is decomposed into two terms as

$$\|D_\mu - D_{\hat{\mu}_n}\|_{\infty, K} \leq \|D_{\mu, \mathcal{S}(\mu)} - D_{\hat{\mu}_n, \mathcal{S}(\mu)}\|_{\infty, K} + \|D_{\hat{\mu}_n, \mathcal{S}(\mu)} - D_{\hat{\mu}_n, \mathbb{X}_n}\|_{\infty, K}.$$

Then, Theorems 8 and 9 are used to upper bound both terms by quantities for which convergence rates can be derived. To do so, the following assumption is used.

**Assumption (A).**

- (i) The support of  $\mu$  is included in a compact subset of  $\mathbb{R}^d$ :  $\mathcal{S}(\mu) \subset K$ .
- (ii)  $\mu$  is an  $(a, b)$ -standard probability measure for some  $a > 0$  and  $b \geq 1$ , i.e., for all  $x \in \mathcal{S}(\mu)$  and  $0 < r \leq a^{-1/b}$ ,

$$\mu(\mathcal{B}(x, r)) \geq ar^b.$$

- (iii)  $\min d_\mu \geq \sigma$  for some  $\sigma > 0$ .

- (iv)  $\mathcal{S}(\mu)$  is connected.

Recall that Assumption (A).(i) is without loss of generality (see Section A.1). Assumption (A).(ii) is the most important assumption and will direct the rate of convergence of the empirical FDTM.  $b$  may be interpreted as an upper bound on the intrinsic dimension of the support of  $\mu$ . Regarding Assumption (A).(iii), Lemma 2 ensures that it is satisfied by any measure without atoms of mass larger than  $m$ . Assumption (A).(iii) and (iv) are technical assumptions that do not have a fundamental impact on the results. We denote  $\mathcal{M}_{K, a, b, \sigma}$  the set of measures satisfying Assumption (A) for given  $K$ ,  $a$ ,  $b$  and  $\sigma$ , and can now state our main result.

**Theorem 10** (Convergence of the empirical FDTM). Assume that  $\mu \in \mathcal{M}_{K, a, b, \sigma}$  satisfies Assumption (A) and that  $m \leq \frac{1}{2}$ . Then for all  $n \geq \frac{1}{m}$ ,

$$\mathbb{E} \left[ \|D_\mu - D_{\hat{\mu}_n}\|_{\infty, K} \right] \lesssim d \text{diam}(K)^{\beta+1} \frac{\log(n)}{n^{\frac{1}{2b}}}$$

where  $\lesssim$  hides a multiplicative constant depending on  $m$ ,  $p$ ,  $\beta$ ,  $a$ ,  $b$  and  $\sigma/\text{diam}(K)$ .

A more detailed statement is provided in Section 4.1. A weaker version may be stated without Assumption (A).(iii) and (iv), which does not require any lower bound on the DTM but features an additional term of order  $n^{-1/2p}$ , possibly achieving a worse rate of convergence when  $p$  is large.

The main strength of Theorem 10 is its generality. The most restrictive assumption is the  $(a, b)$ -standard assumption which remains light compared to regularity conditions on the density or on the support of the measure. Compared to the convergence result provided by [11, Theorem 2.3] for the Fermat distance, Theorem 10 is not reduced to the asymptotic regime and does not involve any normalization constant.

Another idea to achieve a similar result in the case where the Fermat distance cannot be defined do to  $\mu$  not having a density, is to use the original Fermat distance but with a “smoothed” version of  $\mu$ , for instance through a convolution with a Gaussian measure, instead of replacing the density with the DTM in Eq. (1). However by doing so the intrinsic dimension of the measure would default to the ambient dimension, which would affect negatively the convergence properties of the Fermat distance. In Theorem 10, the convergence rate does not involve the ambient dimension.

*Minimax lower bound.* Our final point of interest is to evaluate how close to optimal is our proposed FDTM estimator from a minimax viewpoint. We establish a lower bound on the convergence rate of any estimator of the FDTM in its worst case scenario for the measure  $\mu$ —see Section 4.2.

**Theorem 11** (Minimax lower bound). Fix  $K \subset \mathbb{R}^d$  and  $b \geq 1$  and assume that  $K \subset \mathbb{R}^d$  is not a segment. Then there exist  $x, y \in K$  and constants  $a_0, \sigma_0$  and  $n_0$  depending on  $m, p, \beta$  and  $b$  such that for all  $a \leq a_0, \sigma \leq \sigma_0$  and  $n \geq n_0$ ,

$$\min_{\hat{D} \in \mathcal{D}_n} \max_{\mu \in \mathcal{M}_{K,a,b,\sigma}} \mathbb{E} \left[ \left| \hat{D} - D_\mu(x, y) \right| \right] \gtrsim \frac{\|x - y\|^{\beta+1}}{n^{\frac{1}{b} \wedge \frac{1}{2}}}$$

where  $\mathcal{D}_n$  is the set of possible estimators of  $D_\mu(x, y)$  from  $n$  sample points and  $\gtrsim$  hides a multiplicative constant depending on  $m, p, \beta$  and  $b$ .

The above lower bound features a convergence rate depending on the inverse of the intrinsic dimension, similarly to the upper bound in Theorem 10, although a gap subsists when  $b \geq 2$ . Our estimator is therefore minimax optimal when  $b = 1$ . The minimax optimal convergence rate when  $b \geq 2$  remains unclear and its precise study is beyond the scope of this paper.

### 3 Geodesics and Stability Properties of the FDTM

This sections details how the existence of geodesics for the FDTM is ensured, how to obtain a uniform bound on their Euclidean length, and how this ensures quantitative stability bounds on the FDTM.

#### 3.1 Existence of Geodesics

Let  $\mu$  be a probability measure with support contained in a compact convex set  $K$ ,  $F \subset K$  a closed domain and  $x, y \in \mathbb{R}^d$ . Recall that these assumptions are without loss of generality (see Section A.1). Without domain restrictions, the existence of geodesics can be derived from compactness arguments (see Proposition 2.5.19 in [3]). In general, the FDTM model does not exactly fit the framework for this result due to the fact that the concatenation of two admissible paths may not be admissible if both paths meet outside the domain. In the following we adapt the proof to account for this issue.

*Proof of Theorem 3.* We first treat the case of endpoints  $x, y \in F$ . Note that the topologies of both the Euclidean and the FDTM metrics are equivalent (a rigorous explanation is given in Section A.2). In particular, since  $F$  is a closed subset of  $K$ ,  $(F, \|\cdot\|)$  is a compact metric space and so is  $(F, D_{\mu,F})$ . Moreover, the set of rectifiable curve, i.e., of curves with finite Euclidean length, is the same as the set of curves with finite FDTM length.

Let  $(\gamma_k)_{k \geq 1}$  a sequence of paths in  $\Gamma_F(x, y)$  such that  $\lim_{k \rightarrow +\infty} D_\mu(\gamma_k) = D_{\mu,F}(x, y)$ . This sequence has uniformly bounded FDTM length by definition. Therefore, according to [3, Theorem 2.5.14] and replacing the sequence with a subsequence if needed,  $\gamma_k$  converges uniformly to a rectifiable curve  $\gamma$ . Moreover, [3, Proposition 2.3.4.iv] states that  $\gamma \mapsto D_\mu(\gamma)$  is lower semi-continuous, which implies that

$$D_\mu(\gamma) \leq \liminf_{k \rightarrow +\infty} D_\mu(\gamma_k) = D_{\mu,F}(x, y) .$$

It remains to show that  $\gamma \in \Gamma_F(x, y)$ . It is immediate from the uniform convergence that  $\gamma(0) = x$  and  $\gamma(1) = y$ . Now, given any open interval  $I$  such that  $\gamma(I) \subset \mathbb{R}^d \setminus F$ , it suffices to show that  $\dot{\gamma}$  has constant direction over all segments  $J \subset I$ . First,  $\gamma(J)$  is compact and at positive distance away from  $F$ , so that by uniform convergence  $\gamma_k(J) \subset \mathbb{R}^d \setminus F$  for  $k$  large enough, hence  $\dot{\gamma}_k(t) = \|\dot{\gamma}_k(t)\| u_k$  for all  $t \in J$  where  $u_k$  is a unit vector. Then for all  $s < t \in J$ ,

$$\gamma_k(t) - \gamma_k(s) = \int_s^t \|\dot{\gamma}_k(t)\| u_k dt = \int_s^t \|\dot{\gamma}_k(t)\| dt \cdot u_k .$$

Due to the uniform convergence of  $\gamma_k$  to  $\gamma$ ,  $\gamma_k(t) - \gamma_k(s)$  converges to  $\gamma(t) - \gamma(s)$ .  $u_k$  is a unit vector and therefore

$$\lim_{k \rightarrow +\infty} u_k = \lim_{k \rightarrow +\infty} \frac{\gamma_k(t) - \gamma_k(s)}{\|\gamma_k(t) - \gamma_k(s)\|} = \frac{\gamma(t) - \gamma(s)}{\|\gamma(t) - \gamma(s)\|} .$$

Since this is true for arbitrary  $s, t \in J$ , we get

$$\dot{\gamma}(t) = \|\dot{\gamma}(t)\|u$$

for all  $t$  in the interior of  $J$ , where  $u = \lim_{k \rightarrow +\infty} u_k$ . This concludes the fact that  $\gamma \in \Gamma_F(x, y)$ . Therefore,  $\gamma$  is a geodesic.

If  $x$  does not belong to  $F$ , notice that any path in  $\Gamma_F(x, y)$  is made of a straight line  $[x, z]$  going to  $F$  followed by a path in  $\Gamma_F(z, y)$ , where  $z \in F$  may be equal to  $y$ . Thus

$$D_{\mu, F}(x, y) = \inf_{z \in F} (D_{\mu}([x, z]) + D_{\mu, F}(z, y))$$

is the infimum of a continuous function of  $z$  over the compact set  $F$ , hence a minimum. Therefore, there exists a geodesic. The same reasoning goes for  $y \notin F$  with the addition of the direct path  $[x, y]$  which may not go through  $F$  at all. This concludes the proof in the general case.  $\square$

## 3.2 Bounding the Length of Geodesics

In the particular case where the DTM is lower bounded by a positive quantity, a bound can be obtained on the length of geodesics depending on the lower bound by using elementary arguments. In general, the properties of the DTM allow to obtain a bound that does not require any assumption on  $\mu$ , although this requires a more complex reasoning.

### 3.2.1 Case of a Lower Bounded DTM

When  $\min d_{\mu} > 0$ , the length of geodesics can easily be upper bounded. Indeed, the FDTM between two endpoints  $x$  and  $y$  is at most the FDTM length of the segment  $[x, y]$  and

$$D_{\mu, F}(x, y) \leq D_{\mu}([x, y]) \leq \|x - y\| \max_{[x, y]} d_{\mu}^{\beta}.$$

Then, if  $\gamma \in \Gamma_{\mu, F}^*(x, y)$ , by using the lower bound on the DTM,

$$D_{\mu, F}(x, y) = D_{\mu}(\gamma) \geq |\gamma|(\min d_{\mu})^{\beta}.$$

Therefore,

$$|\gamma| \leq \left( \frac{\max_{[x, y]} d_{\mu}}{\min d_{\mu}} \right)^{\beta} \|x - y\|$$

which proves Proposition 7.

### 3.2.2 General Case (sketch of proof)

We now focus on the general case where the upper bound from Theorem 4 can be derived with no dependency on the measure itself nor on the domain. The study of the Euclidean length of geodesic paths requires to consider the sublevel sets

$$L_{\mu, \delta} \stackrel{\text{def}}{=} \{x \in \mathbb{R}^d : d_{\mu}(x) < \delta\}$$

of the DTM. A key property of the DTM is that  $d_{\mu}(x)$  being small implies that a substantial amount of mass from  $\mu$  lies nearby  $x$ . This remark allows to upper bound the packing and covering numbers of  $L_{\mu, \delta}$ , for which we recall the definitions.

**Definition 12.** Let  $A \subset \mathbb{R}^d$ ,  $\mathcal{F}$  a finite family of points in  $A$  and  $r > 0$ .

- $\mathcal{F}$  is said to be a *r-packing* of  $A$  if points in  $\mathcal{F}$  are at least  $r$  apart from each other, i.e., if  $\mathcal{B}(x, r/2) \cap \mathcal{B}(x', r/2) = \emptyset$  for any distinct points  $x, x' \in \mathcal{F}$ .
- $\mathcal{F}$  is said to be a *r-covering* of  $A$  if all points in  $A$  are at most  $r$  apart from a point in  $\mathcal{F}$ , i.e., if  $A \subset \bigcup_{x \in \mathcal{F}} \mathcal{B}(x, r)$ .

The packing number is then defined as

$$\text{pack}(A, r) = \max \{ n \geq 0 : A \text{ admits a } r\text{-packing of size } n \}$$

and the covering number as

$$\text{cov}(A, r) = \min \{ n \geq 0 : A \text{ admits a } r\text{-covering of size } n \} .$$

A  $r$ -packing  $\mathcal{F}$  of  $A$  is *maximal* if no point from  $A$  can be appended to  $\mathcal{F}$  while preserving the packing property. A maximal  $r$ -packing is thus also a  $r$ -covering. In particular,  $\text{cov}(A, r) \leq \text{pack}(A, r)$ .

**Lemma 13** (Covering property of the DTM sublevels). Let  $\mu$  be a probability measure. Then for all  $\delta > 0$ ,

$$\text{cov}(L_{\mu, \delta}, 4\delta) \leq \text{pack}(L_{\mu, \delta}, 4\delta) \leq \frac{2}{m} .$$

*Proof.* Let  $\delta > 0$ . For all  $x \in L_{\mu, \delta}$ , since  $u \mapsto \delta_{\mu, u}(x)$  is non-decreasing,

$$\delta^p > d_\mu(x)^p = \frac{1}{m} \int_0^m \delta_{\mu, u}(x)^p du \geq \frac{1}{m} \int_{m/2}^m \delta_{\mu, u}(x)^p du \geq \frac{1}{2} \delta_{\mu, m/2}(x)^p ,$$

hence  $\delta_{\mu, m/2}(x) < 2^{1/p} \delta \leq 2\delta$ , which by definition implies that  $\mu(\mathcal{B}(x, 2\delta)) \geq \frac{m}{2}$ . Consider a maximal  $4\delta$ -packing of  $L_{\mu, \delta}$  by points  $(x_i)_{1 \leq i \leq N}$  where  $N = \text{pack}(L_{\mu, \delta}, 4\delta)$ . By definition, the balls  $(\mathcal{B}(x_i, 2\delta))_i$  are pairwise disjoint. Since  $\mu$  is a probability measure it follows that

$$1 \geq \mu \left( \bigsqcup_{i=1}^N \mathcal{B}(x_i, 2\delta) \right) = \sum_{i=1}^N \mu(\mathcal{B}(x_i, 2\delta)) \geq N \frac{m}{2} ,$$

therefore  $\text{pack}(L_{\mu, \delta}, 4\delta) \leq \frac{2}{m}$ . □

The upper bound on the Euclidean length of geodesics stated by Theorem 4 builds upon the Lipschitz property of the DTM (Eq. 2) and Lemma 13. Indeed, one can modify the trajectory of a geodesic path within a given sublevel to create an alternative path that avoids long squiggling in areas of low DTM. Both the Euclidean and FDTM lengths of this alternative path within the sublevel can be upper bounded knowing that the DTM is Lipschitz and that the size of the sublevel is proportional to the level itself.

Before giving more details on this reasoning we introduce some notations. To formalize the decomposition and concatenation of paths, we work in the Abelian free group generated by all paths where the  $+$  operator represents the concatenation in a general sense, that is without assuming that endpoints are shared. Precisely, we use the quotient group that identifies two general paths when they represent the same overall trajectory, e.g.,  $[x, y] + [y, z] = [y, z] + [x, y] = [x, y, z]$ . The Euclidean and FDTM lengths both naturally extend as morphisms to general paths:

$$|\gamma + \gamma'| = |\gamma| + |\gamma'| \quad \text{and} \quad D_\mu(\gamma + \gamma') = D_\mu(\gamma) + D_\mu(\gamma') .$$

This is also true for countable concatenations of paths. Given  $\gamma \in \Gamma_F(x, y)$  and  $\delta > 0$ , we call *sub- $\delta$  sections* of  $\gamma$ , denoted by  $[\gamma]^\delta$ , the connected sections of the path located in the sublevel area  $L_{\mu, \delta}$  of the DTM. Likewise, the remaining part of the path is called *super- $\delta$  sections* of  $\gamma$  and denoted by  $[\gamma]_\delta$ . The part of the path located in  $L_{\mu, \delta} \setminus L_{\mu, \delta'}$  is denoted by

$$[\gamma]_{\delta'}^\delta \stackrel{\text{def}}{=} [\gamma]^\delta - [\gamma]_{\delta'} .$$

Finally, the connected sections of a path that do not belong to the domain  $F$ , which are required to be straight lines by definition, are called *outer sections*. On the other hand, connected sections remaining within the domain are called *inner sections*.

**Lemma 14** (Path Modification). Let  $\gamma \in \Gamma_F(x, y)$  and  $\delta > 0$ . Then there exist constants  $c$  and  $\rho$  depending on  $m$  and  $\beta$  and a decomposition of the path  $\gamma = \eta + \chi + \omega$  along with a modified path  $\tilde{\gamma} = \tilde{\eta} + \tilde{\chi} + \tilde{\omega} \in \Gamma_F(x, y)$  such that

- (i)  $\eta \subset [\gamma]^{\rho\delta}$  and  $D_\mu(\tilde{\eta}) \leq c\delta^{\beta+1}$ .
- (ii)  $\tilde{\chi} \subset \chi$  and  $|\chi^\delta| \leq \frac{1}{2}|\gamma^\delta|^{\rho\delta}$ .
- (iii)  $\tilde{\omega} \subset \omega \subset [\gamma]_\delta$ .

Section B provides more insight on the construction of the alternative path of Lemma 14 (see for instance Figs. 6 and 7) along with the complete proof. Precisely, Lemma 14 is first proven in the case where no constraints are put on the domain, i.e.,  $F = \mathbb{R}^d$ , in Section B.1. The proof in the general case is more technical and is detailed in Section B.2.

Given a geodesic  $\gamma$ , its Euclidean length can then be upper bounded by considering the modifications given by Lemma 14 associated with a well-chosen decreasing sequence of thresholds  $(\delta_k)_{k \geq 0}$  ranging from  $\|d_\mu\|_{\infty, \gamma}$  to 0. Indeed, the part of  $\gamma$  belonging to  $L_{\mu, \delta_k} \setminus L_{\mu, \delta_{k+1}}$  immediately satisfies

$$D_\mu([\gamma]_{\delta_{k+1}}^{\delta_k}) \geq |[\gamma]_{\delta_{k+1}}^{\delta_k}| \cdot \delta_{k+1}^\beta$$

and the left-hand side can be upper bounded by comparing it to the FDTM length of the modification associated with the threshold  $\delta_k$  using the geodesic nature of  $\gamma$ . This reasoning is detailed in Sections B.1 and B.2—in the case of  $F = \mathbb{R}^d$  then in the general case—and eventually yields Eq. (5). Finally, Section B.3 contains the detailed computations to establish Corollary 6.

### 3.3 Stability of the FDTM

The upper bound on the length of geodesics allows to derive stability results in a straightforward manner.

#### 3.3.1 Stability with respect to the Measure

We state a more precise version of Theorem 8. While the Lipschitz stability w.r.t. the Wasserstein distance is interesting in itself, an intermediate bound will be used instead when dealing with convergence rate of the empirical FDTM. Theorem 15 is a direct consequence of the upper bound on the length of geodesics from Corollary 5 and of the Wasserstein stability of the DTM from Eq. (3).

**Theorem 15.** Let  $\mu, \nu \in \mathcal{M}_K$  and  $F \subset K$  a closed domain. Then

$$\|D_{\mu, F} - D_{\nu, F}\|_{\infty, K} \leq \lambda \text{diam}(K) \|d_\mu^\beta - d_\nu^\beta\|_{\infty, K} \leq \frac{\beta\lambda}{m^{1/p}} \text{diam}(K)^\beta W_p(\mu, \nu)$$

where  $\lambda$  is the hidden constant in Eq. (5).

*Proof.* Consider two endpoints  $x$  and  $y$  in  $K$  and let  $\gamma \in \Gamma_{\mu, F}^*(x, y)$  be a geodesic. Since by definition  $D_{\mu, F}(x, y) = D_\mu(\gamma)$  and  $D_{\nu, F}(x, y) \leq D_\nu(\gamma)$ ,

$$D_{\nu, F}(x, y) - D_{\mu, F}(x, y) \leq \int_\gamma |d_\nu^\beta - d_\mu^\beta| \leq |\gamma| \|d_\nu^\beta - d_\mu^\beta\|_{\infty, K} \leq \lambda \text{diam}(K) \|d_\nu^\beta - d_\mu^\beta\|_{\infty, K}$$

where  $\lambda > 0$  is the hidden constant in Corollary 5. Switching the roles of  $\mu$  and  $\nu$  gives the same bound for the opposite quantity. Then, using the  $\beta \text{diam}(K)^{\beta-1}$ -Lipschitz property of  $t \in [0, \text{diam}(K)] \mapsto t^\beta$  along with Eqs. (3) and (4) yields

$$|D_{\mu, F}(x, y) - D_{\nu, F}(x, y)| \leq \beta\lambda \text{diam}(K)^\beta \|d_\mu - d_\nu\|_\infty \leq \frac{\beta\lambda}{m^{1/p}} \text{diam}(K)^\beta W_p(\mu, \nu),$$

which concludes Theorem 15. □

### 3.3.2 Stability with respect to the Domain (sketch of proof)

We state an intermediate version of Theorem 9.

**Theorem 16.** Let  $\mu \in \mathcal{M}_K$  and  $F, G \subset K$  two closed domains. Assume that any point in  $F$  is at distance from  $G$  at most

$$\|d(\cdot, G)\|_{\infty, F} \leq \frac{\text{diam}(K)}{25\beta} .$$

Then for all  $x, y \in K$ ,

$$D_{\mu, G}(x, y) \leq D_{\mu, F}(x, y) + c \text{diam}(K)^{\beta + \frac{1}{2}} \sqrt{\|d(\cdot, G)\|_{\infty, F}}$$

where  $c = 10\sqrt{\beta}\lambda$  and  $\lambda$  is the hidden constant in Corollary 5.

It is straightforward that Theorem 16 directly implies Theorem 9 when the Hausdorff distance  $d_H(F, G)$  is smaller than the given threshold. In the other case, the constant  $\lambda$  given in Section B.2 is larger than one so that being above the threshold implies that

$$\text{diam}(K) \leq 25\beta d_H(F, G) \leq 100\beta\lambda^2 d_H(F, G) .$$

Then, considering the straight path  $[x, y] \in \Gamma_G(x, y)$  gives the desired upper bound on the FDTM by using Eq. (4):

$$D_{\mu, G}(x, y) \leq \|x - y\| \left( \max_K d_\mu \right)^\beta \leq \text{diam}(K)^{\beta+1} \leq 10\sqrt{\beta}\lambda \text{diam}(K)^{\beta + \frac{1}{2}} \sqrt{d_H(F, G)} .$$

We now describe the main reasoning behind Theorem 16. Given a geodesic  $\gamma$  w.r.t. the domain  $F$ ,  $\gamma$  is approximated by another path which is admissible for the domain  $G$ . This is done by cutting  $\gamma$  into small bits, identifying the nearest neighbor in  $G$  at each step and drawing a polygonal path between the resulting points. The presence of long outer sections where points are not in  $F$  requires some caution. Indeed, the points outside  $F$  may not be approximated efficiently by points in  $G$ . In this case however, since the outer section must draw a straight line between two points in  $F$  by definition, the segment connecting their nearest neighbors in  $G$  still provides a good approximation of the geodesic. This analysis is detailed in Section C and eventually shows that if bits are chosen with a typical length  $r$ —barring long outer sections—then the resulting FDTM offset is at most proportional to  $r + \frac{1}{r} \max_{x \in F} d(x, G)$ . Therefore,  $r$  is best chosen to scale as the square root of  $\max_{x \in F} d(x, G)$ , which eventually yields Theorem 16.

## 4 Estimating the FDTM

In this section, we study the special case where the domain is chosen as the support of the measure and the FDTM  $D_\mu = D_{\mu, \mathcal{S}(\mu)}$  is estimated by sampling a set  $\mathbb{X}_n$  of  $n$  points i.i.d. according to  $\mu$  and considering the empirical FDTM  $D_{\hat{\mu}_n} = D_{\hat{\mu}_n, \mathbb{X}_n}$ .

### 4.1 Theoretical Convergence Rate

Applying the previous stability results to compare  $D_{\hat{\mu}_n, \mathbb{X}_n}$  to  $D_{\mu, \mathcal{S}(\mu)}$  allows to obtain explicit convergence rate of the empirical FDTM under Assumption (A). In Assumption (A).(ii),  $b$  may be interpreted as an upper bound on the intrinsic dimension of the support of  $\mu$ . This is true in the sense of the Minkowski dimension, which coincides with standard notion of dimension if  $\mathcal{S}(\mu)$  is a sub-manifold of  $\mathbb{R}^d$ . Conversely, if  $\mu$  is supported on a compact manifold of dimension  $b$  with density bounded by below, then  $\mu$  is  $(a, b)$ -standard where  $a$  depends on the density lower bound and on the geometry of the manifold [12]. Finally, note that if a measure  $\mu$  is  $(a, b)$ -standard, then for all  $b' \geq b$  it is also  $(a', b')$ -standard for small enough  $a'$ .<sup>1</sup> In order to control the convergence rate of the empirical FDTM over  $K$ , we decompose the offset into two terms as

$$\|D_\mu - D_{\hat{\mu}_n}\|_{\infty, K} \leq \|D_{\mu, \mathcal{S}(\mu)} - D_{\hat{\mu}_n, \mathcal{S}(\mu)}\|_{\infty, K} + \|D_{\hat{\mu}_n, \mathcal{S}(\mu)} - D_{\hat{\mu}_n, \mathbb{X}_n}\|_{\infty, K} . \quad (6)$$

<sup>1</sup>By letting  $a' = a^{b'/b}$ , one has  $(a')^{-1/b'} = a^{-1/b}$  and  $ar^b \geq a'r^{b'}$  for  $0 < r \leq (a')^{-1/b'}$ .

Both terms can be further upper bounded using our stability results Theorems 9 and 15, involving  $\|d_\mu^p - d_{\hat{\mu}_n}^p\|_\infty$  and  $d_H(\mathcal{S}(\mu), \mathbb{X}_n)$  respectively. These two quantities then converge with known rate under Assumption (A).(ii), namely

$$\mathbb{E} \left[ \|d_\mu^p - d_{\hat{\mu}_n}^p\|_{\infty, K} \right] \lesssim \frac{\log(n)}{\sqrt{n}} \quad \text{and} \quad \mathbb{E} [d_H(\mathcal{S}(\mu), \mathbb{X}_n)] \lesssim \left( \frac{\log(n)}{n} \right)^{\frac{1}{b}}. \quad (7)$$

The former is directly deduced from [5] whereas the latter is a standard result. Note that the conditions (iii) and (iv) in Assumption (A) are required only to obtain the convergence rate on the empirical DTM. Without these assumptions, one could instead study the convergence of the empirical measure in Wasserstein distance. Indeed, [2] showed that, assuming that  $\mu$  is  $(a, b)$ -standard,

$$\mathbb{E} [W_p(\mu, \hat{\mu}_n)] \lesssim \max \left( \frac{1}{n^{\frac{1}{2p}}}, \frac{1}{n^{\frac{1}{b}}} \right).$$

This would however lead to a term of order  $n^{-1/2p}$  in the final bound, which is slower than the statement below if  $p > b$ . Equations (6) and (7) put together with Theorems 9 and 15 allow for the following convergence rate which is a slightly more detailed version of Theorem 10.

**Theorem 17.** Assume that  $\mu \in \mathcal{M}_{K,a,b,\sigma}$  satisfies Assumption (A) and  $m \leq \frac{1}{2}$ . Then for all  $n \geq \frac{1}{m}$ ,

$$\mathbb{E} \left[ \|D_\mu - D_{\hat{\mu}_n}\|_{\infty, K} \right] \lesssim \text{diam}(K)^{\beta+1} \left( \left( \frac{\text{diam}(K)}{\sigma} \right)^{(p-\beta)\vee 0} \frac{d \log(n)}{\sqrt{n}} + \left( \frac{\log(n)}{n} \right)^{\frac{1}{2b}} \right)$$

where  $\lesssim$  hides a multiplicative constant depending on  $m, p, \beta, a$  and  $b$ .

Details on the above discussion are provided in Section D.

## 4.2 Minimax Lower Bound

In this section we describe the reasoning to get the minimax lower bound stated in Theorem 11, which is based on Le Cam's Lemma (see for instance [14]). This lemma essentially states that if two measures  $\mu_1, \mu_2 \in \mathcal{M}_{K,a,b,\sigma}$  are sufficiently close in total variation distance while maintaining a significant FDTM offset, then any estimator of the FDTM cannot distinguish between both FDTMs up to a certain amount of samples. This provides a worst case lower bound on the convergence rate of any estimator of the FDTM for the class of measure  $\mathcal{M}_{K,a,b,\sigma}$ . We now describe an example of such measures such that Le Cam's Lemma yields the lower bound of order  $n^{-1/b}$  stated in Theorem 11.

*Example 4.1.* Let  $x$  and  $y$  be the basis vectors of  $\mathbb{R}^2$ . Given  $\alpha \in (0, 1)$ ,  $r > 0$  and  $0 < \varepsilon < (1 - \alpha)^{1/b}$ , let

$$\begin{aligned} \mu &= m\alpha\delta_{-ry} + m(1 - \alpha)\delta_{ry} + (1 - m)\rho, \\ \nu &= \mu - m\varepsilon^b\delta_{ry} + m\lambda, \end{aligned}$$

where  $\rho$  is the uniform probability measure over  $[3ry, 4ry]$  and  $\lambda$  has density  $z \mapsto b\|z - (r - \varepsilon)y\|^{b-1}$  with regard to the Lebesgue measure on  $[(r - \varepsilon)y, ry]$ , amounting to a total mass  $\lambda([(r - \varepsilon)y, ry]) = \varepsilon^b$ . Then, there exists a choice of  $\alpha \leq \frac{1}{2}$  and  $r \leq \frac{1}{4}$  depending on  $p$  and  $\beta$  along with positive constants  $a, c$  and  $C$  depending on  $m, p$  and  $\beta$  such that for all  $0 < \varepsilon \leq c$ ,  $\mu$  and  $\nu$  are both  $(a, b)$ -standard, satisfy  $d_{TV}(\mu, \nu) = m\varepsilon^b$  and

$$D_\mu(-x, x) - D_\nu(-x, x) \geq C\varepsilon.$$

Figure 1 illustrates the reasoning behind Example 4.1, which is to add the smallest amount of mass allowed by the  $(a, b)$ -standard assumption to enable a shorter path that drastically changes the FDTM. The presence of a small atom of mass  $\alpha$  is necessary to ensure that the shorter path admissible only for  $\nu$  is indeed shorter FDTM-wise. The  $\rho$  part of the measure does not play a significant role besides simply holding the rest of the mass far away from the setup of the example so that it doesn't interfere with the reasoning. It is not defined as an atom to ensure the assumption that there are no atoms of mass greater than  $m$ .

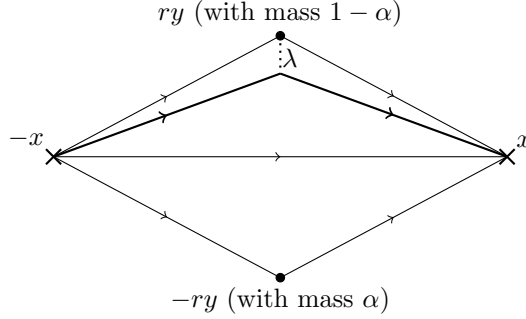


Figure 1: Example 4.1 in the case  $m = 1$  with the main admissible paths. The dotted line represents the addition of a small density in  $\nu$  which allows for a new shorter path (drawn thicker) that is shorter which is the reason for the significant offset in FDTM.

Notice that the measures in Example 4.1 do not have connected support. However, a curve of small enough density may be added to connect all components of the support, taking sufficiently long detours away from the origin so that it does not enable shorter paths. Provided that this curve has a mass small enough that it changes the DTM by a negligible amount, the result still holds with different constants. The technical computations required to establish the inequality in Example 4.1 are omitted here, as they do not involve any particularly deep arguments.

The assumption that  $K$  is not a segment in Theorem 11 ensures due to convexity that one can find a 2-dimensional disc within  $K$  in which one can fit the example measures from Example 4.1. Moreover, the dependency in  $\|x - y\|^{\beta+1}$  comes naturally from the scaling behavior of the FDTM as discussed in Section A.1. Theorem 11 is then obtained using Le Cam's Lemma with Example 4.1 and letting  $\varepsilon$  be of order  $n^{-1/b}$  so that  $d_{\text{TV}}(\mu_1, \mu_2)$  is of order  $n^{-1}$ . In the case where  $b = 1$ , the lower bound of order  $n^{-1/2}$  can be obtained by considering two atoms on a line with a uniform density on the segment between them, and creating a second measure by shifting a small amount of mass from one atom to another.

### 4.3 Practical Computations and Heuristics

Regarding the computational complexity of the estimator, notice that the restriction of  $D_{\hat{\mu}_n}$  to the sample points  $\mathbb{X}_n$  is the metric of a complete weighted graph with vertices  $\mathbb{X}_n$  and weights

$$w_{x,y} = \int_{[x,y]} d_{\hat{\mu}_n} = \|x - y\| \int_0^1 d_{\hat{\mu}_n}((1-t)x + ty)^\beta dt ,$$

where, letting  $k = \lfloor mn \rfloor$  and  $d^i(z, \mathbb{X}_n)$  be the distance from  $z$  to its  $i$ -th neighbor in  $\mathbb{X}_n$ ,

$$d_{\hat{\mu}_n}(z)^p = \frac{1}{m} \int_0^m \delta_u(z)^p du = \frac{1}{m} \left( \sum_{i=1}^k \frac{1}{n} (d^i(z, \mathbb{X}_n))^p + \frac{mn - k}{n} (d^{k+1}(z, \mathbb{X}_n))^p \right) .$$

While theoretically possible, computing such metric is unreasonable in practice, hence we propose a few methods to reduce time complexity while ensuring that the limiting object remains the same.

*Selecting edges.* When considering the complete graph over  $\mathbb{X}_n$ , the quadratic amount of edges affects both the computation time of weights and the shortest path algorithm. In the context of classic Fermat distance, the graph edges may be restricted to a subset of order  $n \log(n)$  by pruning the graph, connecting each vertex to its  $\log(n)$  closest neighbors. It is indeed guaranteed that with high probability such procedure doesn't affect the Fermat distance as  $n$  grows larger [11, Proposition 2.16]. This trick however cannot be performed in the context of FDTM without altering the limiting object as it would effectively restrict paths to remain constrained within the support, which may exclude geodesics. If it is known that no geodesic exits the support, e.g., when the latter is convex, such method may be applied.

A similar method that can be used in general is to use a graph spanner that restrict the edges while preserving some edges to span all directions. This is done by Yao graphs [13], which divide the space from

each vertex into a set of cones and keeps the shortest edges in each cone, effectively keeping one edge in each possible direction. Building a  $\log(n)$ -nearest neighbors graph has complexity  $\mathcal{O}(n \log(n))$ , whereas building a  $\log(n)$ -directions Yao graph has complexity  $\mathcal{O}(n^2)$  in general, but can be computed in  $\mathcal{O}(n \log(n))$  in dimension  $d = 2$  [9].

Both the pruned graph and the Yao graph are sub-graphs with  $\mathcal{O}(n \log(n))$  edges. Yao graphs are an efficient method when  $d = 2$  as it is fast and do not restrict the paths to the support. However, it does not scale well with dimension, hence a nearest neighbors method may be wiser in higher dimension, at the cost of potentially altering the limit.

*Computing weights.* The exact formula for  $w_{x,y}$  can be first simplified by rounding  $m$  to  $\lfloor mn \rfloor / n$  and replacing the integral with the discrete approximation of the segment

$$w_{x,y} = \frac{\|x - y\|}{r} \sum_{t=1}^r \left( \frac{1}{k} \sum_{i=1}^k d^i(x_t, \mathbb{X}_n)^p \right)^{\frac{\beta}{p}}$$

where  $k = \lfloor mn \rfloor$  and  $x_1, \dots, x_r$  is a regular subdivision of the segment  $[x, y]$ . By choosing  $r = \log(n)$  to increase accuracy with  $n$ , each weight has a complexity of  $n \log(n)$  to compute. A faster approximation would be to first compute the DTM on each sample points, which has a complexity of  $\mathcal{O}(n^2)$  but in practice does not appear to be the computational bottleneck for reasonably sized datasets. Then  $w_{x,y}$  may be approximated by the average DTM value  $\frac{1}{2}(d_\mu(x)^\beta + d_\mu(y)^\beta)$ . While this approximation is not accurate on longer edges, it is on shorter ones as the DTM is Lipschitz, which is sufficient if using a nearest neighbors approach which only selects shorter edges.

*Computing geodesics.* Computing all geodesics from one source has a complexity of order  $\mathcal{O}((|E| + n) \log(n))$  using Dijkstra algorithm, where  $|E|$  is the amount of edges. Using the previous edge selection methods, this yields a complexity of  $\mathcal{O}(n \log(n)^2)$ .

## 5 Numerical Illustrations

Finally, we display a few simulations of the FDTM on different setups to illustrate its general behavior. Computations were made using the GUDHI library (see <https://gudhi.inria.fr/>), which provides an efficient way to compute the DTM.

### 5.1 Convergence of the Empirical FDTM on the Unit Circle

Consider  $\mu$  the uniform distribution on the unit circle in  $\mathbb{R}^2$ . In this specific case, the FDTM can be explicitly computed and geodesics between two endpoints on the circle are always made up of finite number of equally sized cords—see Fig. 2. We see in Fig. 3 that the theoretical upper bound from Theorem 10 (in this case with  $b = 1$ ) is coherent with the experiment.

### 5.2 Comparison with the Fermat Distance

Given  $\alpha > 1$ , the sample Fermat distance is the complete graph metric with vertices the sample points and edges weights  $w_{x,y} = \|x - y\|^\alpha$ . It has been shown [11, Theorem 2.3] that the sample Fermat distance converges under appropriate normalization and regularity assumptions on the measure towards the (continuous) Fermat distance defined by Eq. (1). One shortcoming of the Fermat distance is its requirement to have a density supported on a manifold of fixed dimension. As a consequence, while the sample Fermat distance can be defined for any point cloud, there are no guarantees on its convergence if the underlying measure is a mixture of measures of different dimensions.

Figure 4 provides an example of a measure that can be modified to “trick” the sample Fermat distance by adding a shortcut of negligible mass but lower dimensionality, allowing for a more efficient path. As the FDTM is based on mass and not density, it is not affected by this shortcut, which highlights its robustness. In each column, the bottom point cloud is obtained by shifting approximately  $\sqrt{n}$  points (where  $n$  is the total amount of points) among the top point cloud to create a shortcut in the middle. This means that the Total

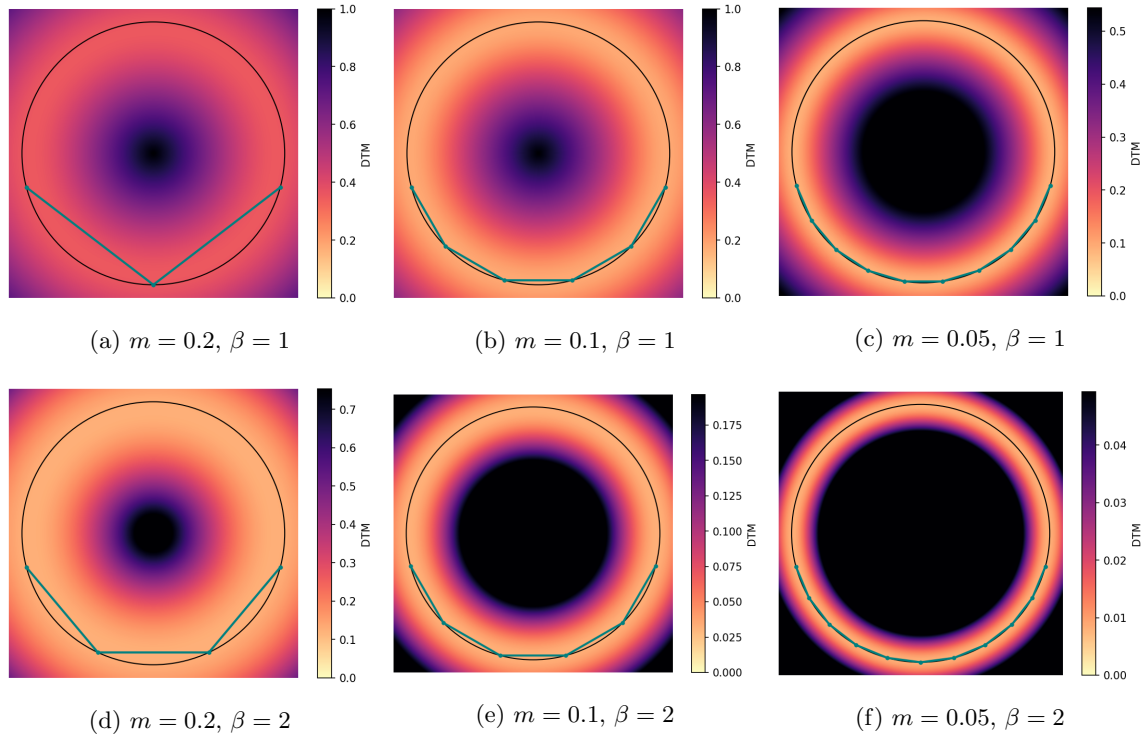


Figure 2: An example geodesic on the unit circle for  $p = 2$  and various values of  $m$  and  $\beta$ . Background colors represent the values of the DTM cropped to a relevant range. The smaller  $m$  or the higher  $\beta$ , the closer geodesics stay to the support circle.

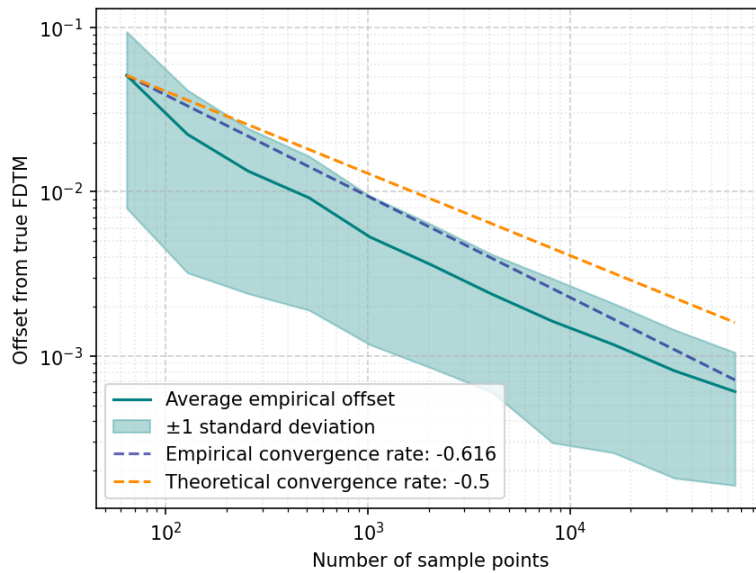


Figure 3: Convergence of the empirical FDTM on the unit circle compared to the theoretical rate  $n^{-0.5}$ . The FDTM parameters are  $m = 0.1$ ,  $p = 2$  and  $\beta = 2$ . The empirical FDTM is averaged over 500 iterations of sampling the points.

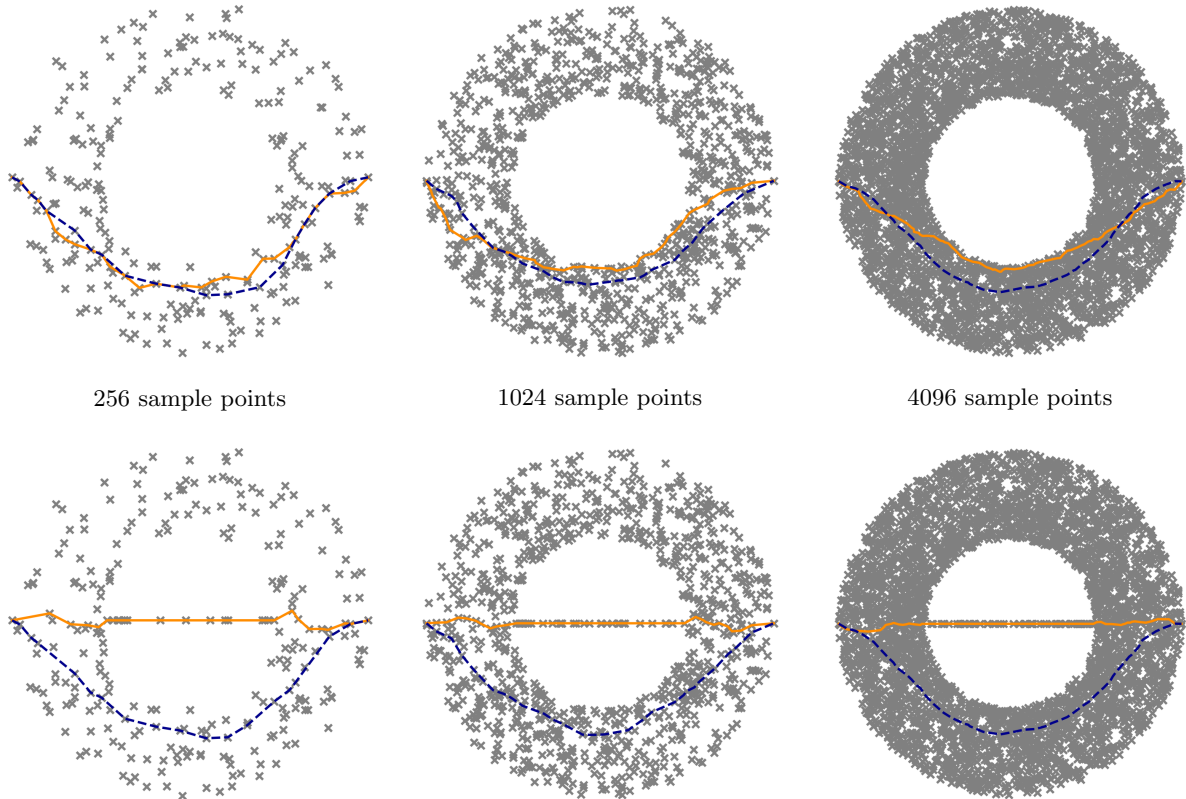


Figure 4: Sample Fermat (orange solid line) and FDTM (blue dashed line) geodesics. The sample Fermat parameter is  $\alpha = 1.1$  and the FDTM parameters are  $m = 0.1$ ,  $p = 2$ ,  $\beta = 2$ .

Variation distance between both point clouds vanishes as  $n$  grows larger, but the typical distance between points is the same on the shortcut as on the ring so that the sample Fermat distance favors the shortcut.

Figure 5 plots the relative offset of the distance caused by the addition of the shortcut, i.e.,  $\frac{|l-l'|}{l}$  where  $l$  is the distance for the point cloud without shortcut and  $l'$  the distance with shortcut. We see that the FDTM offset vanishes as the number of points increases, whereas the Fermat offset remains consistent despite the Total Variation distance between both measures decreasing.

## 6 Conclusion and Future Work

In this work we introduced a new type of density-based metric inspired by the Fermat distance and derived quantitative properties of stability and estimation. The FDTM metric is defined without any restriction on the subject measure and our results hold for minimal assumptions on this measure. Moreover, our proposed estimator of the FDTM is naturally defined as the FDTM of the empirical measure and its convergence is not influenced by the ambient dimension of the data.

Future work may investigate the design of an efficient estimation algorithm for practical applications based on the ideas developed in Section 4.3. Theoretical questions like the optimal minimax bounds for the convergence of the empirical FDTM or the generalization to non-Euclidean spaces are also of interest. Moreover, in the optic of further adapting results coming from the Fermat literature, objects such as the Fermat graph Laplacians studied in [10] may be adapted to the FDTM model, hopefully with analogous results. Finally, as of now the link between Fermat distance and FDTM is mainly reduced to the intuition that they should behave similarly. The relationship between both metrics may be studied more thoroughly, building upon previous work such as [1], especially in the case where  $m$  is chosen close to 0.

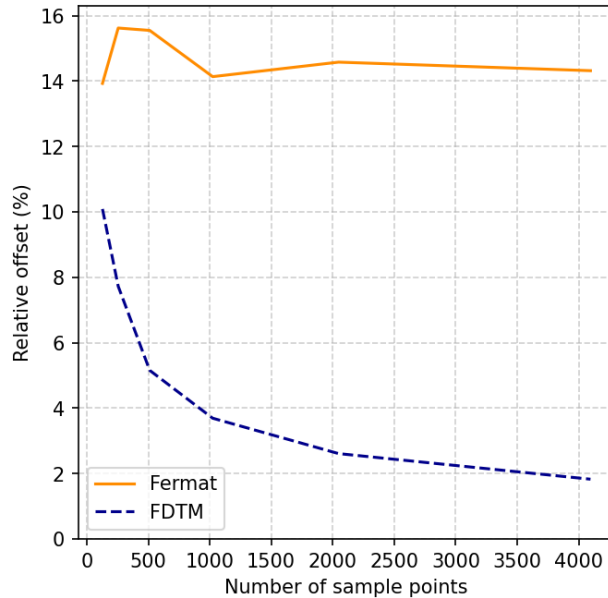


Figure 5: Relative offset in distance due to the addition of a shortcut. Each point is obtained by averaging the offset over 200 random datasets.

**Acknowledgements** This work has been supported by the ANR TopAI chair in Artificial Intelligence (ANR-19-CHIA-0001) and by École Normale Supérieure de Paris. The authors also thank Laure Ferraris for contributing to a preliminary weaker stability result for the FDTM.

## References

- [1] Gérard Biau et al. “A Weighted K-Nearest Neighbor Density Estimate for Geometric Inference”. In: *Electronic Journal of Statistics* 5 (Jan. 2011). ISSN: 1935-7524. DOI: [10.1214/11-EJS606](https://doi.org/10.1214/11-EJS606).
- [2] Emmanuel Boissard and Thibaut Le Gouic. “On the Mean Speed of Convergence of Empirical and Occupation Measures in Wasserstein Distance”. In: *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques* 50.2 (May 2014). ISSN: 0246-0203. DOI: [10.1214/12-AIHP517](https://doi.org/10.1214/12-AIHP517).
- [3] Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A Course in Metric Geometry*. Vol. 33. Graduate Studies in Mathematics. Providence, Rhode Island: American Mathematical Society, June 2001. ISBN: 978-1-4704-1794-9. DOI: [10.1090/gsm/033](https://doi.org/10.1090/gsm/033).
- [4] Frédéric Chazal, David Cohen-Steiner, and Quentin Mérigot. “Geometric Inference for Measures Based on Distance Functions”. In: *Foundations of Computational Mathematics* 11.6 (2011), p. 733. DOI: [10.1007/s10208-011-9098-0](https://doi.org/10.1007/s10208-011-9098-0).
- [5] Frédéric Chazal, Pascal Massart, and Bertrand Michel. “Rates of Convergence for Robust Geometric Inference”. In: *Electronic Journal of Statistics* 10.2 (Jan. 2016), pp. 2243–2286. ISSN: 1935-7524, 1935-7524. DOI: [10.1214/16-EJS1161](https://doi.org/10.1214/16-EJS1161).
- [6] Frédéric Chazal et al. “Convergence Rates for Persistence Diagram Estimation in Topological Data Analysis”. In: *Journal of Machine Learning Research* 16.110 (2015), pp. 3603–3635. ISSN: 1533-7928.
- [7] Antonio Cuevas and Alberto Rodríguez-Casal. “On Boundary Estimation”. In: *Advances in Applied Probability* 36.2 (June 2004), pp. 340–354. ISSN: 0001-8678, 1475-6064. DOI: [10.1239/aap/1086957575](https://doi.org/10.1239/aap/1086957575).
- [8] Ximena Fernández et al. “Intrinsic Persistent Homology via Density-Based Metric Learning”. In: *Journal of Machine Learning Research* 24.1 (Jan. 2023), 75:3341–75:3382. ISSN: 1532-4435.

- [9] Daniel Funke and Peter Sanders. “Efficient Yao Graph Construction”. In: *21st International Symposium on Experimental Algorithms (SEA 2023)*. Vol. 265. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023, 20:1–20:20. ISBN: 978-3-95977-279-2. DOI: [10.4230/LIPIcs.SEA.2023.20](https://doi.org/10.4230/LIPIcs.SEA.2023.20).
- [10] Nicolas Garcia Trillos et al. “Fermat Distances: Metric Approximation, Spectral Convergence, and Clustering Algorithms”. In: *Journal of Machine Learning Research* 25.1 (Jan. 2024), 176:8331–176:8395. ISSN: 1532-4435.
- [11] Pablo Groisman, Matthieu Jonckheere, and Facundo Sapienza. “Nonhomogeneous Euclidean First-Passage Percolation and Distance Learning”. In: *Bernoulli* 28.1 (Feb. 2022), pp. 255–276. ISSN: 1350-7265. DOI: [10.3150/21-BEJ1341](https://doi.org/10.3150/21-BEJ1341).
- [12] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. “Finding the Homology of Submanifolds with High Confidence from Random Samples”. In: *Discrete & Computational Geometry* 39.1 (Mar. 2008), pp. 419–441. ISSN: 1432-0444. DOI: [10.1007/s00454-008-9053-2](https://doi.org/10.1007/s00454-008-9053-2).
- [13] Andrew Chi-Chih Yao. “On Constructing Minimum Spanning Trees in K-Dimensional Spaces and Related Problems”. In: *SIAM Journal on Computing* 11.4 (Nov. 1982), pp. 721–736. ISSN: 0097-5397. DOI: [10.1137/0211059](https://doi.org/10.1137/0211059).
- [14] Bin Yu. “Assouad, Fano, and Le Cam”. In: *SIAM Journal on Computing* (1997).

## A Complementary Discussion

The first part of this section details the discussion about the assumption of considering all objects within a large convex compact subset  $K \subset \mathbb{R}^d$ . The second part details the fact that the FDTM and Euclidean metrics share the same topology over the domain  $F$ .

### A.1 Restriction to a Compact Set

As long as we restrict the study of the FDTM to endpoints that belong to a compact set  $A$ , the fact that  $d_\mu$  is a Lipschitz and proper function ensures that geodesic paths remain in a greater compact set  $B$ . Indeed, for all  $x, y \in A$ ,

$$D_{\mu,F}(x, y) \leq D_\mu([x, y]) \leq \text{diam}(A) \cdot \max_A d_\mu^\beta$$

and any path leaving a compact set  $B$  chosen large enough would achieve a greater FDTM length, making it impossible to be a geodesic. Therefore, replacing  $F$  with the cropped domain  $F \cap B$  does not affect the FDTM within  $A$ . Moreover, the compact subset  $K$  defined as the convex hull of

$$\bigcup_{x \in B} \bar{B}(x, \delta_{\mu,m}(x))$$

contains all points that are relevant w.r.t. the DTM in  $B$ , so that replacing  $\mu$  with the push-forward measure  $(\pi_K)_*\mu$  where  $\pi_K$  is the projection over  $K$  does not affect the DTM within  $B$ , nor does it affect the FDTM within  $A$  by extension. Therefore, the assumption that all objects are included in a great compact convex set  $K$  is without loss of generality.

Now, notice that scaling through  $h : x \mapsto \text{diam}(K)^{-1}x$  sends  $K$  to a subset of unit diameter and that all objects considered behave coherently with this scaling: all distances are multiplied by  $\text{diam}(K)^{-1}$ , including the DTM. Therefore, for all  $x, y \in K$ ,

- $\Gamma_{h(F)}(h(x), h(y)) = \{h \circ \gamma, \gamma \in \Gamma_F(x, y)\}$ .
- $\forall \gamma \in \Gamma_F(x, y), D_{h_*\mu}(h \circ \gamma) = \text{diam}(K)^{-(\beta+1)} D_\mu(\gamma)$ .
- $\Gamma_{h_*\mu, h(F)}^*(x, y) = \{h \circ \gamma, \gamma \in \Gamma_{\mu, F}^*(x, y)\}$ .
- $D_{h_*\mu, h(F)}(h(x), h(y)) = \text{diam}(K)^{-(\beta+1)} D_{\mu, F}(x, y)$ .

As a consequence, any result can be proven with the assumption that  $\text{diam}(K) = 1$ , then extended to the general case by scaling accordingly with  $\text{diam}(K)$ .

Finally, let us detail Eq. (4): Since  $\mathcal{S}(\mu) \subset K$ , all the mass lies entirely within a radius of at most  $\text{diam}(K)$  from any  $x \in K$ , hence  $\delta_{\mu,u}(x) \leq \text{diam}(K)$  for any  $u \in (0, 1)$  and thus  $d_\mu(x) \leq \text{diam}(K)$ . As a consequence, since admissible paths for  $x, y \in K$  are included in the convex hull of  $F \cup \{x, y\}$  and therefore in  $K$ , it follows that  $d_\mu \leq \text{diam}(K)$  over any path  $\gamma \in \Gamma_F(x, y)$ .

### A.2 Equivalent Topologies

Here we detail the following statement from Section 3.1.

**Lemma 18.**  $D_{\mu,F}$  defines a metric over  $F$  and its topology is equivalent to the Euclidean topology.

*Proof.*  $D_\mu$  is immediately non-negative and symmetric. Provided that  $x, y, z \in F$ , if  $\gamma \in \Gamma_F(x, y)$  and  $\gamma' \in \Gamma_F(y, z)$  then the concatenation of both paths belongs to  $\Gamma_F(x, z)$ . The triangular inequality follows from considering the infimum over  $\gamma$  and  $\gamma'$ .

Now, let us show that in  $\mathbb{R}^d$ , any ball for the Euclidean metric contains a ball for the FDTM metric. First, the DTM may be equal to 0 only over a finite set, as  $d_\mu(x) = 0$  if and only if  $\mu$  has an atom of mass at least  $m$  at  $x$ . Then, for all  $x \in \mathbb{R}^d$  and small enough  $r > 0$ ,  $d_\mu$  is lower bounded over the sphere  $\mathcal{S}(x, r)$  by some  $\delta > 0$ . Given any point  $y$  at Euclidean distance larger than  $r$  from  $x$ , any path  $\gamma \in \Gamma_F(x, y)$  must intersect  $\mathcal{S}(x, r)$  and its FDTM length can be lower bounded by some  $\varepsilon = f(r, \delta, \beta) > 0$  by Lipschitz property of the DTM. Therefore, for all small enough  $r > 0$  there exists a FDTM ball of radius  $\varepsilon$  included in the

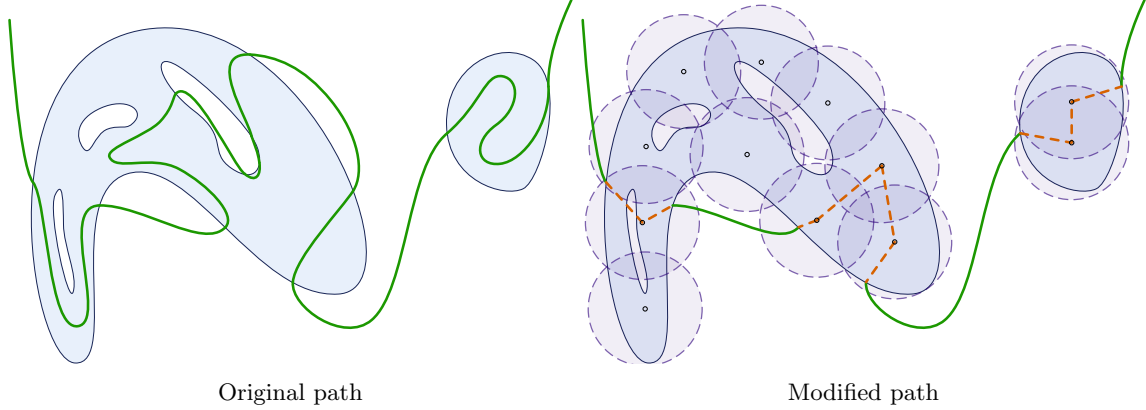


Figure 6: Example of the modification process of Lemma 19. The blue area represents a sublevel and the dashed circles on the right figure represent a covering of this sublevel. The sublevel sections of the original path are replaced with new sections making up  $\tilde{\eta}$  (in orange) on the right figure. Some of the superlevel sections of the original path are removed due to creating loops and the remaining parts make up  $\tilde{\omega}$  (in green).

Euclidean ball of radius  $r$ . This implies that the FDTM between two distinct points is positive—hence the FDTM is a metric—and coincidentally shows that the FDTM topology is finer than the Euclidean topology.

Conversely, the Euclidean topology is finer than the FDTM as a direct consequence of the continuity of  $D_\mu$  w.r.t. the Euclidean topology. This concludes that both topologies are equivalent.  $\square$

## B Upper Bound on the Euclidean Length of Geodesics

This section provides the missing proofs of Section 3.2.2 needed to obtain Theorem 4. We first consider the simpler case where the domain  $F$  is the entire space  $\mathbb{R}^d$  in Section B.1 to showcase the main reasoning without too many technicalities. Section B.2 then generalizes the reasoning to work with any  $F$ . Section B.3 finally details the computations required to obtain Corollary 6.

### B.1 Details without Domain Constraints

We provide in this section the details omitted in Section 3.2.2 in the case where  $F = \mathbb{R}^d$ , i.e., disregarding any constraint on the domain. The resulting bound on the length of geodesics is smaller than in the general case. The first step is Lemma 14, which we rewrite below in the specific case where no domain constraints are considered.

**Lemma 19.** Let  $\gamma \in \Gamma(x, y)$  and  $\delta > 0$ . Then there exists a modified path  $\tilde{\gamma} = \tilde{\eta} + \tilde{\omega} \in \Gamma(x, y)$  such that

- (i)  $\tilde{\eta}$  is a modification of the sub- $\delta$  sections  $[\gamma]^\delta$  and satisfies

$$D_\mu(\tilde{\eta}) \leq c\delta^{\beta+1} \quad (8)$$

where  $c = \frac{16 \cdot 5^\beta}{m}$ .

- (ii)  $\tilde{\omega} \subset [\gamma]_\delta$  is a subset of the super- $\delta$  sections of  $\gamma$ .

*Proof.* Denote  $\eta_1 \in \Gamma(x_1, y_1), \dots, \eta_i \in \Gamma(x_i, y_i), \dots$  the connected paths making up  $[\gamma]^\delta$ . Consider a minimal  $4\delta$ -covering  $\mathcal{F}$  of the sublevel  $L_{\mu, \delta}$ . Then according to Lemma 13,  $|\mathcal{F}| \leq \frac{2}{m}$ . For a given  $i \geq 0$ ,  $x_i$  and  $y_i$  belong to the same connected component of  $L_{\mu, \delta}$  since they are connected by  $\eta_i$ , and therefore can be joined by “hopping” between points in  $\mathcal{F}$ : There exist pairwise distinct points  $z_{i,1}, \dots, z_{i,k_i} \in \mathcal{F}$  such that

$\|x_i - z_{i,1}\| \leq 4\delta$ ,  $\|y_i - z_{i,k_i}\| \leq 4\delta$  and for all  $1 \leq k \leq k_i - 1$ ,  $\|z_{i,k} - z_{i,k+1}\| \leq 8\delta$ . We then define for all  $i$ ,  $\bar{\eta}_i = [x_i, z_{i,1}, z_{i,2}, \dots, z_{i,k_i-1}, z_{i,k_i}, y_i] \in \Gamma_F(x_i, y_i)$  the corresponding polygonal path from  $x_i$  to  $y_i$ , along with

$$\bar{\gamma} = [\gamma]_\delta + \sum_i \bar{\eta}_i = \gamma + \sum_i (\bar{\eta}_i - \eta_i) \in \Gamma(x, y)$$

a new path replacing subsections of  $\gamma$  with these new paths. Finally, we define  $\tilde{\gamma}$  as the result of deleting all loops from  $\bar{\gamma}$  occurring in  $\mathcal{F}$ , so that each point  $z \in \mathcal{F}$  is visited at most once by  $\tilde{\gamma}$ . Moreover, we write  $\tilde{\gamma} = \tilde{\eta} + \tilde{\omega}$  where  $\tilde{\eta} \subset \sum_i \bar{\eta}_i$  and  $\tilde{\omega} \subset [\gamma]_\delta$  are the remaining parts after the removal of the loops. This concludes the construction of the modified path  $\tilde{\gamma} \in \Gamma(x, y)$ .

It remains to upper bound  $D_\mu(\tilde{\eta})$ . Breaking down the straight lines  $[z_{i,k}, z_{i,k+1}]$  into two halves in the definition of the paths  $\bar{\eta}_i$  shows that  $\tilde{\eta}$  is the sum of at most  $2|\mathcal{F}|$  paths of the form  $[w, z]$  or  $[z, w]$  where  $z \in \mathcal{F} \subset L_{\mu, \delta}$  and  $w \in \bar{\mathcal{B}}(z, 4\delta)$  is either an endpoint  $x_i$  or  $y_i$ , or a point halfway to another  $z' \in \mathcal{F}$ . It follows immediately that  $|\tilde{\eta}| \leq 8|\mathcal{F}|\delta$ . Moreover,  $\tilde{\eta}$  remains at any point at distance at most  $4\delta$  from  $\mathcal{F} \subset L_{\mu, \delta}$ , hence  $\tilde{\eta} \subset L_{\mu, 5\delta}$  by Lipschitz property of  $d_\mu$ . Thus,

$$D_\mu(\tilde{\eta}) \leq |\tilde{\eta}| \cdot (5\delta)^\beta \leq \frac{16 \cdot 5^\beta}{m} \delta^{\beta+1},$$

which concludes the proof.  $\square$

We are now able to bound the length of any geodesic.

*Proof of Theorem 4.* Let  $\gamma \in \Gamma_\mu^*(x, y)$  be a geodesic and  $\delta_0 > 0$ . For all  $k \geq 0$  let  $\delta_k = \rho^{-k} \delta_0$  with  $\rho = 2^{1/\beta}$ . Consider  $\tilde{\gamma}_k = \tilde{\eta}_k + \tilde{\omega}_k$  the modified path associated with the sub- $\delta_k$  level given by Lemma 19. Then

$$\begin{aligned} \delta_{k+1}^\beta |[\gamma]_{\delta_{k+1}}^{\delta_k}| &\leq D_\mu([\gamma]_{\delta_{k+1}}^{\delta_k}) && (d_\mu^\beta \geq \delta_{k+1}^\beta \text{ over } [\gamma]_{\delta_{k+1}}^{\delta_k}) \\ &\leq D_\mu(\gamma) - D_\mu([\gamma]_{\delta_k}) \\ &\leq D_\mu(\tilde{\gamma}_k) - D_\mu(\tilde{\omega}_k) && (\gamma \text{ is optimal and } \tilde{\omega}_k \subset [\gamma]_{\delta_k}) \\ &= D_\mu(\tilde{\eta}_k) \\ &\leq c\delta_k^{\beta+1}. && (\text{Eq. 8}) \end{aligned}$$

Therefore,

$$|[\gamma]^{\delta_0}| = \sum_{k=0}^{+\infty} |[\gamma]_{\delta_{k+1}}^{\delta_k}| \leq c \sum_{k=0}^{+\infty} \frac{\delta_k^{\beta+1}}{\delta_{k+1}^\beta} = c\rho^\beta \sum_{k=0}^{+\infty} \rho^{-k} \delta_0 = c \frac{2^{1+1/\beta}}{2^{1/\beta} - 1} \delta_0 \leq \frac{4c\beta}{\log(2)} \delta_0.$$

The inequality  $(2^{1/\beta} - 1)^{-1} \leq \beta/\log(2)$  was used at the last step. In particular, choosing  $\delta_0 = \|d_\mu\|_{\infty, \gamma}$  provides an upper bound on the length of the entire geodesic and concludes the proof.  $\square$

## B.2 Details with Domain Constraints

We now provide the details omitted in Section 3.2.2 in the general case. First, we state an intermediate result to upper bound the length of any segment within a sublevel of the DTM.

**Lemma 20.** Let  $A$  be a measurable subset of  $\mathbb{R}^d$  and  $x, y \in \mathbb{R}^d$ . Then for all  $r > 0$ ,

$$|[x, y] \cap A| \leq \text{cov}(A, r) \cdot 2r.$$

*Proof.* Denote  $l = \|x - y\|$  and  $\gamma : [0, l] \rightarrow \mathbb{R}^d$  the arc length parameterization of  $[x, y]$ . Let  $\mathcal{F}$  be a minimal  $r$ -covering of  $A$ . Then

$$|[x, y] \cap A| = \int_0^l \mathbb{1}_A(\gamma(t)) dt \leq \int_0^l \mathbb{1}_{(\cup_{a \in \mathcal{F}} \mathcal{B}(a, r))}(\gamma(t)) dt \leq \sum_{a \in \mathcal{F}} \int_0^l \mathbb{1}_{\mathcal{B}(a, r)}(\gamma(t)) dt.$$

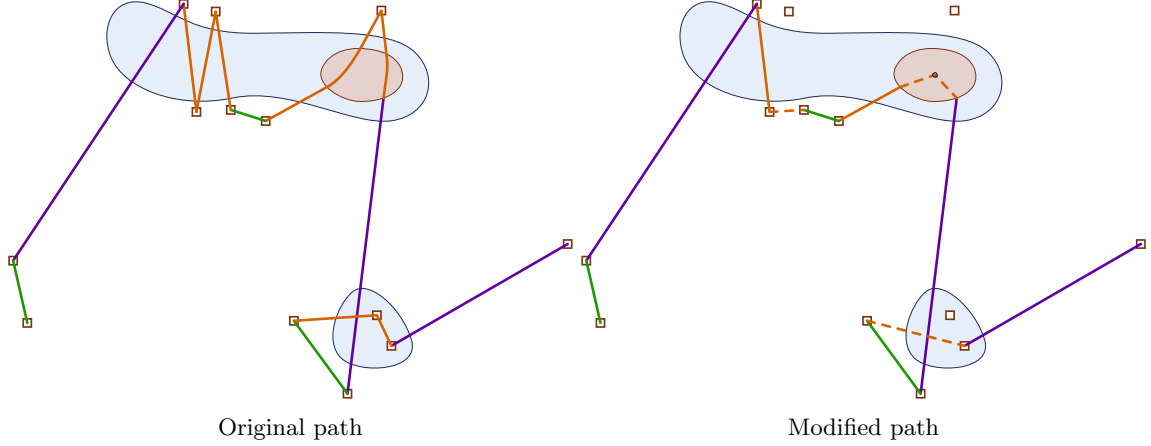


Figure 7: Example of the modification process of Lemma 22. The blue area represents a sublevel, the brown area and squares represent the domain. The  $\eta$  sections of the original path (left figure, orange) are partially replaced (right figures, orange dashed lines) to make up  $\tilde{\eta}$  (right figure, orange). The  $\chi$  and  $\omega$  sections (purple and green respectively) are left untouched in this example as no loop is removed.

If  $a \in \mathcal{F}$  and  $\gamma(t), \gamma(s) \in \mathcal{B}(a, r)$ , then  $t - s = \|\gamma(t) - \gamma(s)\| \leq 2r$ . Therefore,

$$|[x, y] \cap A| \leq \sum_{a \in \mathcal{F}} (\sup\{t : \gamma(t) \in \mathcal{B}(a, r)\} - \inf\{t : \gamma(t) \in \mathcal{B}(a, r)\}) \leq \text{cov}(A, r) \cdot 2r .$$

□

Together, Lemmas 13 and 20 yield the following result.

**Corollary 21.** Let  $\delta > 0$  and  $x, y \in \mathbb{R}^d$ . Then

$$|[x, y] \cap L_{\mu, \delta}| \leq \frac{16}{m} \delta .$$

Recall Lemma 14 in the general case, here with precise constants.

**Lemma 22.** Let  $\gamma \in \Gamma_F(x, y)$  and  $\delta > 0$ . Then there exists a decomposition of the path  $\gamma = \eta + \chi + \omega$  along with a modified path  $\tilde{\gamma} = \tilde{\eta} + \tilde{\chi} + \tilde{\omega} \in \Gamma_F(x, y)$  such that

- (i)  $\eta \subset [\gamma]^{\rho\delta}$  contains the inner sections of  $[\gamma]^\delta$  and the outer sections of  $\gamma$  that intersect  $L_{\mu, \delta}$  while not being too long outside this sublevel, with  $\rho = 1 + \frac{32}{m}$ .  $\tilde{\eta}$  is a modification of  $\eta$  satisfying

$$D_\mu(\tilde{\eta}) \leq c\delta^{\beta+1} \tag{9}$$

where  $c = \frac{16 \cdot 5^\beta + 1152 \cdot 165^\beta}{m^{2+\beta}}$ .

- (ii)  $\chi$  contains the remaining outer sections of  $\gamma$  intersecting  $L_{\mu, \delta}$  and satisfies

$$|[\chi]^\delta| \leq \frac{1}{2} |[\gamma]_\delta^{\rho\delta}| . \tag{10}$$

$\tilde{\chi} \subset \chi$  is a subset of these outer sections.

- (iii)  $\omega$  contains the rest of  $\gamma$  and  $\tilde{\omega} \subset \omega \subset [\gamma]_\delta$  is a subset of this remaining part.

When dealing with the domain constraints to modify the sublevel of a path, the main issue is that if a long outer section crosses the sublevel, then it is not possible to modify only the sublevel part of this outer section. Indeed, doing so would break the segment at a point that is outside the domain  $F$  and the new path would no longer belong to  $\Gamma_F(x, y)$ . To account for this issue, we choose to modify the “problematic” outer

sections when their length outside the sublevel is small enough similarly as we do with inner sections of the sublevel. All these sections make up  $\eta$ . The remaining outer sections make up  $\chi$  and have the property that their length inside the sublevel is smaller than a fraction of their length outside, which allows to bound the overall length by other means.

Figure 7 illustrates the process described in Lemma 22. Notice that a loop subsists in the modified path, which cannot be removed as it occurs at a point outside the domain. This example illustrates how it not always possible to modify the path with limited length within the sublevel proportional to its covering number, as  $\chi$  sections may not be reliably modified due to their arbitrarily large length, even when they create loops.

*Proof.* Denote  $\eta_1 \in \Gamma_F(x_1, y_1), \dots, \eta_i \in \Gamma_F(x_i, y_i), \dots$  the connected paths making up  $[\gamma]^\delta \cap F$  and  $[a_1, b_1], \dots, [a_j, b_j], \dots$  the outer sections of  $\gamma$  that intersect with the sublevel  $L_{\mu, \delta}$  and such that any part of the segment outside  $L_{\mu, \delta}$  is of length at most  $\theta = \frac{32}{m}\delta$ , which in particular implies that all  $a_j$  and  $b_j$  belong to  $L_{\mu, \delta + \theta}$  by Lipschitz property of  $d_\mu$ . Denote  $\eta = \sum_i \eta_i + \sum_j [a_j, b_j]$ ,  $\chi$  the remaining outer sections of  $\gamma$  intersecting with  $L_{\mu, \delta}$  and finally  $\omega = \gamma - \eta - \chi \subset [\gamma]^\delta$  the rest of the path.

Consider a minimal  $4\delta$ -covering  $\mathcal{F}$  of  $L_{\mu, \delta} \cap F$ . According to Lemma 13,  $|\mathcal{F}| \leq \frac{2}{m}$ . For a given  $i \geq 0$ ,  $x_i$  and  $y_i$  belong to the same connected component of  $L_{\mu, \delta} \cap F$  since they are connected by  $\eta_i$ . Therefore, there exist pairwise distinct points  $z_{i,1}, \dots, z_{i,k_i} \in \mathcal{F}$  such that  $\|x_i - z_{i,1}\| \leq 4\delta$ ,  $\|y_i - z_{i,k_i}\| \leq 4\delta$  and for all  $1 \leq k \leq k_i - 1$ ,  $\|z_{i,k} - z_{i,k+1}\| \leq 8\delta$ . We then define for all  $i$ ,  $\bar{\eta}_i^1 = [x_i, z_{i,1}, z_{i,2}, \dots, z_{i,k_i-1}, z_{i,k_i}, y_i] \in \Gamma_F(x_i, y_i)$  the corresponding polygonal path from  $x_i$  to  $y_i$ .

Consider a minimal  $4(\delta + \theta)$ -covering  $\mathcal{G}$  of  $\{a_1, b_1, \dots, a_j, b_j, \dots\} \subset L_{\mu, \delta + \theta}$ . According to Lemma 13,  $|\mathcal{G}| \leq \frac{2}{m}$ . For all  $j \geq 0$ , define  $\bar{\eta}_j^2 = [a_j, z'_{j,1}, z'_{j,2}, b_j]$  where  $z'_{j,1}$  and  $z'_{j,2}$  are respectively the closest neighbors of  $a_j$  and  $b_j$  in  $\mathcal{G}$ , and finally let

$$\bar{\gamma} = \sum_i \bar{\eta}_i^1 + \sum_j \bar{\eta}_j^2 + \chi + \omega = \gamma + \sum_i (\bar{\eta}_i^1 - \eta_i) + \sum_j (\bar{\eta}_j^2 - [a_j, b_j]) .$$

Since  $\bar{\gamma}$  is the result of replacing some sections of  $\gamma$  with straight lines between points in  $F$ ,  $\bar{\gamma} \in \Gamma_F(x, y)$ . Finally, we define  $\tilde{\gamma}$  as the result of deleting all loops from  $\bar{\gamma}$  occurring at points in  $\mathcal{F} \cup \mathcal{G} \subset F$ , so that  $\tilde{\gamma} \in \Gamma_F(x, y)$  and each point  $z \in \mathcal{F} \cup \mathcal{G}$  is visited at most once by  $\tilde{\gamma}$ . Moreover, we write  $\tilde{\gamma} = \tilde{\eta} + \tilde{\chi} + \tilde{\omega}$  where  $\tilde{\eta} = \tilde{\eta}^1 + \tilde{\eta}^2$  and  $\tilde{\eta}^1 \subset \sum_i \bar{\eta}_i^1$ ,  $\tilde{\eta}^2 \subset \sum_j \bar{\eta}_j^2$ ,  $\tilde{\chi} \subset \chi$  and  $\tilde{\omega} \subset \omega \subset [\gamma]^\delta$  are the remaining parts after the removal of the loops. This concludes the construction of the modified path  $\tilde{\gamma}$ .

It remains to establish Eqs. (9) and (10). With the exact same reasoning as for Eq. (8),  $|\tilde{\eta}^1| \leq 8|\mathcal{F}|\delta$ ,  $\tilde{\eta}^1 \subset L_{\mu, 5\delta}$  and thus

$$D_\mu(\tilde{\eta}^1) \leq |\tilde{\eta}^1| \cdot (5\delta)^\beta \leq \frac{16 \cdot 5^\beta}{m} \delta^{\beta+1} . \quad (11)$$

On the other hand,  $\tilde{\eta}^2$  is the sum of at most  $|\mathcal{G}|$  paths  $\bar{\eta}_j^2$  (either whole or just part of them). For all  $j$ ,  $[a_j, b_j]$  has length at most  $\theta$  outside  $L_{\mu, \delta}$  by assumption and at most  $\frac{16}{m}\delta$  inside  $L_{\mu, \delta}$  due to Corollary 21. Moreover, by definition of  $\mathcal{G}$ ,  $\|a_j - z'_{j,1}\| \leq 4(\delta + \theta)$ ,  $\|b_j - z'_{j,2}\| \leq 4(\delta + \theta)$  and

$$\|z'_{j,1} - z'_{j,2}\| \leq \|z'_{j,1} - a_j\| + \|a_j - b_j\| + \|b_j - z'_{j,2}\| \leq 8(\delta + \theta) + \theta + \frac{16}{m}\delta ,$$

so that

$$|\bar{\eta}_j^2| = \|a_j - z'_{j,1}\| + \|z'_{j,1} - z'_{j,2}\| + \|z'_{j,2} - b_j\| \leq 16(\delta + \theta) + \theta + \frac{16}{m}\delta \leq \frac{32}{m}\delta + 17\theta = 18\theta$$

and finally

$$|\tilde{\eta}^2| \leq |\mathcal{G}| \cdot 18\theta \leq \frac{1152}{m^2} \delta .$$

Then, for all  $j$ , any point of  $\bar{\eta}_j^2$  is at distance at most  $4(\delta + \theta)$  from  $[a_j, b_j] \subset L_{\mu, \delta + \theta}$ , hence  $\tilde{\eta}^2 \subset L_{\mu, 5(\delta + \theta)}$  and

$$D_\mu(\tilde{\eta}^2) \leq |\tilde{\eta}^2| \cdot (5(\delta + \theta))^\beta \leq \frac{1152}{m^2} \left( \frac{160}{m} + 5 \right)^\beta \delta^{\beta+1} \leq \frac{1152 \cdot 165^\beta}{m^{2+\beta}} \delta^{\beta+1} . \quad (12)$$

Put together, Eqs. (11) and (12) yield

$$D_\mu(\tilde{\eta}) = D_\mu(\tilde{\eta}^1) + D_\mu(\tilde{\eta}^2) \leq \frac{16 \cdot 5^\beta + 1152 \cdot 165^\beta}{m^{2+\beta}} \delta^{\beta+1}.$$

Now, denote  $(\bar{\chi}_k)_k$  the different segments making up  $\chi$ . For each segment  $\bar{\chi}_k$ , according to Corollary 21,

$$|[\bar{\chi}_k]^\delta| \leq \frac{16}{m} \delta = \frac{1}{2} \theta \leq \frac{1}{2} |[\bar{\chi}_k]_{\delta}^{\delta+\theta}|. \quad (13)$$

The last inequality is due to the following: If  $[\bar{\chi}_k]_\delta \subset L_{\mu, \delta+\theta}$  then it follows immediately from the assumption that  $|[\bar{\chi}_k]_\delta| > \theta$ . Else,  $\bar{\chi}_k$  intersects both  $L_{\mu, \delta}$  and  $\mathbb{R}^d \setminus L_{\mu, \delta+\theta}$ , which are separated by a distance of at least  $\theta$  by Lipschitz property of  $d_\mu$ , so that  $|[\bar{\chi}_k]_{\delta}^{\delta+\theta}| \geq \theta$ . Then, summing Eq. (13) over all  $\bar{\chi}_k$  yields

$$|\chi|^\delta \leq \frac{1}{2} |[\chi]_{\delta}^{\rho\delta}| \leq \frac{1}{2} |[\gamma]_{\delta}^{\rho\delta}|$$

where  $\rho = \frac{\delta+\theta}{\delta} = 1 + \frac{32}{m}$ , which concludes the proof.  $\square$

We are now able to bound the length of any geodesic.

*Proof of Theorem 4.* Let  $\gamma \in \Gamma_{\mu, F}^*(x, y)$  be a geodesic and  $\delta_0 > 0$ . For all  $k \geq 0$  let  $\delta_k = \rho^{-k} \delta_0$  with  $\rho = 1 + \frac{32}{m}$ . Consider  $\gamma = \eta_k + \chi_k + \omega_k$  and  $\tilde{\gamma}_k = \tilde{\eta}_k + \tilde{\chi}_k + \tilde{\omega}_k$  the decomposition and modified path associated with the sub- $\delta_k$  level given by Lemma 22. Recall that by construction  $[\gamma]_{\delta_k}^{\delta_k} = \eta_k + [\chi_k]_{\delta_k}^{\delta_k}$ , thus

$$|[\gamma]_{\delta_0}^{\delta_0}| = \sum_{k=0}^{+\infty} |[\gamma]_{\delta_{k+1}}^{\delta_{k+1}}| = \sum_{k=0}^{+\infty} |[\eta_k]_{\delta_{k+1}}^{\delta_{k+1}} + [\chi_k]_{\delta_{k+1}}^{\delta_{k+1}}| \leq \sum_{k=0}^{+\infty} |[\eta_k]_{\delta_{k+1}}^{\delta_{k+1}}| + \sum_{k=0}^{+\infty} |[\chi_k]_{\delta_{k+1}}^{\delta_{k+1}}|. \quad (14)$$

On one hand,  $|[\chi_k]_{\delta_k}^{\delta_k}| \leq \frac{1}{2} |[\gamma]_{\delta_k}^{\delta_{k-1}}|$  for all  $k \geq 0$  by Eq. (10), hence

$$\sum_{k=0}^{+\infty} |[\chi_k]_{\delta_k}^{\delta_k}| \leq \frac{1}{2} \sum_{k=0}^{+\infty} |[\gamma]_{\delta_k}^{\delta_{k-1}}| = \frac{1}{2} |[\gamma]_{\delta_0}^{\rho\delta_0}| \quad (15)$$

On the other hand, for all  $k \geq 0$ ,

$$\begin{aligned} \delta_{k+1}^\beta |[\eta_k]_{\delta_{k+1}}^{\delta_{k+1}}| &\leq D_\mu([\eta_k]_{\delta_{k+1}}^{\delta_{k+1}}) && (d_\mu^\beta \geq \delta_{k+1}^\beta \text{ over } [\eta_k]_{\delta_{k+1}}^{\delta_{k+1}}) \\ &\leq D_\mu(\eta_k) \\ &= D_\mu(\gamma) - D_\mu(\chi_k + \omega_k) \\ &\leq D_\mu(\tilde{\gamma}_k) - D_\mu(\tilde{\chi}_k + \tilde{\omega}_k) && (\gamma \text{ is optimal and } \tilde{\chi}_k + \tilde{\omega}_k \subset \chi_k + \omega_k) \\ &= D_\mu(\tilde{\eta}_k) \\ &\leq c \delta_k^{\beta+1}. && (\text{Eq. 9}) \end{aligned}$$

Therefore

$$\sum_{k=0}^{+\infty} |[\eta_k]_{\delta_{k+1}}^{\delta_{k+1}}| \leq c \sum_{k=0}^{+\infty} \frac{\delta_k^{\beta+1}}{\delta_{k+1}^\beta} = c \rho^\beta \sum_{k=0}^{+\infty} \rho^{-k} \delta_0 = c \frac{\rho^{\beta+1}}{\rho-1} \delta_0. \quad (16)$$

It follows from Eqs. (14) to (16) that

$$|[\gamma]_{\delta_0}^{\delta_0}| \leq c \frac{\rho^{\beta+1}}{\rho-1} \delta_0 + \frac{1}{2} |[\gamma]_{\delta_0}^{\rho\delta_0}|$$

and then

$$|[\gamma]_{\delta_0}^{\delta_0}| \leq 2c \frac{\rho^{\beta+1}}{\rho-1} \delta_0 + |[\gamma]_{\delta_0}^{\rho\delta_0}|.$$

In particular, choosing  $\delta_0 = \|d_\mu\|_{\infty, \gamma}$  provides an upper bound on the length of the entire geodesic and concludes the proof with

$$|\gamma| \leq 2c \frac{\rho^{\beta+1}}{\rho-1} \|d_\mu\|_{\infty, \gamma}.$$

$\square$

### B.3 Local Bound

Finally, we provide the proof of Corollary 6. The first step is to upper bound the DTM over a geodesic using its values at the endpoints and the Euclidean distance between them.

**Lemma 23.** Let  $x, y \in \mathbb{R}^d$  and  $\gamma \in \Gamma_{\mu, F}^*(x, y)$  be a geodesic path. Then

$$\|d_\mu\|_{\infty, \gamma} \leq \max(d_\mu(x), d_\mu(y)) + \|x - y\| .$$

*Proof.* Denote  $r = \|x - y\|$  and  $\delta_0 = \max(d_\mu(x), d_\mu(y))$ . For all  $t \in [0, 1]$ , by Lipschitz property of  $d_\mu$ ,  $d_\mu((1-t)x + ty) \leq d_\mu(x) + t\|x - y\| \leq \delta_0 + tr$ , so that

$$D_\mu([x, y]) \leq r \int_0^1 (\delta_0 + tr)^\beta dt = \frac{1}{\beta + 1} \left( (\delta_0 + r)^{\beta+1} - \delta_0^{\beta+1} \right) . \quad (17)$$

Let  $\gamma \in \Gamma_{\mu, F}^*(x, y)$  be an arc length parameterization, that is,  $\gamma : [0, |\gamma|] \rightarrow \mathbb{R}^d$  with  $\|\dot{\gamma}(t)\| = 1$  almost everywhere. Denote  $\|d_\mu\|_{\infty, \gamma} = \delta$ . If  $\delta < \delta_0$  then the result is immediate. Else, assume that  $\delta \geq \delta_0$  and denote  $\tau \in [0, |\gamma|]$  such that  $d_\mu(\gamma(\tau)) = \delta$ . The Lipschitz property of  $d_\mu$  ensures that  $\delta - d_\mu(x) \leq \tau$  and the FDTM length of the section of  $\gamma$  between times  $\tau - \delta + d_\mu(x)$  and  $\tau$  can be lower bounded as

$$\int_{\tau - \delta + d_\mu(x)}^\tau d_\mu(\gamma(t))^\beta dt \geq \int_0^{\delta - d_\mu(x)} (\delta - t)^\beta dt = \frac{1}{\beta + 1} (\delta^{\beta+1} - d_\mu(x)^{\beta+1}) .$$

Doing the same reasoning for the integral between  $\tau$  and  $\tau + \delta - d_\mu(y)$  then implies that

$$D_{\mu, F}(\gamma) \geq \frac{1}{\beta + 1} (\delta^{\beta+1} - d_\mu(x)^{\beta+1}) + \frac{1}{\beta + 1} (\delta^{\beta+1} - d_\mu(y)^{\beta+1}) \geq \frac{2}{\beta + 1} (\delta^{\beta+1} - \delta_0^{\beta+1}) . \quad (18)$$

Now, the fact that  $\gamma$  is a geodesic together with Eqs. (17) and (18) imply that

$$\frac{2}{\beta + 1} (\delta^{\beta+1} - \delta_0^{\beta+1}) \leq D_\mu(\gamma) \leq D_\mu([x, y]) \leq \frac{1}{\beta + 1} \left( (\delta_0 + r)^{\beta+1} - \delta_0^{\beta+1} \right) ,$$

hence

$$\delta^{\beta+1} \leq \frac{1}{2} \left( (\delta_0 + r)^{\beta+1} + \delta_0^{\beta+1} \right) \leq (\delta_0 + r)^{\beta+1}$$

and finally

$$\|d_\mu\|_{\infty, \gamma} = \delta \leq \delta_0 + r = \max(d_\mu(x), d_\mu(y)) + \|x - y\| .$$

□

*Proof of Corollary 6.* Let  $\gamma$  be an arc length parameterization of a geodesic. From Theorem 4 and Lemma 23 we know that

$$|\gamma| \leq \lambda \left( \max(d_\mu(x), d_\mu(y)) + \|x - y\| \right)$$

where  $\lambda$  is the hidden constant in Eq. (5). If  $\max(d_\mu(x), d_\mu(y)) \leq 2^{\beta+2}\|x - y\|$  then the result follows immediately as

$$|\gamma| \leq (2^{\beta+2} + 1)\lambda\|x - y\| .$$

Else, assume without loss of generality that  $d_\mu(x) > 2^{\beta+2}\|x - y\|$  and notice that by Lipschitz property of  $d_\mu$ ,

$$D_\mu(\gamma) \leq D_\mu([x, y]) \leq \|x - y\| \left( d_\mu(x) + \frac{d_\mu(x)}{2^{\beta+2}} \right)^\beta \leq 2\|x - y\| d_\mu(x)^\beta . \quad (19)$$

On the other hand,

$$D_\mu(\gamma) = \int_0^{|\gamma|} d_\mu(\gamma(t))^\beta dt \geq \int_0^{|\gamma| \wedge d_\mu(x)} (d_\mu(x) - t)^\beta dt . \quad (20)$$

Let us show that  $|\gamma| < \frac{1}{2}d_\mu(x)$ . Indeed, assuming that  $|\gamma| \geq \frac{1}{2}d_\mu(x)$  implies that

$$D_\mu(\gamma) \geq \int_0^{\frac{1}{2}d_\mu(x)} (d_\mu(x) - t)^\beta dt \geq \frac{d_\mu(x)}{2} \left( d_\mu(x) - \frac{d_\mu(x)}{2} \right)^\beta = \left( \frac{d_\mu(x)}{2} \right)^{\beta+1} ,$$

hence with Eq. (19),

$$d_\mu(x)^{\beta+1} \leq 2^{\beta+1} D_\mu(\gamma) \leq 2^{\beta+2} \|x - y\| d_\mu(x)^\beta$$

and

$$d_\mu(x) \leq 2^{\beta+2} \|x - y\| ,$$

leading to a contradiction. Now, the fact that  $|\gamma| \geq \frac{1}{2}d_\mu(x)$  allows to further lower bound Eq. (20) as

$$D_\mu(\gamma) \geq \int_0^{|\gamma|} (d_\mu(x) - t)^\beta dt \geq |\gamma| \left( \frac{d_\mu(x)}{2} \right)^\beta .$$

Together with Eq. (19) this finally implies that

$$|\gamma| \leq \left( \frac{2}{d_\mu(x)} \right)^\beta \cdot 2 \|x - y\| d_\mu(x)^\beta \leq 2^{\beta+1} \|x - y\| .$$

In either case it always holds that

$$|\gamma| \leq (2^{\beta+2} + 1) \|x - y\| ,$$

which concludes the proof.  $\square$

## C Stability with regard to the Domain

This section is devoted to the proof of Theorem 16. We fix  $K \subset \mathbb{R}^d$ ,  $F \subset K$  and  $\mu \in \mathcal{M}_K$  and assume without loss of generality that  $\text{diam}(K) = 1$  (see Section A.1), which allows to upper bound  $d_\mu \leq 1$  over  $K$  (see Eq. 4). In particular, since  $d_\mu$  is 1-Lipschitz and  $t \in [0, 1] \mapsto t^\beta$  is  $\beta$ -Lipschitz, it follows that  $d_\mu^\beta$  is  $\beta$ -Lipschitz over  $K$ .

### C.1 Preliminary Bounds

First, we provide various upper bounds that will be used to compare the FDTM length of slightly offset paths.

**Lemma 24.** Let  $x, y \in K$ . Then

$$D_\mu([x, y]) \leq D_{\mu, F}(x, y) + \frac{\beta}{2} \|x - y\|^2 .$$

*Proof.* Let  $\gamma \in \Gamma_{\mu, F}^*(x, y)$  be a geodesic and  $\bar{\gamma} = [x, y]$  the straight path from  $x$  to  $y$ , both with an arc length parameterization. First, cut the integral in two halves as

$$D_{\mu, F}([x, y]) = \int_0^{|\bar{\gamma}|/2} d_\mu(\bar{\gamma}(t))^\beta dt + \int_{|\bar{\gamma}|/2}^{|\bar{\gamma}|} d_\mu(\bar{\gamma}(t))^\beta dt .$$

Let us compare each half to its equivalent half in  $D_{\mu,F}(x, y) = D_{\mu}(\gamma)$ . Recall that  $d_{\mu}^{\beta}$  is  $\beta$ -Lipschitz over  $K$ . Then

$$\int_0^{|\bar{\gamma}|/2} d_{\mu}(\bar{\gamma}(t))^{\beta} dt \leq \int_0^{|\bar{\gamma}|/2} d_{\mu}(\gamma(t))^{\beta} dt + \beta \int_0^{|\bar{\gamma}|/2} \|\bar{\gamma}(t) - \gamma(t)\| dt .$$

The first term is less than the first half of the integral  $D_{\mu}(\gamma)$  as  $|\bar{\gamma}| = \|x - y\| \leq |\gamma|$ . As for the second term, both  $\bar{\gamma}$  and  $\gamma$  are 1-Lipschitz and coincide at 0 so that  $\|\bar{\gamma}(t) - \gamma(t)\| \leq 2t$ . Then

$$\int_0^{|\bar{\gamma}|/2} d_{\mu}(\bar{\gamma}(t))^{\beta} dt \leq \int_0^{|\gamma|/2} d_{\mu}(\gamma(t))^{\beta} dt + \frac{\beta}{4} \|x - y\|^2 .$$

With the exact same reasoning, a similar bound holds for the end of the paths as

$$\int_{|\bar{\gamma}|/2}^{|\bar{\gamma}|} d_{\mu}(\bar{\gamma}(t))^{\beta} dt \leq \int_{|\gamma|/2}^{|\gamma|} d_{\mu}(\gamma(t))^{\beta} dt + \frac{\beta}{4} \|x - y\|^2 .$$

Finally, summing both inequalities concludes that

$$D_{\mu,F}([x, y]) \leq D_{\mu,F}(x, y) + \frac{\beta}{2} \|x - y\|^2 .$$

□

**Lemma 25.** Let  $x, y, x', y' \in K$  such that  $\|x - x'\| \leq \varepsilon$  and  $\|y - y'\| \leq \varepsilon$ . Then

$$D_{\mu}([x', y']) \leq D_{\mu}([x, y]) + \beta \|x - y\| \varepsilon + 2\varepsilon . \quad (21)$$

Moreover, if  $x, y \in F$ ,

$$D_{\mu}([x', y']) \leq D_{\mu,F}(x, y) + \frac{\beta}{2} (\|x - y\| + 2\varepsilon)^2 + 2\varepsilon . \quad (22)$$

*Proof.* First, by triangular inequality,

$$\|x' - y'\| \leq \|x' - x\| + \|x - y\| + \|y - y'\| \leq \|x - y\| + 2\varepsilon \quad (23)$$

and

$$\|((1-t)x + ty) - ((1-t)x' + ty')\| \leq (1-t)\|x - x'\| + t\|y - y'\| \leq \varepsilon . \quad (24)$$

Recall that  $d_{\mu}^{\beta}$  is  $\beta$ -Lipschitz over  $K$ . Then

$$\begin{aligned} D_{\mu}([x', y']) &= \|x' - y'\| \int_0^1 d_{\mu}((1-t)x' + ty')^{\beta} dt \\ &\leq (\|x - y\| + 2\varepsilon) \int_0^1 d_{\mu}((1-t)x' + ty')^{\beta} dt && \text{(Eq. 23)} \\ &\leq \|x - y\| \int_0^1 d_{\mu}((1-t)x' + ty')^{\beta} dt + 2\varepsilon && (d_{\mu} \leq 1 \text{ over } K) \\ &\leq \|x - y\| \int_0^1 (d_{\mu}((1-t)x + ty)^{\beta} + \beta\varepsilon) dt + 2\varepsilon && \text{(Eq. 24)} \\ &= D_{\mu}([x, y]) + \beta \|x - y\| \varepsilon + 2\varepsilon \end{aligned}$$

which is Eq. (21). Alternatively, if  $x, y \in F$ , then  $[x', x] + [x, y] + [y, y'] \in \Gamma_F(x', y')$  and

$$\begin{aligned} D_{\mu}([x', y']) &\leq D_{\mu,F}(x', y') + \frac{\beta}{2} \|x' - y'\|^2 && \text{(Lemma 24)} \\ &\leq D_{\mu,F}(x, y) + \frac{\beta}{2} \|x' - y'\|^2 + D_{\mu,F}(x', x) + D_{\mu,F}(y, y') && \text{(triangular inequality)} \\ &\leq D_{\mu,F}(x, y) + \frac{\beta}{2} (\|x - y\| + 2\varepsilon)^2 + 2\varepsilon && \text{(Eq. 23) and } d_{\mu} \leq 1 \end{aligned}$$

which is Eq. (22). □

## C.2 Main Proof

In order to establish the stability result we show that for all  $r > 0$  any path can be decomposed into sections of length close to  $r$ , excluding long outer sections that cannot be broken down, and with the exception of a possible small section between two long outer sections. This decomposition will be applied to a geodesic  $\gamma \in \Gamma_{\mu, F}^*(x, y)$  to construct a path in  $\Gamma_G(x, y)$  that approximates  $\gamma$ . From now on, the convention used for path parameterization is to use an arc length parameterization:  $\gamma : [0, |\gamma|] \rightarrow \mathbb{R}^d$  with  $\|\dot{\gamma}(t)\| = 1$  almost everywhere.

**Lemma 26** (Path Decomposition). Let  $\gamma \in \Gamma_F(x, y)$ . For all  $r > 0$  one can construct a sequence of parameters  $0 = t_0, \dots, t_N = |\gamma|$  and associated points  $x_i = \gamma(t_i)$  such that

- For all  $i \in \{0, \dots, N\}$ ,  $x_i \in F \cup \{x, y\}$ .
- For all  $i \in \{0, \dots, N-1\}$ , the section  $[t_i, t_{i+1}]$  satisfies one of the following:
  - (a)  $t_{i+1} - t_i \geq r$  and  $[t_i, t_{i+1}]$  is an outer section of  $\gamma$  w.r.t. the domain  $F$ .
  - (b)  $\frac{1}{2}r \leq t_{i+1} - t_i \leq 2r$ .
  - (c)  $t_{i+1} - t_i \leq \frac{1}{2}r$  and nearby sections  $[t_{i-1}, t_i]$  and  $[t_{i+1}, t_{i+2}]$  (if defined) both satisfy (a).

Furthermore, denote  $\{0, \dots, N-1\} = A \sqcup B \sqcup C$  with  $A$  (resp.  $B$  and  $C$ ) a set of indices  $i$  such that  $[t_i, t_{i+1}]$  satisfies (a) (resp. (b) and (c)). Assuming that  $\gamma$  is not composed of a single section of type (c), it holds that

$$|\gamma| \geq \frac{1}{4} \left( 2 \sum_{i \in A} (t_{i+1} - t_i) + 2|B|r + |C|r \right). \quad (25)$$

Note that the case of a single section of type (c) is excluded from Eq. (25) for technical reasons and correspond to the case where the length of the path is negligible compared to the parameter  $r$ , which will later be chosen proportional to  $\sqrt{\max_{x \in F} d(x, G)}$ . Theorem 16 is easy to derive in this case regardless, which will be done at the end of the section.

*Proof.* First, identify all the outer sections  $[s, t]$  of  $\gamma$  such that  $(t - s) \geq r$ . These sections all satisfy (a), and to conclude it suffices to show that the remaining connected parts of  $\gamma$  can each be decomposed into sections of type (b) or (c).

Consider such section  $[t_0, t_{-1}]$ . By assumption, no outer section  $[s, t] \subset [t_0, t_{-1}]$  has length  $t - s \geq r$ . In other words, if  $t - s \geq r$ , then  $\gamma([s, t])$  intersects  $F$ . If  $|\gamma| \leq \frac{1}{2}r$ , then the whole section satisfies (c) right away and no decomposition is needed. Else, we construct the decomposition iteratively. Assuming  $t_0, \dots, t_i$  are constructed and  $t_{-1} - t_i \geq \frac{1}{2}r$ —which is true for  $i = 0$ —we do the following:

- If  $t_{-1} - t_i \leq 2r$ , set  $t_{i+1} = t_{-1}$ . Then  $[t_i, t_{i+1}]$  satisfies (b), which concludes the construction.
- Else, set  $t_{i+1} = \inf\{t \geq t_i + \frac{1}{2}r : \gamma(t) \in F\}$ . Then:
  - $t_{i+1} - t_i \leq \frac{3}{2}r$ . Indeed, else  $[t_i + \frac{1}{2}r, t_i + \frac{3}{2}r]$  would be an outer section of length at least  $r$ , leading to a contradiction. Therefore,  $[t_i, t_{i+1}]$  satisfies (b).
  - We have

$$t_{-1} - t_{i+1} = t_{-1} - t_i - (t_i - t_{i+1}) \geq 2r - \frac{3}{2}r = \frac{1}{2}r,$$

which allows to continue the construction.

Then each section of this decomposition of  $[t_0, t_{-1}]$  satisfy (b), which concludes the construction.

We now proceed to prove Eq. (25). By definition of the decomposition, after a type (c) section comes a type (a) section or the end of the path, hence  $|C| \leq |A| + 1$ , which in turns implies  $|C| \leq 2|A|$  provided that

$|C| = 0$  or  $|A| \geq 1$ , which is true as long as the path is not composed of a unique type (c) section. In this case,

$$\begin{aligned}
|\gamma| &= \sum_{i \in A \cup B \cup C} (t_{i+1} - t_i) \\
&\geq \sum_{i \in A} (t_{i+1} - t_i) + \frac{1}{2}|B|r && (t_{i+1} - t_i \geq r/2 \text{ if } i \in B) \\
&\geq \frac{1}{2} \sum_{i \in A} (t_{i+1} - t_i) + \frac{1}{2}|A|r + \frac{1}{2}|B|r && (t_{i+1} - t_i \geq r \text{ if } i \in A) \\
&\geq \frac{1}{4} \left( 2 \sum_{i \in A} (t_{i+1} - t_i) + 2|B|r + |C|r \right) && (|A| \geq |C|/2)
\end{aligned}$$

which concludes the proof.  $\square$

Given a geodesic  $\gamma$ , Theorem 16 is proven by constructing a path  $\bar{\gamma}$  admissible for the domain  $G$  as close as possible to  $\gamma$ . This is done by decomposing  $\gamma$  as in Lemma 26 and forming the polygonal path going through the closest neighbors in  $G$  of each intermediate step. Upper bounds from Lemma 25 are then used to compare the FDTM distance between each section of  $\gamma$  and the corresponding segment in  $\bar{\gamma}$ .

*Proof of Theorem 16.* Let  $\gamma \in \Gamma_{\mu,F}^*(x, y)$  be a geodesic and denote

$$\varepsilon = \max_{x \in F} d(x, G) \leq \frac{1}{25\beta} \quad \text{and} \quad r = \sqrt{\frac{\varepsilon}{\beta}}.$$

In particular,  $\varepsilon \leq \frac{r}{5}$ . Consider the decomposition of  $\gamma$  given by Lemma 26 with such  $r$  and assume that we are not in the case of a single type (c) section. Let  $x'_0 = x$ ,  $x'_N = y$  and for each  $1 \leq i \leq N-1$ ,  $x'_i \in G$  such that  $\|x_i - x'_i\| \leq \varepsilon$ . The polygonal path  $\bar{\gamma} = [x'_0, x'_1, \dots, x'_N]$  belongs to  $\Gamma_G(x, y)$  as it is made of segments between points of  $G \cup \{x, y\}$ . Then, upper bounding the difference between  $D_\mu([x'_i, x'_{i+1}])$  and  $D_{\mu,F}(x_i, x_{i+1})$  and summing over  $i$  allows to obtain the desired result.

The following provides bounds for all three cases. They are obtained by using Lemma 25, the bounds on  $\varepsilon$  and by upper bounding complex numerical fractions by “cleaner” constants. If  $i \in A$ , the geodesic  $\gamma$  draws a straight line between  $x_i$  and  $x_{i+1}$ , which implies that this line itself is a geodesic and that  $\|x_i - x_{i+1}\| = t_{i+1} - t_i \geq r$ . Then, the first inequality in Lemma 25 eventually yields

$$D_\mu([x'_i, x'_{i+1}]) - D_{\mu,F}(x_i, x_{i+1}) \leq (\beta\sqrt{\varepsilon} + 2\sqrt{\beta})(t_{i+1} - t_i)\sqrt{\varepsilon} \leq 5\sqrt{\beta}(t_{i+1} - t_i)\sqrt{\varepsilon}. \quad (26)$$

If  $i \in B$ ,  $\|x_i - x_{i+1}\| \leq 2r$  and the second inequality in Lemma 25 eventually yields

$$D_\mu([x'_i, x'_{i+1}]) - D_{\mu,F}(x_i, x_{i+1}) \leq \frac{\beta}{2}(2r + 2\varepsilon)^2 + 2\varepsilon \leq 5\sqrt{\beta}r\sqrt{\varepsilon}. \quad (27)$$

If  $i \in C$ ,  $\|x_i - x_{i+1}\| \leq \frac{r}{2}$  and the second inequality in Lemma 25 eventually yields

$$D_\mu([x'_i, x'_{i+1}]) - D_{\mu,F}(x_i, x_{i+1}) \leq \frac{\beta}{2} \left( \frac{r}{2} + 2\varepsilon \right)^2 + 2\varepsilon \leq \frac{5}{2}\sqrt{\beta}r\sqrt{\varepsilon}. \quad (28)$$

Upper bounding the FDTM on the path  $\bar{\gamma}$  according to Eqs. (26) to (28) then yields

$$\begin{aligned}
D_{\mu,G}(x, y) &\leq D_\mu(\bar{\gamma}) \\
&= \sum_{i \in A \cup B \cup C} D_\mu([x'_i, x'_{i+1}]) \\
&\leq \sum_{i \in A \cup B \cup C} D_\mu(x_i, x_{i+1}) + \frac{5}{2}\sqrt{\beta} \left( 2 \sum_{i \in A} (t_{i+1} - t_i) + 2|B|r + |C|r \right) \sqrt{\varepsilon} \\
&\leq D_\mu(\gamma) + \frac{5}{2}\sqrt{\beta} \cdot 4|\gamma| \cdot \sqrt{\varepsilon} && (\text{Eq. 25}) \\
&= D_{\mu,F}(x, y) + 10\sqrt{\beta}|\gamma|\sqrt{\varepsilon},
\end{aligned}$$

which concludes the proof since  $|\gamma|$  is upper bounded by a constant according to Corollary 5. It now remains to treat the case of a single type (c) section in the decomposition of  $\gamma$ . In this case,  $\|x - y\| \leq \frac{\varepsilon}{2} = \frac{1}{2}\sqrt{\varepsilon/\beta}$  along with  $\|x - y\| \leq |\gamma|$  and we have by Lemma 24

$$D_{\mu,G}(x, y) - D_{\mu,F}(x, y) \leq \frac{\beta}{2}\|x - y\|^2 \leq \frac{\beta}{2}|\gamma|\frac{1}{2}\sqrt{\frac{\varepsilon}{\beta}} \leq 10\sqrt{\beta}|\gamma|\sqrt{\varepsilon}.$$

□

## D Convergence of the Empirical FDTM

This section is devoted to proving Theorem 17 which we recall below.

**Theorem 17.** Assume that  $\mu \in \mathcal{M}_{K,a,b,\sigma}$  satisfies Assumption (A) and  $m \leq \frac{1}{2}$ . Then for all  $n \geq \frac{1}{m}$ ,

$$\mathbb{E} \left[ \|D_\mu - D_{\hat{\mu}_n}\|_{\infty, K} \right] \lesssim \text{diam}(K)^{\beta+1} \left( \left( \frac{\text{diam}(K)}{\sigma} \right)^{(p-\beta)\vee 0} \frac{d \log(n)}{\sqrt{n}} + \left( \frac{\log(n)}{n} \right)^{\frac{1}{2b}} \right)$$

where  $\lesssim$  hides a multiplicative constant depending on  $m, p, \beta, a$  and  $b$ .

This result derives from convergence results regarding  $|d_\mu^p - d_{\hat{\mu}_n}^p|$  and  $d_{\text{H}}(\mathcal{S}(\mu), \mathbb{X}_n)$ . The following briefly detail these results and how they imply Theorem 17 using Theorems 9 and 15. In the following, assume without loss of generality that  $\text{diam}(K) = 1$  (see Section A.1), which allows to upper bound  $d_\mu \leq 1$  over  $K$  (see Eq. 4). First, we state the convergence of the empirical DTM.

**Proposition 27.** Assume that  $\mu$  is  $(a, b)$ -standard,  $\mathcal{S}(\mu)$  is connected and that  $m \leq \frac{1}{2}$ . Then there exists a constant  $c > 0$  depending on  $m, p, a$  and  $b$  such that for all  $n \geq \frac{1}{m}$ ,

$$\mathbb{E} \left[ \|d_\mu^p - d_{\hat{\mu}_n}^p\|_{\infty, K} \right] \leq cd \frac{\log(n)}{\sqrt{n}}.$$

Proposition 27 is due to [5, Theorem 3] which provides this convergence rate<sup>2</sup> with a constant depending on the modulus of continuity of  $u \mapsto \delta_{\mu,u}(x)^p$ . Under our regularity assumptions, this modulus of continuity can be upper bounded according to [5, Lemma 3] by the function

$$\omega : v \in [0, 1] \mapsto p \left( \frac{v}{a} \right)^{\frac{1}{b}}.$$

We now state the convergence of the sample point cloud to the support in Hausdorff distance.

**Proposition 28.** Assume that  $\mu$  is  $(a, b)$ -standard. Then there exists a constant  $c > 0$  depending on  $a$  and  $b$  such that for all  $n \geq 1$ ,

$$\mathbb{E} [d_{\text{H}}(\mathcal{S}(\mu), \mathbb{X}_n)] \leq c \left( \frac{\log(n)}{n} \right)^{\frac{1}{b}}.$$

Proposition 28 may be obtained following the same overall reasoning as in [7] or [6]. Intuitively, the standard assumption implies that any point  $x$  in the support should be at distance of order  $(an)^{-1/b}$  from  $\mathbb{X}_n$  at most. Indeed,  $\mu(\mathcal{B}(x, (an)^{-1/b})) \geq \frac{1}{n}$ , so that this ball is expected to contain at least 1 point in  $\mathbb{X}_n$ . Moreover, the standard assumption implies that the covering number of the support when using balls of radius  $n^{-1/b}$  is of order  $n$ . Therefore, the Hausdorff distance can be upper bounded by the covering radius plus the maximal distance from any of the  $n$  centers to  $\mathbb{X}_n$ , which eventually leads to an upper bound on the expectation of order  $(\log(n)/n)^{1/b}$  as stated. The precise computations are omitted here and we now move on to the proof of Theorem 17.

<sup>2</sup>[5] only discusses the case where  $m = \frac{k}{n}$  for some integer  $k$ . The general statement for any  $m \in (0, 1)$  is interpolated with a slightly larger constant  $c$ .

*Proof of Theorem 17.* Recall that we assume  $\text{diam}(K) = 1$  without loss of generality. In order to use Proposition 27, assume that  $m \leq \frac{1}{2}$  and  $n \geq \frac{1}{m}$ . Decompose the offset into the offset w.r.t. the DTM and the offset w.r.t. the domain as

$$\|D_\mu - D_{\hat{\mu}_n}\|_{\infty, K} \leq \|D_{\mu, \mathcal{S}(\mu)} - D_{\hat{\mu}_n, \mathcal{S}(\mu)}\|_{\infty, K} + \|D_{\hat{\mu}_n, \mathcal{S}(\mu)} - D_{\hat{\mu}_n, \mathbb{X}_n}\|_{\infty, K}. \quad (29)$$

Recall that  $\sigma \leq d_\mu \leq 1$  and  $d_{\hat{\mu}_n} \leq 1$  almost surely (a.s.) over  $K$  by assumptions. If  $\beta \geq p$  then  $t \in [0, 1] \mapsto t^{\beta/p}$  is  $\frac{\beta}{p}$ -Lipschitz hence

$$|d_\mu^\beta - d_{\hat{\mu}_n}^\beta| \leq \frac{\beta}{p} |d_\mu^p - d_{\hat{\mu}_n}^p| \quad \text{a.s.}$$

Else if  $\beta \leq p$ , for all  $t > 0$  and  $s \geq 0$ ,  $t^{\beta-p} s^p \leq s^\beta \leq t^\beta = t^{\beta-p} t^p$  if  $s \leq t$  and  $t^{\beta-p} s^p \geq s^\beta \geq t^\beta = t^{\beta-p} t^p$  if  $s \geq t$ , so that either way

$$|t^\beta - s^\beta| \leq t^{\beta-p} |t^p - s^p|.$$

In particular, with  $t = d_\mu$  and  $s = d_{\hat{\mu}_n}$ ,

$$|d_\mu^\beta - d_{\hat{\mu}_n}^\beta| \leq d_\mu^{\beta-p} |d_\mu^p - d_{\hat{\mu}_n}^p| \leq \sigma^{\beta-p} |d_\mu^p - d_{\hat{\mu}_n}^p| \quad \text{a.s.}$$

Therefore, whether  $\beta$  is larger or smaller than  $p$ , it always holds that

$$|d_\mu^\beta - d_{\hat{\mu}_n}^\beta| \leq \max\left(\frac{\beta}{p}, \sigma^{\beta-p}\right) |d_\mu^p - d_{\hat{\mu}_n}^p| \quad \text{a.s.} \quad (30)$$

The first term in Eq. (29) can thus be bounded as follows:

$$\begin{aligned} \mathbb{E} \left[ \|D_{\mu, \mathcal{S}(\mu)} - D_{\hat{\mu}_n, \mathcal{S}(\mu)}\|_{\infty, K} \right] &\lesssim \mathbb{E} \left[ \|d_\mu^\beta - d_{\hat{\mu}_n}^\beta\|_{\infty, K} \right] && \text{(Theorem 15)} \\ &\leq \max\left(\frac{\beta}{p}, \sigma^{\beta-p}\right) \mathbb{E} \left[ \|d_\mu^p - d_{\hat{\mu}_n}^p\|_{\infty, K} \right] && \text{(Eq. 30)} \\ &\lesssim \frac{d \log(n)}{\sigma^{(p-\beta) \vee 0} \sqrt{n}} && \text{(Proposition 27)} \end{aligned} \quad (31)$$

where  $\lesssim$  hides a multiplicative constant depending on  $m, p, \beta, a$  and  $b$ . Regarding the second term in Eq. (29),

$$\begin{aligned} \mathbb{E} \left[ \|D_\mu - D_{\mu, \mathbb{X}_n}\|_{\infty, K} \right] &\lesssim \mathbb{E} \left[ \sqrt{d_{\text{H}}(\mathcal{S}(\mu), \mathbb{X}_n)} \right] && \text{(Theorem 9)} \\ &\leq \sqrt{\mathbb{E} [d_{\text{H}}(\mathcal{S}(\mu), \mathbb{X}_n)]} && \text{(Jensen's inequality)} \\ &\lesssim \left( \frac{\log(n)}{n} \right)^{\frac{1}{2b}}. && \text{(Proposition 28)} \end{aligned} \quad (32)$$

Putting together Eqs. (29), (31) and (32) concludes the proof of Theorem 17. When  $\text{diam}(K) \neq 1$ , the dependency with  $\text{diam}(K)$  appears naturally in the inequality as the FDTM is homogeneous to a distance at power  $\beta + 1$  and  $\sigma$  is homogeneous to a distance.  $\square$