Estimation of Conformal Metrics

Jérôme Taupin
Université Paris-Saclay, France
INRIA Saclay, Palaiseau, France

Abstract

We study deformations of the geodesic distances on a domain of $\mathbb{R}^N$ induced by a function called conformal factor. We show that under a positive reach assumption on the domain (not necessarily a submanifold) and mild assumptions on the conformal factor, geodesics for the conformal metric have good regularity properties in the form of a lower bounded reach. This regularity allows for efficient estimation of the conformal metric from a random point cloud with a relative error proportional to the Hausdorff distance between the point cloud and the original domain. We then establish convergence rates of order $n^{-1/d}$ that are close to sharp when the intrinsic dimension d of the domain is large, for an estimator that can be computed in $\mathcal{O}(n^2)$ time. Finally, this paper includes a useful equivalence result between ball graphs and nearest-neighbors graphs when assuming Ahlfors regularity of the sampling measure, allowing to transpose results from one setting to another.

2012 ACM Subject Classification Theory of computation → Computational geometry

Keywords and phrases Geometric inference, metric estimation, conformal metric, geodesics, sets of positive reach

Digital Object Identifier 10.4230/LIPIcs.SoCG.2026.92

Related Version *Full Version*: `https://hal.science/hal-05516466v1` [17]

1 Introduction

This paper studies metrics over subsets of the Euclidean space $(\mathbb{R}^N, \|\cdot\|)$ obtained by conformal deformation of the shortest-path metric via a positive function. We are particularly interested in the regularity of such metrics and in their estimation via i.i.d. sampling of points.

▶ **Definition 1.1.** *Let $M \subset \mathbb{R}^N$ be a closed path-connected* domain *and $f : M \to \mathbb{R}_+^*$ be a* conformal factor. *For all $x, y \in M$ the* conformal distance *between x and y over M via f is defined as*

$$D_{M,f}(x,y) \overset{def}{=} \inf_{\gamma \in \Gamma_M(x,y)} \int_I f(\gamma(t))\|\dot{\gamma}(t)\|\mathrm{d}t \tag{1}$$

where $\Gamma_M(x,y)$ is the set of all Lipschitz paths $\gamma : I \to M$ where $I = [a,b]$ is a nontrivial segment, $\gamma(a) = x$ and $\gamma(b) = y$. The quantity minimized over all paths γ is denoted

$$|\gamma|_f \overset{def}{=} \int_I f(\gamma(t))\|\dot{\gamma}(t)\|\mathrm{d}t$$

and referred to as the conformal length of the curve γ via f. In the case where $f = 1$, we write $D_M(x,y) = D_{M,1}(x,y)$ and $|\gamma| = \int_I \|\dot{\gamma}\|$, respectively the distance between x and y induced by the ambient metric over M and the Euclidean length of a curve γ.

In the following, to ensure regularity of the conformal metric, the domain M is assumed to have positive reach τ_M, which we recall is defined as the supremum of all $r > 0$ such that any point in the offset $M^r = \{x \in \mathbb{R}^N : d(x, M) < r\}$ has a unique projection onto M – where $d(x, M) = \inf_{y \in M} \|x - y\|$ denotes the distance from point x to subset M – see [10]. Moreover, the conformal factor f is assumed to be κ-Lipschitz and lower bounded by $f_{\min} > 0$. These are the sole assumptions used in this paper regarding M and f.

42nd International Symposium on Computational Geometry (SoCG 2026).
Editors: Hee-Kap Ahn, Michael Hoffmann, and Amir Nayyeri; Article No. 92; pp. 92:1–92:15
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Notice that f is defined over M in Definition 1.1 as paths are constrained to the domain. Since any Lipschitz function defined over a subset of $\mathbb{R}^N$ can be extended to the whole space preserving its Lipschitz property and lower bound (see [16, Theorem 1] for instance), one may assume without loss of generality that f is defined over the entire space $\mathbb{R}^N$, which is useful to estimate the conformal metric efficiently. The connectedness and positive reach of M ensure that for any endpoints $x, y \in M$, there exists a Lipschitz path from x to y in M, so that $\Gamma_M(x, y)$ is always nonempty and $D_{M,f}(x, y)$ is always finite. This can be deduced from Lemma 2.1 below. Most of the time, a path γ is chosen to be parameterized with constant velocity, i.e., $\|\dot\gamma\|$ is constant over I, with I being either $[0, 1]$ or $[0, |\gamma|]$. The latter is referred to as an *arc-length* parameterization, where $\|\dot\gamma\| = 1$ almost everywhere.

The *induced metric* $D_M = D_{M,1}$ is the metric induced by the ambient space $\mathbb{R}^N$ onto M. If M is a $\mathcal{C}^1$ submanifold of $\mathbb{R}^N$ and f is $\mathcal{C}^1$, $D_{M,f}$ is a Riemannian metric with tensor $f(x)^2 \cdot g_x$ at point $x \in M$ where g_x is the tensor of the induced metric D_M at x. The term "conformal" used in this paper is borrowed from the Riemannian literature. Finally, if μ is a measure over a submanifold M with density ρ with regard to (w.r.t.) the volume form of M and f is a negative power of the density ρ, $D_{M,f}$ has already been studied and is sometimes referred to as the *Fermat distance* [13]. This particular case is one of the main motivations for this article as the Fermat distance has been used for various practical applications, see for instance [12] and the references therein. Particular examples of the conformal factor f are discussed in Section 5.

Conformal metrics are included in the more general class of length-spaces, for which it is known that the infimum in Definition 1.1 is in fact a minimum [7, Theorem 2.5.23]. That is, for all $x, y \in M$ there exists a path $\gamma \in \Gamma_M(x, y)$ such that $D_{M,f}(x, y) = |\gamma|_f$, called a *minimizing geodesic* – and shortened to geodesic in this paper for brevity. Denote $\Gamma^\star_{M,f}(x, y)$ the set of such geodesics between two endpoints x and y along with $\Gamma^\star_{M,f} = \bigcup_{x\neq y\in M} \Gamma^\star_{M,f}(x, y)$ the set of all geodesics w.r.t. the conformal metric. We discuss in Section 2 the regularity of geodesics w.r.t. $D_{M,f}$ and show in Proposition 2.4 that under the aforementioned assumptions on M and f, geodesics have positive reach that is lower bounded by an explicit constant depending on the reach of the domain and on the conformal factor. In particular, any geodesic may be parameterized as a $\mathcal{C}^{1,1}$ curve with an explicit upper bound on the Lipschitz constant of its first derivative.

We then show in Section 3 that the conformal metric can be approached using polygonal paths on a weighted graph built from a point cloud $X \subset M$, provided that the graph only contains edges of length at most some threshold r and that X is close to M in Hausdorff distance. The small edges condition ensures that paths on the graph cannot venture too far outside the domain. This kind of construction is the same as the one used for the Isomap algorithm [5], although we allow more generality by adapting the weights to the conformal factor f. When $f = 1$ and M is a geodesically convex $\mathcal{C}^2$ submanifold, [5] provides a relative error bound of order $\frac{r^2}{\tau^2} + \frac{\rho}{r}$ where τ is the minimum radius of curvature of M and ρ denotes the Hausdorff distance between M and X. The term depending on ρ can be made quadratic as shown in [2], where an assumption called geodesic smoothness that is slightly weaker than that of a positive reach allows to obtain a bound of order $Ar + \frac{\rho^2}{r^2}$ where A depends on the assumption. This assumption however lacks precision when comparing distances at a local scale which results in the first term being of order r instead of r^2. By using a positive reach assumption instead, the upper bound can be further improved to $\frac{r^2}{\tau^2} + \frac{\rho^2}{r^2}$ according to [3] where τ is the reach of the domain, although the setup also implicitly assumes that the domain M is a smooth manifold isometric to a convex domain. Going back to our setup, we adapt these prior works to fit any conformal change. Under some condition on the weight function used, that in particular always holds for the induced metric, we establish in Theorem 3.5 the same upper bound on the approximation error as in [3], that is $\frac{r^2}{\tau^2} + \frac{\rho^2}{r^2}$ where τ is the explicit lower bound on the reach of any geodesic provided by Proposition 2.4. This result emphasizes the fact that the only assumption on the domain needed for this level of precision is that of a positive reach. The approximation error can thus be made proportional to the Hausdorff distance between the point cloud and the domain by choosing appropriately $r \asymp \sqrt{\tau\,\rho}$. When the conformal factor f is unknown, replacing it with an estimate g does not alter the results provided that g is close enough to f, as shown in Lemma 3.6.

Assuming that $X = X_n$ is the outcome of $n \geq 1$ i.i.d. samples from a d-standard measure μ over M, we study in Section 4 the estimator of $D_{M,f}$ built on X_n following the construction of Section 3. This estimator is shown in Theorem 4.2 to converge to $D_{M,f}$ at a rate of $n^{-1/d}$ provided that f is known or can be estimated with same rate. This rate follows from Theorem 3.5 and the fact that X_n is known to converge to M in Hausdorff distance at a rate of $n^{-1/d}$. In particular, this allows for efficient estimation of the induced metric with the sole assumption that M has positive reach. This convergence is shown using ball graphs. However, practical estimation is made easier by using nearest-neighbors graphs instead, in part because it does not require to know the intrinsic dimension d to obtain an optimal convergence speed. For this reason, we show in Theorem 4.5 that it is possible to retrieve the same convergence speed when replacing the ball graph in the estimator with a k-nearest-neighbors graph with $k \asymp \sqrt{n}$. To do so, we establish an equivalence with high probability between nearest-neighbors graphs and ball graphs when their respective parameters are properly scaled and the underlying measure is d-Ahlfors, see Proposition 4.4. Finally, the nearest-neighbors estimator of the conformal distance between two fixed points may be computed in $\mathcal{O}(n^2)$ if the time complexity of evaluating f is considered constant. Under the stronger assumption that M is a $\mathcal{C}^k$ submanifold of dimension d with $k \geq 2$, the minimax optimal convergence rate for the induced metric is known to be of order $n^{-k/d}$ [1]. When only assuming positive reach and no differential structure on M, the discussion of the optimal minimax rate proves to be harder as we are not able to match the upper bound of $n^{-1/d}$ and obtain a lower bound of order $n^{-1/(d-1/2)}$ instead, see Theorem 4.6. We discuss with more details the comparison between our setup and the $\mathcal{C}^k$ case in Section 4.4.

Technical proofs are deferred to the appendix which is available in the full version [17].

2 Regularity of the geodesics for the conformal metric

Recall that the term "geodesic" refers in this paper to a curve that achieves a global minimum of the conformal length. In this section we study the regularity of the geodesics w.r.t. $D_{M,f}$. We show that geodesics are $\mathcal{C}^{1,1}$ curve with reach bounded from below by a constant depending only on τ_M, κ and $f_{\min}$. This regularity property is crucial to obtain a good approximation of the geodesics by polygonal paths. The reach of a set can be characterized through the local distortion of the induced metric compared to the Euclidean metric.

▶ **Lemma 2.1** ([6, Theorem 1]). *The reach of M may be expressed as*

$$\tau_M = \sup\left\{r > 0 \;:\; \forall x, y \in M,\; \|x-y\| < 2r \Rightarrow D_M(x,y) \leq 2r \arcsin\left(\frac{\|x-y\|}{2r}\right)\right\} .$$

Given a path $\gamma \in \Gamma_M$ parameterized over an interval I and without self-intersection, denote τ_γ the reach of the curve $\gamma(I) \subset \mathbb{R}^N$ along with $D_\gamma = D_{\gamma(I)}$ the induced metric over the curve for short. $D_\gamma(x,y)$ is nothing more than the length of the curve $\gamma(I)$ between two intermediate point x and y. In particular if x and y are the endpoints of γ then $D_\gamma(x,y) = |\gamma|$.

One consequence of Lemma 2.1 is that the length of a geodesic between two points at most $2\tau_M$ apart is upper bounded by the length of an arc of radius τ_M between both points. That is, if $x, y \in M$ and $\|x - y\| < 2\tau_M$ then

$$D_M(x, y) \leq 2\tau_M \arcsin\left(\frac{\|x - y\|}{2\tau_M}\right) . \tag{2}$$

Moreover, if γ is a geodesic, it then induces a geodesic between any of the points it goes through, which implies that D_γ is exactly the restriction of D_M to the curve γ. Then, another consequence of the characterization of Lemma 2.1 is that the reach of M is the minimal reach of any geodesic, that is

$$\tau_M = \inf_{\gamma \in \Gamma^\star_M} \tau_\gamma .$$

We now introduce a notion of reach associated with the conformal deformation of M by f using the same point of view of the metric and its geodesics.

▶ **Definition 2.2.** *The* conformal reach *of M via f is defined as the minimal reach of any geodesic w.r.t. the conformal metric, that is*

$$\tau_{M,f} \overset{def}{=} \inf_{\gamma \in \Gamma^\star_{M,f}} \tau_\gamma .$$

In the case of the induced metric this notion coincides with the usual notion of reach, that is $\tau_{M,1} = \tau_M$. The characterization given by Lemma 2.1 also holds for the conformal reach.

▶ **Proposition 2.3.** *The conformal reach of M via f may be expressed as*

$$\tau_{M,f} = \sup \left\{ \begin{array}{c} r > 0 \ : \ \forall x, y \in M, \forall \gamma \in \Gamma^\star_{M,f}(x, y), \\ \|x - y\| < 2r \Rightarrow |\gamma| \leq 2r \arcsin\left(\frac{\|x-y\|}{2r}\right) \end{array} \right\} . \tag{3}$$

Beware that Equation (3) involves geodesic w.r.t. the conformal metric $D_{M,f}$, but compares their Euclidean length – not conformal – to the one of an arc of radius r. The proof is straightforward and included in the full version [17]. Using Proposition 2.3, we are able to lower bound the reach of any geodesic w.r.t. $D_{M,f}$.

▶ **Proposition 2.4.** *Assume that M has positive reach $\tau_M > 0$ and that f is κ-Lipschitz and lower bounded by $f_{\min} > 0$. Then any conformal geodesic γ w.r.t. $D_{M,f}$ has positive reach $\tau_\gamma > 0$. Precisely, the conformal reach of M via f is lower bounded as follows.*

$$\tau_{M,f} \geq \mathcal{T}_{M,f} \quad \textit{where} \quad \mathcal{T}_{M,f} \overset{def}{=} \min\left(\frac{\tau_M}{2} \ , \ \frac{f_{\min}}{8\kappa}\right) .$$

The reasoning behind Proposition 2.4 is the following. If the reach of a geodesic γ w.r.t. $D_{M,f}$ is small compared to τ_M, then according to Proposition 2.3 there exists a path in M significantly shorter than γ Euclidean-wise. If τ_γ is also small compared to $f_{\min}/\kappa$, then this path is shown to also have a shorter conformal length than γ due to the properties of f, which implies a contradiction with the geodesic nature of γ. This reasoning is detailed in [17, Appendix B.2]. Now, consider an arc-length parameterized geodesic $\gamma : I \to M$ w.r.t. $D_{M,f}$, i.e., such that $\|\dot{\gamma}\| = 1$ almost everywhere over I. Being a geodesic, γ has no self-intersection. Then, stating that γ has positive reach is equivalent to stating that γ is a $\mathcal{C}^{1,1}$ curve, i.e., that $\dot{\gamma}$ is Lipschitz w.r.t. the angular distance. Precisely, for all $t, s \in I$,

$$\angle(\dot{\gamma}(t), \dot{\gamma}(s)) \leq \frac{|t - s|}{\tau_\gamma} . \tag{4}$$

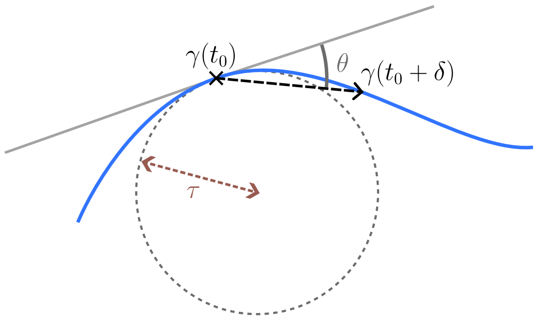

Figure 1 Bounding the angular velocity of a path with positive reach.

See [10, Remark 4.20] and [15, Theorem 4] for references. Equation (4) allows to obtain a finer control on the approximation error of the polygonal paths on the graph by upper bounding efficiently the difference between small successive steps $\gamma(t)-\gamma(t-\delta)$ and $\gamma(t+\delta)-\gamma(t)$ of a geodesic path.

▶ **Lemma 2.5.** *Let $\gamma : [0,|\gamma|] \to \mathbb{R}^N$ be an arc-length parameterized curve without self-intersection and with positive reach τ_γ. Then for all $t_0 \in [0,|\gamma|]$ and $\delta \in (0, \frac{\pi}{2}\tau_\gamma]$ such that $[t_0-\delta, t_0+\delta] \subset [0,|\gamma|]$, the angle between the velocity vector of γ at t_0 and the direction from $\gamma(t_0)$ to $\gamma(t_0+\delta)$ is upper bounded as follows:*

$$\angle\big(\gamma(t_0+\delta)-\gamma(t_0), \dot\gamma(t_0)\big) \le \frac{\delta}{2\tau_\gamma} \ . \tag{5}$$

Moreover, denoting $v^- = \gamma(t_0)-\gamma(t_0-\delta)$ and $v^+ = \gamma(t_0+\delta)-\gamma(t_0)$, the difference between small steps of the path on both side from t_0 is upper bounded as follows:

$$\left\| \frac{v^+}{\|v^+\|} - \frac{v^-}{\|v^-\|} \right\| \le \frac{1}{\tau_\gamma} \min\left(\|v^+\|, \|v^-\|\right) \ . \tag{6}$$

Lemma 2.5 is key to obtain an approximation error of the conformal metric proportional to the Hausdorff distance between the domain and a point cloud. A proof is provided in [17, Appendix A.1] and Equation (5) is illustrated by Figure 1.

3 Polygonal approximation of the conformal metric

Consider a point cloud $X \subset M$ of points sampled from the domain. If the point cloud approximates well the domain, it is possible to estimate the conformal distance over M through weighted polygonal paths built over X within a small margin of error. Assuming that f is known, the edges of the polygonal path are weighted using f to approximate the conformal distance between their endpoints. The degree of approximation of M by X is measured using the Hausdorff distance

$$d_{\mathrm{H}}(M,X) \overset{\text{def}}{=} \max\left(\sup_{x\in M} d(x,X)\ ,\ \sup_{x\in X} d(x,M)\right)$$

which simplifies to $d_{\mathrm{H}}(M,X) = \sup_{x\in M} d(x,X)$ assuming that $X \subset M$. By definition, any point in M is at distance at most $d_{\mathrm{H}}(M,X)$ from a point in X.

The idea behind the polygonal approximation is the following. Consider a geodesic $\gamma \in \Gamma^\star_{M,f}(x,y)$ between two points $x,y \in M$. If $d_{\mathrm{H}}(M,X)$ is small, there exists a polygonal path over X that follows closely the trajectory of γ. Assuming that the edges of this path are equipped with appropriate weights to simulate the conformal distance between their endpoints, the total weight of the polygonal path should be close to the conformal length $|\gamma|_f$. Moreover, since γ is a geodesic, polygonal paths should not be able to have a weight much smaller than $|\gamma|_f$. This is ensured by using only short edges, which allows the weights to approximate well the conformal length by preventing shortcuts outside the domain. In the end, the shortest weighted path on the graph is expected to retrieve $|\gamma|_f = D_{M,f}(x, y)$.

3.1 Weighted graphs

Let us now describe formally this construction. The polygonal approximation of $D_{M,f}$ is defined as the metric of an appropriate weighted graph on a point cloud X.

▶ **Definition 3.1.** *Let $X \subset \mathbb{R}^N$ be a finite point cloud.*

- *For all $r > 0$, the r-*ball graph *over X is denoted $G_r(X)$ and is the graph of vertex set X and with an edge between points x and y if and only if $\|x - y\| \leq r$.*
- *For all integer $k \geq 1$, the k-*nearest-neighbors graph *(or k-NN graph) over X is denoted $\mathcal{G}_k(X)$ and is the graph of vertex set X with an edge between two points x and y if and only if x (or y) is among the k nearest points of X to y (or x) excluding self.*

In both cases, the parameter r or k is referred to as the threshold *of the graph.*

Ball graphs are more convenient to study whereas NN graphs are more practical, see Section 4.2. We define the polygonal metric in both contexts. The graph used is equipped with a weight function to approximate the conformal length of its edges.

▶ **Definition 3.2.** *Given a function $f : \mathbb{R}^N \to \mathbb{R}^*_+$, consider the weight functions*

$$w_{f,q}(x, y) \overset{def}{=} \frac{\|x - y\|}{2(q-1)} \left(f(x) + 2\sum_{k=2}^{q-1} f\left(\frac{q-k}{q-1}x + \frac{k-1}{q-1}y\right) + f(y) \right)$$

defined for any $q \geq 2$ and $x, y \in \mathbb{R}^N$, along with

$$w_{f,\infty}(x, y) \overset{def}{=} \|x - y\| \int_0^1 f\big((1-t)x + ty\big)\mathrm{d}t \ .$$

The parameter $q \in \{2, \ldots, \infty\}$ is referred to as the resolution *of the weights.*

The weight function $w_{f,q}$ is meant to approximate the conformal distance between the endpoints of an edge. Proposition 3.3 below shows that it is indeed the case when the endpoints are close, which is the reason why graphs with small edges are used. Recall that we have assumed without loss of generality that f is defined over the whole space, allowing its evaluation outside M when $q \neq 2$. As q grows, the weight $w_{f,q}$ becomes more accurate and converges to the weight with infinite resolution $w_{f,\infty}$, the latter being exactly the conformal length of the straight path. However, the computational cost is linear with q and there is no closed-form formula for $w_{f,\infty}$ in general. On the other hand,

$$w_{f,2}(x, y) = \|x - y\| \frac{f(x) + f(y)}{2}$$

is the simplest choice for the weights and may be more suitable if evaluating f is possible only at points in X. Note that the weight $w_{f,q}$ is the optimal approximation of $w_{f,\infty}$ using only q samples of the Lipschitz function f. Under stronger assumptions on f such as $\mathcal{C}^k$ regularity, other definitions would be better suited.

▶ **Proposition 3.3.** *For all $x, y \in M$ such that $\|x - y\| \leq \tau_{M,f}$,*

$$\left|1 - \frac{w_{f,q}(x,y)}{D_{M,f}(x,y)}\right| \leq \frac{\kappa}{4f_{\min}} \frac{\|x-y\|}{q-1} + \frac{\|x-y\|^2}{16\mathcal{T}_{M,f}^2} \ . \tag{7}$$

and we denote $\delta_q(\|x-y\|)$ this upper bound. The first term in Equation (7) *is to be interpreted as* 0 *if $q = \infty$ and represents the offset between $w_{f,q}$ and $w_{f,\infty}$, whereas the second term represents the offset between $w_{f,\infty}$ and $D_{M,f}$.*

Proposition 3.3 shows that the conformal metric $D_{M,f}$ may be approximated locally by the weights $w_{f,q}$. Note that for fixed resolution q in the r-ball graph, the distortion between the weights and the conformal distances is linear in r. However, by choosing q to be inversely proportional to r, the distortion becomes quadratic in r. The proof for Proposition 3.3 is provided in [17, Appendix B.3]. We now introduce the polygonal approximation of the conformal metric as the metric of the weighted graph built on X.

▶ **Definition 3.4.** *Consider a point cloud $X \subset \mathbb{R}^N$, function $f : \mathbb{R}^N \to \mathbb{R}_+^*$ and parameters $r > 0$ or $k \geq 1$ and $q \in \{2, \ldots, \infty\}$. The polygonal metric associated with these parameters is defined between two points $x, y \in \mathbb{R}^N$ as*

$$\widehat{D}_{X,f}(x,y) \overset{def}{=} \min_{(x_0,\ldots,x_K)} \sum_{k=0}^{K-1} w_{f,q}(x_k, x_{k+1}) \tag{8}$$

where the minimum is taken over the set of paths $(x_0, \ldots, x_K)$ such that $x_0 = x$ and $x_k = y$ in the graph G that is chosen either as the r-ball graph $G_r(X \cup \{x,y\})$ or the k-NN graph $\mathcal{G}_k(X \cup \{x,y\})$.

The choice of a ball graph or a NN graph along with the threshold r or k and the resolution q of the weights are left implicit when writing $\widehat{D}_{X,f}$ to avoid heavy notations. Note that when the endpoints x and y do not belong to the point cloud, the distance $\widehat{D}_{X,f}(x,y)$ is computed by adding them to the graph. As a result, $\widehat{D}_{X,f}$ induces a distance over X but not over M. Indeed, when considering endpoints outside X, $\widehat{D}_{X,f}$ may not satisfy the triangular inequality due to the set of possible paths depending on the endpoints.

3.2 Metric approximation

To evaluate how efficient is the approximation of a metric, the error between the true distances and their estimation may be measured using either of the multiplicative loss functions

$$\ell_{\infty,M}\big(D', D\big) \overset{\text{def}}{=} \sup_{x \neq y \in M} \left|\frac{D'(x,y) - D(x,y)}{D'(x,y) \vee D(x,y)}\right| \quad \text{and} \quad l_{\infty,M}\big(D' | D\big) \overset{\text{def}}{=} \sup_{x \neq y \in M} \left|1 - \frac{D'(x,y)}{D(x,y)}\right| \ .$$

In the case where D' takes infinite values – which happens for instance when D' is the metric of a non-connected graph – we let $\ell_{\infty,M}(D', D) = 1$ and $l_{\infty,M}(D'|D) = +\infty$. The inequality

$$\ell_{\infty,M}\big(D', D\big) \leq l_{\infty,M}\big(D'|D\big) \leq \frac{\ell_{\infty,M}\big(D', D\big)}{1 - \ell_{\infty,M}\big(D', D\big)} \tag{9}$$

holds in general, so that both losses are equivalent. The loss $\ell_{\infty,M}$ is however easier to manipulate in some situations as it is symmetric and upper bounded by 1. Thanks to the regularity of geodesics stated by Proposition 2.4 and the local approximation of the conformal metric by the weights $w_{f,q}$ stated by Proposition 3.3, we are able to show that the conformal metric $D_{M,f}$ is approximated by the polygonal metric from Definition 3.4 using the r-ball graph.

▶ **Theorem 3.5.** *Let $X \subset M$ be a point cloud, $r > 0$ and $q \in \{2, \ldots, \infty\}$ two parameters. Assume that $4d_H(M, X) \leq r \leq \mathcal{T}_{M,f}$. Then the approximation $\widehat{D}_{X,f}$ defined in Definition 3.4 using the r-ball graph with resolution q satisfies*

$$l_{\infty,M}\big(\widehat{D}_{X,f}|D_{M,f}\big) \leq \frac{1}{32\mathcal{T}_{M,f}}\frac{r}{q-1} + \frac{r^2}{8\mathcal{T}_{M,f}^2} + 56\frac{d_H(M,X)^2}{r^2} \ . \tag{10}$$

The first term in Equation (10) *is to be interpreted as* 0 *if* $q = \infty$.

Let us describe the reasoning behind Theorem 3.5. Given $x \neq y \in X$, paths in $G_r(X)$ cannot create significant shortcuts outside M as $r \leq \tau_{M,f}$. This is illustrated by Proposition 3.3 from which it can be deduced that

$$\widehat{D}_{X,f}(x,y) \geq \big(1 - \delta_q(r)\big) D_{M,f}(x,y) \tag{11}$$

and, assuming that $\|x - y\| \leq r$,

$$\widehat{D}_{X,f}(x,y) \leq \big(1 + \delta_q(r)\big) D_{M,f}(x,y) \ . \tag{12}$$

As for when $\|x - y\| > r$, consider a geodesic $\gamma \in \Gamma^\star_{M,f}(x,y)$. Decompose γ into sections of length at most $r - 2d_{\mathrm{H}}(M,X)$ and for each intermediate point select a point in X at distance at most $d_{\mathrm{H}}(M,X)$. This process draws a polygonal path in $G_r(X \cup \{x,y\})$ as it uses edges of length at most r. Summing the approximation error from Proposition 3.3 between each edge of this path and the corresponding section of γ eventually yields the upper bound

$$\widehat{D}_{X,f}(x,y) \leq \left(1 + \delta_q(r) + c_1\frac{d_{\mathrm{H}}(M,X)}{\mathcal{T}_{M,f}} + c_2\frac{d_{\mathrm{H}}(M,X)^2}{r^2}\right) D_{M,f}(x,y) \tag{13}$$

when $\|x - y\| > r$ and where c_1 and c_2 are universal constants. The positive reach of γ and Lemma 2.5 play a crucial role in obtaining terms of order $d_{\mathrm{H}}(M,X)/\mathcal{T}_{M,f}$ and $d_{\mathrm{H}}(M,X)^2/r^2$ instead of $d_{\mathrm{H}}(M,X)/r$. Together, Equations (11)–(13) eventually imply Equation (10). The details of this reasoning are provided in [17, Appendix B.4]. The upper bound in Equation (10) is made proportional to $d_{\mathrm{H}}(M,X)$ by setting

$$r \asymp \sqrt{\mathcal{T}_{M,f}\ d_{\mathrm{H}}(M,X)} \quad \text{and} \quad q \asymp \sqrt{\frac{\mathcal{T}_{M,f}}{d_{\mathrm{H}}(M,X)}} \ .$$

In the case of the induced metric with $f = 1$, the first term in the right-hand side of Equation (10) disappears, and the error is of order $r^2/\tau_M^2 + d_{\mathrm{H}}(M,X)^2/r^2$ similarly to the one obtained in [3]. In the general conformal case, we keep this magnitude of error by setting the resolution q to be at least of order $1/r$. If it is not possible to evaluate f outside the point cloud X, setting $q = 2$ instead yields an error of order $r/\mathcal{T}_{M,f} + d_{\mathrm{H}}(M,X)^2/r^2$ which is slightly worse. A similar upper bound was achieved in [2] for the induced metric and the term linear in r was due to the geodesically smooth assumption used in the paper being slightly weaker than a positive reach assumption. In particular, the latter allows for an efficient local estimation in the form of Equation (12). The main takeaway of Theorem 3.5 is that the crucial assumption to achieve an approximation error proportional to $d_{\mathrm{H}}(M,X)$ is the positive reach. In particular, the domain need not be $\mathcal{C}^2$, or even a submanifold.

3.3 Unknown conformal factor

In general, f may not be known and needs to be estimated from the data. The polygonal metric $\widehat{D}_{X,g}$ from Definition 3.4 can be defined for any function $g : \mathbb{R}^N \to \mathbb{R}_+^*$ without the need of Lipschitz and lower bound assumptions. Then, the approximation error between $\widehat{D}_{X,g}$ and $D_{M,f}$ is upper bounded by the sum of the error w.r.t. the conformal factor and the error w.r.t. the domain.

▶ **Lemma 3.6.** *Let $f : \mathbb{R}^N \to \mathbb{R}_+$ a function lower bounded by $f_{\min}$, $X \subset M$ a point cloud and $g : \mathbb{R}^N \to \mathbb{R}_+^*$ such that $\|g - f\|_\infty \le \frac{1}{2} f_{\min}$. Then*

$$\ell_{\infty,M}\left(\widehat{D}_{X,g}, D_{M,f}\right) \le \frac{2}{f_{\min}} \|g - f\|_\infty + 2\ell_{\infty,M}\left(\widehat{D}_{X,f}, D_{M,f}\right) . \tag{14}$$

This result holds regardless of the type of graph, threshold and resolution as long as they are shared between $\widehat{D}_{X,f}$ and $\widehat{D}_{X,g}$.

The proof of Lemma 3.6 consists in straightforward computations and is deferred to [17, Appendix B.5]. As a consequence, in the context of estimation of the domain M by a point cloud X_n as in Section 4 and of the conformal factor f by a function $f_n : \mathbb{R}^N \to \mathbb{R}_+$, the rate of convergence of $\widehat{D}_{X_n,f_n}$ towards $D_{M,f}$ is the slowest rate of convergence among that of $\widehat{D}_{X_n,f}$ towards $D_{M,f}$ and that of f_n towards f. Recall that f_n is not required to satisfy the Lipschitz and lower bounded assumptions for the polygonal metric $\widehat{D}_{X_n,f_n}$ to be defined and for Lemma 3.6 to hold. It may also be defined only over X_n if the resolution is set to $q = 2$.

4 Estimation from random samples

In this section we transpose Theorem 3.5 to a probabilistic setup where the point cloud X is replaced with a random point cloud X_n consisting of n i.i.d. samples from a probability measure μ with support M. The following assumptions on μ are needed to ensure that X_n covers M efficiently as n grows, allowing to deduce explicit convergence rates for the estimator $\widehat{D}_{X_n,f}$ from Theorem 3.5.

▶ **Definition 4.1.** *Consider a probability measure μ with support $M \subset \mathbb{R}^N$ and $d \ge 2$.*

- *The measure μ is d-standard with lower constant $c_\mu > 0$ if for all $x \in M$ and $r > 0$,*

  $$\mu\big(B(x,r)\big) \ge c_\mu r^d \wedge 1$$

  where $B(x,r)$ denotes the open ball centered at x with radius r.
- *The measure μ is d-Ahlfors with lower and upper constants $c_\mu > 0$ and $C_\mu > 0$ if for all $x \in M$ and $r > 0$,*

  $$c_\mu r^d \wedge 1 \le \mu\big(B(x,r)\big) \le C_\mu r^d .$$

If μ is d-standard with lower constant c_μ, we also denote $L_\mu = c_\mu^{-1/d}$ which is related to the size of the support. The definition ensures that $\mu\big(\mathrm{B}(x, L_\mu)\big) = 1$ for all $x \in M$, hence $L_\mu \ge \mathrm{diam}(M)$. Moreover, d acts as the intrinsic dimension of the support M of μ. For instance, any measure with a density ρ w.r.t. the volume measure of a submanifold of $\mathbb{R}^N$ of dimension d is d-Ahlfors if ρ is bounded above and below. The d-standard assumption ensures that the random point cloud X_n converges to M in Hausdorff distance. Ahlfors regularity is a stronger assumption and is needed to show the equivalence between ball graphs and NN graphs in Proposition 4.4.

4.1 Convergence of the ball graph estimator

We first discuss the case of a ball graph estimator. It is known that if μ is d-standard, then $d_{\mathrm{H}}(M, X_n)$ is of order $L_\mu(\log(n)/n)^{1/d}$ at most, see [17, Appendix A.2]. This convergence combined with Theorem 3.5 allows to derive convergence rates for the estimation of $D_{M,f}$.

▶ **Theorem 4.2.** *Assume that X_n is the result of n i.i.d. samples from a d-standard probability measure μ with support M. Consider the estimator $\widehat{D}_{X_n,f}$ from Definition 3.4 using the r-ball graph with resolution q where r and q are specified below. Then there exists a constant n_0 depending on L_μ, $\mathcal{T}_{M,f}$ and d such that for all $n \geq n_0$ the following holds.*

- *If $r = 8\sqrt{L_\mu \mathcal{T}_{M,f}} \left(\frac{\log(n)}{n}\right)^{\frac{1}{2d}}$ and $q \geq 1 + 4\frac{\mathcal{T}_{M,f}}{r}$, then*

$$\mathbb{E}\left[\ell_{\infty,M}\left(\widehat{D}_{X_n,f}, D_{M,f}\right)\right] \leq 16\frac{L_\mu}{\mathcal{T}_{M,f}}\left(\frac{\log(n)}{n}\right)^{\frac{1}{d}} . \tag{15}$$

- *If $r = 8L_\mu^{2/3}\mathcal{T}_{M,f}^{1/3}\left(\frac{\log(n)}{n}\right)^{\frac{2}{3d}}$ and $q = 2$, then*

$$\mathbb{E}\left[\ell_{\infty,M}\left(\widehat{D}_{X_n,f}, D_{M,f}\right)\right] \leq 8\left(\frac{L_\mu}{\mathcal{T}_{M,f}}\right)^{\frac{2}{3}}\left(\frac{\log(n)}{n}\right)^{\frac{2}{3d}} . \tag{16}$$

The first case in Theorem 4.2 includes $q = \infty$. There is no use setting q to be greater than the indicated threshold of order $\mathcal{T}_{M,f}/r$, as the term in the upper bound depending on q becomes negligible past this threshold. Note that the optimal choice for r and q requires the knowledge of L_μ, $\mathcal{T}_{M,f}$ and most importantly d. Theorem 4.2 is stated with these choices of parameters to highlight the optimal theoretical dependency in these constants. For practical purposes, the knowledge of L_μ and $\mathcal{T}_{M,f}$ is not actually needed as, in general, any choice of $r \asymp (\log(n)/n)^{1/2d}$ and $q \gtrsim 1/r$ yields a convergence rate of $(\log(n)/n)^{1/d}$. Likewise, any choice of $r \asymp (\log(n)/n)^{2/3d}$ and $q = 2$ yields a convergence rate of $(\log(n)/n)^{2/3d}$. It is not possible however to get rid of the dependency in d for the range parameter r without altering the convergence rate. To circumvent this issue, one may use the NN graph instead, for which the optimal choice of the parameter is shown in Section 4.2 to be $k = \sqrt{n}$ which does not depend on d. Theorem 4.2 is derived directly from Theorem 3.5 and the convergence of X_n to M in Hausdorff distance. The precise computations are deferred to [17, Appendix C.1]. In the case of the induced metric, the weight function $w_{f,q}$ becomes the Euclidean distance regardless of the resolution and $\mathcal{T}_{M,f}$ is replaced with τ_M.

▶ **Corollary 4.3.** *Under the same assumptions as in Theorem 4.2 and when $f = 1$, there exists a constant n_0 depending on L_μ, τ_M and d such that for all $n \geq n_0$, setting $r = 8\sqrt{L_\mu \tau_M}(\log(n)/n)^{1/2d}$, the estimator $\widehat{D}_{X_n,1}$ satisfies*

$$\mathbb{E}\left[\ell_{\infty,M}\left(\widehat{D}_{X_n,1}, D_M\right)\right] \leq 16\frac{L_\mu}{\tau_M}\left(\frac{\log(n)}{n}\right)^{\frac{1}{d}} .$$

Recall that in the case of a submanifold of class $\mathcal{C}^k$ with $k \geq 2$ and dimension d, the optimal convergence rate for the estimation of the induced metric is $n^{-k/d}$ [1]. Corollary 4.3 extends the upper bound to any set of positive reach with the convergence rate $n^{-1/d}$.

4.2 Convergence of the nearest-neighbors graph estimator

We now treat the case of a NN graph estimator, by observing that for adequate choice of r and k, the r-ball graph and k-NN graph are similar. Indeed, assuming that μ is d-Ahlfors regular, a ball of radius r is expected to contain between $c_\mu nr^d$ and $C_\mu nr^d$ points from X_n.

▶ **Proposition 4.4.** *Let μ be a d-Ahlfors measure as defined in Definition 4.1 and X_n a point cloud sampled i.i.d. from μ. Let $k \geq 1$, $\varepsilon \in (0,1)$ and define*

$$r_- = \left((1-\varepsilon)\frac{k}{C_\mu(n-1)}\right)^{\frac{1}{d}} \quad \text{and} \quad r_+ = \left(\frac{1}{1-\varepsilon}\frac{k}{c_\mu(n-1)}\right)^{\frac{1}{d}} .$$

Then with probability at least $1 - 2ne^{-\varepsilon^2 k/2}$*, the* k*-NN graph* $\mathcal{G}_k(X_n)$ *is enclosed between two ball graphs* $G_{r_-}(X_n)$ *and* $G_{r_+}(X_n)$*, that is*

$$\mathbb{P}\left(G_{r_-}(X_n) \subset \mathcal{G}_k(X_n) \subset G_{r_+}(X_n)\right) \geq 1 - 2n \exp\left(-\frac{\varepsilon^2}{2}k\right)$$

where the inclusion $\subset$ *refers to the inclusion of edge sets.*

Proposition 4.4 shows that if r and k are chosen such that $k \asymp nr^d$, the k-NN graph and the r-ball graph are very similar with high probability. This result stems from a common intuition and a detailed proof is provided in [17, Appendix A.3] for completeness. Theorem 4.2 is then extended to k-NN graphs by choosing $k \asymp \sqrt{n \log(n)}$, which does not depend on d.

▶ **Theorem 4.5.** *Assume that* X_n *is the result of* n *i.i.d. samples from a* d*-Ahlfors probability measure* μ *with support* M*. Consider the estimator* $\widehat{D}_{X_n,f}$ *from Definition 3.4 using the* k*-NN graph with parameters* $k = \lceil \sqrt{n \log(n)}\, \rceil$ *and* $q = \lceil n^{1/4} \rceil$*. Then there exists a constant* n_0 *depending on* L_μ*,* $\mathcal{T}_{M,f}$ *and* d *such that for all* $n \geq n_0$ *the following holds.*

$$\mathbb{E}\left[\ell_{\infty,X_n}\left(\widehat{D}_{X_n,f}, D_{M,f}\right)\right] \leq C \left(\frac{\log(n)}{n}\right)^{\frac{1}{d}}$$

where C *is a constant depending on* L_μ *and* $\mathcal{T}_{M,f}$*.*

Notice that in Theorem 4.5 the loss is over X_n instead of M like in Theorem 4.2. The same result could be stated over M, although the proof would be more tedious and is therefore omitted here. Recall that no knowledge on either d, c_μ, C_μ, τ_M, κ or $f_{\min}$ is necessary to compute the estimator, so that it can be used in practice. Setting the resolution q to be $(n/\log(n))^{1/2d}$ would be sufficient to achieve the same upper bound, however this choice requires the knowledge of d. On the other hand, when setting the resolution to $q = 2$ the optimal choice of k can be shown similarly to be $k = n^{1/3} \log(n)^{2/3}$ and to yield a convergence rate of order $(\log(n)/n)^{2/3d}$. Theorem 4.5 is a direct consequence of Theorem 4.2 and Proposition 4.4 and the precise computations are provided in [17, Appendix C.2]. Regarding the case of the induced metric, the same statement as in Corollary 4.3 holds for NN graphs. Namely, the estimator $\widehat{D}_{X_n,1}$ defined over the k-NN graph with $k = \lceil \sqrt{n \log(n)}\, \rceil$ converges to D_M at rate $(\log(n)/n)^{1/d}$.

Let us now discuss the algorithmic complexity of the k-NN estimator with resolution q. In general, building the k-NN graph over a point cloud X of size n is done in $\mathcal{O}(n^2 N)$ time when the ambient dimension N is large. Now, consider two points $x, y \in \mathbb{R}^N$ and assume that f can be evaluated at any point with cost c_f. Computing the edges of the graph $\mathcal{G}_k(X \cup \{x, y\})$ takes $\mathcal{O}(nkqc_f)$ time as the amount of edges in the graph is $\mathcal{O}(nk)$ and the weight of each edge uses q evaluations of f. Finally, computing the infimum that defines $D_{X,f}(x, y)$ in Definition 3.4 using Dijkstra's algorithm takes $\mathcal{O}(n \log(n) + nk)$ time. Overall, if $k = \lceil \sqrt{n \log(n)}\, \rceil$ and $q = \lceil n^{1/4} \rceil$, the time complexity is $\mathcal{O}(n^2 N + n^{7/4} \log(n)^{1/2} c_f)$.

4.3 Minimax lower bound

We now study the worst case performance of any estimator of the induced metric D_M.

▶ **Theorem 4.6.** *Denote* $\mathfrak{D}_n$ *the set of all estimators* $\widehat{D}$ *of the induced metric based on* n *samples, that given any point cloud* X *of* n *points in* $\mathbb{R}^N$ *provides a function* $\widehat{D}_X : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}_+$*. Let* $2 \leq d \leq N$*,* $L, \tau > 0$ *and* $\mathcal{M}(d, L, \tau)$ *be the set of all* d*-standard measures* μ *with* $L_\mu \leq L$ *and support* $M_\mu \subset \mathbb{R}^N$ *that has positive reach lower bounded by* τ*. Then there exists two constants* $C > 0$ *and* n_0 *depending on* d*,* L *and* τ *such that for all* $n \geq n_0$*,*

$$\inf_{\hat{D} \in \mathfrak{D}_n} \sup_{\mu \in \mathcal{M}(d,L,\tau)} \mathbb{E}_{X \sim \mu^{\otimes n}}\left[\ell_{\infty,M_\mu}\left(\widehat{D}_X, D_{M_\mu}\right)\right] \geq C \left(\frac{1}{n}\right)^{\frac{1}{d-1/2}} . \tag{17}$$

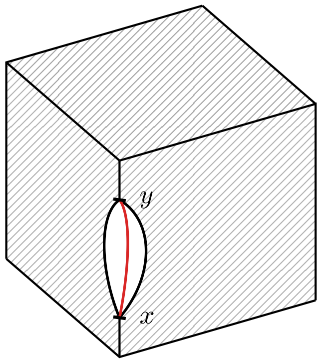

Figure 2 Carving an edge of the cube in $\mathbb{R}^3$.

Notice that Theorem 4.6 only addresses the case of the induced metric. This is not a loss of generality and in fact highlights the fact that the conformal change does not make the problem any harder, as we have established in Section 2 that geodesics have the same regularity as in the case of the induced metric. Given any other function f satisfying the assumptions for a conformal metric, the same lower bound may be obtained with a similar reasoning as what follows, albeit with more technicalities. The rate $n^{-1/(d-1/2)}$ is faster than the rate $n^{1/d}$ obtained in Theorems 4.2 and 4.5, although the difference becomes negligible when d is large. The nature of the minimax convergence rate remains therefore open. Under stronger assumptions on the domain, this question is already solved – see [1] – which we discuss in Section 4.4.

Theorem 4.6 is based on Le Cam's method [19], for which a statement adapted to our setup is given in [17, Lemma C.1]. The method consists in finding two measures μ_1 and μ_2 in $\mathcal{M}(d, L, \tau)$ that are at most $1/n$ apart in total variation distance and such that the relative difference between D_{M_1} and D_{M_2} is of order at least $n^{-1/(d-1/2)}$.

Consider μ_1 the uniform probability on the cube $M_1 = [-\alpha L, \alpha L]^d \times \{0\}^{N-d} \subset \mathbb{R}^N$ where $\alpha > 0$. Let $0 < \varepsilon \leq \alpha L \wedge \tau$ and consider M_2 the result of removing from M_1 its intersection with a ball of radius τ centered at $(0, t, t, \dots, t)$ for some $t > \alpha L$ such that the ball intersects the edge from $(-\alpha L, \alpha L, \alpha L, \dots, \alpha L)$ to $(\alpha L, \alpha L, \alpha L, \dots, \alpha L)$ at two points x and y that are 2ε apart, as pictured in Figure 2. Denote μ_2 the uniform probability over M_2, which has reach exactly τ due to the carved area. By choosing α small enough, which depends only on d, μ_1 and μ_2 are d-standard with lower constant L^{-d}. Then μ_1 and μ_2 both belong to $\mathcal{M}(d, L, \tau)$. The volume of the carved area is of order $\varepsilon(\varepsilon^2)^{d-1}$ as it spans a length 2ε between x and y and a length of order ε^2 for every other dimension of the cube. This implies that the total variation distance between μ_1 and μ_2 is of order ε^{2d-1}. Moreover, The distance from x to y goes from 2ε in M_1 to $2\tau \arcsin(\varepsilon/\tau)$ in M_2, following the red arc of radius τ in Figure 2. This implies that the relative difference between both distances is of order ε^2. Choosing $\varepsilon \asymp n^{-1/2d-1}$ so that the total variation distance is at most $1/n$ eventually yields the desired bound. This reasoning is detailed in [17, Appendix C.3] and the reason why it fails to achieve a bound of order $n^{-1/d}$ is discussed more in depth in the full version [17].

4.4 Case of a smooth manifold

Recall that the minimax convergence rate for the induced metric of a $\mathcal{C}^k$ manifold of dimension d is $n^{-k/d}$ [1]. The methods that achieve such rates are however not computationally feasible in practice as they rely on manifold reconstruction via non-discrete sets. For instance, the optimal convergence rate in the $\mathcal{C}^2$ case was achieved by [3] using the tangential Delaunay complex. Precisely, these methods build a $n^{-k/d}$-approximation of the manifold w.r.t. the Hausdorff distance, then state that the induced metric over this reconstruction is an estimation of the original metric with an error proportional to the Hausdorff distance.

In order to get a concrete estimator, one may use such manifold reconstruction, then sample a fine $n^{-k/d}$-net over it which approximates the original domain with the same Hausdorff error. Our work then implies that the polygonal metric over this net would be an estimation of the original metric with minimax optimal error of order $n^{-k/d}$. Such construction would however be very costly as the point cloud size would grow from n to n^k.

5 Practical examples of conformal factors

We discuss two examples of conformal factors associated with a measure from the literature.

5.1 Density as a conformal factor

If μ is a measure with density ρ w.r.t. the volume measure over a submanifold M, the conformal change via $f = \rho^{-\beta}$ for some parameter $\beta > 0$ is sometimes referred to as the *Fermat distance* due to the parallel with the Fermat principle in optics – that may however be observed with any conformal change. This metric has been applied to various topological data analysis and learning problems, e.g., in [11, 12], as it tends to accentuate features and disparities in the data. Moreover, the metric can be estimated in a simple fashion by using the same overall reasoning as ours but using weights of the form $w(x, y) = \|x - y\|^\alpha$ where $\alpha > 1$ is a parameter depending on β and d [13, 14]. However, this kind of estimation does not feature any known convergence rate. Using our estimator instead provides an alternative method with quantitative guarantees, granted that an estimator of the density is available. To this extent, density estimation is a well-studied problem with many propositions in the literature. For instance, [4] provides a kernel density estimator ρ_n that converges in L^p norm towards ρ at a rate of $n^{-1/d+1}$ under our assumptions and provided that M is a $\mathcal{C}^1$ submanifold and that ρ is also $\mathcal{C}^1$. Under these assumptions, setting $f_n = \rho_n^{-\beta}$ yields a convergence rate of $n^{-1/d}$ according to Lemma 3.6, at the cost of a more complex computation than the usual discrete Fermat distance studied in [13].

5.2 Distance-to-measure as a conformal factor

In general, given a parameter $m \in (0, 1)$, the *distance-to-measure* $d_{\mu,m} : \mathbb{R}^N \to \mathbb{R}_+$ introduced in [8] is defined for any measure μ over $\mathbb{R}^N$. It is 1-Lipschitz and lower bounded by a positive value as long as μ has no atom, hence satisfies the assumption for our work. A slightly different setup where paths are allowed to leave the domain under a specific constraint is studied in [18], where it is argued that this metric should behave similarly to the conformal metric associated with the density but with more stability w.r.t. the measure. The distance-to-measure is shown [9] to be estimated from n i.i.d. samples of the underlying measure with convergence rate $n^{-1/2}$, which is faster than $n^{-1/d}$ hence does not impact the convergence speed of the estimator of the conformal metric according to Lemma 3.6.

6 Conclusion

In this work, we have shown that under a reach assumption on the domain and Lipschitz lower bounded assumptions on the conformal change, the conformal metric has the same regularity as the induced metric in the sense of geodesics having positive reach. We have also shown that positive reach is a sufficient assumption to ensure metric estimation from a polygonal metric with error proportional to the Hausdorff distance between the point cloud and the original domain. This leads to a convergence rate of order $n^{-1/d}$ for the estimation of the conformal metric using n i.i.d. samples from a d-standard measure, which in particular applies to the induced metric of any set with positive reach.

References

1. Eddie Aamari, Clément Berenfeld, and Clément Levrard. Optimal reach estimation and metric learning. *The Annals of Statistics*, 51(3):1086–1108, 2023-06. doi:10.1214/23-AOS2281.
2. Catherine Aaron and Olivier Bodart. Convergence rates for estimators of geodesic distances and Frèchet expectations. *Journal of Applied Probability*, 55(4):1001–1013, 2018-12. doi:10.1017/jpr.2018.66.
3. Ery Arias-Castro, Adel Javanmard, and Bruno Pelletier. Perturbation Bounds for Procrustes, Classical Scaling, and Trilateration, with Applications to Manifold Learning. *Journal of Machine Learning Research*, 21(1):498–534, 2020. URL: http://jmlr.org/papers/v21/18-720.html.
4. Clément Berenfeld and Marc Hoffmann. Density estimation on an unknown submanifold. *Electronic Journal of Statistics*, 15(1):2179–2223, 2021-01. doi:10.1214/21-EJS1826.
5. Mira Bernstein, Vin Silva, John Langford, and Joshua Tenenbaum. Graph Approximations to Geodesics on Embedded Manifolds, 2000-12-20. URL: https://api.semanticscholar.org/CorpusID:10751942.
6. Jean-Daniel Boissonnat, André Lieutier, and Mathijs Wintraecken. The reach, metric distortion, geodesic convexity and the variation of tangent spaces. *Journal of Applied and Computational Topology*, 3(1):29–58, 2019-06-01. doi:10.1007/s41468-019-00029-8.
7. Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A Course in Metric Geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, 2001-06-12. doi:10.1090/gsm/033.
8. Frédéric Chazal, David Cohen-Steiner, and Quentin Mérigot. Geometric Inference for Measures based on Distance Functions. *Foundations of Computational Mathematics*, 11(6):733, 2011. doi:10.1007/s10208-011-9098-0.
9. Frédéric Chazal, Pascal Massart, and Bertrand Michel. Rates of convergence for robust geometric inference. *Electronic Journal of Statistics*, 10(2):2243–2286, 2016-01. doi:10.1214/16-EJS1161.
10. Herbert Federer. Curvature measures. *Transactions of the American Mathematical Society*, 93(3):418–491, 1959. doi:10.1090/S0002-9947-1959-0110078-1.
11. Ximena Fernández, Eugenio Borghini, Gabriel Mindlin, and Pablo Groisman. Intrinsic persistent homology via density-based metric learning. *Journal of Machine Learning Research*, 24(1):75:3341–75:3382, 2023-01-01. URL: https://jmlr.org/papers/v24/21-1044.html.
12. Nicolas Garcia Trillos, Anna Little, Daniel McKenzie, and James M. Murphy. Fermat Distances: Metric Approximation, Spectral Convergence, and Clustering Algorithms. *Journal of Machine Learning Research*, 25(1):176:8331–176:8395, 2024-01-01. URL: http://jmlr.org/papers/v25/23-0939.html.
13. Pablo Groisman, Matthieu Jonckheere, and Facundo Sapienza. Nonhomogeneous Euclidean first-passage percolation and distance learning. *Bernoulli*, 28(1):255–276, 2022-02. doi:10.3150/21-BEJ1341.
14. Sung Jin Hwang, Steven B. Damelin, and Alfred O. Hero Iii. Shortest path through random points. *The Annals of Applied Probability*, 26(5):2791–2823, 2016-10. doi:10.1214/15-AAP1162.
15. André Lieutier and Mathijs Wintraecken. Manifolds of positive reach, differentiability, tangent variation, and attaining the reach, 2024-12-06. doi:10.48550/arXiv.2412.04906.

16 E. J. McShane. Extension of range of functions. *Bulletin of the American Mathematical Society*, 40(12):837–842, 1934. doi:10.1090/S0002-9904-1934-05978-0.

17 Jérôme Taupin. Estimation of Conformal Metrics.

18 Jérôme Taupin and Frédéric Chazal. Fermat Distance-to-Measure: A robust Fermat-like metric, 2025-04-03. arXiv:2504.02381.

19 Bin Yu. Assouad, Fano, and Le Cam. *Research Papers in Probability and Statistics*, 1997. doi:10.1007/978-1-4612-1880-7_29.