arXiv:1508.00674v1 [math.DG] 4 Aug 2015

RECONSTRUCTION AND INTERPOLATION OF MANIFOLDS I: THE GEOMETRIC WHITNEY PROBLEM

CHARLES FEFFERMAN, SERGEI IVANOV, YAROSLAV KURYLEV, MATTI LASSAS, HARIHARAN NARAYANAN

Abstract. We study the geometric Whitney problem on how a Riemannian manifold (M, g) can be constructed to approximate a metric space (X, d_X). This problem is closely related to manifold interpolation (or manifold learning) where a smooth n-dimensional surface $S \subset \mathbb{R}^m$, $m > n$ needs to be constructed to approximate a point cloud in $\mathbb{R}^m$. These questions are encountered in differential geometry, machine learning, and in many inverse problems encountered in applications. The determination of a Riemannian manifold includes the construction of its topology, differentiable structure, and metric.

We give constructive solutions to the above problems. Moreover, we characterize the metric spaces that can be approximated, by Riemannian manifolds with bounded geometry: We give sufficient conditions to ensure that a metric space can be approximated, in the Gromov-Hausdorff or quasi-isometric sense, by a Riemannian manifold of a fixed dimension and with bounded diameter, sectional curvature, and injectivity radius. Also, we show that similar conditions, with modified values of parameters, are necessary.

Moreover, we characterise the subsets of Euclidean spaces that can be approximated in the Hausdorff metric by submanifolds of a fixed dimension and with bounded principal curvatures and normal injectivity radius.

The above interpolation problems are also studied for unbounded metric sets and manifolds. The results for Riemannian manifolds are based on a generalisation of the Whitney embedding construction where approximative coordinate charts are embedded in $\mathbb{R}^m$ and interpolated to a smooth surface. We also give algorithms that solve the problems for finite data.

Keywords: Whitney's extension problem, Riemannian manifolds, machine learning, inverse problems.

Contents

1. Introduction and the main results 2
1.1. Geometrization of Whitney's extension problem 2
1.2. Manifold reconstruction and inverse problems 6
1.3. Interpolation of manifolds in Hilbert spaces 9
1.4. Surface interpolation and Machine Learning 11
2. Approximation of metric spaces 14
2.1. Gromov-Hausdorff approximations 14
2.2. Almost intrinsic metrics 14
2.3. Verifying GH closeness to the disc 16
2.4. Learning the subspaces that approximate the data locally 18
3. Proof of Theorem 2 20
4. Proof of Theorem 1 28
4.1. Approximate charts 28
4.2. Approximate Whitney embedding 30
4.3. The manifold M 35
4.4. Riemannian metric and quasi-isometry 39

Date: August 4, 2015.

5. Algorithms and proof of Corollary 1.8 42
6. Appendix: Curvature and injectivity radius 44
References 47

1. Introduction and the main results

1.1. **Geometrization of Whitney's extension problem.** In this paper we develop a geometric version of Whitney's extension problem. Let $f : K \to \mathbb{R}$ be a function defined on a given (arbitrary) set $K \subset \mathbb{R}^n$, and let $m \geq 1$ be a given integer. The classical Whitney problem is the question whether f extends to a function $F \in C^m(\mathbb{R}^n)$ and if such an F exists, what is the optimal C^m norm of the extension. Furthermore, one is interested in the questions if the derivatives of F, up to order m, at a given point can be estimated, or if one can construct extension F so that it depends linearly on f.

These questions go back to the work of H. Whitney [76, 77, 78] in 1934. In the decades since Whitney's seminal work, fundamental progress was made by G. Glaeser [42], Y. Brudnyi and P. Shvartsman [13, 14, 15, 16, 17, 18] and [69, 70, 71], and E. Bierstone-P. Milman-W. Pawluski [7]. (See also N. Zobin [82, 83] for the solution of a closely related problem.)

The above questions have been answered in the last few years, thanks to work of E. Bierstone, Y. Brudnyi, C. Fefferman, P. Milman, W. Pawluski, P. Shvartsman and others, (see [7, 12, 13, 15, 16, 18, 28, 29, 30, 31, 32, 33].) Along the way, the analogous problems with $C^m(\mathbb{R}^n)$ replaced by $C^{m,\omega}(\mathbb{R}^n)$, the space of functions whose m^{th} derivatives have a given modulus of continuity ω, (see [32, 33]), were also solved.

The solution of Whitney's problems has led to a new algorithm for interpolation of data, due to C. Fefferman and B. Klartag [35, 36], where the authors show how to compute efficiently an interpolant $F(x)$, whose C^m norm lies within a factor C of least possible, where C is a constant depending only on m and n.

In recent years, the focus of attention in this problem has moved to the direction when the measurements $\widetilde{f} : K \to \mathbb{R}$ on the function f are given with errors bounded by $\varepsilon > 0$. Then, the task is to find a function $F : \mathbb{R}^n \to \mathbb{R}$ such that $\sup_{x\in K} |F(x) - \widetilde{f}(x)| \leq \varepsilon$. Since the solution is not unique, one wants to find the extension that has the optimal norm in $C^m(\mathbb{R}^n)$, see e.g. [35, 36]. Finding F can be considered as the task of finding a graph $\Gamma(F) = \{(x, F(x)) :\ x \in \mathbb{R}^n\} \subset \mathbb{R}^{n+1}$ of a function in $C^m(\mathbb{R}^n)$ that approximates the points $\{(x, \widetilde{f}(x)) :\ x \in K\}$. To formulate the above problems in geometric (i.e. coordinates invariant) terms, instead of a graph set $\Gamma(F)$, we aim to construct a general surface or a Riemannian manifold that approximates the given data. Also, instead of the $C^m(\mathbb{R}^n)$-norms, we will measure the optimality of the solution in terms of invariant bounds for the curvature and the injectivity radius.

In this paper we consider the following two geometric Whitney problems:

A. Let E be a separable Hilbert space, e.g. $\mathbb{R}^N$, and assume that we are given a set $X \subset E$. When can one construct a smooth n-dimensional surface $M \subset E$ that approximates X with given bounds for the geometry of M and the Hausdorff distance between M and X? How can the surface M can be efficiently constructed when X is given?
B. Let (X, d_X) be a metric space. When there exists a Riemannian manifold (M, g) that has given bounds for geometry and approximates well X? How can the manifold (M, g) be constructed when X is given?

In Question B, by 'approximation' we mean Gromov-Hausdorff or quasi-isometric approximation, see definitions in Def. 1.3 and Section 2.1.

We answer the Question A in Theorem 2 below, by showing that if $X \subset E$ is locally (i.e., at a certain small scale) close to affine n-dimensional planes, see Def. 1.9, there is a surface $M \subset E$ such that the Hausdorff distance of X and M is small and the second fundamental form and the normal injectivity radius of M are bounded.

The answer to the Question B is given in Theorem 1 below. Roughly speaking, it asserts that the following natural conditions on X are necessary and sufficient: locally, X should be close to $\mathbb{R}^n$, and globally, the metric of X should be almost intrinsic.

The conditions in Theorem 1 are optimal, up to multiplying the obtained bounds by a constant factor depending on n. Theorem 1 gives sufficient conditions for metric spaces that approximate smooth manifolds. In Corollary 1.4 we show that similar conditions, with modified values of parameters, are necessary.

The result of Theorem 2 is optimal, up to multiplication the obtained bounds by a constant factor depending on n.

The proofs of Theorems 1 and 2 are constructive and give raise to algorithms when X is a finite set. Moreover, we give algorithms that verify if a finite data set X satisfies the characterisations given in Theorems 1 and 2.

Next we formulate the definitions needed to state the results rigorously.

Notation. For a metric space X and sets $A, B \subset X$, we denote by $d_H^X(A,B)$, or just by $d_H(A,B)$, the Hausdorff distance between A and B in X.

By $d_{GH}(X,Y)$ we denote the Gromov-Hausdorff (GH) distance between metric spaces X and Y. For the reader's convenience, we collect definitions and elementary facts about the GH distance in section 2.1. For more detailed account of the topic, see e.g. [20, 61, 68]. In most cases we work with *pointed* GH distance between pointed metric spaces (X, x_0) and (Y, y_0), where $x_0 \in X$ and $y_0 \in Y$ are distinguished points. For the definition of pointed GH distance, see [61, §1.2 in Ch. 10]) or section 2.1.

For a metric space X, $x \in X$ and $r > 0$, we denote by $B_r^X(x)$ or $B_r(x)$ the ball of radius r centered at x. For $X = \mathbb{R}^n$, we use notation $B_r^n(x) = B_r^{\mathbb{R}^n}(x)$ and $B_r^n = B_r^n(0)$. For a set $A \subset X$ and $r > 0$, we denote by $U_r^X(A)$ or $U_r(A)$ the metric neighborhood of A of radius r, that is the set points within distance r from A.

When speaking about GH distance between metric balls $B_r^X(x)$ and $B_r^Y(y)$, we always mean the pointed GH distance where the centers x and y are distinguished points of the balls. We abuse notation and write $d_{GH}(B_r^X(x), B_r^Y(y))$ to denote this pointed GH distance.

For a Riemannian manifold M, we denote by Sec_M its sectional curvature and by inj_M its injectivity radius.

Small metric balls in a Riemannian manifold are GH close to Euclidean balls. More precisely, let M be a Riemannian n-manifold with $|\mathrm{Sec}_M| < K$ and $\mathrm{inj}_M > 2\rho_0$ where K and ρ_0 are positive constants, and $0 < r \le \min\{K^{-1/2}, \rho_0\}$. Then the metric ball $B_r^M(x)$ in M and the Euclidean ball $B_r^n = B_r^{\mathbb{R}^n}(0)$ satisfy

$$d_{GH}(B_r^M(x), B_r^n) \le Kr^3. \tag{1.1}$$

For a proof of this estimate, see section 6.

If M is a submanifold of $\mathbb{R}^N$, one can write a similar estimate for the Hausdorff distance in $\mathbb{R}^N$. Namely if the principal curvatures of M are bounded by $\kappa > 0$, then M deviates from its tangent space by at most $\frac{1}{2}\kappa r^2$ within a ball of radius r. Thus the Hausdorff distance between r-ball $B_r^M(x)$ in M and the ball $B_r^{T_xM}(x) =$

$B_r^N(x) \cap T_xM$ of the affine tangent space of M at x satisfy

$$d_H(B_r^M(x), B_r^{T_xM}(x)) \le \tfrac{1}{2}\kappa r^2. \tag{1.2}$$

Note the different order of the above estimates for the intrinsic distances (1.1) and the extrinsic distances (1.2).

With (1.1) in mind, we give the following definition.

Definition 1.1. Let X be a metric space, $r > \delta > 0$, $n \in \mathbb{N}$. We say that X is *δ-close to $\mathbb{R}^n$ at scale r* if, for any $x \in X$,

$$d_{GH}(B_r^X(x), B_r^n) < \delta. \tag{1.3}$$

Condition (1.3) can be effectively verified, up to a constant factor, see *Algorithm GHDist* below. The condition can be also formulated for finite subsets: If sequences $(y_j)_{j=1}^N \subset B_r^n$ and $(x_j)_{j=1}^N \subset B_r^X(x)$ are $\frac{\delta}{4}$-nets such that $|d_{\mathbb{R}^n}(y_j, y_k) - d_X(x_j, x_k)| < \frac{\delta}{4}$ for all $j, k = 1, 2, \dots, N$, then (1.3) is valid by [17, Prop. 7.3.16 and Cor. 7.3.28]. On the other hand, if X is $\frac{\delta}{16}$-close to $\mathbb{R}^n$ at scale r, then such $\frac{\delta}{4}$-nets exists.

In a Riemannian manifold, large-scale distances are determined by small-scale ones through the lengths of paths. However Definition 1.1 does not impose any restrictions on distances larger that $2r$ in X. To rectify this, we need to make the metric 'almost intrinsic' as explained below.

Definition 1.2. Let $X = (X, d)$ be a metric space and $\delta > 0$. A *δ-chain* in X is a finite sequence $x_1, x_2, \dots, x_N \in X$ such that $d(x_i, x_{i+1}) < \delta$ for all $1 \le i \le N-1$. A sequence $x_1, x_2, \dots, x_N \in X$ is said to be *δ-straight* if

$$d(x_i, x_j) + d(x_j, x_k) < d(x_i, x_k) + \delta \tag{1.4}$$

for all $1 \le i < j < k \le N$. We say that X is *δ-intrinsic* if for every pair of points $x, y \in X$ there is a δ-straight δ-chain $x_1, \dots, x_N$ with $x_1 = x$ and $x_N = y$.

Clearly every Riemannian manifold (more generally, every length space) is δ-intrinsic for any $\delta > 0$. Moreover, if X lies within GH distance δ from a length space, then X is $C\delta$-intrinsic. In fact, this property characterizes δ-intrinsic metrics, see Lemma 2.2.

If a metric space $X = (X, d)$ is δ-close to $\mathbb{R}^n$ at scale $r > \delta$ (see Definition 1.1), then one can change 'large' distances in X so that the resulting metric is $C\delta$-intrinsic and coincides with d within balls of radius r. The new distances are measured along 'discrete shortest paths' in X. For details, see (2.2) and Lemma 2.3 in section 2.2.

In order to conveniently compare metric spaces at both small scale and large scale, we need the notion of quasi-isometry.

Definition 1.3. Let X, Y be metric spaces, $\varepsilon > 0$ and $\lambda \ge 1$. A (not necessarily continuous) map $f\colon X \to Y$ is said to be a *(λ, ε)-quasi-isometry* if the image $f(X)$ is an ε-net in Y and

$$\lambda^{-1} d_X(x, y) - \varepsilon < d_Y(f(x), f(y)) < \lambda d_X(x, y) + \varepsilon \tag{1.5}$$

for all $x, x' \in X$, where d_X and d_Y denote the distances in X and Y, resp.

Unlike the use of quasi-isometries in e.g. geometric group theory, in this paper we consider quasi-isometries with parameters $\varepsilon \approx 0$ and $\lambda \approx 1$. The quasi-isometry relation is almost symmetric: if there is a (λ, ε)-quasi-isometry from X to Y, then there exists a $(\lambda, C\lambda\varepsilon)$-quasi-isometry from Y to X, where C is a universal constant. We say that metric spaces X and Y are *(λ, ε)-quasi-isometric* if there is a (λ, ε)-quasi-isometry in either direction.

The existence of (λ, ε)-quasi-isometry $f\colon X \to Y$ implies that

$$d_{GH}(X,Y) < 2(\lambda - 1)\,\mathrm{diam}(X) + 2\varepsilon. \tag{1.6}$$

If X and Y are intrinsic (or ε-intrinsic), a similar estimate holds for metric balls:

$$d_{GH}(B_R^X(x), B_R^Y(f(x))) < C(\lambda - 1)R + C\varepsilon \tag{1.7}$$

for every $x \in X$ and $R > 0$. See section 2.1 for the proof.

Now we formulate our main result.

Theorem 1. *For every $n \in \mathbb{N}$ there exist $\sigma_1 = \sigma_1(n) > 0$ and $C = C(n) > 0$ such that the following holds. Let X be a metric space, $r > 0$ and*

$$0 < \delta < \sigma_1 r. \tag{1.8}$$

Suppose that X is δ-intrinsic and δ-close to $\mathbb{R}^n$ at scale r, see Definitions 1.1 and 1.2. Then there exists a complete n-dimensional Riemannian manifold M such that

1. *X and M are $(1 + C\delta r^{-1}, C\delta)$-quasi-isometric and therefore*

$$d_{GH}(X,M) < C\delta r^{-1}\,\mathrm{diam}(X). \tag{1.9}$$

2. *The sectional curvature Sec_M of M satisfies $|\mathrm{Sec}_M| \le C\delta r^{-3}$.*
3. *The injectivity radius of M is bounded below by $r/2$.*

The estimate (1.9) follows from the existence of a $(1+C\delta r^{-1}, C\delta)$-quasi-isometry from X to M due to (1.6) and the fact that $\mathrm{diam}(X) > r$. The proof of Theorem 1 is given in Section 4.

The quasi-isometry parameters and sectional curvature bound in Theorem 1 are optimal up to constant factors depending only on n, see Remark 4.20.

Furthermore, Theorem 1 gives a characterisation result for metric spaces that GH approximate smooth manifolds with certain geometric bounds. The precise formulation is the following.

Let $\mathcal{M}(n, K, i_0, D)$ denote the class of n-dimensional compact Riemannian manifolds M satisfying $|\mathrm{Sec}_M| \le K$, $\mathrm{inj}_M \ge i_0$, and $\mathrm{diam}(M) \le D$. Denote by $\mathcal{M}_\varepsilon(n, K, i_0, D)$ the class of metric spaces X such that $d_{GH}(X,M) < \varepsilon$ for some $M \in \mathcal{M}(n, K, i_0, D)$. Also, let $\mathcal{X}(n, \delta, r, D)$ denote the class of metric spaces X that are δ-intrinsic and δ-close to $\mathbb{R}^n$ at scale r, and satisfy $\mathrm{diam}(X) \le D$. Theorem 1 has the following corollary that concerns neighbourhoods of smooth manifolds and the class of metric spaces that satisfy a weak δ-flatness condition in the scale of injectivity radius and a strong δ-flatness condition in a small scale r.

Corollary 1.4. *For every $n \in \mathbb{N}$ there exist $\sigma_2 = \sigma_2(n) > 0$ and $C = C(n) > 0$ such that the following holds. Let $K, i_0, D > 0$ and assume that $i_0 < \sqrt{\sigma_2/K}$. Let $\delta_0 = K i_0^3$, $0 < \delta < \delta_0$, and $r = (\delta/K)^{\frac{1}{3}}$. Let $\mathcal{X}$ be the class of metric spaces defined by*

$$\mathcal{X} := \mathcal{X}(n, \delta, r, D) \cap \mathcal{X}(n, \delta_0, i_0, D).$$

Then

$$\mathcal{M}_{\varepsilon_1}(n, K/2, 2i_0, D - \delta) \subset \mathcal{X} \subset \mathcal{M}_{\varepsilon_2}(n, CK, i_0/4, D) \tag{1.10}$$

where $\varepsilon_1 = \delta/6$ and $\varepsilon_2 = CDK^{1/3}\delta^{2/3}$.

The optimal values of ε_1 and ε_2 in Corollary 1.4 remains an open question. The proof of Corollary 1.4 is given at the end of section 4. It is based on Theorem 1 and Proposition 1.7 below.

In Corollary 1.4, the first inclusion in (1.10) means that $X \in \mathcal{X}$ is a necessary condition that a metric space X approximates a smooth manifold $M \in \mathcal{M}(n, K/2, 2i_0, D - \delta)$ with accuracy ε_1. Likewise, the second inclusion in (1.10) implies that $X \in \mathcal{X}$ is a sufficient condition that a metric space X approximates a smooth manifold $M \in M(n, CK, i_0/4, D)$ with accuracy ε_2.

We note that an algorithm based on Theorem 1, that also summarises some of the main objects used in the proof of the theorem, is given in Section 5, see also Fig. 4.2.

In the proof of Theorem 1, M is constructed as a submanifold of a separable Hilbert space E, which is either $\mathbb{R}^N$ with a large N (in case when X is bounded) or ℓ^2 endowed with the the standard $\|\cdot\|_{\ell^2}$ norm. However the Riemannian metric on M is different from the one inherited from E.

Here is the idea of the proof of Theorem 1. Since the r-balls in X are GH close to the Euclidean ball B^n_r, they admit nice maps (2δ-isometries) to B^n_r. These maps can be used as a kind of coordinate charts for X, allowing us to argue about X as if it were a manifold. In particular, we can mimic the proof of Whitney Embedding Theorem (on classical Whitney embeddings, see [79, 80]). If X were a manifold, this would give us a diffeomorphic submanifold of a higher-dimensional Euclidean space E. In our case we get a set $\Sigma \subset E$ which is a Hausdorff approximation of a submanifold $M \subset E$. In order to prove this, we use Theorem 2 (see subsection 1.3 below) which characterizes sets approximable by (nice) submanifolds. We emphasize that the resulting submanifold $M \subset E$ is the image of a Whitney embedding but not of a Nash isometric embedding [54, 55]. As the last step of the construction (see section 4.4), we construct a Riemannian metric g on M so that a natural map from X to (M, g) is almost isometric at scale r. The construction is explicit and can be performed in an algorithmic manner, see section 5. Then, with the assumption that X is δ-intrinsic, it is not hard to show that X and (M, g) are quasi-isometric with small quasi-isometry constants.

Convention. Here and later we fix the notation n for the dimension of a (sub)manifold in question. Throughout the paper we denote by c, C, C_1, etc., various constants depending only on n and, when dealing with derivative estimates, on the order of the derivative involved. To indicate dependence on other parameters, we use notation like $C(M, k)$ or $C_{M,k}$ for numbers depending on manifold M and number k. The same letter C can be used to denote different constants, even within one formula.

1.2. **Manifold reconstruction and inverse problems.** Theorem 1 and Corollary 1.4 give quantitative estimates on how one can use discrete metric spaces as models of Riemannian manifolds, for example for the purposes of numerical analysis. With this approach, a data set representing a Riemannian manifold is just a matrix of distances between points of some δ-net. Naturally, the distances can be measured with some error. In fact, only 'small scale' distances need to be known, see Corollary 1.8 below.

The statement of Theorem 1 provides a verifiable criterion to tell whether a given data set approximates any Riemannian manifold (with certain bounds for curvature and injectivity radius). See section 2.3 for an explicit algorithm.

The proof of Theorem 1 is constructive. It provides an algorithm, although a rather complicated one, to *construct* a Riemannian manifold approximated by a given discrete metric space X. See section 5 for an outline of the algorithm.

Next we formulate results that describe properties of the manifold M constructed from data X that approximates some smooth manifold $\widetilde{M}$ and discuss how this result is used in inverse problems.

1.2.1. *Reconstructions with data that approximate a smooth manifold.* When dealing with inverse problems, it is assumed that the data set X comes from some unknown Riemannian manifold $\widetilde{M}$, and moreover some *a priori* bounds on the geometry of this manifold are given. Applying Theorem 1 to this data set yields another manifold M which is $(1+C\delta r^{-1}, C\delta)$-quasi-isometric to $\widetilde{M}$. One naturally asks what information about the original manifold $\widetilde{M}$ can be recovered. An answer is given by the following proposition.

Proposition 1.5 (cf. Theorem 8.19 in [43])**.** *There exist $\sigma_0 = \sigma_0(n) > 0$ and $C = C(n) > 0$ such that the following holds. Let M and $\widetilde{M}$ be complete Riemannian n-manifolds with $|\operatorname{Sec}_M| \le K$ and $|\operatorname{Sec}_{\widetilde{M}}| \le K$, where $K > 0$.*

Let $0 < \sigma < \sigma_0$ and assume that M and $\widetilde{M}$ are $(1+\sigma, \sigma r)$-quasi-isometric, where $r < \min\{(\sigma/K)^{1/2}, \operatorname{inj}_M, \operatorname{inj}_{\widetilde{M}}\}$.

Then M and $\widetilde{M}$ are diffeomorphic. Moreover there exists a bi-Lipschitz diffeomorphism between M and $\widetilde{M}$ with bi-Lipschitz constant bounded by $1 + C\sigma$.

We do not prove Proposition 1.5 because it is essentially the same as Theorem 8.19 in [43] except that the approximation is quasi-isometric rather than GH. To prove Proposition 1.5 one can apply the same arguments as in [43, 8.19] using coordinate neighborhoods of size r. The estimates are not given explicitly in [43] but they follow from the argument. These results can be regarded as quantitative versions of Cheeger's Finiteness Theorem [23], see [61, Ch. 10] and [60] for different proofs.

Remark 1.6. Using results of [1] one can show that M and $\widetilde{M}$ in Proposition 1.5 are close to each other in $C^{1,\alpha}$ topology. However we do not know explicit estimates in this case.

1.2.2. *An improved estimate for the injectivity radius.* The injectivity radius estimate provided by Theorem 1 is not good enough in the context of manifold reconstruction. Indeed, in order to obtain a good approximation one has to begin with a small r. (Recall that for Theorem 1 to work, δ should be of order Kr^3 where K is the curvature bound.) However Theorem 1 guarantees only a lower bound of order r for inj_M, so a priori one could end up with an approximating manifold M with a very small injectivity radius. In order to rectify this we need the following result.

Proposition 1.7. *There exists $C = C(n) > 0$ such that the following holds. Let $K > 0$ and let $M, \widetilde{M}$ be complete n-dimensional Riemannian manifolds with $|\operatorname{Sec}_M| \le K$ and $|\operatorname{Sec}_{\widetilde{M}}| \le K$.*

1. Let $x \in M$, $\widetilde{x} \in \widetilde{M}$, and $0 < \rho \le \min\{\operatorname{inj}_{\widetilde{M}}(\widetilde{x}), \frac{\pi}{\sqrt{K}}\}$. Then

$$\operatorname{inj}_M(x) \ge \rho - C \cdot d_{GH}(B_\rho^M(x), B_\rho^{\widetilde{M}}(\widetilde{x})). \tag{1.11}$$

2. Suppose that M and $\widetilde{M}$ are $(1 + \delta r^{-1}, \delta)$-quasi-isometric where $\delta > 0$ and

$$0 < r \le \min\{\operatorname{inj}_{\widetilde{M}}(\widetilde{x}), \tfrac{\pi}{\sqrt{K}}\}. \tag{1.12}$$

Then

$$\operatorname{inj}_M \ge (1 - C\delta r^{-1}) \min\{\operatorname{inj}_{\widetilde{M}}, \tfrac{\pi}{\sqrt{K}}\}. \tag{1.13}$$

The situation described in the second part of Proposition 1.7 occurs when M and $\widetilde{M}$ are two manifolds approximating the same metric space X as in Theorem 1. or when M is a reconstruction of $\widetilde{M}$ as in Corollary 1.8 below. The proof of Proposition 1.7 is given in section 6.

1.2.3. *An approximation result with only one parameter.* We summarize the manifold reconstruction features of Theorem 1 in the following corollary where all approximations, errors in data, as well as the errors in the reconstruction are given in terms of a single parameter $\widehat{\delta}$. Essentially, the corollary tells that a manifold N can be approximately reconstructed from a $\widehat{\delta}$-net X of N and the information about local distances between points of X containing small errors. This type of results are useful e.g. in inverse problems discussed below.

Corollary 1.8. *Let $K > 0$, $n \in \mathbb{Z}_+$ and N be a compact n-dimensional manifold with sectional curvature bounded by $|\operatorname{Sec}_N| \le K$. There exists $\delta_0 = \delta_0(n, K)$ such that if $0 < \widehat{\delta} < \delta_0$ then the following holds:*

Let $r = (\widehat{\delta}/K)^{1/3}$ and suppose that the injectivity radius inj_N of N satisfies $\operatorname{inj}_N > 2r$. Also, let $X = \{x_j : j = 1, 2, \dots, J\} \subset N$ be a $\widehat{\delta}$-net of N and $\widetilde{d}\colon X \times X \to \mathbb{R}_+ \cup \{0\}$ be a function that satisfies for all $x, y \in X$

$$|\widetilde{d}(x, y) - d_N(x, y)| \le \widehat{\delta}, \quad \text{if } d_N(x, y) < r, \tag{1.14}$$

and

$$\widetilde{d}(x, y) > r - \widehat{\delta}, \quad \text{if } d_N(x, y) \ge r.$$

Given the set X and the function $\widetilde{d}$, one can effectively construct a compact n-dimensional Riemannian manifold (M, g) such that:

(1) *There is a diffeomorphism $F : M \to N$ satisfying*

$$\frac{1}{L} \le \frac{d_N(F(x), F(y))}{d_M(x, y)} \le L, \quad \text{for all } x, y \in M, \tag{1.15}$$

where $L = 1 + CK^{1/3}\widehat{\delta}^{\,2/3}$.

(2) *There is $C_1 = C_1(n) > 0$ such that the sectional curvature Sec_M of M satisfies $|\operatorname{Sec}_M| \le C_1 K$.*

(3) *The injectivity radius inj_M of M satisfies*

$$\operatorname{inj}_M \ge \min\{(C_1 K)^{-1/2}, (1 - CK^{1/3}\widehat{\delta}^{\,2/3}) \operatorname{inj}_N\}.$$

The proof of Corollary 1.8 is given in the end of Section 5.

We call the function $\widetilde{d}\colon X \times X \to \mathbb{R}_+ \cup \{0\}$, defined on the $\widehat{\delta}$-net X and satisfying the assumptions of Corollary 1.8, an approximate local distance function with accuracy $\widehat{\delta}$. Many inverse problems can be reduced to a setting where one can determine the distance function $d_N(x_j, x_k)$, with measurement errors $\epsilon_{j,k}$, in a discrete set $\{x_j\}_{j \in J} \subset N$. Thus, if the set $\{x_j\}_{j \in J}$ is $\widehat{\delta}$-net in N, the errors $\epsilon_{j,k}$ satisfy conditions (1.14), and $\widehat{\delta}$ is small enough, then the diffeomorphism type of the manifold can be uniquely determined by Corollary 1.8. Moreover, the bi-Lipschitz condition (1.15) means that also the distance function can be determined with small errors. We emphasize that in (1.14) one needs to approximately know only the distances smaller than $r = (\widehat{\delta}/K)^{1/3}$. The larger distances can be computed as in (2.2).

1.2.4. *Manifold reconstructions in imaging and inverse problems.* Recently, geometric models have became an area of focus of research in inverse problems. As an example of such problems, one may consider an object with a variable speed of wave propagation. The travel time of a wave between two points defines a natural non-Euclidean distance between the points. This is called the travel time metric and it corresponds to the distance function of a Riemannian metric. In many topical inverse problem the task is to determine the Riemannian metric inside an object from external measurements, see e.g. [50, 51, 57, 58, 72, 74]. These problems are the idealizations of practical imaging tasks encountered in medical imaging or in

Earth sciences. Also, the relation of discrete and continuous models for these problems is an active topic of research, see e.g. [6, 9, 10, 49]. In these results discrete models have been reconstructed from various types of measurement data. However, a rigorously analyzed technique to construct a smooth manifold from these discrete models to complete the construction has been missing until now.

In practice the measurement data contains always measurement errors and is limited. This is why the problem of the approximate reconstruction of a Riemannian manifold and the metric on it from discrete or noisy data is essential for several geometric inverse problems. Earlier, various regularization techniques have been developed to solve noisy inverse problems in the PDE-setting, see e.g. [27, 53], but most of such methods depend on the used coordinates and, therefore, are not invariant. One of the purposes of this paper is to provide invariant tools for solving practical imaging problems.

An example of problems with limited data is an inverse problem for the heat kernel, where the information about the unknown manifold (M, g) is given in the form of discrete samples $(h_M(x_j, y_k, t_i))_{j,k\in J,i\in I}$ of the heat kernel $h_M(x, y, t)$, satisfying

$$\begin{aligned}&(\partial_t - \Delta_g)h_M(x, y, t) = 0, \quad \text{on } (x, t) \in M \times \mathbb{R}_+,\\ &h_M(x, y, 0) = \delta_y(x),\end{aligned}$$

where the Laplace operator Δ_g operates in the x variable, see e.g. [48]. Here $y_j = x_j$, where $\{x_j :\ j \in J\}$ is a finite ε-net in an open set $\Omega \subset M$, while $\{t_i :\ i \in I\}$ is in ε-net of the time interval (t_0, t_1). It is also natural to assume that one is given measurements $h^{(m)}_M(x_j, y_k, t_i)$ of the heat kernel with errors satisfying $|h^{(m)}_M(x_j, y_k, t_i) - h_M(x_j, y_k, t_i)| < \varepsilon$. Several inverse problems for wave equation lead to a similar problem for the wave kernel $G_M(x, y, t)$ satisfying

$$\begin{aligned}&(\partial_t^2 - \Delta_g)G_M(x, y, t) = \delta_0(t)\delta_y(x), \quad \text{on } (x, t) \in M \times \mathbb{R},\\ &G_M(x, y, t) = 0, \quad \text{for } t < 0,\end{aligned}$$

see e.g. [45, 48, 56]. In the case of complete data (corresponding to the case when ε vanishes), the inverse problem for heat kernel and wave kernel are equivalent to the inverse interior spectral problem, see [47]. In this problem one considers the eigenvalues λ_k of $-\Delta_g$, counted by their multiplicity, and the corresponding $L^2(M)$-orthonormal eigenfunctions, $\varphi_k(x)$ that satisfy

$$-\Delta_g\varphi_k(x) = \lambda_k\varphi_k(x), \quad x \in M.$$

In the inverse interior spectral problem one assumes that we are given the first N smallest eigenvalues, λ_k, $k = 1, 2, \ldots, N$, and values $\varphi^{(m)}_k(x_j)$ at the ε-net points $\{x_j;\ j \in J\} \subset \Omega$, where $|\varphi^{(m)}_k(x_j) - \varphi_k(x_j)| < \varepsilon$ and $\Omega \subset M$ is open. It is shown in [2, 49] that these data determine a metric space (X, d_X) which is a δ GH-approximation to the unknown manifold M, where $\delta = \delta(\varepsilon, N; \Omega)$ tends to 0 as $\varepsilon \to 0$ and $N \to \infty$. It should be noted that the above works deal with the case of manifolds with boundary and the Laplace operators with some classical boundary conditions, however, the constructions used there can be immediately extended to the case of a compact Riemannian manifold with data in a bounded open subdomain. Theorem 1 completes the solution of the above inverse problems by constructing a smooth manifold that approximates M.

1.3. **Interpolation of manifolds in Hilbert spaces.** As already mentioned, in the proof of Theorem 1 we need to approximate a set in a Hilbert space by an n-dimensional submanifold (with bounded geometry). At small scale, the set in question should be close to affine subspaces in the following sense.

Definition 1.9. Let E be a Hilbert space, $X \subset E$, $n \in \mathbb{N}$ and $r, \delta > 0$. We say that X is *δ-close to n-flats at scale r* if for any $x \in X$, there exists an n-dimensional affine space $A_x \subset E$ through x such that

$$d_H(X \cap B_r^E(y),\, A_x \cap B_r^E(x)) \le \delta. \tag{1.16}$$

To formulate our result for the sets in Hilbert spaces, we recall some definitions. By a *closed submanifold* of a Hilbert space E we mean a finite-dimensional smooth submanifold which is a closed subset of E. One can show that a finite-dimensional submanifold $M \subset E$ is closed if and only if it is properly embedded, that is the inclusion $M \hookrightarrow E$ is a proper map.

Let $M \subset E$ be a closed submanifold. The *normal injectivity radius* of M is the supremum of all $r > 0$ such that normal exponential map of M is diffeomorphic in the r-neighborhood of the null section of the normal bundle of M. Let r be the normal injectivity radius of M and assume that $r > 0$. Then for every $x \in U_r(M)$ there exists a unique nearest point in M. We denote this nearest point by $P_M(x)$ and refer to the map $P_M : U_r(M) \to M$ as the *normal projection*. Note that the normal projection is a smooth map.

Theorem 2. *For every $n \in \mathbb{N}$ there exists $\sigma_0 = \sigma_0(n) > 0$ such that the following holds. Let E be a separable Hilbert space, $X \subset E$, $r > 0$ and*

$$0 < \delta < \sigma_0 r. \tag{1.17}$$

Suppose that X is δ-close to n-flats at scale r (see Def. 1.9). Then there exists a closed n-dimensional smooth submanifold $M \subset E$ such that:

(1) $d_H(X, M) \le 5\delta$.
(2) *The second fundamental form of M at every point is bounded by $C\delta r^{-2}$.*
(3) *The normal injectivity radius of M is at least $r/3$.*
(4) *The normal projection $P_M : U_{r/3}(M) \to M$ is globally C-Lipschitz, i.e.*

$$|P_M(x) - P_M(y)| \le C|x - y| \tag{1.18}$$

for all $x, y \in U_{r/3}(M)$, and satisfies

$$\|d_x^k P_M\| < C_{n,k}\delta r^{-k} \tag{1.19}$$

for all $k \ge 2$ and $x \in U_{r/3}(M)$.

The proof of Theorem 2 is given in Section 3.

We note that an algorithm based on Theorem 2, that summarises also the main objects used in its proof, is given in Section 5, see also Fig. 3.

In Remark 3.14 below we show that the bounds in claims (2) and (3) in Theorem 2 are optimal, up to constant factors depending on n. Thus Theorem 2 gives necessary and sufficient conditions (up to multiplication of the bounds by a constant factor) for a set $X \subset E$ to approximate a smooth submanifold with given geometric bounds.

Notation. In (1.19) and throughout the paper, d_x^m denotes the mth differential of a smooth map. The norm of the mth differential is derived from the Euclidean norm on E in the standard way. We extend this notation to the case $m = 0$ by setting $d_x^0 f = f(x)$ for any map f. As usual, we define the C^m-norm of a map f defined on an open set $U \subset E$, by

$$\|f\|_{C^m(U)} = \sup_{x \in U} \max_{0 \le k \le m} \|d_x^k f\|.$$

In order to approximate a submanifold M as in Theorem 2, the set X must contain as many points as a $C\delta$-net in M. This is an unreasonably large number of points when δ is small. The following corollary allows one to reconstruct M from a smaller approximating set. It involves two parameters ε and δ where ε is a 'density' of a net and δ is a 'measurement error'. Note that δ may be much smaller than ε. A similar generalization is possible for Theorem 1 but we omit these details.

Corollary 1.10. *For every $n \in \mathbb{N}$ there exists $\sigma_0 = \sigma_0(n) > 0$ such that the following holds. Let E be a Hilbert space, $X \subset E$, $0 < \varepsilon < r/10$ and $0 < \delta < \sigma_0 r$. Suppose that for every $x \in X$ there exists an n-dimensional affine subspace $A_x \subset E$ such that the set $X \cap B_r(x)$ is within Hausdorff distance δ from an ε-net of the affine n-ball $A_x \cap B_r(x)$.*

Then there exists a closed n-dimensional submanifold $M \subset E$ satisfying properties 2–4 of Theorem 2 and such that X is within Hausdorff distance $C\delta$ from an ε-net of M.

Proof sketch. Below, the symbol $\angle$ denotes the angle between n-dimensional affine subspaces of E.

Consider the set $X' = \bigcup_{x \in X}(A_x \cap B_r(x)) \subset E$. A suitably modified version of Lemma 3.2 implies that $\angle(A_x, A_y) < C\delta r^{-1}$ for all $x, y \in X$ such that $|x - y| < r$. It then follows that X' is $C\delta$-close to n-flats at scale $r - C\delta$. Now the corollary follows from Theorem 2 applied to X'. □

1.4. **Surface interpolation and Machine Learning.** The results of this paper solve some classical problems in Machine Learning. Next we give a short review on existing methods and discuss how Theorem 2 is applied for problems of Manifold Learning.

1.4.1. *Literature on Manifold Learning.* The following methods aim to transform data lying near a d-dimensional manifold in an N dimensional space into a set of points in a low dimensional space close to a d-dimensional manifold. During transformation all of them try to preserve some geometric properties, such as appropriately measured distances between points of the original data set. Usually the Euclidean distance to the 'nearest' neighbours of a point is preserved. In addition some of the methods preserve, for points farther away, some notion of geodesic distance capturing the curvature of the manifold.

Perhaps the most basic of such methods is 'Principal Component Analysis' (PCA), [59, 44] where one projects the data points onto the span of the d eigenvectors corresponding to the top d eigenvalues of the $(N \times N)$ covariance matrix of the data points.

An important variation is the 'Kernel PCA' [67] where one defines a feature map $\Phi(\cdot)$ mapping the data points into a Hilbert space called the feature space. A 'kernel matrix' K is built whose $(i, j)^{th}$ entry is the dot product $\langle\Phi(x_i), \Phi(x_j)\rangle$ between the data points x_i, x_j. From the top d eigenvectors of this matrix, the corresponding eigenvectors of the covariance matrix of the image of the data points in the feature space can be computed. The data points are projected onto the span of these eigenvectors of this covariance matrix in the feature space.

In the case of 'Multi Dimensional Scaling' (MDS) [38], only pairwise distances between points are attempted to be preserved. One minimizes a certain 'stress function' which captures the total error in pairwise distances between the data points and between their lower dimensional counterparts. For instance, a raw stress function could be $\Sigma(\|x_i - x_j\| - \|y_i - y_j\|)^2$, where x_i are the original data points, y_i, the transformed ones, and $\|x_i - x_j\|$, the distance between x_i, x_j.

'Isomap' [73] attempts to improve on MDS by trying to capture geodesic distances between points while projecting. For each data point a 'neighbourhood graph' is constructed using its k neighbours (k could be varied based on various criteria), the edges carrying the length between points. Now shortest distance between points is computed in the resulting global graph containing all the neighbourhood graphs using a standard graph theoretic algorithm such as Dijkstra's. Let $D = [d_{ij}]$ be the $n \times n$ matrix of graph distances. Let $S = [d_{ij}^2]$ be the $n \times n$ matrix of squared graph distances. Form the matrix $A = \frac{1}{2}HSH$, where $H = I - n^{-1}\mathbf{1}\mathbf{1}^T$. The matrix A is of rank $t < n$, where t is the dimension of the manifold. Let $A^Y = \frac{1}{2}HS^YH$, where $[S^Y]_{ij} = \|y_i - y_j\|^2$. Here the y_i are arbitrary t-dimensional vectors. The embedding vectors $\widehat{y}_i$ are chosen to minimize $\|A - A^Y\|$. The optimal solution is given by the eigenvectors $v_1, \ldots, v_t$ corresponding to the t largest eigenvalues of A. The vertices of the graph G are embedded by the $t \times n$ matrix

$$\widehat{Y} = (\widehat{y}_1, \ldots, \widehat{y}_n) = (\sqrt{\lambda_1}v_1, \ldots, \sqrt{\lambda_t}v_t)^T.$$

'Maximum Variance Unfolding' (MVU) [75] also constructs the neighbourhood graph as in the case of Isomap but tries to maximize distance between projected points keeping distance between nearest points unchanged after projection. It uses semidefinite programming for this purpose.

In 'Diffusion Maps' [37], a complete graph on the data points is built and each edge is assigned a weight based on a gaussian: $w_{ij} \equiv e^{\frac{\|x_i - x_j\|^2}{\sigma^2}}$. Normalization is performed on this matrix so that the entries in each row add up to 1. This matrix is then used as the transition matrix P of a Markov chain. P^t is therefore the transition probability between data points in t steps. The d nontrivial eigenvalues λ_i and their eigenvectors v_i of P^t are computed and the data is now represented by the matrix $[\lambda_1 v_1, \cdots, \lambda_d v_d]$, with the row i corresponding to data point x_i.

The following are essentially local methods of manifold learning in the sense that they attempt to preserve local properties of the manifold around a data point.

'Local Linear Embedding' (LLE) [64] preserves solely local properties of the data. Let N_i be the neighborhood of x_i, consisting of k points. Find optimal weights $\widehat{w}_{ij}$ by solving $\widehat{W} := \arg\min_W \sum_{i=1}^n \|x_i - \sum_{j=1}^n w_{ij}x_j\|^2$, subject to the constraints (i) $\forall i, \sum_j w_{ij} = 1$, (ii) $\forall i, j, w_{ij} \geq 0$, (iii) $w_{ij} = 0$ if $j \notin N_i$. Once the weight matrix $\widehat{W}$ is found a spectral embedding is constructed using it. More precisely, a matrix $\widehat{Y}$ is is a $t \times n$ matrix constructed satisfying $\widehat{Y} = \arg\min_Y Tr(YMY^T)$, under the constraints $Y\mathbf{1} = 0$ and $YY^T = nI_t$, where $M = (I_n - \widehat{W})^T(I_n - \widehat{W})$. $\widehat{Y}$ is used to get a t-dimensional embedding of the initial data.

In the case of the 'Laplacian Eigenmap' [3], [46] again, a nearest neighbor graph is formed. The details are as follows. Let n_i denote the neighborhood of i. Let $W = (w_{ij})$ be a symmetric $(n \times n)$ weighted adjacency matrix defined by (i) $w_{ij} = 0$ if j does not belong to the neighborhood of i; (ii) $w_{ij} = \exp(\|x_i - x_j\|^2/2\sigma^2)$, if x_j belongs to the neighborhood of x_i. Here σ is a scale parameter. Let G be the corresponding weighted graph. Let $D = (d_{ij})$ be a diagonal matrix whose i^{th} entry is given by $(W\mathbf{1})_i$. The matrix $L = D - W$ is called the Laplacian of G. We seek a solution in the set of $t \times n$ matrices $\widehat{Y} = \arg\min_{Y:YDY^T=I_t} \mathrm{Tr}(YLY^T)$. The rows of $\widehat{Y}$ are given by solutions of the equation $Lv = \lambda Dv$.

Hessian LLE (HLLE) (also called Hessian Eigenmaps) [40] and 'Local Tangent Space Alignment' (LTSA) [81] attempt to improve on LLE by also taking into consideration the curvature of the higher dimensional manifold while preserving the local pairwise distances. We describe LTSA below.

LTSA attempts to compute coordinates of the low dimensional data points and align the tangent spaces in the resulting embedding. It starts with computing bases for the approximate tangent spaces at the datapoints x_i by applying PCA on the neighboring data points. The coordinates of the low dimensional data points are computed by carrying out a further minimization $min_{Y_i,L_i} \Sigma_i \|Y_i J_k - L_i \Theta_i\|^2$. Here Y_i has as its columns, the lower dimensional vectors, J_k is a 'centering' matrix,

Θ_i has as its columns the projections of the k neighbors onto the d eigenvectors obtained from the PCA and L_i maps these coordinates to those of the lower dimensional representation of the data points. The minimization is again carried out through suitable spectral methods.

The alignment of local coordinate mappings also underlies some other methods such as 'Local Linear Coordinates' (LLC) [65] and 'Manifold Charting' [11].

There are also methods which map higher dimensional data points to lower dimensional piecewise linear manifolds (as opposed to smooth manifolds). Under this restriction these methods produce optimal manifolds. The manifold is a simplicial complex in the case of Cheng et al [25] and a witness complex in the case of Boissonnat et al [8].

Each of the algorithms is based on strong domain based intuition and in general performs well in practice atleast for the domain for which it was originally intended. PCA is still competitive as a general method.

Some of the algorithms are known to perform correctly under the hypothesis that data lie on a manifold of a specific kind. In Isomap and LLE, the manifold has to be an isometric embedding of a convex subset of Euclidean space. In the limit as number of data points tends to infinity, when the data approximate a manifold, then one can recover the geometry of this manifold by computing an approximation of the Laplace-Beltrami operator. Laplacian Eigenmaps and Diffusion maps rest on this idea. LTSA works for parameterized manifolds and detailed error analysis is available for it.

1.4.2. *Theorems 1 and 2 and the problems of machine learning.* The Theorem 1 addresses the fundamental question, when a given metric space (X, d_X), corresponding to data points and their 'abstract' mutual distances, approximate a Riemannian manifold with a bounded sectional curvature and injectivity radius. In the context of Theorem 1, the distances are measured in intrinsic sense in M and X.

Theorem 2 deals with approximating a subset of a Hilbert space E satisfying certain local constraints by a manifold having a bounded second fundamental form and normal injectivity radius. In the context of Theorem 2, the distances are measured in extrinsic sense in E. Such approximations have extensively been considered in machine learning or, more precisely, manifold learning and non-linear dimensionality reduction, where the goal is to approximate the set of data lying in a high-dimensional space like E by a submanifold in E of a low enough dimension in order to visualize these data, see e.g. references of Section 1.4.1.

The results of this paper provide for the observed data an abstract low-dimensional representation of the intrinsic manifold structure that the data may possess. In particular, the topology of the manifold structure is determined, assuming that the sampling density has been sufficient. As described in Section 3, the proof of Theorem 2 is of a constructive nature and provides an algorithm to perform such visualisation. Note that this algorithm starts with tangent-type planes which makes it distantly similar to the LTSA method in machine learning, see e.g. [52, 81]. In paper [34], the authors provide a method of visualization of a given data using a probabilistic setting. In comparison, Theorem 2 helps us visualize data in a deterministic setting.

Another application of the results of the paper to machine learning deals with the spectral techniques to perform dimensionality reduction, see e.g. [3, 4, 5]. Using the constructions of Section 4.2, we can associate with the data set not only the metric structure but also point measures and use the constructions of [22] to find approximately the eigenvalues and eigenfunctions of the sought for manifold.

The results of this paper also have implications for a probabilistic model of the data. Thus both Theorem 1, which involves an abstract manifold, and Theorem

2 which involves an embedded manifold have implications for Machine Learning. We would be assuming that data is drawn using independent and identically distributed (i.i.d.) samples from a probability distribution supported on a manifold with random, e.g. gaussian, noise. As the amount of data increases, with high probability, the Hausdorff distance of the set of corrupted samples to the manifold first decreases (if the noise is sufficiently small) and then increases. We stop drawing data at the point where the high-probability bound on the Hausdorff distance begins to increase, and fit a manifold to this data.

2. Approximation of metric spaces

In this section we collect preliminaries about GH and quasi-isometric approximation of metric spaces. In subsections 2.3 and 2.4 we present algorithms that can be used to verify the assumptions of Theorems 1 and 2.

2.1. **Gromov-Hausdorff approximations.** Let X be a metric space. Recall that the Hausdorff distance between sets $A, B \subset X$ is defined by

$$d_H(A,B) = \inf\{r > 0 : A \subset U_r(B) \text{ and } B \subset U_r(A)\}$$

where U_r denotes the r-neighborhood of a set.

The Gromov-Hausdorff (GH) distance $d_{GH}(X,Y)$ between metric spaces X and Y is defined as follows: for every $\varepsilon > 0$, one has $d_{GH}(X,Y) < \varepsilon$ if and only if there exists a metric space Z and subsets $X', Y' \subset Z$ isometric to X and Y, resp., such that $d_H(X',Y') < \varepsilon$. One can always assume that Z is the disjoint union of X and Y with a metric extending those of X and Y.

The pointed GH distance between pointed metric spaces (X, x_0) and (Y, y_0) is defined in the same way with an additional requirement that $d_Z(x_0, y_0) < \varepsilon$. See e.g. [61, §1.2 in Ch. 10] for details.

Example 2.1 (Distorted net). Recall that a subset S of a metric space X is called an ε-net if $U_\varepsilon(S) = X$. Let S be an ε-net in X and imagine that we have measured the distances between points of S with an absolute error ε. That is, we have a distance function d' on $S \times S$ such that $|d'(x,y) - d(x,y)| < \varepsilon$ for all $x, y \in S$. Then the GH distance between X and (S, d') is bounded by 2ε. This follows from the fact that the inclusion $S \hookrightarrow X$ is an ε-isometry from (S, d') to (X, d), see below.

Strictly speaking, the 'measurement errors' in this example may break the triangle inequality so that (S, d') is no longer a metric space. This can be fixed by adding 3ε to all d'-distances.

Let X, Y be metric spaces and $\varepsilon > 0$. A (not necessarily continuous) map $f\colon X \to Y$ is called an ε-isometry if $f(X)$ is an ε-net in Y and

$$|d_Y(f(x), f(y)) - d_X(x,y)| < \varepsilon$$

for all $x, y \in X$. Equivalently, an ε-isometry is a $(1,\varepsilon)$-quasi-isometry (cf. Definition 1.3). If $d_{GH}(X,Y) < \varepsilon$ then there exists a 2ε-isometry from X to Y, and conversely, if there is an ε-isometry from X to Y then $d_{GH}(X,Y) < 2\varepsilon$, see e.g. [20, Corollary 7.3.28]. Throughout the paper we use these facts without explicit reference.

Clearly a (λ, ε)-quasi-isometry from a bounded space X is a $((\lambda - 1)D + \varepsilon)$-isometry where $D = \mathrm{diam}(X)$. This implies (1.6) and (1.7).

2.2. **Almost intrinsic metrics.** Here we discuss properties of δ-intrinsic metrics and related notions from Definition 1.2. First observe that, if $x_1, x_2, \dots, x_N$ is a δ-straight sequence, then its 'length' satisfies

$$\sum_{i=1}^{N-1} d(x_i, x_{i+1}) \le d(x_1, x_N) + (N-2)\delta. \tag{2.1}$$

This follows by induction from (1.4) and the triangle inequality.

The next lemma characterizes almost intrinsic metrics as those that are GH close to Riemannian manifolds. However manifolds provided by this lemma may have extremely large curvatures and tiny injectivity radii.

Lemma 2.2. *Let X be a metric space and $\delta > 0$.*

1. If there exists a length space Y such that $d_{GH}(X, Y) < \delta$, then X is 6δ-intrinsic.

2. Conversely, if X is δ-intrinsic, then there exists a two-dimensional Riemannian manifold M such that $d_{GH}(X, M) < C\delta$, where C is the universal constant.

Proof. 1. By the definition of the GH distance, there exists a metric d on the disjoint union $Z := X \sqcup Y$ such that d extends d_X and d_Y and $d_H(X, Y) < \delta$ in (Z, d). Let $x, x' \in X$. Since $d_H(X, Y) < \delta$, there exist $y, y' \in Y$ such that $d(x, y) < \delta$ and $d(x', y') < \delta$. Connect y to y' by a minimizing geodesic and let $y = y_1, y_2, \dots, y_N = y'$ be a sequence of points along this geodesic such that $d(y_i, y_{i+1}) < \delta$ for all i. For each $i = 2, \dots, N-1$, choose $x_i \in X$ such that $d(x_i, y_i) < \delta$. Then $x, x_2, \dots, x_{N-1}, x'$ is a 6δ-straight 3δ-chain connecting x and x'. Since x and x' are arbitrary points of X, the claim follows.

2. Since we do not use this claim, we do not give a detailed proof of it. Here is a sketch of the construction. First, arguing as in [20, Proposition 7.7.5], one can approximate X by a metric graph. If X is δ-intrinsic, the graph can be made GH $C\delta$-close to X. Consider a piecewise-smooth arcwise isometric embedding of the graph into $\mathbb{R}^3$ and let M be a smoothed boundary of a small neighborhood of the image. Then M is a two-dimensional Riemannian manifold which can be made arbitrarily close to the graph and hence $C\delta$-close to X. □

Now we describe a construction that makes a $C\delta$-intrinsic metric out of a metric which is δ-close to $\mathbb{R}^n$ at scale r (see Definition 1.1). More generally, let $X = (X, d)$ be a metric space in which every ball of radius r is δ-intrinsic, where $r > \delta > 0$. For $x, y \in X$, define the new distance $d'(x, y)$ by

$$d'(x, y) = \inf_{\{x_i\}} \left\{ \sum_{i=1}^{N-1} d(x_i, x_{i+1}) : \ x_1 = x, \ x_N = y \right\} \tag{2.2}$$

where the infimum is taken over all finite sequences $x_1, \dots, x_N$ connecting x to y and such that every pair of subsequent points x_i, x_{i+1} is contained in a ball of radius r in (X, d).

In order to avoid infinite d'-distances, we need to assume that any two points can be connected by such a sequence. If this is not the case, X divides into components separated from one another by distance at least r. For our purposes, such components are unrelated to one another just like connected components of a manifold.

Lemma 2.3. *Under the above assumptions, the function d' given by (2.2) is a $C\delta$-intrinsic metric on X, where C is a universal constant. Furthermore, d and d' coincide within any ball of radius r.*

Proof. The triangle inequality for d implies that d' is a metric, $d' \geq d$, and $d'(x, y) = d(x, y)$ if x and y belong to an r-ball in (X, d). It remains to verify that (X, d') is $C\delta$-intrinsic. Let $x, y \in X$ and let $x = x_1, \dots, x_N = y$ be a sequence almost realizing the infimum in (2.2). Every pair x_i, x_{i+1} belongs to an r-ball in (X, d). Since this ball is δ-intrinsic, there exists a δ-straight δ-chain connecting x_i to x_{i+1}. Joining such chains together yields a δ-chain connecting x to y. Using the triangle inequality, one can easily verify that this chain is 10δ-straight. □

The next lemma shows that if a map is almost isometric at small scale then it is a quasi-isometry with small constants. It is used in the proof of Theorem 1.

Lemma 2.4. *Let $r > 5\delta > 0$. Let X and Y be δ-intrinsic metric spaces and $f\colon X \to Y$ a map such that $f(X)$ is a δ-net in Y and*

$$|d_Y(f(x), f(y)) - d_X(x, y)| < \delta \tag{2.3}$$

for all $x, y \in X$ such that

$$\min\{d_X(x, y), d_Y(f(x), f(y))\} < r.$$

Then f is a $(1 + Cr^{-1}\delta, C\delta)$-quasi-isometry, where C is a universal constant.

Proof. Let $x, x' \in X$ and $D = d_X(x, x')$. We have to verify that

$$(1 + Cr^{-1}\delta)^{-1}D - C\delta < d_Y(f(x), f(x')) < (1 + Cr^{-1}\delta)D + C\delta \tag{2.4}$$

for a suitable C. Fix $x, x' \in X$ and connect them by a δ-straight δ-chain (see Definition 1.2). This chain contains a subsequence $x = x_1, x_2, \dots, x_N = x'$ such that $r - \delta < d_X(x_i, x_{i+1}) < r$ for all $i = 1, \dots, N - 2$ and $d_X(x_{N-1}, x_N) < r$. Since the subsequence is also δ-straight, by (2.1) we have

$$\sum d_X(x_i, x_{i+1}) < D + (N - 2)\delta. \tag{2.5}$$

Since $d_X(x_i, x_{i+1}) > r - \delta$ for each $i \le N - 2$, the left-hand side is bounded below by $(N - 2)(r - \delta)$. Hence

$$N \le (r - 2\delta)^{-1}D + 2 < 2r^{-1}D + 2. \tag{2.6}$$

By (2.3) we have $d_Y(f(x_i), f(x_{i+1})) < d_X(x_i, x_{i+1}) + \delta$ for all i. Therefore

$$\sum d_Y(f(x_i), f(x_{i+1})) < \sum d_X(x_i, x_{i+1}) + (N - 1)\delta < D + (2N - 3)\delta.$$

by (2.5). Thus

$$d_Y(f(x), f(x')) < D + (2N - 3)\delta < D + (4r^{-1}D + 1)\delta = (1 + 4r^{-1}\delta)D + \delta$$

where the second inequality follows from (2.6). This proves the second inequality in (2.4). To prove that first one, apply the same argument to an 'almost inverse' map $g\colon Y \to X$ constructed as follows: for each $y \in Y$, let $g(y)$ be an arbitrary point from the set $f^{-1}(B_\delta(y))$. □

2.3. **Verifying GH closeness to the disc.** Here we present an algorithm that can be used to verify the main assumption of Theorem 1. Namely, given a discrete metric space X, $n \in \mathbb{N}$ and $r > 0$, one can approximately (i.e., up to a factor $C = C(n)$) find the smallest δ such that X is δ-close to $\mathbb{R}^n$ at scale r (see Definition 1.1). Due to rescaling it suffices to handle the case $r = 1$.

Thus the problem boils down to the following: given a point $x_0 \in X$, find approximately the (pointed) GH distance between the metric ball $B_1^X(x_0) \subset X$ of radius 1 centered at x_0 and the Euclidean unit ball $B_1^n \subset \mathbb{R}^n$. The following algorithm solves this problem.

Algorithm GHDist: Assume that we are given n, the point $x_0 \in X$ and the ball $X_0 = B_1^X(x_0) \subset X$. We regard X_0 as a metric space with metric $d = d_X|_{X_0 \times X_0}$. We implement the following steps:

(1) Let $x_1 \in X_0$ be a point that minimizes $|1 - d(x_0, x)|$ over all $x \in X_0$.
(2) Given $x_1, x_2, \dots x_m$ for $m \le n$, we define the coordinate function

$$f_m(x) = \tfrac{1}{2}\big(d(x, x_0)^2 - d(x, x_m)^2 + d(x_0, x_m)^2\big) \tag{2.7}$$

(3) Given $x_1, x_2, \dots x_m$ and coordinate functions $f_1(x), f_2(x), \dots, f_m(x)$ for $m \le n - 1$, choose x_{m+1} that is the solution of the minimization problem

$$\min_{x \in X_0} K_m(x), \quad K_m(x) = \max(|1 - d(x_0, x)|, |f_1(x)|, \dots, |f_m(x)|).$$

(4) When $x_1, x_2, \dots x_n$ and coordinate functions $f_1(x), f_2(x), \dots, f_n(x)$ are determined, we compute for $f(x) = (f_m(x))_{m=1}^n$

$$\delta_1 = \sup_{x,x'\in X_0} \Big| d(x',x) - |f(x') - f(x)| \Big|, \quad \delta_2 = \sup_{y\in B_1^n} \inf_{x\in X_0} |f(x) - y|,$$
$$\delta_a = \max(\delta_1, \delta_2).$$

Then the algorithms outputs the value of δ_a and the map f.

Lemma 2.5. *Suppose there exists a δ-isometry $h\colon X_0 \to B_1^n \subset \mathbb{R}^n$ satisfying $h(x_0) = 0$. Then*

(1) *The output value δ_a of the above algorithm satisfies $\delta_a < C\delta$.*
(2) *The map $f\colon X_0 \to \mathbb{R}^n$ obtained in the algorithm is a $C\delta$-isometry from X_0 to the Euclidean ball $B_{1+\delta_a}^n$.*
(3) *Moreover there exists an orthogonal map $U\colon \mathbb{R}^n \to \mathbb{R}^n$ such that*

$$|f(x) - U(h(x))| < C\delta$$

for all $x \in X_0$.

Here C is a constant depending on n.

Proof. It follows from the definition of δ_a that $f : X_0 \to \mathbb{R}^n$ is a δ_a-isometry from X_0 to $B_{1+\delta}^n$. Thus, the second claim follows from the first and third claims. Let us proceed with their proofs.

As h is a δ-isometry, for any $y_1 \in \partial B_1^n$ there is $x_1' \in X_0$ such that $|h(x_1') - y_1| < \delta$ and hence $|h(x_1')| > 1 - \delta$. Using again the fact that h is a δ-isometry, we see that $d(x_1', x_0) > 1 - 2\delta$. Hence, the point x_1 chosen in the algorithm satisfies $d(x_1, x_0) > 1 - 2\delta$ and $|h(x_1)| > 1 - 3\delta$.

Assume now that we have constructed, using the algorithm, the points $x_1, \dots, x_m$, $m < n$, the corresponding $f_i(x), i = 1, \dots, m$, see (2.7), and

$$h_i(x) := \langle h(x), h(x_i)\rangle = \tfrac{1}{2}(|h(x)|^2 - |h(x) - h(x_i)|^2 + |h(x_i)|^2), \quad i = 1, \dots, m,$$

where $\langle\cdot,\cdot\rangle$ is the inner product of $\mathbb{R}^n$. As h is a δ-isometry, we have, for some $C = C(n)$,

$$|h_m(x) - f_m(x)| \le C\delta, \quad x \in X_0. \tag{2.8}$$

Moreover, assume next that for $i, k \in \{1, 2, \dots, m\}$, $i \ne k$, we have

$$|h(x_i)| > 1 - C\delta, \quad |h_i(x_k)| = |\langle h(x_i), h(x_k)\rangle| \le C\delta. \tag{2.9}$$

Then, let $y_{m+1} \in \partial B_1^n \cap \{h(x_1), \dots, h(x_m)\}^\perp$. Then there is $x_{m+1}' \in X_0$ such that $|h(x_{m+1}') - y_{m+1}| < \delta$. This yields that

$$|h(x_{m+1}')| > 1 - \delta. \tag{2.10}$$

Moreover,

$$|h_i(x_{m+1}')| = |\langle h(x_i), h(x_{m+1}')\rangle| \le C\delta, \quad i = 1, \dots, m.$$

Due to (2.8), (2.10), the above inequality implies that $K_{m+1}(x_{m+1}') < C\delta$. This implies that the minimizer x_{m+1} of K_{m+1} also satisfies $K_{m+1}(x_{m+1}) < C\delta$. As h is a δ-isometry, it follows from (2.7) that (2.8) remains valid for $i = m+1$. In turn, these imply that (2.9) is valid also for $i, k \in \{1, 2, \dots, m+1\}$, $i \ne k$. By induction, (2.9) is valid for all $i, k \in \{1, 2, \dots, n\}$, $i \ne k$.

Applying Gram-Schmidt algorithm to vectors $h(x_j)$ and formula (2.9), we see there is an orthonormal basis $(w_j)_{j=1}^n$ of $\mathbb{R}^n$ such that $|w_j - h(x_j)| < C\delta$. Hence, using formula (2.8), we see that $Ay = \sum_{m=1}^n \langle y, h(x_m)\rangle w_m$ is a linear operator

$A:\mathbb{R}^n\to\mathbb{R}^n$ that satisfies $\|A-I\|\le C\delta$. Let $U\colon\mathbb{R}^n\to\mathbb{R}^n$ be the orthogonal linear map which maps $(w_m)_{m=1}^n$ to the standard basis $(e_m)_{m=1}^n$ of $\mathbb{R}^n$. By (2.8),

$$|UA(h(x))-f(x)|=\big|\sum_{m=1}^n(h_m(x)-f_m(x))e_m\big|\le C\delta$$

for all $x\in X_0$. The 3rd claim of the lemma follows.

As h is a δ-isometry, this also proves that f is a $C\delta$-isometry from X_0 to a ball B^n_ρ where $\rho=1+C\delta$. The 1st claim of the lemma follows. □

The above lemma implies that the (pointed) Gromov-Hausdorff distance between X_0 and B^n_1 satisfies

$$C^{-1}\delta_a\le d_{GH}(X_0,B^n_1)\le 2\delta_a.$$

Thus the algorithm *GHDist* gives the Gromov-Hausdorff distance of X_0 and B^n_1 up to a constant factor C depending only on dimension n.

Remark 2.6. As a by-product of Lemma 2.5, we get the following fact: if $h_1,h_2\colon X_0\to B^n_1$ are two δ-isometries, then there exists an orthogonal map $U\colon\mathbb{R}^n\to\mathbb{R}^n$ such that $|h_1(x)-U(h_2(x))|\le C\delta$ for all $x\in X_0$. The map U can be constructed by a modification of the algorithm *GHDist*, see the proof of Lemma 2.5.

2.4. **Learning the subspaces that approximate the data locally.** Let X be a finite set of points in $E=\mathbb{R}^N$ and $X\cap B_1(x):=\{x,\widetilde{x}_1,\dots,\widetilde{x}_s\}$ be a set of points within a Hausdorff distance δ of some (unknown) unit n-dimensional disc $D_1(x)$ centered at x. Here $B_1(x)$ is the set of points in $\mathbb{R}^N$ whose distance from x is less or equal to 1. We give below a simple algorithm that finds a unit n-disc centered at x within a Hausdorff distance $Cn\delta$ of $X_0:=X\cap B_1(x)$, where C is an absolute constant.

The basic idea is to choose a near orthonormal basis from X_0 where x is taken to be the origin and let the span of this basis intersected with $B_1(x)$ be the desired disc.

Algorithm FindDisc:

(1) Let x_1 be a point that minimizes $|1-|x-x'||$ over all $x'\in X_0$.
(2) Given $x_1,\dots x_m$ for $m\le n-1$, choose x_{m+1} such that

$$\max(|1-|x-x'||,|\langle x_1/|x_1|,x'\rangle|,\dots,|\langle x_m/|x_m|,x'\rangle|)$$

is minimized among all $x'\in X_0$ for $x'=x_{m+1}$.

Let $\widetilde{A}_x$ be the affine n-dimensional subspace containing $x,x_1,\dots,x_n$, and the unit n-disc $\widetilde{D}_1(x)$ be $\widetilde{A}_x\cap B_1(x)$. Recall that for two subsets A,B of $\mathbb{R}^N$, $d_H(A,B)$ represents the Hausdorff distance between the sets. The same letter C can be used to denote different constants, even within one formula.

Lemma 2.7. *Suppose there exists an n-dimensional affine subspace A_x containing x such that $D_1(x)=A_x\cap B_1(x)$ satisfies $d_H(X_0,D_1(x))\le\delta$. Suppose $0<\delta<\frac{1}{2n}$. Then $d_H(X_0,\widetilde{D}_1(x))\le Cn\delta$, where C is an absolute constant.*

Proof. Without loss of generality, let x be the origin. Let $d(x,y)$ be used to denote $|x-y|$. We will first show that for all $m\le n-1$,

$$\max\left(|1-d(x,x_{m+1})|,\left|\left\langle\frac{x_1}{|x_1|},x_{m+1}\right\rangle\right|,\dots,\left|\left\langle\frac{x_m}{|x_m|},x_{m+1}\right\rangle\right|\right)<\delta.$$

To this end, we observe that the minimum over $D_1(x)$ of

$$\max\left(|1-d(x,y)|,\left|\left\langle\frac{(x_1)}{|x_1|},y\right\rangle\right|,\dots,\left|\left\langle\frac{(x_m)}{|x_m|},y\right\rangle\right|\right)\tag{2.11}$$

is 0, because the dimension of $D_1(x)$ is n and there are only $m \le n-1$ linear equality constraints. Also, the radius of $D_1(x)$ is 1, so $|1 - d(x, z_{m+1})|$ has a value of 0 where a minimum of (2.11) occurs at $y = z_{m+1}$. Since the Hausdorff distance between $D_1(x)$ and X_0 is less than δ there exists a point $y_{m+1} \in X_0$ whose distance from z_{m+1} is less than δ. For this point y_{m+1}, we have δ greater than

$$\max\left(|1-d(x,y_{m+1})|, \left|\left\langle \frac{(x_1)}{|x_1|}, y_{m+1}\right\rangle\right|, \dots, \left|\left\langle \frac{(x_m)}{|x_m|}, y_{m+1}\right\rangle\right|\right). \tag{2.12}$$

Since

$$\max\left(|1-d(x,x_{m+1})|, \left|\left\langle \frac{(x_1)}{|x_1|}, x_{m+1}\right\rangle\right|, \dots, \left|\left\langle \frac{(x_m)}{|x_m|}, x_{m+1}\right\rangle\right|\right)$$

is no more than the corresponding quantity in (2.12), we see that for each $m+1 \le n$,

$$\max\left(|1-d(x,x_{m+1})|, \left|\left\langle \frac{(x_1)}{|x_1|}, x_{m+1}\right\rangle\right|, \dots, \left|\left\langle \frac{(x_m)}{|x_m|}, x_{m+1}\right\rangle\right|\right) < \delta.$$

Let $\widetilde{V}$ be an $N \times n$ matrix whose i^{th} column is the column x_i. Let the operator 2-norm of a matrix Z be denoted $\|Z\|$. For any distinct i, j we have $|\langle x_i, x_j\rangle| < \delta$, and for any i, $|\langle x_i, x_i\rangle - 1| < 2\delta$, because $0 < 1-\delta < |x_i| < 1$. Therefore,

$$\|\widetilde{V}^t\widetilde{V} - I\| \le C_1 n\delta.$$

Therefore, the singular values of $\widetilde{V}$ lie in the interval

$$I_C = (\exp(-Cn\delta), \exp(Cn\delta)) \supseteq (1 - C_1 n\delta, 1 + C_1 n\delta).$$

For each $i \le n$, let x_i' be the nearest point on $D_1(x)$ to the point x_i. Since the Hausdorff distance of X_0 to $D_1(x)$ is less than δ, this implies that $|x_i' - x_i| < \delta$ for all $i \le n$. Let $\widehat{V}$ be an $N \times n$ matrix whose i^{th} column is x_i'. Since for any distinct i, j $|\langle x_i', x_j'\rangle| < 3\delta + \delta^2$, and for any i, $|\langle x_i', x_i'\rangle - 1| < 4\delta$,

$$\|\widehat{V}^t\widehat{V} - I\| \le Cn\delta.$$

This means that the singular values of $\widehat{V}$ lie in the interval I_C.

We shall now proceed to obtain an upper bound of $Cn\delta$ on the Hausdorff distance between X_0 and $\widetilde{D}_1(x)$. Recall that the unit n-disc $\widetilde{D}_1(x)$ is $\widetilde{A}_x \cap B_1(x)$. By the triangle inequality, since the Hausdorff distance of X_0 to $D_1(x)$ is less than δ, it suffices to show that the Hausdorff distance between $D_1(x)$ and $\widetilde{D}_1(x)$ is less than $Cn\delta$.

Let x' denote a point on $D_1(x)$. We will show that there exists a point $z' \in \widetilde{D}_1(x)$ such that $|x' - z'| < Cn\delta$.

Let $\widehat{V}\alpha = x'$. By the bound on the singular values of $\widehat{V}$, we have $|\alpha| < \exp(Cn\delta)$. Let $y' = \widetilde{V}\alpha$. Then, by the bound on the singular values of $\widetilde{V}$, we have $|y'| \le \exp(Cn\delta)$. Let $z' = y'/|y'|$. By the preceding two lines, z' belongs to $\widetilde{D}_1(x)$. We next obtain an upper bound on $|x' - z'|$

$$|x' - z'| \le |x' - y'| \tag{2.13}$$
$$+|y' - z'|. \tag{2.14}$$

We examine the term in (2.13)

$$|x' - y'| = |\widehat{V}\alpha - \widetilde{V}\alpha| \le \sup_i |x_i - x_i'| (\sum_i |\alpha_i|) \le \delta n \exp(Cn\delta).$$

We next bound the term in (2.14).

$$|y' - z'| \le |y'|(1 - \exp(-Cn\delta)) \le Cn\delta.$$

Together, these calculations show that

$$|x' - z'| < Cn\delta.$$

A similar argument shows that if z'' belongs to $\widetilde{D}_1(x)$ then there is a point $p' \in D_1(x)$ such that $|p' - z''| < Cn\delta$; the details follow. Let $\widehat{V}\beta = z''$. From the bound on the singular values of $\widehat{V}$, $|\beta| < \exp(Cn\delta)$. Let $q' := \widetilde{V}\beta$. Let $p' := \frac{q'}{|q'|}$.

$$\begin{aligned} |p' - z''| &\leq |q' - z''| + |p' - q'| \\ &\leq |\widetilde{V}\beta - V\beta| + |1 - \widetilde{V}\beta| \\ &\leq \sup_i |x_i - x_i'| (\sum_i |\beta_i|) + C\delta n \\ &\leq \delta n \exp(Cn\delta) + C\delta n \\ &\leq C\delta n. \end{aligned}$$

This proves that the Hausdorff distance between X_0 and $\widetilde{D}_1(x)$ is bounded above by $Cn\delta$ where C is a universal constant.

□

3. Proof of Theorem 2

The statement of Theorem 2 is scale invariant: it does not change if one multiplies r and δ by $\lambda > 0$ and applies a λ-homothety to all subsets of E. Hence it suffices to prove the theorem only for $r = 1$. When $r = 1$, the theorem turns into the following proposition (where σ_0 is renamed to δ_0):

Proposition 3.1. *There exists $\delta_0 = \delta_0(n) > 0$ such that the following holds. Let E be a separable Hilbert space, $X \subset E$ and $0 < \delta < \delta_0$. Suppose that for every $x \in X$ there is an n-dimensional affine subspace $A_x \subset E$ through x such that*

$$d_H(X \cap B_1(x), A_x \cap B_1(x)) < \delta. \tag{3.1}$$

Then there is a closed n-dimensional smooth submanifold $M \subset E$ such that

1. *$d_H(X, M) \leq 5\delta$.*
2. *The second fundamental form of M at every point is bounded by $C\delta$.*
3. *The normal injectivity radius of M is at least $1/3$.*
4. *The normal projection $P_M : U_{1/3}(M) \to M$ satisfies*

$$|P_M(x) - P_M(y)| \leq C|x - y| \tag{3.2}$$

and

$$\|d_x^k P_M\| < C_{n,k}\delta \tag{3.3}$$

for all $x, y \in U_{1/3}(M)$ and all $k \geq 2$.

The proof of Proposition 3.1 occupies the rest of this section. Let X and $\{A_x\}_{x\in X}$ be as in the proposition. Let $P_{A_x} : E \to A_x$ be the orthogonal projection to A_x. By $\vec{A}_x$ we denote the linear subspace parallel to A_x. For $x \in X$ and $\rho > 0$, we define $B_\rho^X(x) = X \cap B_\rho(x)$ and $D_\rho(x) = A_x \cap B_\rho(x)$. In this notation, (3.1) takes the form

$$d_H(B_1^X(x), D_1(x)) < \delta, \qquad x \in X. \tag{3.4}$$

In the sequel we assume that δ is sufficiently small so that the inequalities arising throughout the proof are valid. The required bound for δ depends only on n.

Lemma 3.2. *Let $p, q \in X$ be such that $|p - q| < 1$. Then $\operatorname{dist}(q, A_p) < \delta$ and $\angle(A_p, A_q) < 5\delta$.*

Proof. Since $q \in B_1^X(p)$, we have

$$\operatorname{dist}(q, A_p) \le \operatorname{dist}(q, D_1(p)) \le d_H(B_1^X(p), D_1(p)) < \delta$$

by (3.4). It remains to prove the second claim of the lemma.

Let $z = P_{A_p}(\frac{p+q}{2})$. Then $|z-p| < \frac{1}{2}$ and $|z-q| < \frac{1}{2} + \delta$ by the triangle inequality. Define $B = A_p \cap B_{1/2-2\delta}(z)$. We claim that $\operatorname{dist}(y, A_q) < 2\delta$ for every $y \in B$. Indeed, let $y \in B$. Then $|y-q| < 1-\delta$ and $|y-p| < 1-2\delta$. The latter implies that $y \in D_1(p)$, hence by (3.4) there exists $x \in X$ such that $|x-y| < \delta$. By the triangle inequality we have $x \in B_1^X(q)$, hence (3.4) implies that $\operatorname{dist}(x, A_q) < \delta$. Therefore $\operatorname{dist}(y, A_q) \le |y-x| + \operatorname{dist}(x, A_q) < 2\delta$ as claimed.

Define a function $h\colon \vec{A}_p \to \mathbb{R}_+$ by $h(v) = \operatorname{dist}(z+v, A_q)^2$. As shown above, $h(v) \le 4\delta^2$ for all $v \in \vec{A}_p$ such that $|v| \le \frac{1}{2} - 2\delta$. The function h is polynomial of degree 2, i.e., $h(v) = Q(v) + L(v) + h_0$ where Q is a (nonnegative) quadratic form, L is a linear function and $h_0 = h(0)$. Furthermore,

$$Q(v) = \sin^2 \angle(v, \vec{A}_q) \cdot |v|^2$$

for all $v \in \vec{A}_p$. Let $\alpha = \angle(A_p, A_q)$, and let $v_0 \in \vec{A}_p$ be such that $\angle(v_0, \vec{A}_q) = \alpha$ and $|v_0| = \frac{1}{2} - 2\delta$. Then

$$Q(v_0) = \frac{h(v_0) + h(-v_0)}{2} - h(0) \le 4\delta^2$$

since $h(\pm v_0) \le 4\delta^2$ and $h(0) \ge 0$. Thus $\sin^2(\alpha) \cdot |v_0|^2 \le 4\delta^2$, or, equivalently,

$$\sin\alpha \le 2\delta(\tfrac{1}{2} - 2\delta)^{-1} = 4\delta(1-4\delta)^{-1}.$$

If δ is sufficiently small, this implies the desired inequality $\alpha < 5\delta$. □

Let X_0 be a maximal (with respect to inclusion) $\frac{1}{100}$-separated subset of X. Note that X_0 is a $\frac{1}{100}$-net in X and X_0 is at most countable. Let $X_0 = \{q_i\}_{i=1}^{|X_0|}$. For brevity, we introduce notation $A_i = A_{q_i}$ and $P_i = P_{A_{q_i}}$.

Throughout the argument below we assume that $|X_0| = \infty$, i.e. X_0 is a countably infinite set. In the case when X_0 is finite the proof is the same except that ranges of some indices should be restricted.

Assuming that $\delta < \frac{1}{300}$, there is a number $N = N(n)$ such that every set of the form $X_0 \cap B_1(q_i)$ contains at most N points. This follows from the fact that this set is $\frac{1}{100}$-separated and contained in the δ-neighborhood of a unit n-dimensional ball $D_1(q_i)$.

Fix a smooth function $\mu\colon \mathbb{R}_+ \to [0,1]$ such that $\mu(t) = 1$ for all $t \in [0, \frac{1}{3}]$ and $\mu(t) = 0$ for all $t \ge \frac{1}{2}$. For each $i \ge 1$ define a function $\mu_i\colon E \to [0,1]$ by

$$\mu_i(x) = \mu(|x - q_i|).$$

Clearly μ_i is smooth and $\|d_x^k \mu_i\|$ is bounded (by a constant depends only on n and k) for every $k \ge 1$. Let $\varphi_i\colon E \to E$ be a map given by

$$\varphi_i(x) = \mu_i(x) P_i(x) + (1 - \mu_i(x))x. \tag{3.5}$$

Now define a map $f_i\colon E \to E$ by

$$f_i = \varphi_i \circ \varphi_{i-1} \circ \ldots \circ \varphi_1 \tag{3.6}$$

for all $i \ge 1$, and let $f_0 = id_E$.

For $x \in E$ and $i \ge 1$ we have $f_i(x) = f_{i-1}(x)$ if $|f_{i-1}(x) - q_i| \ge \frac{1}{2}$. This follows from the relation $f_i = \varphi_i \circ f_{i-1}$ and the fact that φ_i is the identity outside the ball $B_{1/2}(q_i)$.

Let $U = U_{1/4}(X_0) \subset E$. We are going to show that for every $x \in U$ the sequence $\{f_i(x)\}$ stabilizes and hence a map $f = \lim_{i\to\infty} f_i$ is well-defined on U.

Define $B_m = B_{1/4}(q_m)$ for $m = 1, 2, \ldots$. Note that $U = \bigcup_m B_m$.

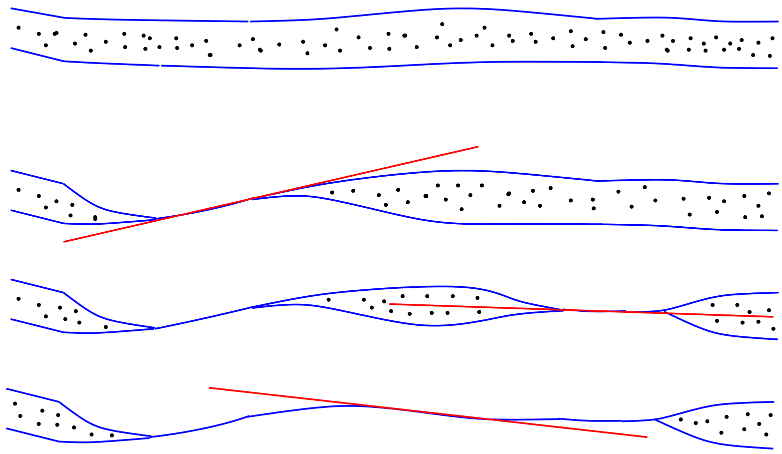

Figure 1. A schematic visualisation of the interpolation algorithm Algorithm 'SurfaceInterpolation' based on Theorem see Section 5. In the figure on top, the black data points $X \subset E = \mathbb{R}^m$ have a δ-neighbourhood $U = U_\delta(X)$. The boundary of U is marked by blue. In the figures below, we determine, near points $x_i \in X$, $i = 1, 2, 3$ the approximating n-dimensional planes A_i, marked by red lines. Then we map the set U by applying to it iteratively functions $\varphi_i : E \to E$, defined in (3.5). The maps φ_i are convex combinations of the projector P_{A_i}, onto A_i, and the identity map. Figures 2,3, and 4 from the top show the sets $\varphi_1(U)$, $\varphi_2(\varphi_1(U))$ and $\varphi_3(\varphi_2(\varphi_1(U)))$, respectively. The limit of these sets converge to the n-dimensional surface $M \subset E$.

Lemma 3.3. *If $x \in B_m$ then $|f_i(x) - q_m| < \frac{1}{3}$ for all i.*

Proof. Suppose the contrary and let

$$i_0 = \min\{i : |f_i(x) - q_m| \geq \tfrac{1}{3}\}.$$

Let $i \leq i_0$ be such that $|q_i - q_m| < 1$ and let $z = f_{i-1}(x)$. Since $i - 1 < i_0$, we have $|z - q_m| < \frac{1}{3}$. Lemma 3.2 applied to $p = q_i$ and $q = q_m$ implies that $|P_i(z) - P_m(z)| < 6\delta$. Since P_m is the orthogonal projection to a subspace containing q_m, we have $|P_m(z) - q_m| \leq |z - q_m|$, therefore

$$|P_i(z) - q_m| \leq |P_m(z) - q_m| + |P_i(z) - P_m(z)| \leq |z - q_m| + 6\delta$$

and hence the point

$$f_i(x) = \varphi_i(z) = \mu_i(z)P_i(z) + (1 - \mu_i(z))z$$

satisfies

$$|f_i(x) - q_m| \leq \mu_i(z)|P_i(z) - q_m| + (1 - \mu_i(z))|z - q_m| \leq |z - q_m| + 6\delta.$$

Thus

$$|f_i(x) - q_m| \leq |f_{i-1}(x) - q_m| + 6\delta \tag{3.7}$$

for all $i \leq i_0$ such that $|q_i - q_m| < 1$. For indices $i \leq i_0$ such that $|q_i - q_m| \geq 1$, we have

$$|f_{i-1}(x) - q_i| \geq 1 - |f_{i-1}(x) - q_m| > 1 - \tfrac{1}{3} > \tfrac{1}{2}$$

and hence $f_i(x) = f_{i-1}(x)$. Since there are at most $N = N(n)$ indices $i \le i_0$ such that $|q_i - q_m| < 1$, by (3.7) it follows that

$$|f_{i_0}(x) - q_m| \le |x - q_m| + 6N\delta < |x - q_m| + \tfrac{1}{20} < \tfrac{1}{3},$$

provided that $\delta < 1/120N$. This contradicts the choice of i_0. □

Lemma 3.3 implies that there exists only finitely many indices i such that $f_i|_{B_m} \ne f_{i-1}|_{B_m}$. Indeed, if $f_i(x) \ne f_{i-1}(x)$ for some $x \in B_m$, then $|q_i - q_m| < 1$ because $|f_{i-1}(x) - q_m| < \frac{1}{3}$ by Lemma 3.3 and $|f_{i-1}(x) - q_i| < \frac{1}{2}$ (since φ_i is the identity outside $B_{1/2}(q_i)$). Thus the sequence $\{f_i|_{B_m}\}_{i=1}^{\infty}$ stabilized and hence the map

$$f(x) = \lim_{i\to\infty} f_i(x) \tag{3.8}$$

is well-defined and smooth on B_m. Since m is arbitrary, f is well-defined and smooth on $U = \bigcup_m B_m$.

Lemma 3.4. *For every $q_m \in X_0$ and every $k \ge 0$ we have*

$$\|f_i - P_m\|_{C^k(B_m)} \le C\delta \qquad \textit{for all } i \ge m, \tag{3.9}$$

and therefore

$$\|f - P_m\|_{C^k(B_m)} \le C\delta. \tag{3.10}$$

Proof. Let $I_m = \{i : |q_i - q_m| < 1\}$ and let $j_1 < \dots < j_{N_m}$ be all elements of I_m. Recall that $N_m = |I_m| \le N = N(n)$. As shown above, Lemma 3.3 implies that φ_i is the identity on $f_{i-1}(B_m)$ for $i \notin I_m$. Therefore for every i we have

$$f_i|_{B_m} = \varphi_{j_{l(i)}} \circ \varphi_{j_{l(i)-1}} \circ \ldots \circ \varphi_{j_1}|_{B_m} \tag{3.11}$$

where $l(i) = \max\{k : j_k \le i\}$.

We compare φ_i and f_i with maps $\widehat{\varphi}_i$ and $\widehat{f}_i$ defined by

$$\widehat{\varphi}_i = \mu_i(x)P_m(x) + (1 - \mu_i(x))x,$$

and

$$\widehat{f}_i = \widehat{\varphi}_i \circ \widehat{\varphi}_{i-1} \circ \ldots \circ \widehat{\varphi}_1.$$

By induction one easily sees that

$$\widehat{f}_i(x) = \lambda_i(x)P_m(x) + (1 - \lambda_i(x))x \tag{3.12}$$

for some $\lambda_i(x) \in [0,1]$, $\lambda_1(x) \le \lambda_2(x) \le \dots$. Therefore $\widehat{f}_i(B_m) \subset B_m$ for all i. Similarly to the case of f_i this implies that

$$\widehat{f}_i|_{B_m} = \widehat{\varphi}_{j_{l(i)}} \circ \widehat{\varphi}_{j_{l(i)-1}} \circ \ldots \circ \widehat{\varphi}_{j_1}|_{B_m} \tag{3.13}$$

By Lemma 3.2, for every $i \in I_m$ we have

$$\|P_i(x) - P_m(x)\| \le C\delta, \qquad \|d_x P_i - d_x P_m\| \le C\delta$$

for all $x \in B_1(q_m)$ and therefore

$$\|\widehat{\varphi}_i - \varphi_i\|_{C^k(B_1(q_m))} = \|\mu_i \cdot (P_m - P_i)\|_{C^k(B_1(q_m))} \le C\delta.$$

This estimate, (3.11), (3.13) and the fact that $l(i) \le |I_m| \le N(n)$ imply that

$$\|f_i - \widehat{f}_i\|_{C^k(B_m)} \le C\delta \tag{3.14}$$

for all i and $k \ge 0$. Observe that $\widehat{\varphi}_m|_{B_m} = P_m|_{B_m}$ since $\mu_m = 1$ on B_m. This fact and relations (3.12) imply that $\widehat{f}_i|_{B_m} = P_m|_{B_m}$ for all $i \ge m$. Therefore for $i \ge m$ the estimate (3.14) turns into (3.9) and the lemma follows. □

Lemma 3.5. $f_m(B_m) \subset D_{1/3}(q_m)$.

Proof. Let $x \in B_m$ and $y = f_{m-1}(x)$, then $f_m(x) = \varphi_m(y)$. By Lemma 3.3, $|y - q_m| < \frac{1}{3}$. Therefore $\mu_m(y) = 1$ and hence $\varphi_m(y) = P_m(y)$. Thus $f_m(x) = P_m(y) \in D_{1/3}(q_m)$. □

By definition, $f = g \circ f_m$ for some smooth map $g\colon E \to E$. Therefore $f(B_m)$ is contained in an image of an n-dimensional disc $D_{1/3}(q_m)$ under a smooth map g.

Lemma 3.6. *$f(B_m) \subset U_{4\delta}(D_{1/3}(q_m))$ for every m, and $f(U) \subset U_{5\delta}(X)$.*

Proof. Let $x \in B_m$. By Lemma 3.3 we have $f_i(x) \in B_{1/3}(q_m)$ for all i. Let us show that $f_i(x) \in U_{4\delta}(A_m)$ for all $i \geq m$. This is true for $i = m$ since $f_m(x) \in D_{1/3}(q_m) \subset A_m$ by Lemma 3.5. Arguing by induction, let $i > m$ and assume that $y = f_{i-1}(x) \in U_{4\delta}(A_m)$. If $|y - q_i| \geq \frac{1}{2}$, then $f_m(x) = y \in U_{4\delta}(A_m)$, so we assume that $|y - q_i| < \frac{1}{2}$. Note that

$$|q_i - q_m| \leq |q_m - y| + |y - q_i| < \tfrac{1}{3} + \tfrac{1}{2} < 1.$$

By definition, the point $f_i(x) = \varphi_i(y)$ belongs to the line segment $[yz]$ where $z = P_i(y)$. Since $z \in A_i$ and $|q_i - z| \leq |q_i - y| < \frac{1}{2}$, we have

$$\operatorname{dist}(z, A_m) \leq \operatorname{dist}(q_i, A_m) + \tfrac{1}{2}\sin\angle(A_i, A_m) < \delta + \tfrac{5}{2}\delta < 4\delta$$

where the second inequality follows from Lemma 3.2. Thus $z \in U_{4\delta}(A_m)$. Since $f_i(x) \in [yz]$, both y and z belong to $U_{4\delta}(A_m)$ and $U_{4\delta}(A_m)$ is a convex set, $f_i(x) \in U_{4\delta}(A_m)$ as claimed.

Thus $f_i(x) \in U_{4\delta}(A_m) \cap B_{1/3}(q_m)$ for all $x \in B_m$ and all $i \geq m$. This implies the first claim of the lemma. To prove the second one, recall that $D_1(q_m) \subset U_\delta(X)$ by (3.4). Hence $f(B_m) \subset U_{4\delta}(D_{1/3}(q_m)) \subset U_{5\delta}(X)$. Since m is arbitrary, the second assertion of the lemma follows. □

Now define

$$M = f(U_{1/5}(X_0)). \tag{3.15}$$

We are going to show that M is a desired submanifold.

Lemma 3.7. *For every $y \in M$ there exists $q_m \in X_0$ such that $|y - q_m| < \frac{1}{100} + 5\delta$ and*

$$M \cap B_{1/100}(y) \subset f(D_{1/10}(q_m)).$$

In particular, $M = \bigcup_m f(D_{1/10}(q_m))$.

Proof. By Lemma 3.6, $y \in U_{5\delta}(X)$. Since X_0 is a $\frac{1}{100}$-net in X, there is point $q_m \in X_0$ such that $|y - q_m| < \frac{1}{100} + 5\delta$. Let us show that this point satisfies the requirements of the lemma. Let $W = M \cap B_{1/100}(y)$ and $D = D_{1/10}(q_m)$. We are to show that $W \subset f(D)$. Fix a point $z \in W$. Observe that

$$|z - q_m| \leq |z - y| + |y - q_m| < \tfrac{1}{100} + \tfrac{1}{100} + C\delta = \tfrac{1}{50} + C\delta.$$

Since $z \in M$, we have $z = f(x)$ for some $x \in U_{1/5}(X_0)$. Let $p \in X_0$ be such that $|x - p| < \frac{1}{5}$. Then $|z - P_{A_p}(x)| < C\delta$ by Lemma 3.4. On the other hand,

$$|x - P_{A_p}(x)| \leq |x - p| < \tfrac{1}{5}.$$

Therefore

$$|x - q_m| \leq |x - P_{A_p}(x)| + |z - P_{A_p}(x)| + |z - q_m| < \tfrac{1}{5} + C\delta + \tfrac{1}{50} + C\delta < \tfrac{1}{4},$$

thus $x \in B_m$.

By Lemma 3.4 it follows that $|z - P_m(x)| = |f(x) - P_m(x)| < C\delta$ and $|f_m(x) - P_m(x)| < C\delta$. Therefore $|f_m(x) - z| < C\delta$ and hence

$$|f_m(x) - q_m| \leq |f_m(x) - z| + |z - q_m| < \tfrac{1}{50} + C\delta.$$

By Lemma 3.5 we have $f_m(x) \in A_m$, hence $f_m(x) \in D_{1/50 + C\delta}(q_m)$.

Now consider the map $f_m|_D$. By Lemma 3.5, its image $f_m(D)$ is contained in A_m. By Lemma 3.4, $f_m|_D$ is $C\delta$-close to the projection $P_m|_D$, which equals id_D since $D \subset A_m$. Thus $f_m|_D$ is $C\delta$-close to the identity and maps D to a subset of the n-dimensional subspace A_m. By topological reasons, see [62, Thm. 1.2.6], this implies that $f_m(D)$ contains an n-ball $D_{1/10-C\delta}(q_m)$. Since $f_m(x) \in D_{1/50+C\delta}(q_m) \subset D_{1/10-C\delta}(q_m)$, it follows that there exists a point $x' \in D$ such that $f_m(x') = f_m(x)$. Since f factors through f_m, this implies that $f(x') = f(x) = z$. Thus $z \in f(D)$. Since z is an arbitrary point of W, the lemma follows. □

The next lemma shows that M is a submanifold and provides bounds for derivatives of a parametrization of M.

Lemma 3.8. *M is a closed n-dimensional smooth submanifold of E. Moreover for every $y \in M$ there exists a smooth map $\varphi\colon V \to E$, where $V = B^n_{1/10}$ is the ball of radius $\frac{1}{10}$ in $\mathbb{R}^n$, such that $y \in \varphi(V) \subset M$ and φ is $C\delta$-close to an affine isometric embedding in the C^k-topology for any $k \geq 0$.*

Proof. Pick $y \in M$ and let $q_m \in X_0$ be as in Lemma 3.7. Let $D = D_{1/10}(q_m)$. By Lemma 3.4, $f|_D$ is $C\delta$-close to the identity in the C^k-topology. In particular, $f(D)$ is an embedded smooth n-dimensional submanifold. By Lemma 3.7,

$$f(D) \cap B_{1/100}(y) = M \cap B_{1/100}(y).$$

Hence $M \cap B_{1/100}(y)$ is a submanifold for every $y \in M$ and therefore M is a submanifold.

To see that M is closed, recall that $|y - q_m| < \frac{1}{100} + 5\delta$. Since $f|_D$ is $C\delta$-close to identity, this implies that the f-image of the boundary of D is separated away from y by distance at least $\frac{1}{10} - \frac{1}{100} - C\delta > \frac{1}{100}$. Therefore $M \cap B_{1/100}(y)$ is contained in a compact subset of the submanifold $f(D)$. Since this holds within a uniform radius $\frac{1}{100}$ from any $y \in M$, it follows that M is a closed set in E.

To construct the desired local parametrization φ, just compose $f|_D$ with an affine isometry between D and an appropriate ball in $\mathbb{R}^n$. □

Note that the existence of local parametrizations that are $C\delta$-close to affine isometries (in the C^2-topology) implies that the second fundamental form of M is bounded by $C\delta$. Let us verify the remaining assertions of Proposition 3.1. The first one is the following lemma.

Lemma 3.9. *$d_H(M, X) \leq 5\delta$.*

Proof. By Lemma 3.6 we have $M \subset U_{5\delta}(X)$. It remains to prove the inclusion $X \subset U_{5\delta}(M)$. Fix $x \in X$ and let $q_m \in X_0$ be such that $|q_m - x| \leq \frac{1}{100}$. Consider the map $P_m \circ f|_{D_{1/5}(q_m)}$ from $D_{1/5}(q_m) \subset A_m$ to A_m. By Lemma 3.4, this map is $C\delta$-close to the identity. Therefore its image contains the n-disc $D_{1/5-C\delta}(q_m)$. This disc contains the point $P_m(x)$ because

$$|P_m(x) - q_m| \leq |x - q_m| \leq \tfrac{1}{100} < \tfrac{1}{5} - C\delta.$$

Hence $P_m(x) \in P_m(f(D_{1/5}(q_m)))$. This means that there exists $y \in D_{1/5}(q_m)$ such that $P_m(f(y)) = P_m(x)$. By Lemma 3.6, we have $\mathrm{dist}(f(y), A_m) < 4\delta$ and therefore

$$|f(y) - P_m(x)| = |f(y) - P_m(f(y))| < 4\delta.$$

By (3.4) we have $\mathrm{dist}(x, A_m) \leq \delta$ and therefore $|x - P_m(x)| \leq \delta$. Hence

$$|f(y) - x| \leq |f(y) - P_m(x)| + |x - P_m(x)| < 4\delta + \delta = 5\delta.$$

Observe that $f(y) \in M$ since $y \in D_{1/5}(q_m) \subset U_{1/5}(X_0)$. This and the above inequality imply that $x \in U_{5\delta}(M)$. Since x is an arbitrary point of X, we have shown that $X \subset U_{5\delta}(M)$. The lemma follows. □

Remark 3.10. We observe that

$$M = f(U_\delta(X)) \tag{3.16}$$

(compare with (3.15)). Indeed, we have $M \subset \bigcup_m f(D_{1/10}(q_m))$ by Lemma 3.7 and $D_{1/10}(q_m) \subset U_\delta(X)$ by (3.4).

One can think of (3.16), (3.15) and the last claim of Lemma 3.7 as various reconstruction procedures for M.

Lemma 3.11. *$|f(y) - y| < C\delta$ for every $y \in U_\delta(X)$.*

Proof. Since $y \in U_\delta(X)$, there is $x \in X$ such that $|x - y| < \delta$. Pick $q_m \in X_0$ such that $|x - q_m| < \frac{1}{100}$. Then $y \in B_m$ and hence $|f(y) - P_m(y)| < C\delta$ by Lemma 3.4. By (3.1) we have $\text{dist}(x, A_m) < \delta$ and hence

$$|y - P_m(y)| = \text{dist}(y, A_m) < 2\delta.$$

Therefore $|f(y) - y| \le |f(y) - P_m(y)| + |y - P_m(y)| < C\delta + 2\delta$. □

Now we are in position to prove the third assertion of Proposition 3.1. We are going to show that the normal injectivity radius of M is no less than $\frac{2}{5} > \frac{1}{3}$ (in fact, any explicit constant smaller than $\frac{1}{2}$ would do). Suppose the contrary, i.e., that the normal injectivity radius of M is less than $\frac{2}{5}$.

First we observe that any relatively small part of M has large normal injectivity radius. More precisely, let κ be an upper bound for the principal curvatures of M and let Σ be a ball of radius $\frac{1}{2}\kappa^{-1}$ centered at $x \in M$ with respect to the intrinsic metric of M. Then the normal injectivity radius of Σ is greater or equal to $(C\kappa)^{-1}$. Indeed, for any point $y \in \Sigma$ we have $\angle(T_y\Sigma, T_x\Sigma) < 1/2$. It follows that Σ is a graph of a smooth function over a region in $T_x\Sigma$, the first derivatives of this function are bounded by 1, and its second derivatives are bounded by $C\kappa^{-1}$. One easily sees that this implies the lower bound $(C\kappa)^{-1}$ for the normal injectivity radius of Σ. Taking into account that $\kappa < C\delta$, we conclude the normal injectivity radius of any intrinsic ball of radius $(C\delta)^{-1}$ in M is bounded below by $(C\delta)^{-1}$.

Hence the non-injectivity of the normal exponential map within normal distance $\frac{2}{5}$ can occur only for points $x, y \in M$ separated by intrinsic distance at least $(C\delta)^{-1}$. Thus there are points $x, y \in M$ such that $|x - y| < \frac{4}{5}$ and $d_M(x, y) > (C\delta)^{-1} > 1$ where d_M is the intrinsic distance in M. We are going to show that these two inequalities contradict each other.

Let $x, y \in M$ be as above. Then by (3.16) there are points $x', y' \in U_\delta(X)$ such that $f(x') = x$ and $f(y') = y$. By Lemma 3.11 we have $|x - x'| < C\delta$ and $|y - y'| < C\delta$, hence $|x' - y'| < \frac{4}{5} + C\delta$ by the triangle inequality. Let $x'', y'' \in X$ be such that $|x' - x''| < \delta$ and $|y' - y''| < \delta$. Then

$$|x'' - y''| \le |x' - y'| + 2\delta < \tfrac{4}{5} + C\delta < 1.$$

Hence $y'' \in B_1^X(x'')$. This and (3.4) imply that $y'' \in U_\delta(D_1(x''))$. Therefore both x' and y' and hence the line segment $[x', y']$ are contained in the 2δ-neighborhood of the affine n-disc $D_1(x'')$. By (3.4), this neighborhood is contained in $U_{3\delta}(X)$. Thus $[x', y']$ is contained in $U_{3\delta}(X)$ and hence in the domain of f. Consider the f-image of the line segment $[x', y']$. It is a smooth path in M connecting x and y. Lemma 3.4 for $k = 1$ implies that f is locally Lipschitz with Lipschitz constant $1 + C\delta$. Therefore

$$\text{length}(f([x', y'])) \le (1 + C\delta)|x' - y'| < (1 + C\delta)(\tfrac{4}{5} + C\delta) < 1.$$

Hence $d_M(x, y) < 1$, a contradiction. This contradiction proves the third claim of Proposition 3.1.

It remains to prove the fourth assertion of Proposition 3.1. By Lemma 3.8, M admits local parametrizations that are $C\delta$-close (in any C^k-topology) to affine isometric embeddings. This and the fact that the normal injectivity radius is bounded below by $\frac{2}{5}$ imply that at every point $x \in U_{2/5}(M)$ the normal projection P_M is well-defined and its derivatives of any order are $C\delta$-close to those of the orthogonal projection to an affine subspace. This implies (3.3).

In order to prove (3.2), consider the first derivative $d_x P_M$ where $x \in U_{2/5}M$. As shown above, it is $C\delta$-close to an orthogonal projection and hence is Lipschitz with Lipschitz constant close to 1. It follows that (3.2) holds (with $C \approx 1$) whenever the line segment $[x,y]$ is contained in $U_{2/5}(M)$. This argument handles all pairs $x,y \in U_{1/3}(M)$ with $|x-y| < \frac{2}{5}-\frac{1}{3} = \frac{1}{15}$. For $x,y \in U_{1/3}(M)$ such that $|x-y| \geq \frac{1}{15}$, (3.2) follows from the fact that $|x-P_M(x)|$ and $|y-P_M(y)|$ are bounded by $\frac{1}{3}$ and therefore

$$|P_M(x)-P_M(y)| \leq |x-y| + \tfrac{2}{3} \leq 11|x-y|.$$

This finishes the proof of Proposition 3.1. As explained in the beginning of this section, of Theorem 2 follows via a rescaling argument.

Remark 3.12. The subspaces A_x from Definition 1.9 approximate the tangent spaces of the submanifold M constructed in Theorem 2. More precisely, if $x \in X$ and $y = P_M(x)$, then the angle of A_x and T_yM satisfies

$$\angle(A_x, T_yM) < C\delta r^{-1} \tag{3.17}$$

where T_yM is the tangent space to M at y.

To prove (3.17), let $A = T_yM$ and consider the intrinsic ball $D = B^M_{r/2}(y)$. Due to the bound $C\delta r^{-2}$ on the second fundamental form, D is $C\delta$-close to the ball $B^A_{r/2}(y)$ in the tangent space. On the other hand, since M is 5δ-close to X and $X \cap B_r(X)$ is δ-close to A_x, D is 6δ-close to A_x. This implies that $B^A_{r/2}(y)$ is $C\delta$-close to A_x and (3.17) follows.

Remark 3.13. Lemma 3.4 and the above arguments about P_M imply that

$$\|f - P_M\|_{C^k(U_{1/5}(X))} < C\delta$$

for all k. Thus, for computation purposes, the explicitly constructed map f is as good as the normal projection P_M.

Remark 3.14. Let us show that the constants in Theorem 2 are optimal, up to constant factors. Let $M \subset E$ be a closed n-dimensional submanifold whose second fundamental form is bounded by $\kappa_{\delta,r} = \frac{1}{2}\delta r^{-2}$, with $0 < \delta < r < 1$, and normal injectivity radius is bounded from below by $2r$. Let $x \in M$. Using formula (1.2) we see that

$$d_H(B^M_{2r}(x), B^{T_xM}_{2r}(x)) \leq \delta. \tag{3.18}$$

Here $B^M_{2r}(x)$ is the intrinsic ball in M of radius $2r$ centered at x.

Our assumptions on M imply that the normal projection P_M is well-defined and 2-Lipschitz in the ball $B^E_r(x)$. Hence for any $z \in M \cap B^E_r(x)$ the projection $P_M([x,z])$ of the line segment $[x,z]$ is a curve of length at most $2r$. Therefore $z = P_M(z) \in B^M_{2r}(x)$. Thus $M \cap B^E_r(x) \subset B^M_{2r}(x)$. Also note that $B^M_r(x) \subset M \cap B^E_r(x)$. These relations, (3.18) and (1.2) imply that $d_H(M \cap B^E_r(x), B^{T_xM}_r(x)) \leq \delta$. As x above is an arbitrary point of M, we have that M is δ-close to n-flats at scale r. This shows that in Theorem 2 the bounds in claims (2) and (3) on the second fundamental form and the normal injectivity radius are optimal, up to multiplying these bounds by constant factors depending on n.

4. Proof of Theorem 1

Similarly to the proof of Theorem 2, we first observe that the the statement of Theorem 1 is scale invariant and it suffices to prove it for $r = 1$. When $r = 1$, Theorem 1 is equivalent to the following proposition with $\delta_0 = \sigma_0 > 0$.

Proposition 4.1. *For every positive integer n there exists $\delta_0 = \delta_0(n) > 0$ such that the following holds. Let $0 < \delta < \delta_0$ and let X be a metric space which is δ-intrinsic and δ-close to $\mathbb{R}^n$ at scale 1. Then there exists a complete n-dimensional Riemannian manifold M such that*

1. *There is a $(1 + C\delta, C\delta)$-quasi-isometry from X to M.*
2. *The sectional curvature Sec_M of M satisfies $|\mathrm{Sec}_M| \le C\delta$.*
3. *The injectivity radius of M is bounded below by $1/2$.*

The proof of Proposition 4.1 occupies the rest of this section, which is split into several subsections. We recycle the letter r for use in other notation. We fix n and assume that a metric space X satisfies the assumption of the proposition for a sufficiently small $\delta > 0$.

Fix a maximal $\frac{1}{100}$-separated set $X_0 \subset X$. We say that two points $x, y \in X_0$ are *adjacent* if $d_X(x, y) < 1$ and say that they are *neighbors* if $d_X(x, y) < \frac{1}{2}$.

The adjacency relation defines a graph which we refer to as the *adjacency graph*. The set of vertices of this graph is X_0 and the edges are between all pairs of adjacent points. We need the following properties of this graph.

Lemma 4.2. *1. The adjacency graph is connected.*
2. Its vertex degrees are bounded by a constant depending only on n.

Proof. 1. Let $x, y \in X_0$. Since X is δ-intrinsic, there is a δ-chain $x_1, \dots, x_N \in X$ with $x_1 = x$ and $x_N = y$. For each x_i, there is a point $x'_i \in X_0$ with $d_X(x_i, x'_i) \le \frac{1}{100}$. By the triangle inequality, $d_X(x'_i, x'_{i+1}) < 2\delta + \frac{1}{50} < 1$ for all i, and we may assume that $x'_1 = x$ and $x'_N = y$. Then the sequence $x'_1, \dots, x'_N$ is a path connecting x to y in the adjacency graph.

2. Let $q \in X_0$. Since $d_H(B_1(q), B_1^n) < \delta$, there exists a 2δ-isometry $f \colon B_1(q) \to B_1^n$. Let $Y = X_0 \cap B_1(q)$ be the set of points adjacent to q. Since Y is $\frac{1}{100}$-separated, its image $f(Y)$ is a $(\frac{1}{100} - 2\delta)$-separated subset of B_1^n. We may assume that δ is so small that $\frac{1}{100} - 2\delta > \frac{1}{200}$. Then the cardinality of Y is no greater than the maximum possible number of $\frac{1}{200}$-separated points in B_1^n. $\square$

Lemma 4.2 implies that the set X_0 is at most countable. In the sequel we assume that X_0 is countably infinite, $X_0 = \{q_i\}_{i=1}^\infty$. In the case when X_0 is finite, the proof is the same except that the indices are restricted to a finite set.

4.1. Approximate charts.

Fix a collection of points $\{p_i\}_{i=1}^\infty$ in $\mathbb{R}^n$ such that the Euclidean unit balls $D_i := B_1(p_i)$ are disjoint. For $r > 0$, we denote by D_i^r the Euclidean ball $B_r(p_i) \subset \mathbb{R}^n$.

Recall that $X_0 = \{q_i\}_{i=1}^\infty$. For each $i \in \mathbb{N}$ we have $d_{GH}(B_1(q_i), D_i) < \delta$ since D_i is isometric to B_1^n. Recall that here we are dealing with pointed GH distance between between balls where the centers are distinguished points. Hence there exists a 2δ-isometry $f_i \colon B_1(q_i) \to D_i$ such that $f_i(q_i) = p_i$.

We fix 2δ-isometries $f_i \colon B_1(q_i) \to D_i$, $i \in \mathbb{N}$, for the rest of the proof. The balls D_i and the maps f_i play the role of coordinate charts in X. The next lemma provides a kind of transition maps between charts.

Lemma 4.3. *For each pair of adjacent points $q_i, q_j \in X_0$, there exists an affine isometry $A_{ij} \colon \mathbb{R}^n \to \mathbb{R}^n$ such that*

$$|A_{ij}(f_i(x)) - f_j(x)| < C\delta \tag{4.1}$$

for every $x \in B_1(q_i) \cap B_1(q_j)$.

Proof. Let $Y = B_1(q_i) \cap B_1(q_j)$. Since $d_{GH}(B_1(q_i), B_1^n) < \delta$ and $q_j \in B_1(q_i)$, there exists $x_0 \in Y$ such that

$$\max\{d_X(x_0, q_i), d_X(x_0, q_j)\} < \tfrac{1}{2} d_X(q_i, q_j) + 2\delta.$$

The map $f_i|_Y$ is a $C\delta$-approximation from Y to intersection of Euclidean balls $Z := D_i \cap B_{1+2\delta}(f_i(q_j)) \subset \mathbb{R}^n$. By the choice of x_0, Z contains the ball of radius $\frac{1}{3}$ centered at $f_i(x_0)$. Consider the map $h_1 \colon Y \to \mathbb{R}^n$ defined by $h_1(x) = f_i(x) - f_i(x_0)$. It is a $C\delta$-isometry from Y to the set Z_1 obtained from Z by the parallel translation by $-f_i(x_0)$. Observe that $B_{1/3}(0) \subset Z_1 \subset B_2(0)$. Similarly, the map $h_2 \colon Y \to \mathbb{R}^n$ defined by $h_2(x) = f_j(x) - f_j(x_0)$ is a $C\delta$-isometry from Y to a set $Z_2 \subset \mathbb{R}^n$ with similar properties. Note that $h_1(x_0) = h_2(x_0) = 0$.

Arguing as in Lemma 2.5 (cf. Remark 2.6) we see that there exists an orthogonal map $U \colon \mathbb{R}^n \to \mathbb{R}^n$ such that

$$|U(h_1(x)) - h_2(x)| < C\delta \tag{4.2}$$

for all $x \in Y$. Now define $A_{ij} \colon \mathbb{R}^n \to \mathbb{R}^n$ by

$$A_{ij}(y) = U(y - f_i(x_0)) + f_j(x_0), \qquad y \in \mathbb{R}^n.$$

This definition and (4.2) implies (4.1). □

We fix maps A_{ij} constructed in Lemma 4.3 for the rest of the proof. We may assume that $A_{ji} = A_{ij}^{-1}$ for all i, j and A_{ii} is the identity map.

Lemma 4.4. *Let $q_i, q_j, q_k \in X_0$ be three pairwise adjacent points. Then*

$$|A_{jk}(A_{ij}(x)) - A_{ik}(x)| < C\delta \tag{4.3}$$

for all $x \in D_i$.

Proof. Consider the set $Y = B_1(q_i) \cap B_1(q_j) \cap B_1(q_k) \subset X$. The map $f_i|_Y$ is a $C\delta$-isometry from Y to the intersection of Euclidean balls

$$Z := D_i \cap B_{1+2\delta}(a) \cap B_{1+2\delta}(b) \subset \mathbb{R}^n,$$

where $a = f_i(q_j)$ and $b = f_i(q_k)$. Let $x \in Z$. Then there exists $p \in Y$ such that $f_i(p)$ is $C\delta$-close to x. Let $y = f_j(p)$ and $z = f_k(p)$. Then by (4.1) we have $|A_{ij}(x) - y| < C\delta$, $|A_{jk}(y) - z| < C\delta$ and $|A_{ik}(x) - z| < C\delta$ and therefore

$$|A_{jk}(A_{ij}(x)) - A_{ik}(x)| < C\delta.$$

Thus (4.3) holds for every $x \in Z$. Since $Z \subset D_i$ and Z contains a ball of radius $\frac{1}{3}$, it follows that (4.3) holds for all $x \in D_i$. □

Lemma 4.5. *Let $q_i, q_j, q_k \in X_0$. Then*

1. *If q_i and q_j are adjacent, then*

$$\Big| |A_{ij}(p_i) - p_j| - d_X(q_i, q_j) \Big| < C\delta.$$

2. *If q_k is adjacent to both q_i and q_j, then*

$$\Big| |A_{ik}(p_i) - A_{jk}(p_j)| - d_X(q_i, q_j) \Big| < C\delta$$

Proof. The first assertion follows from the second one by setting $k = j$ (recall that A_{jj} is the identity map). Let us prove the second assertion.

Since $p_i = f_i(q_i)$, (4.1) implies that $A_{ik}(p_i)$ is $C\delta$-close to $f_k(q_i)$. Similarly, $A_{jk}(p_j)$ is $C\delta$-close to $f_k(q_j)$. Hence the distance $|A_{ik}(p_i) - A_{jk}(p_j)|$ differs from $|f_k(q_i) - f_k(q_j)|$ by at most $C\delta$. In its turn, the distance $|f_k(q_i) - f_k(q_j)|$ differs from $d_X(q_i, q_j)$ by at most 2δ because f_j is a 2δ-isometry. Thus $|A_{ik}(p_i) - A_{jk}(p_j)|$ differs from $d_X(q_i, q_j)$ by at most $C\delta$ and the lemma follows. □

Lemma 4.6. *For every $i \in \mathbb{N}$ and every $x \in D_i^{1/3}$ there exist $j \in \mathbb{N}$ such that q_i and q_j are neighbors and $A_{ij}(x) \in D_j^{1/50}$.*

Proof. Since f_i is a 2δ-isometry from $B_1(q_i)$ to D_i, there exists $y \in B_1(q_i) \subset X$ such that $|f_i(y) - x| \le 2\delta$. Since X_0 is a $\frac{1}{100}$-net in X, there is a point $q_j \in X_0$ such that $d_X(y, q_j) \le \frac{1}{100}$. For this point q_j we have

$$|x - f_i(q_j)| < |f_i(y) - f_i(q_j)| + 2\delta < d_X(y, q_j) + 4\delta \le \tfrac{1}{100} + 4\delta$$

since f_i is a 2δ-isometry. This and the fact that $x \in D_i^{1/3}$ imply that

$$|p_i - f_i(q_j)| < \tfrac{1}{3} + \tfrac{1}{100} + 4\delta.$$

Since $p_i = f_i(q_i)$ and f_i is a 2δ-isometry, it follows that

$$d_X(q_i, q_j) < \tfrac{1}{3} + \tfrac{1}{100} + 6\delta < \tfrac{1}{2}.$$

Thus q_i and q_j are neighbors, in particular there is a well-defined map A_{ij}. Since A_{ij} is an isometry, we have

$$|A_{ij}(x) - A_{ij}(f_i(q_j))| = |x - f_i(q_j)| < \tfrac{1}{100} + 4\delta.$$

By (4.1) we have $|A_{ij}(f_i(q_j)) - f_j(q_j)| < C\delta$, hence

$$|A_{ij}(x) - p_j| = |A_{ij}(x) - f_j(q_j)| < \tfrac{1}{100} + C\delta < \tfrac{1}{50}$$

provided that δ is sufficiently small. Thus $A_{ij}(x) \in D_j^{1/50}$ as claimed. □

4.2. **Approximate Whitney embedding.** At this point we essentially forget about the original metric space X and use the collection of balls $D_i \subset \mathbb{R}^n$ and maps A_{ij} from the previous section for the rest of the construction. Let $\Omega = \bigcup D_i$.

Let S be the unit sphere in $\mathbb{R}^{n+1}$ centered at e_{n+1}, where $e_1, \dots, e_{n+1}$ is the standard basis of $\mathbb{R}^n$. Note that S contains the points 0 and $2e_{n+1}$. For every $r > 0$ we denote by S_r the set of points in S lying at distance less than r from the 'north pole' $2e_{n+1}$. That is, $S_r = S \cap B_r(2e_{n+1})$.

Fix a smooth map $\varphi\colon \mathbb{R}^n \to S$ with the following properties:

(1) $\varphi(x) = 0$ for all $x \in \mathbb{R}^n \setminus B_{1/5}(0)$.
(2) $\varphi|_{B_{1/5}(0)}$ is a diffeomorphism onto $S \setminus \{0\}$.
(3) $\varphi|_{B_{1/10}(0)}$ is a diffeomorphism onto the spherical cap S_1.
(4) $\varphi|_{B_{1/50}(0)}$ is a diffeomorphism onto the spherical cap $S_{1/10}$.

For each i let $\varphi_i(x) = \varphi(x - p_i)$ and define a map $F_i\colon \Omega \to S \subset \mathbb{R}^{n+1}$ as follows. If a point $x \in \Omega$ belongs to a ball D_j, put

$$F_i(x) = \begin{cases} \varphi_i(A_{ji}(x)), & \text{if } D_j \text{ is adjacent to } D_i \\ 0, & \text{otherwise.} \end{cases} \tag{4.4}$$

In particular $F_i(x) = \varphi_i(x)$ if $x \in D_i$.

Lemma 4.7. *If $F_i(x) \neq 0$ for some $x \in D_j^{1/5}$, then q_i and q_j are neighbors.*

Proof. The assumption $F_i(x) \neq 0$ implies that q_i and q_j are adjacent and therefore $F_i(x) = \varphi_i(A_{ji}(x))$. Thus $\varphi_i(A_{ji}(x)) \neq 0$ and hence $|A_{ji}(x) - p_i| < \frac{1}{5}$. Since A_{ji} is an isometry and $|p_j - x| < \frac{1}{5}$, we have

$$|A_{ji}(p_j) - p_i| \le |p_j - x| + |A_{ji}(x) - p_i| < \tfrac{2}{5}.$$

This and Lemma 4.5(2) imply that $d_X(q_i, q_j) < \frac{2}{5} + C\delta < \frac{1}{2}$, hence q_i and q_j are neighbors. □

Let E be the space of square-summable sequences $(u_i)_{i=1}^\infty$ in $\mathbb{R}^{n+1}$ equipped with the norm defined by $|u|^2 = \sum |u_i|^2$ for $u = (u_i)_{i=1}^\infty$. This is a Hilbert space naturally isomorphic to ℓ^2. Define a map $F\colon \Omega \to E$ by

$$F(x) = (F_i(x))_{i=1}^\infty \tag{4.5}$$

Lemma 4.2 implies that for every $x \in U$ there are only finitely many indices i such that $F_i(x) \neq 0$. Therefore the sequence $F(x) \in (\mathbb{R}^{n+1})^\infty$ is finite and hence indeed belongs to E.

Lemma 4.8. *1. F is smooth and moreover*

$$\|F\|_{C^k(\Omega)} \le C_k \tag{4.6}$$

for all $k \ge 0$.

2. For every $i \in \mathbb{N}$ the restriction $F|_{D_i^{1/10}}$ is uniformly bi-Lipschitz, that is,

$$C^{-1}|x-y| \le |F(x) - F(y)| \le C|x-y| \tag{4.7}$$

for all $x, y \in D_i^{1/10}$.

Proof. 1. Let $x \in D_i$. By Lemma 4.2, there is at most C indices j such that $F_j|_{D_i} \neq 0$. For every such j we have $\|d_x^k F_j\| \le \|\varphi\|_{C^k(\mathbb{R}^n)}$, therefore $\|d_x^k F\| \le C \cdot \|\varphi\|_{C^k(\mathbb{R}^n)} = C_k$.

2. The second inequality in (4.7) follows from (4.6). To prove the first one, observe that $|F(x) - F(y)| \ge |F_i(x) - F_i(y)| \ge C^{-1}|x-y|$ since the ith coordinate projection from E to $\mathbb{R}^n$ does not increase distances and $F_i|_{D_i^{1/10}} = \varphi_i|_{D_i^{1/10}}$ is bi-Lipschitz. □

Eq. (4.7) implies that the first derivative of F is uniformly bi-Lipschitz, i.e.,

$$C^{-1}|v| \le |d_x F(v)| \le C|v| \tag{4.8}$$

for all $x \in D_i^{1/10}$ and $v \in \mathbb{R}^n$.

Lemma 4.8 implies that for each i the image $\Sigma_i := F(D_i^{1/10})$ is a smooth submanifold of E. Moreover this submanifold has bounded geometry (e.g., bounded curvatures, normal injectivity radius, etc.) We are going to apply Theorem 2 to the union $\Sigma = \bigcup_i \Sigma_i$ in E. As the first step, we show that these submanifolds lie close to one another.

Lemma 4.9. *Suppose that q_i and q_j are neighbors and let $x \in D_i^{1/5}$. Then $A_{ij}(x) \in D_j$ and*

$$|F(x) - F(A_{ij}(x))| < C\delta.$$

Moreover,

$$\|d_x^m(F - F \circ A_{ij})\| < C_m \delta \tag{4.9}$$

for all $m \ge 0$.

Proof. By Lemma 4.5,

$$|A_{ij}(p_i) - p_j| < d_X(q_i, q_j) + C\delta < \tfrac{1}{2} + C\delta.$$

Since A_{ij} is an isometry, $|A_{ij}(x) - A_{ij}(p_i)| = |x - p_i| < \frac{1}{5}$. Therefore

$$|A_{ij}(x) - p_j| \le |A_{ij}(x) - A_{ij}(p_i)| + |A_{ij}(p_i) - p_j| < \tfrac{1}{2} + \tfrac{1}{5} + C\delta < 1,$$

hence $A_{ij}(x) \in D_j$. Since x is an arbitrary point of $D_i^{1/5}$, we have shown that $A_{ij}(D_i^{1/5}) \subset D_j$.

Recall that the number of indices k such that F_k does not vanish on $D_i \cup D_j$ is bounded by a constant depending only on n. Hence in order to verify (4.9) it suffices to show that

$$\|d_x^m(F_k - F_k \circ A_{ij})\| < C_m\delta \tag{4.10}$$

for every fixed k. Consider four cases.

Case 1: q_k is adjacent to both q_i and q_j. In this case

$$F_k|_{D_i^{1/5}} = \varphi_k \circ A_{ik}|_{D_i^{1/5}}$$

and

$$F_k \circ A_{ij}|_{D_i^{1/5}} = \varphi_k \circ A_{jk} \circ A_{ij}|_{D_i^{1/5}}.$$

Now (4.10) follows follows from the fact that the affine isometries A_{ik} and $A_{jk} \circ A_{ij}$ are $C\delta$-close on D_i by Lemma 4.4.

Case 2: q_k is not adjacent to q_i and q_j. This case is trivial because $F_k|_{D_i}$ and $F_k \circ A_{ij}|_{D_i}$ both vanish by definition.

Case 3: q_k is adjacent to q_j but not to q_i. In this case $F_k|_{D_i} = 0$ by definition. Let us show that $F_k \circ A_{ij}|_{D_i^{1/5}}$ also vanishes. Since $d_X(q_k, q_i) \geq 1$, Lemma 4.5 implies that $|A_{kj}(p_k) - A_{ij}(p_i)| > 1 - C\delta$. Hence for every $y \in D_i^{1/5}$,

$$|A_{kj}(p_k) - A_{ij}(y)| > 1 - \tfrac{1}{5} - C\delta > \tfrac{1}{5}.$$

Since $A_{kj} = A_{jk}^{-1}$ and A_{kj} is an isometry, this implies that $|p_k - A_{jk} \circ A_{ij}(y)| > \frac{1}{5}$ and hence

$$F_k \circ A_{ij}(y) = \varphi_k \circ A_{jk} \circ A_{ij}(y) = 0$$

for every $y \in D_i^{1/5}$.

Case 4: q_k is adjacent to q_i but not to q_j. In this case $F_k \circ A_{ij}|_{D_i^{1/5}} = 0$, so it suffices to prove that $F_k|_{D_i^{1/5}} = 0$. Suppose the contrary, then Lemma 4.7 implies that q_k and q_i are neighbors. Since q_i and q_j are also neighbors, it follows that q_k and q_j are adjacent. This contradiction proves the claim. □

We introduce the following notation for some important subsets of E. For every $i \in \mathbb{N}$ define

$$\Sigma_i = F(D_i^{1/10}) \quad \text{and} \quad \Sigma_i^0 = F(D_i^{1/50}).$$

Let $\Sigma = \bigcup_i \Sigma_i$ and $\Sigma^0 = \bigcup_i \Sigma_i^0$.

Recall that Σ_i is a smooth n-dimensional submanifold of E. For a point $x \in \Sigma_i$, we denote by $T_x\Sigma_i$ the tangent space of Σ_i at x realized as an affine subspace of E containing x. That is, $T_x\Sigma_i$ is the n-dimensional affine subspace of E tangent to Σ_i at x.

Lemma 4.10. *For every $x \in \Sigma_i$ there exist $j \in \mathbb{N}$ and $y \in \Sigma_j^0$ such that*

$$|x - y| < C\delta \tag{4.11}$$

and

$$\angle(T_x\Sigma_i, T_y\Sigma_j) < C\delta. \tag{4.12}$$

Proof. Since $x \in \Sigma_i$, we have $x = F(z)$ for some $z \in D_i^{1/10}$. By Lemma 4.6 there exists j such that q_i and q_j are neighbors and $A_{ij}(z) \in D_i^{1/50}$. Let $y = F(A_{ij}(z))$, then $y \in \Sigma_j^0$. Lemma 4.9 for $m = 0$ implies that

$$|x - y| = |F(z) - F(A_{ij}(z))| < C\delta$$

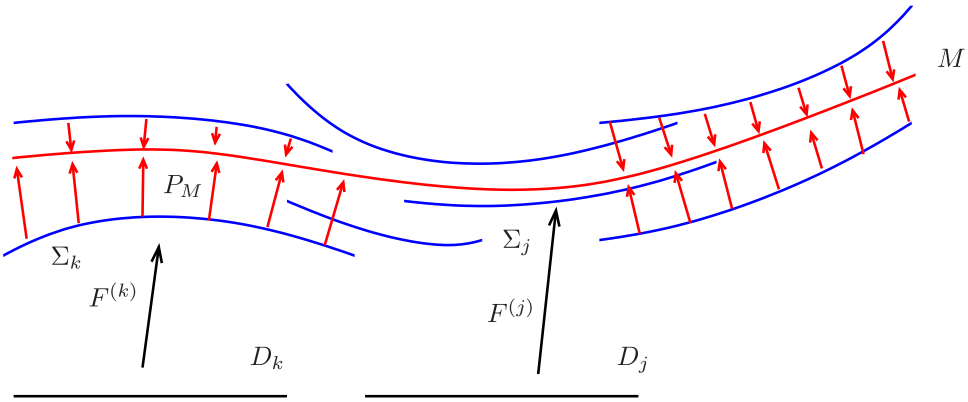

FIGURE 2. A schematic visualisation of the interpolation algorithm 'ManifoldConstruction' based on Theorem 1, see Section 5. Assume that a finite metric space (X, d_X) is given. Then, we construct local coordinate charts $D_i^r \subset \mathbb{R}^n$ approximating the r-balls $B_r^X(x_i) \subset X$ in the data space X. We embed these local charts to an Euclidean space $E = \mathbb{R}^m$ using a Whitney-type embeddings $F^{(i)} = F|_{D_i} : D_i^{1/10} \to \Sigma_i$. Surfaces $\Sigma_i \subset E$ are denoted by blue curves. Using the algorithm SurfaceInterpolation, the union $\bigcup_i \Sigma$ is interpolated to a red surface $M \subset E$. When P_M is the normal projector onto M, denoted by the red arrows, we can determine a metric tensor g_i on $P_M(\Sigma_i)$ by pushing forward the Euclidean metric from D_i to $P_M(\Sigma_i)$ by the map $P_M \circ F|_{D_i}$. The metric tensor g on M is obtained by computing a smooth weighted average of tensors g_i.

proving (4.11). To prove (4.12), observe that $T_x\Sigma_i$ and $T_y\Sigma_j$ are parallel to the images of the derivatives d_zF and $d_{A_{ij}(z)}F$, resp. The image of $d_{A_{ij}(z)}F$ coincides with the image of $d_z(F \circ A_{ij})$. By Lemma 4.9 for $m = 1$ we have

$$\|d_zF - d_z(F \circ A_{ij})\| < C\delta.$$

This and (4.8) imply (4.12). □

We use general metric space notation for subsets of E. In particular, for a set $Z \subset E$ and $r > 0$ we denote by $U_r(Z)$ the r-neighborhood of Z in E.

Lemma 4.11. *$\Sigma \cap U_{1/2}(\Sigma_i^0) \subset U_{C\delta}(\Sigma_i)$ for every $i \in \mathbb{N}$.*

Proof. Let $q \in \Sigma \cap U_{1/2}(\Sigma_i^0)$. Since $q \in U_{1/2}(\Sigma_i^0)$, there exists $y \in D_i^{1/50}$ such that $|q - F(y)| < \frac{1}{2}$. Since $q \in \Sigma$, we have $q = F(z)$ where $z \in D_j^{1/10}$ for some j. Since the ith coordinate projection from E to $\mathbb{R}^{n+1}$ does not increase distances,

$$|F_i(z) - F_i(y)| \le |F(z) - F(y)| = |q - F(y)| < \tfrac{1}{2}.$$

Recall that $F_i(y) = \varphi_i(y)$ because $y \in D_i$. Since $y \in D_i^{1/50}$, the point $\varphi_i(y)$ belongs to the spherical cap $S_{1/10}$. Hence $|F_i(y) - 2e_{n+1}| < \frac{1}{10}$. Therefore

$$|F_i(z) - 2e_{n+1}| \le |F_i(z) - F_i(y)| + |F_i(y) - 2e_{n+1}| < \tfrac{1}{2} + \tfrac{1}{10} < 1.$$

Thus $F_i(z)$ belongs to the spherical cap $S_1 \subset S \subset \mathbb{R}^{n+1}$, in particular $F_i(z) \neq 0$. Hence $F_i(z) = \varphi_i(A_{ji}(z))$ and therefore $A_{ji}(z) \in \varphi_i^{-1}(S_1) = D_i^{1/10}$.

Since $F_i(z) \neq 0$, Lemma 4.7 implies that q_i and q_j are neighbors. Now by Lemma 4.9 (for $m = 0$) we have

$$|q - F(A_{ji}(z))| = |F(z) - F(A_{ji}(z))| < C\delta.$$

Since $A_{ji}(z) \in D_i^{1/10}$, this inequality implies that

$$q \in U_{C\delta}(F(D_i^{1/10})) = U_{C\delta}(\Sigma_i).$$

Since q is an arbitrary point from the set $\Sigma \cap U_{1/2}(\Sigma_i^0)$, the lemma follows. □

Lemma 4.12. *For every $q \in \Sigma_i^0$ and every $r > 0$,*

$$d_H(\Sigma_i \cap B_r(q), T_q\Sigma_i \cap B_r(q)) < Cr^2.$$

Proof. By Lemma 4.8, $\Sigma_i = F(D_i^{1/10})$ is a surface parametrized by a uniformly bi-Lipschitz smooth map $F|_{D_i^{1/10}}$. We may assume that $r < \frac{1}{50C_0}$ where C_0 is the bi-Lipschitz constant in (4.7). Let $q = F(x)$ where $x \in D_i^{1/50}$. Then every point $q' \in \Sigma_i \cap B_r(q)$ is the image of some $x' \in B_{r/C_0}(x) \subset B_{1/50}(x) \subset D_i^{1/10}$. Hence

$$\mathrm{dist}(q', T_q\Sigma_i) \leq Cr^2,$$

where C_2 is the uniform bound of the second derivatives of $F|_{D_i^{1/10}}$, see (4.6). This means that Σ_i deviates from its tangent space $T_q\Sigma_i$ within the r-ball $B_r(q)$ by distance at most Cr^2.

In addition, the point $q \in \Sigma_i^0 = F(D_i^{1/50})$ is separated by a distance at least $\frac{1}{20C_0} > 2r$ from the boundary of Σ_i. Therefore, for each point from $T_q\Sigma_i \cap B_r(q)$ there exists a point in Σ_i within distance C_2r^2. □

The next lemma essentially says that the $\Sigma \subset E$ is $C\delta$-close to affine spaces in E at a scale of order $\delta^{1/2}$.

Lemma 4.13. *For every $x \in \Sigma_i$ and every $r \geq C\delta^{1/2}$,*

$$d_H(\Sigma \cap B_r(x), T_x\Sigma_i \cap B_r(x)) < Cr^2.$$

Proof. By Lemma 4.10, there exists $j \in \mathbb{N}$ and $q \in \Sigma_j^0$ such that $|x - q| < C\delta$ and $\angle(T_x\Sigma_i, T_q\Sigma_j) < C\delta$. Let $A = T_q\Sigma_j$. Observe that the Hausdorff distance between the affine balls $T_x\Sigma_i \cap B_r(x)$ and $B_r^A(q) = A \cap B_r(q)$ is bounded by

$$|x - q| + r \sin\angle(T_x\Sigma_i, A) < C\delta + Cr\delta < Cr^2$$

since $\delta \leq cr^2$. Hence it suffices to verify that $d_H(\Sigma \cap B_r(x), B_r^A(q)) < Cr^2$. By the definition of the Hausdorff distance, this is equivalent to the following pair of inclusions:

$$\Sigma \cap B_r(x) \subset U_{Cr^2}(B_r^A(q)) \tag{4.13}$$

and

$$B_r^A(q) \subset U_{Cr^2}(\Sigma \cap B_r(x)). \tag{4.14}$$

Since $|x - q| < C\delta$, we have $B_r(x) \subset B_{r+C\delta}(q)$ and therefore

$$\Sigma \cap B_r(x) \subset \Sigma \cap B_{r+C\delta}(q) \subset \Sigma \cap U_{r+C\delta}(\Sigma_j^0) \subset U_{C\delta}(\Sigma_j)$$

where the last inclusion follows from Lemma 4.11. Hence

$$\Sigma \cap B_r(x) \subset U_{C\delta}(\Sigma_j) \cap B_{r+C\delta}(q) \subset U_{C\delta}(\Sigma_j) \cap B_{r+2C\delta}(q)) \subset U_{Cr^2}(B_{r+2C\delta}^A(q))$$

where the last inclusion follows from Lemma 4.12 and the assumption that $\delta \leq r^2$. Since $B_{r+2C\delta}^A(q) \subset U_{2C\delta}(B_r^A(q))$, this implies (4.13).

It remains to verify (4.14). Since $|x - q| < C\delta$ and $r \geq \delta^{1/2}$, we may assume that $r > |x - q|$. Let $r_1 = r - |x - q|$. By Lemma 4.12,

$$B^A_{r_1}(q) \subset U_{Cr^2}(\Sigma \cap B_{r_1}(q)) \subset U_{Cr^2}(\Sigma \cap B_r(x)).$$

Since $B^A_r(q) \subset U_{r-r_1}(B^A_{r_1}(q))$ and $r - r_1 < C\delta < Cr^2$, this implies (4.14) and the lemma follows. □

4.3. **The manifold M.** We choose a positive constant $r_0 < 1$ such that

$$C_0 r_0 < \sigma_0 \tag{4.15}$$

where C_0 is the constant C from Lemma 4.13 and σ_0 is the constant from Theorem 2. Some additional requirements on r_0 arise in the course of the argument below, but the final value of r_0 depends only on n.

We may assume that the constant δ_0 in Proposition 4.1 satisfies $\delta_0 < cr_0^2$, where $c = C^{-2}$ with C being the constant from Lemma 4.13. Then, for $\delta < \delta_0$, Lemma 4.13 implies that

$$d_H(\Sigma \cap B_{r_0}(x), T_x\Sigma_i \cap B_{r_0}(x)) < Cr_0^2 \tag{4.16}$$

for every $x \in \Sigma_i$. This and (4.15) imply that the assumptions of Theorem 2 are satisfied for Σ in place of X, r_0 in place of r, Cr_0^2 in place of δ, and $T_x\Sigma_i$ in place of A_x (for $x \in \Sigma_i$). The conclusion of Theorem 2 with these settings is the following lemma.

Lemma 4.14. *Let Σ satisfy (4.16). If r_0 is sufficiently small and $\delta < cr_0^2$, then there exists a closed n-dimensional smooth submanifold $M \subset E$ such that*
1. $d_H(\Sigma, M) < Cr_0^2 < \frac{1}{10} r_0 < \frac{1}{10}$.
2. *The second fundamental form of M at every point is bounded by C.*
3. *The normal injectivity radius of M is at least $r_0/3$.*
4. *The normal projection $P_M : U_{r_0/3}(M) \to M$ is C-Lipschitz and satisfies $\|d^m_x P_M\| < C_m r_0^{2-m}$ for all $m \geq 2$ and $x \in U_{r_0/3}(M)$.*
5. $\angle(T_x\Sigma_i, T_{P_M(x)}M) < Cr_0$ *for every $x \in \Sigma_i$.* □

In Lemma 4.14(1), the first inequality follows from Theorem 2 and the subsequent ones follow from the assumption that r_0 is sufficiently small. The inequality $d_H(\Sigma, M) < \frac{1}{10} r_0$ ensures that Σ lies 'deep inside' the domain of P_M. The last assertion of Lemma 4.14 comes from Remark 3.12.

Let $M \subset E$ be a submanifold from Lemma 4.14. The fourth assertion of Lemma 4.14 for $m = 2$ implies that

$$\|d_x P_M - d_{P_M(x)} P_M\| \leq C \operatorname{dist}(x, M) < Cr_0 \tag{4.17}$$

for every $x \in U_{r_0/3}(M)$. For $x \in M$, the map $d_x P_M$ is the orthogonal projector in T_xE onto T_xM so that $\|d_x P_M\| \leq 1$. This and (4.17) yield that $\|d_x P_M\| \leq 1 + C \operatorname{dist}(x, M)$ for $x \in U_{r_0/3}(M)$. Hence, $x \mapsto P_M(x)$ is locally Lipschitz in $x \in U_{r_0/3}(M)$ with the Lipschitz constant $1 + Cr_0$. Below, we assume that r_0 is chosen so that this Lipschitz constant satisfies $1 + Cr_0 < 2$.

Recall that the set $\Sigma = \bigcup_i \Sigma_i = \bigcup_i F(D_i^{1/10})$ is contained in the domain of P_M. For each i, define a map $\psi_i : D_i^{1/10} \to M$ by

$$\psi_i = P_M \circ F|_{D_i^{1/10}}$$

and let V_i be the image of ψ_i, that is

$$V_i = P_M(F(D_i^{1/10})) = P_M(\Sigma_i).$$

Observe that

$$|\psi_i(x) - F(x)| \leq d_H(\Sigma, M) < Cr_0^2 < \tfrac{1}{10} \tag{4.18}$$

for every $x \in D_i^{1/10}$. This follows from Lemma 4.14(1) and the fact that $\psi_i(x)$ is the nearest point in M to $F(x)$.

The next lemma shows that the maps ψ_i provide a nice family of coordinate charts for M.

Lemma 4.15. *If r_0 is sufficiently small and $\delta < cr_0^2$, then*

1. ψ_i is uniformly bi-Lipschitz, that is,

$$C^{-1}|x-y| \le |\psi_i(x) - \psi_i(y)| \le C|x-y|$$

for all $x, y \in D_i^{1/10}$. In particular, V_i is an open subset of M and ψ_i is a diffeomorphism between $D_i^{1/10}$ and V_i.

2. $\bigcup_i \psi_i(D_i^{1/30}) = M$.

3. If $i, j \in \mathbb{N}$ are such that $V_i \cap V_j \neq \emptyset$, then q_i and q_j are neighbors.

Proof. 1. Since P_M and $F|_{D_i^{1/10}}$ are uniformly Lipschitz, so is their composition ψ_i. It remains to prove that

$$|\psi_i(x) - \psi_i(y)| \ge C^{-1}|x-y| \tag{4.19}$$

for all $x, y \in D_i^{1/10}$. Lemma 4.14(5) and (4.17) imply that for every $x \in \Sigma_i$ the restriction of $d_x P_M$ to $T_x\Sigma_i$ is Cr_0-close to a linear isometry between $T_x\Sigma_i$ and $T_{P_M(x)}M$. This fact and (4.8) imply that there is $C = C(n)$ such that

$$|d_x\psi_i(v)| \ge C^{-1}|v| \tag{4.20}$$

for all $x \in D_i^{1/10}$ and $v \in \mathbb{R}^n$. By Lemma 4.14(4), the derivatives of P_M up to the second order are bounded by a constant not depending on r_0. By (4.6) it follows that the second derivatives of ψ_i are uniformly bounded. This, (4.20) and a quantitative version of the inverse function theorem imply that (4.19) holds whenever the distance $|x-y|$ is no greater than some constant c_0 depending only on n.

To handle the case when $|x-y| > c_0$, observe that

$$|\psi_i(x) - \psi_i(y)| > |F(x) - F(y)| - Cr_0^2$$

by (4.18). Since $F|_{D_i}$ in uniformly bi-Lipschitz (by Lemma 4.8), it follows that

$$|\psi_i(x) - \psi_i(y)| \ge C^{-1}|x-y| - Cr_0^2 \tag{4.21}$$

for all $x, y \in D_i^{1/10}$. If $|x-y| > c_0$ and r_0 is so small that $C_1 r_0^2 < \frac{1}{2}C_1^{-1}c_0$ where C_1 is the constant C from (4.21), then the right-hand side of (4.21) is bounded below by $\frac{1}{2}C_1^{-1}|x-y|$. Thus (4.19) holds for all $x, y \in D_i^{1/10}$ and the first claim of the lemma follows.

2. Let $x \in M$. By Lemma 4.14(1) there exists $z \in \Sigma$ such that $|x-z| < Cr_0^2$. By Lemma 4.10 there exists $i \in \mathbb{N}$ and $y \in \Sigma_i^0$ such that $|y-z| < C\delta$. Then

$$|x-y| < Cr_0^2 + C\delta < Cr_0^2 < r_0/3$$

where in the last inequality we assume that r_0 is sufficiently small. We are going to show that $x \in F(D_i^{1/30})$.

Since $x \in M$ and $|x-y| < r_0/3$, the straight line segment $[x, y]$ is contained in the domain of P_M. Let γ be the image of this segment under P_M. Then γ is a smooth curve in M connecting x to the point $P_M(y) \in P_M(\Sigma_i^0) = \psi_i(D_i^{1/50})$. Since P_M is locally 2-Lipschitz, we have $\text{length}(\gamma) \le 2|x-y| < Cr_0^2$. We parametrize γ by $[0,1]$ in such a way that $\gamma(0) = P_M(y)$ and $\gamma(1) = x$. Suppose that $x \notin \psi_i(D_i^{1/30})$ and let

$$t_0 = \min\{t \in [0,1] : \gamma(t) \notin \psi_i(D_i^{1/30})\}.$$

This minimum exists since $\psi_i(D_i^{1/30})$ is an open subset of M. Define $\widetilde{\gamma}(t) = \psi_i^{-1}(\gamma(t))$ for all $t \in [0, t_0)$. Note that $t_0 > 0$ and $\widetilde{\gamma}(0) \in D_i^{1/50}$ because $P_M(y) \in \psi_i(D_i^{1/50})$. Since ψ_i is a diffeomorphism onto its image, $\widetilde{\gamma}$ is a smooth curve in D_i. Moreover, since ψ_i is uniformly bi-Lipschitz, we have

$$\operatorname{length}(\widetilde{\gamma}) \le C \operatorname{length}(\gamma) < Cr_0^2.$$

Hence the limit point $p = \lim_{t \to t_0} \widetilde{\gamma}(t)$ exists and satisfies

$$|p - \widetilde{\gamma}(0)| \le \operatorname{length}(\widetilde{\gamma}) < Cr_0^2.$$

We may assume that r_0 is so small that the right-hand side of this inequality is smaller than $\frac{1}{30} - \frac{1}{50}$. Since $\widetilde{\gamma}(0) \in D_i^{1/50}$, it follows that $z \in D_i^{1/50}$. Hence $\gamma(t_0) = \psi_i(p) \in \psi_i(D_i^{1/30})$, contrary to the choice of t_0. This contradiction shows that $x \in \psi_i(D_i^{1/30})$. Since x is an arbitrary point of M, the second claim of the lemma follows.

3. Assume that $V_i \cap V_j \neq \emptyset$. Then there exist $x \in D_i^{1/10}$ and $y \in D_j^{1/10}$ such that $\psi_i(x) = \psi_j(y)$. This equality and (4.18) imply that $|F(x) - F(y)| < \frac{1}{5}$, hence

$$|F_i(x) - F_i(y)| < \tfrac{1}{5} \tag{4.22}$$

(recall that $F_i \colon \Omega \to \mathbb{R}^{n+1}$ is the ith coordinate projection of F). Since $x \in D_i^{1/10}$, the point $F_i(x) \in \mathbb{R}^{n+1}$ belongs to the spherical cap S_1 and therefore $|F_i(x)| > 1$. This and (4.22) imply that $F_i(y) \neq 0$ and hence q_i and q_j are neighbors by Lemma 4.7. □

Note that Lemma 4.15(3) and Lemma 4.2(2) imply that the sets V_i cover M with bounded multiplicity, that is, for every $x \in M$ the number of indices i such that $x \in V_i$ is bounded by a constant depending only on n.

Now we can fix the value of r_0 such that Lemma 4.14 and Lemma 4.15 work. Since r_0 is yet another constant depending only on n, we omit the dependence on r_0 in subsequent estimates and just use the generic notation C. In particular, the fourth assertion of Lemma 4.14 now implies that

$$\|P_M\|_{C^m(U_{r_0/3}(M))} < C_m \tag{4.23}$$

for all $m \ge 0$. This and (4.6) imply that

$$\|\psi_i\|_{C^m(D_i^{1/10})} < C_m \tag{4.24}$$

for all $m \ge 0$.

Lemma 4.16. *If $x \in D_i^{1/10}$, $y \in D_j^{1/10}$ and $\psi_i(x) = \psi_j(y)$, then*

$$|F(x) - F(y)| < C\delta. \tag{4.25}$$

Proof. Applying Lemma 4.10 to the point $F(x) \in \Sigma_i$ yields that there exists $k \in \mathbb{N}$ and a point $z \in D_k^{1/50}$ such that $|F(x) - F(z)| < C\delta$. Since P_M is uniformly Lipschitz, it follows that

$$|\psi_i(x) - \psi_k(z)| < C\delta \tag{4.26}$$

and (since $\psi_i(x) = \psi_j(y)$)

$$|\psi_j(y) - \psi_k(z)| < C\delta. \tag{4.27}$$

This and (4.18) imply that $|F(y) - F(z)| < \frac{1}{5} + C\delta < \frac{1}{2}$, hence $F(y) \in U_{1/2}(\Sigma_k^0)$. By Lemma 4.11 it follows that $F(y) \in U_{C\delta}(\Sigma_k)$. This means that there exists $z' \in D_k^{1/10}$ such that

$$|F(z') - F(y)| < C\delta. \tag{4.28}$$

Then

$$|\psi_k(z') - \psi_j(y)| = |P_M(F(z')) - P_M(F(y))| < C\delta.$$

since P_M is uniformly Lipschitz. This and (4.27) imply that $|\psi_k(z) - \psi_k(z')| < C\delta$. Since ψ_k is uniformly bi-Lipschitz by the first claim of the lemma 4.15, it follows that

$$|z - z'| \le C|\psi_i(z) - \psi_i(z')| < C\delta$$

and hence $|F(z) - F(z')| < C\delta$ by Lipschitz continuity of F. This and (4.28) imply that $|F(y) - F(z)| < C\delta$.

Thus we have shown that (4.27) implies that $|F(y) - F(z)| < C\delta$. Similarly (4.26) implies that $|F(x) - F(z)| < C\delta$ and (4.25) follows. □

We are going to restrict our coordinate maps ψ_i to smaller balls $D_i^{1/15}$. Let $V_i' = \psi_i(D_i^{1/15})$ and $U_{ij} = \psi_i^{-1}(V_i' \cap V_j')$. The set $U_{ij} \subset D_i^{1/15}$ is the natural domain of the transition map $\psi_j^{-1} \circ \psi_i$ between the restricted coordinate charts.

Lemma 4.17. *Let $i, j \in \mathbb{N}$ be such that $V_i' \cap V_j' \neq \emptyset$. Then*

$$\|\psi_j^{-1} \circ \psi_i - A_{ij}\|_{C^m(U_{ij})} < C_m\delta \tag{4.29}$$

for all $m \ge 0$.

Proof. Note that q_i and q_j are neighbors by Lemma 4.15(3). By Lemma 4.9 it follows that $A_{ij}(D_i^{1/10}) \subset D_j$. Consider the map $G\colon D_i^{1/10} \to E$ defined by $G = F \circ A_{ij}|_{D_i^{1/10}}$. By Lemma 4.9 we have

$$\|G - F\|_{C^m(D_i^{1/10})} < C_m\delta. \tag{4.30}$$

This and Lemma 4.14(1) imply that the image of G is contained in the domain of P_M, so we can consider a map $\widetilde{\psi}_i\colon D_i^{1/10} \to M$ defined by $\widetilde{\psi}_i = P_M \circ G$. The relations (4.30) and (4.23) imply that

$$\|\widetilde{\psi}_i - \psi_i\|_{C^m(D_i^{1/10})} < C_m\delta.$$

If δ is sufficiently small, this and Lemma 4.15(1) imply that $\widetilde{\psi}_i$ is a diffeomorphism onto its image, the image of $\widetilde{\psi}_i$ contains V_i', and the composition $\widetilde{\psi}_i^{-1} \circ \psi_i$ is $C\delta$-close to the identity, more precisely,

$$\|\widetilde{\psi}_i^{-1} \circ \psi_i - \mathrm{id}\|_{C^m(D_i^{1/15})} < C_m\delta. \tag{4.31}$$

Let us show that $A_{ij}(U_{ij}) \subset D_j^{1/10}$. Let $x \in U_{ij}$ and $z = A_{ij}(x)$. Then $|F(x) - F(z)| < C\delta$ by Lemma 4.9. Let $y \in U_{ji}$ be such that $\psi_j(y) = \psi_i(x)$. Then $|F(x) - F(y)| < C\delta$ by Lemma 4.16. Therefore $|F(y) - F(z)| < C\delta$. Since $F|_{D_j}$ is uniformly bi-Lipschitz by Lemma 4.8(2), it follows that

$$|y - z| < C|F(y) - F(z)| < C\delta < \tfrac{1}{10} - \tfrac{1}{15},$$

if δ is sufficiently small. Since $y \in U_{ji} \subset D_j^{1/15}$, this implies that $z \in D_j^{1/10}$.

Thus we have shown that $A_{ij}(U_{ij}) \subset D_j^{1/10}$. This implies that

$$\widetilde{\psi}_i|_{U_{ij}} = P_M \circ F \circ A_{ij}|_{U_{ij}} = \psi_j \circ A_{ij}|_{U_{ij}}$$

and therefore

$$\widetilde{\psi}_i^{-1}|_{V_i' \cap V_j'} = A_{ij}^{-1} \circ \psi_j^{-1}|_{V_i' \cap V_j'}.$$

Now (4.31) implies that

$$\|A_{ij}^{-1} \circ \psi_j^{-1} \circ \psi_i - \mathrm{id}\|_{C^m(U_{ij})} < C_m\delta.$$

and (4.29) follows. □

4.4. **Riemannian metric and quasi-isometry.** Now we are going to equip M with a Riemannian metric g such that the resulting Riemannian manifold (M, g) satisfies the assertions of Proposition 4.1. (The metric induced from E is not suitable for this purpose. One of the reasons is that its curvature is bounded by C but not by $C\delta$.)

First we observe that there exists a smooth partition of unity $\{u_i\}$ on M subordinate to the covering $\{V_i'\}$ and such that

$$\|u_j \circ \psi_i\|_{C^m(D_i^{1/15})} < C_m \tag{4.32}$$

for all $i, j \in \mathbb{N}$ and all $m \geq 0$. To construct such a partition of unity, fix a smooth function $h\colon \mathbb{R}^n \to \mathbb{R}_+$ which equals 1 within the ball $B_{1/30}(0)$ and 0 outside the ball $B_{1/15}(0)$. Then define $\widetilde{u}_i\colon M \to \mathbb{R}_+$ by

$$\widetilde{u}_i(x) = \begin{cases} h(\psi_i^{-1}(x) - p_i), & \text{if } x \in V_i' \\ 0, & \text{otherwise.} \end{cases}$$

Finally, let $u = \sum_i \widetilde{u}_i$ and $u_i = \widetilde{u}_i / u$. Lemma 4.17 implies that

$$\|\widetilde{u}_j \circ \psi_i\|_{C^m(D_i^{1/15})} < C_m$$

for all $i, j \in \mathbb{N}$ and all $m \geq 0$. Since the sets V_i' cover M with bounded multiplicity, it follows from Lemma 4.15(2) that a similar estimate holds for $u \circ \psi_i$ and (4.32) follows.

For every $i \in \mathbb{N}$, define a Riemannian metric g_i on V_i by $g_i = (\psi_i^{-1})^* g_E$ where g_E is the standard Euclidean metric in $D_i^{1/10} \subset \mathbb{R}^n$ and the star denotes the pull-back of the metric by a map. In the other words, g_i is the unique Riemannian metric on V_i such that ψ_i is an isometry between $D_i^{1/10}$ and (V_i, g_i). Then Lemma 4.17 implies that

$$\|\psi_j^* g_i - g_E\|_{C^m(U_{ij})} < C_m \delta \tag{4.33}$$

for all $m \geq 0$ and $i, j \in \mathbb{N}$ such that $V_i' \cap V_j' \neq \emptyset$. Define a metric g on M by $g = \sum_i u_i g_i$. The pull-back $\psi_j^* g$ of this metric by a coordinate map ψ_j has the form

$$\psi_j^* g = \sum\nolimits_i (u_i \circ \psi_j) \cdot \psi_j^* g_i. \tag{4.34}$$

By (4.32) and (4.33) it follows that

$$\|\psi_j^* g - g_E\|_{C^m(D_j^{1/15})} < C_m \delta. \tag{4.35}$$

So in the local coordinates defined by ψ_j on V_j' the metric tensor is $C\delta$-close to the Euclidean one and its derivatives up to the second order are bounded by $C\delta$. So are the sectional curvatures of the metric. Thus (M, g) satisfies the second assertion of Proposition 4.1.

Let $d_g\colon M \times M \to \mathbb{R}_+$ be the distance induced by g. The estimate (4.35) implies that the coordinate maps ψ_i are almost isometries between the Euclidean metric on $D_i^{1/15}$ and the metric g on V_i'. More precisely, ψ_i distorts the lengths of tangent vectors by a factor of at most $1 + C\delta$. Therefore

$$(1 + C\delta)^{-1} < \frac{d_g(\psi_i(x), \psi_i(y))}{|x - y|} < 1 + C\delta, \tag{4.36}$$

for all $x, y \in D_i^{1/30}$. (The ball $D_i^{1/30}$ here is twice smaller than the domain where ψ_i is almost isometric. This adjustment is needed because the d_g-distance between points in V_i' can be realized by paths that leave V_i'.)

Now we construct a $(1 + C\delta, C\delta)$-quasi-isometry $\Psi\colon X \to M$. Recall that $X_0 = \{q_i\}_{i=1}^{\infty}$ is a $\frac{1}{100}$-net in our original metric space X and for each $i \in \mathbb{N}$ we have a 2δ-isometry $f_i\colon B_1(q_i) \to D_i$ such that $f_i(q_i) = p_i$. We construct $\Psi\colon X \to M$ as follows. For every $x \in X$, pick a point $q_j \in X_0$ such that $d_X(x, q_j) \le \frac{1}{100}$ and define $\Psi(x) = \psi_j(f_j(x))$. The next lemma shows that the choice of q_j does not make much difference.

Lemma 4.18. *Let $x \in X$ and $q_i \in X_0$ be such that $d_X(x, q_i) < \frac{1}{20}$. Then $f_i(x) \in D_i^{1/15}$ and*

$$d_g(\Psi(x), \psi_i(f_i(x))) < C\delta. \tag{4.37}$$

Proof. Let q_j be the point of X_0 chosen for x in the construction of Ψ. Then $d_X(x, q_j) \le \frac{1}{100}$ and $\Psi(x) = \psi_j(f_j(x))$. By the triangle inequality,

$$d_X(q_i, q_j) < \frac{1}{20} + \frac{1}{100} < \frac{1}{2},$$

hence q_i and q_j are neighbors. Observe that $|f_i(x) - p_i| < \frac{1}{20} + C\delta$ since $p_i = f_i(q_i)$ and f_i is a 2δ-isometry. Similarly, $|f_j(x) - p_j| < \frac{1}{100} + C\delta$. Hence $f_i(x) \in D_i^{1/15}$ and $f_j(x) \in D_j^{1/50}$. By (4.1), the point $f_j(x)$ is $C\delta$-close to $A_{ij}(f_i(x))$, hence $\Psi(x)$ is $C\delta$-close to $\psi_j(A_{ij}(f_i(x)))$. By Lemma 4.9 (for $m = 0$) and Lipschitz continuity of P_M, the latter is $C\delta$-close to $\psi_i(f_i(x))$. Thus

$$|\Psi(x) - \psi_i(f_i(x))| < C\delta.$$

This implies that

$$d_M(\Psi(x), \psi_i(f_i(x))) < C\delta \tag{4.38}$$

where d_M is the intrinsic metric of M induced from E. Indeed, the points $a = \Psi(x)$ and $b = \psi_i(f_i(x))$ can be connected in M by the P_M-image of the line segment $[a, b]$, and the length of this path is bounded by $C\delta$. By construction, the metric d_g on M is bi-Lipschitz equivalent to d_M (with bi-Lipschitz constant depending only on n). Hence (4.38) implies (4.37). □

Now let us show that $\Psi(X)$ is a $C\delta$-net in (M, d_g). Let $Y = \bigcup_i \psi_i(B^{1/20}(q_i))$. It follows from Lemma 4.18 that Y is contained in a $C\delta$-neighborhood of $\Psi(X)$ in (M, d_g). Hence it suffices to prove that Y is a $C\delta$-net in (M, d_g). Since f_i is a 2δ-isometry, the set $f_i(B_{1/20}(q_i))$ is a $C\delta$-net in the ball $D_i^{1/20+C\delta}$. The ψ_i-images of these balls cover M by Lemma 4.15(2). Since each ψ_i almost preserves the metric tensor, it follows that Y, and hence $\Psi(X)$, is a $C\delta$-net in (M, d_g).

Lemma 4.19. *For all $x, y \in X$ such that $d_X(x, y) < \frac{1}{100}$ or $d_g(\Psi(x), \Psi(y)) < \frac{1}{100}$, one has*

$$|d_g(\Psi(x), \Psi(y)) - d_X(x, y)| < C\delta. \tag{4.39}$$

Proof. Let $x \in X$ and q_i be the point of X_0 chosen for x in the construction of Ψ, so that $d_X(x, q_i) \le \frac{1}{100}$. Then $\Psi(x) = \psi_i(f_i(x))$. Note that $|f_i(x) - p_i| < \frac{1}{100} + C\delta < \frac{1}{30}$ since $p_i = f_i(q_i)$ and f_i is a 2δ-isometry. (Recall the definitions in Section 4.1.)

First, we consider the case when $y \in X$ is such that $d_X(y, q_i) < \frac{3}{100}$. Since f_i is a 2δ-isometry, $|f_i(y) - p_i| < \frac{3}{100} + C\delta < \frac{1}{30}$ and the distance $|f_i(x) - f_i(y)|$ differs from $d_X(x, y)$ by at most 2δ. The above and (4.36) imply that

$$|d_g(\psi_i(f_i(x)), \psi_i(f_i(y))) - d_X(x, y)| < C\delta.$$

This and Lemma 4.18 prove (4.39) when $d_X(y, q_i) < \frac{3}{100}$.

In particular, this proves the claim of the lemma in the case when $d_X(x, y) < \frac{1}{100}$ as then by the triangle inequality we have $d_X(y, q_i) < \frac{1}{100} + \frac{1}{100} < \frac{3}{100}$.

Second, we consider the case when $y \in X$ is such that $d_g(\Psi(x), \Psi(y)) < \frac{1}{100}$. For every $r > 0$, denote by $B_i(r)$ the ball of radius r in M with respect to d_g centered at $\psi_i(p_i)$. Since ψ_i almost preserves the metric tensor, we have

$$B_i(\tfrac{1}{15} - C\delta) \subset V_i' = \psi_i(D_i^{1/15}) \subset B_i(\tfrac{1}{15} + C\delta).$$

Since $|f_i(x) - p_i| < \frac{1}{100} + C\delta$, it follows that the point $\Psi(x) = \psi_i(f_i(x))$ belongs to $B(\frac{1}{100} + C\delta)$ and hence $\Psi(y) \in B_i(\frac{1}{100} + \frac{1}{100} + C\delta) = B_i(\frac{1}{50} + C\delta) \subset V_i'$.

Let q_j be the point of X_0 chosen for y when defining Ψ that satisfies $d_X(y, q_j) \leq \frac{1}{100}$. Since $\Psi(y) \in V_i'$, the point $z := \psi_i^{-1}(\Psi(y)) = \psi_i^{-1} \circ \psi_j(f_j(y))$ is well-defined. Moreover, z lies within distance $\frac{1}{50} + C\delta$ from p_i since $\Psi(y) \in B_i(\frac{1}{50} + C\delta)$. By Lemma 4.17, z is $C\delta$-close to $A_{ij}(f_j(y))$ and the latter is $C\delta$-close to $f_i(y)$ by (4.1). Hence $|f_i(y) - p_i| < \frac{1}{50} + C\delta$. Since f_i is a 2δ-isometry, it follows that $d_X(y, q_i) < \frac{1}{50} + C\delta < \frac{3}{100}$. Thus, (4.39) follows from the first part of the proof.

□

Lemma 4.19 and the fact that $\Psi(X)$ is a $C\delta$-net in (M, d_g) imply follows that Ψ satisfies the assumptions of Lemma 2.4 with $r = \frac{1}{100}$ and $C\delta$ in place of δ. Lemma 2.4 implies that Ψ is a $(1 + C\delta, C\delta)$-quasi-isometry from X to (M, d_g) and the first claim of Proposition 4.1 follows. The second claim is already proven above. It remains to prove the third claim of Proposition 4.1. Since Ψ is a $(1+C\delta, C\delta)$-quasi-isometry, every unit ball in (M, d_g) is GH $C\delta$-close to a unit ball in X and hence in $\mathbb{R}^n$. Therefore one can apply Proposition 1.7 with $\widetilde{M} = \mathbb{R}^n$, $\rho = 1$, $K = C\delta$, and $d_{GH}(B_1^n(0), B_1^M(x)) < C\delta$, where $x \in M$. This yields that $\mathrm{inj}_M > 1 - C\delta > \frac{1}{2}$. This finishes the proof of Proposition 4.1 and the proof of Theorem 1.

Remark 4.20. The quasi-isometry parameters in Theorem 1 are optimal up to constant factors. To see this, assume that a metric space X is $(1 + \delta r^{-1}, \delta)$-quasi-isometric to an n-dimensional manifold M with $|\,\mathrm{Sec}_M| \leq \delta r^{-3}$ and $|\,\mathrm{inj}_M| \geq 2r$. Then by (1.7) the r-balls in X are GH $C\delta$-close to r-balls in M. Furthermore, by (1.1) the r-balls in M are GH $C\delta$-close to r-balls in $\mathbb{R}^n$. Hence X is $C\delta$-close to $\mathbb{R}^n$ at scale r.

Thus the assumption of Theorem 1 that X is δ-close to $\mathbb{R}^n$ at scale r is necessary, up to multiplication of the parameters by a constant factor depending on n. The assumption that X is δ-intrinsic could be weakened, but it is not really restrictive due to Lemma 2.3.

Proof of Corollary 1.4. First we prove the first inclusion in (1.10). Let X be a metric space from the class $\mathcal{M}_{\delta/6}(n, K/2, 2i_0, D - \delta)$. Then there exists a manifold $M \in \mathcal{M}(n, K/2, 2i_0, D - \delta)$ such that $d_{GH}(M, X) < \frac{\delta}{6}$. Hence every ρ-ball in X is GH $\frac{\delta}{2}$-close to a ρ-ball in M. Take $\rho = r = (\delta/K)^{1/3} < i_0$, then by (1.1) we have $d_{GH}(B_r^M(x), B_r^n) < \frac{1}{2}Kr^3 = \frac{\delta}{2}$ for every $x \in M$. Hence every r-ball in X is GH δ-close to B_r^n. Thus X is δ-close to $\mathbb{R}^n$ at scale r. Similarly X is δ_0-close to $\mathbb{R}^n$ at scale r_0. Since $d_{GH}(M, X) < \frac{\delta}{6}$ and the Riemannian manifold M is a length space, Lemma 2.2(1) implies that X is δ-intrinsic. We also have $\mathrm{diam}(X) \leq \mathrm{diam}(M) + 2d_{GH}(X, M) \leq D$. Thus $X \in \mathcal{X}$, proving the first inclusion in (1.10).

Now let us prove the second inclusion in (1.10). Let $X \in \mathcal{X}$. Recall that $\delta = Kr^3$, $\delta_0 = Ki_0^3$ and $\delta < \delta_0$. Therefore $r < i_0$ and $\delta r^{-1} < \delta_0 i_0^{-1} < \sigma_2$. If σ_2 is sufficiently small then by Theorem 1 there is a manifold M which is $(1 + C\delta r^{-1}, C\delta)$-quasi-isometric to X and has $|\,\mathrm{Sec}_M| \leq C\delta r^{-3} = CK$. Let us show that $\mathrm{inj}_M > i_0/3$. To see this, apply Theorem 1 to i_0 and δ_0 in place of r and δ. This yields a manifold M_0 which is $(1 + C\delta_0 i_0^{-1}, C\delta_0)$-quasi-isometric to X and has $|\,\mathrm{Sec}_{M_0}| \leq C\delta_0 i_0^{-3} = CK$ and $\mathrm{inj}_{M_0} > i_0/2$. Since $\delta < \delta_0$ and $\delta r^{-1} < \delta_0 i_0^{-1}$, both M and M_0 are $(1 + C\delta_0 i_0^{-1}, C\delta_0)$-quasi-isometric to X. Hence they are $(1 + C\delta_0 i_0^{-1}, C\delta_0)$-quasi-isometric to each other. This fact and Proposition 1.7(2) imply that

$$\mathrm{inj}_M \geq (1 - C\delta_0 i_0^{-1}) \min\{\mathrm{inj}_{M_0}, \tfrac{\pi}{\sqrt{CK}}\} \geq (1 - C\delta_0 i_0^{-1})\frac{i_0}{2} \geq \frac{i_0}{3}$$

provided that σ_2 is sufficiently small.

By (1.9) we have $d_{GH}(X,M) \le C\delta r^{-1}D$. Therefore $\operatorname{diam}(M) \le D(1+C\delta r^{-1})$. Let M_1 be the result of rescaling M by the factor $(1+C\delta r^{-1})^{-1}$ where C is the constant from the above diameter estimate. Then $\operatorname{diam}(M_1) \le D$ and $d_{GH}(M,M_1) \le C\delta r^{-1}D$. Hence

$$d_{GH}(X,M_1) \le d_{GH}(X,M) + d_{GH}(M,M_1) \le C\delta r^{-1}D = CDK^{1/3}\delta^{2/3}. \tag{4.40}$$

We may assume that σ_2 is so small that the above scale factor between M and M_1 is greater than $\frac{3}{4}$. Then $\operatorname{inj}_{M_1} \ge \frac{3}{4}\operatorname{inj}_M \ge i_0/4$ and therefore $M_1 \in \mathcal{M}(n, CK, i_0/4, D)$. This and (4.40) imply the second inclusion in (1.9) and Corollary 1.4 follows. □

5. Algorithms and proof of Corollary 1.8

The constructive proofs of Theorems 1 and 2 yield algorithms that can be used to produce surfaces or manifolds from finite data sets. We give only the sketches of the algorithms. The algorithms use the sub-algorithms *FindDisc* and *GHDist* given in Sections 2.3 and 2.4. In the description of the algorithm we assume that the data set X is finite.

First we outline the algorithm based on Theorem 2.

Algorithm SurfaceInterpolation: Assume that we are given the dimension n, the scale parameter r, and a finite set points $X \subset E = \mathbb{R}^N$. We suppose that X is δr-close to n-flats at scale r where δ is sufficiently small. Our aim is to construct a surface $M \subset E$ that approximates the points of X. We implement the following steps:

(1) We rescale X by the factor $1/r$. After this scaling, the problem is reduced to the case when $r = 1$.
(2) We choose a maximal $\frac{1}{100}$-separated set $X_0 \subset X$ and enumerate the point of X_0 as $\{q_i\}_{i=1}^J$. We apply the algorithm *FindDisc* to every point $q_i \in X_0$ to find an affine subspace A_i through q_i such that the unit n-disc $A_i \cap B_1(q_i)$ lies within Hausdorff distance $C\delta$ from the set $X \cap B_1(q_i)$. We construct the orthogonal projectors $P_i \colon E \to E$ onto A_i.
(3) We construct the functions $\varphi_i : E \to E$, defined in (3.5), that are convex combinations of the projector P_i and the identity map. Then we iterate these maps to construct $f \colon E \to E$, $f = \varphi_J \circ \varphi_{J-1} \circ \ldots \circ \varphi_1$, see (3.6).
(4) We construct the image $M = f(U_\delta(X))$ of the δ-neighborhood of the set X in the map f, see Remark 3.10.

The output of the algorithm SurfaceInterpolation is the n-dimensional surface $M \subset E$.

The algorithm based on Theorem 1 is the following.

Algorithm ManifoldConstruction: Assume that we are given the dimension n, the scale parameter r, and a finite metric space (X,d). Our aim is to construct a smooth n-dimensional Riemannian manifold (M,g) approximating (X,d). We implement the following steps:

(1) We multiply all distances by $1/r$. After this scaling, the problem is reduced to the case when $r = 1$.
(2) For each $x \in X$, we apply the algorithm *GHDist* to the ball $B_1(x) \subset X$ to find the value $\delta_a(x)$. Define $\delta_a = \max_{x \in X} \delta_a(x)$.

    Note that, by Lemma 2.5, the values $\delta_a(x)$ estimate the Gromov-Hausdorff distance between the ball $B_1(x)$ and B_1^n. Thus X is $2\delta_a$-close to $\mathbb{R}^n$ (see Definition 1.1). We require that δ_a is smaller than the constant $\delta_0(n)/2$ given in Proposition 4.1. If this is not valid, we stop the algorithm and give the output that the data does not satisfy the needed assumptions.

(3) We select a subset a maximal $\frac{1}{100}$-separated set $X_0 \subset X$ and enumerate the points of X_0 as $\{q_i\}_{i=1}^N$. We choose a set $\{p_i\}_{i=1}^N$ such that the unit balls $D_i = B_1^n(p_i) \subset \mathbb{R}^n$ are disjoint. For every $q_i \in X_0$, we apply the algorithm *GHDist* to find a δ_a-isometry $f_i : B_1(q_i) \to D_i$.
(4) For all $q_i, q_j \in X_0$ such that $d(q_i, q_j) < 1$, we construct the affine transition maps $A_{ij} : \mathbb{R}^n \to \mathbb{R}^n$, see Lemma 4.3 and Remark 2.6.
(5) Denote $\Omega_0 = \bigcup_{i=1}^N D_i^{1/10}$, where $D_i^{1/10} = B_{1/10}(p_i) \subset \mathbb{R}^n$, and $E = \mathbb{R}^{(n+1)N}$. We construct a Whitney embedding-type map
$$F : \Omega_0 \to E, \quad F(x) = (F_i(x))_{i=1}^N$$
where $F_i : \Omega_0 \to \mathbb{R}^{n+1}$ are given by (4.4).
(6) We construct the local patches $\Sigma_i = F(D_i^{1/10})$ and maximal σ_0-separated subsets $\{y_{i,k}\}_{k=1}^{K_i}$ of Σ_i, where σ_0 is the constant from Proposition 1.5.
(7) We apply algorithm *SurfaceInterpolation* for the points $\{y_{i,k};\ 1 \le i \le N,\ 1 \le k \le K_j\}$ to obtain a surface $M \subset E$. We construct the normal projector $P_M : U_{2/5}(M) \to M$ for the surface M. We note that in this algorithm P_M can be replaced by the map f constructed in the step 3 of the algorithm *SurfaceInterpolation*, see Remark 3.13.
(8) We construct maps $\psi_i = P_M \circ F|_{D_i^{1/10}} : D_i^{1/10} \to P_M(\Sigma_j) \subset M$.
(9) We construct metric tensors on sets $P_M(\Sigma_i) \subset M$ by pushing forward the Euclidean metric g^e on Ω_0 to the sets $P_M(\Sigma_i)$ using the maps ψ_i. Then metric g on M is constructed by using a partition of unity to compute a weighted average of the obtained metric tensors, see (4.34).

The output of the algorithm is the surface $M \subset E$ and the metric g on it. Note that the algorithm uses only the distances within r-balls in X,

Remark 5.1. The manifold M given by the algorithm *ManifoldConstruction* can be represented using local coordinate charts. The algorithm gives sets $U_i = D_i^{1/30} \subset \mathbb{R}^n$, that can be considered as local coordinate charts of M, the metric tensors $g_{jk}^{(i)}(x)$ on these charts, and the set $\mathcal{N}$ of the pairs (i,j) such that $\psi_i(U_i) \cap \psi_j(U_j) \neq \emptyset$. Moreover, the algorithm gives for all $(i,j) \in \mathcal{N}$ the relations $\{(x,x') \in U_i \times U_j : \psi_i(x) = \psi_j(x')\}$ that are the graphs of the transition functions $\eta_{ji} = \varphi_j^{-1} \circ \varphi_i$ that map $V_{ij} = \varphi_i^{-1}(\varphi_i(U_i) \cap \varphi_j(U_j))$ to $V_{ji} = \varphi_j^{-1}(\varphi_i(U_i) \cap \varphi_j(U_j))$. The collection of U_i, $g^{(i)} : U_i \to \mathbb{R}^{n\times n}$, and $\eta_{ij} : V_{ij} \to V_{ji}$, $(i,j) \in \mathcal{N}$ is a representation of the Riemannian manifold M in local coordinate charts. Using this representation we can determine the images of a geodesic $\gamma_{x_0,\xi_0}(s)$, emanating from $(x_0,\xi_0) \in TM$, on several coordinate charts U_i and determine the metric tensor in normal coordinates, [61]. Thus, for practical imaging purposes, the algorithm *ManifoldConstruction* can be continued with the following steps

(10) For given $x_0 \in M$, determine the metric tensor g in the normal coordinates given by the map $\exp_{x_0} : \{\xi \in T_{x_0}M :\ \|\xi\|_g < \rho\}$, where $\rho < \mathrm{inj}_M$.
(11) For given $x_0 \in M$ and independent vectors $\xi_1, \xi_2 \in T_{x_0}M$, visualise the properties of the metric g, e.g. the determinant of the metric in the normal coordinates, using in the set $\|s_1\xi_1 + s_2\xi_2\|_g < \rho$ the map $s = (s_1,s_2) \mapsto \det(g(\exp_{x_0}(s_1\xi_1 + s_2\xi_2)))$, that produces images of two-dimensional slices.

Finally, we prove Corollary 1.8.

Proof of Corollary 1.8. Let us consider $\widehat{\delta} < \delta_0$, where $\delta_0 = \delta_0(n,K)$ is chosen later in the proof, and $r = (\widehat{\delta}/K)^{1/3}$. Then $r < r_0$, where $r_0 = (\delta_0/K)^{1/3}$. By (1.1), the manifold N is $\widehat{\delta}$-close to $\mathbb{R}^n$ at scale $r/2$ provided that above $r_0 \leq$

$\min\{K^{-1/2}, \frac{1}{2}\operatorname{inj}_N\}$. Hence the set X with the approximate distance function $\widetilde{d}$ is $C\delta$-close to $\mathbb{R}^n$ at scale $r/2$. As in Lemma 2.3, we can replace $\widetilde{d}$ by a $C\widehat{\delta}$-intrinsic metric d' on X. This can be done with standard algorithms for finding shortest paths in graphs. By Lemma 2.4, (X, d') is $(1 + C\widehat{\delta}r^{-1}, C\widehat{\delta})$-quasi-isometric to N.

The metric space (X, d') is $C_0\widehat{\delta}$-close to $\mathbb{R}^n$ at scale $r/2$ for some absolute constant C_0. We may assume that $\delta_0 = \delta_0(n, K)$ satisfies $\delta_0 < K^{-1/2}\sigma_1^{3/2}$, where $\sigma_1 = \sigma_1(n)$ is given in Theorem 1. Then $\delta_0 < \sigma_1 r_0$.

As in the above algorithm *ManifoldConstruction*, using the given data one can construct a manifold $M = (M, g)$ which is $(1 + C\widehat{\delta}r^{-1}, C\widehat{\delta})$-quasi-isometric to X and has $|\operatorname{Sec}_M| \le C_1K$. Since both M and N are quasi-isometric to X with these parameters, they are $(1+C\widehat{\delta}r^{-1}, C\widehat{\delta})$-quasi-isometric to each other. By Proposition 1.5 it follows that there exists a bi-Lipschitz diffeomorphism between M and N with bi-Lipschitz constant $1+C\widehat{\delta}r^{-1} = 1+CK^{1/3}\widehat{\delta}^{\,2/3}$. Thus M satisfies the statements 1 and 2 of Corollary 1.8.

To verify the last statement of Corollary 1.8, assume that $\delta_0 = \delta_0(n, K)$ is chosen to be so small that $r_0 = (\delta_0/K)^{1/3} < (C_1K)^{-1/2}$. Then Proposition 1.7(2) applies to M and $\widetilde{M} = N$ with $C\widehat{\delta}$ in place of δ and C_1K in place of K. It implies that

$$\operatorname{inj}_M \ge (1 - C\widehat{\delta}r^{-1}) \min\{\operatorname{inj}_N, \pi(C_1K)^{-1/2}\}.$$

We may assume that δ_0 is so small that the term $1 - C\widehat{\delta}r^{-1} = 1 - CK^{1/3}\widehat{\delta}^{\,2/3}$ in this estimate is greater than $\frac{1}{2}$. Then the last statement of Corollary 1.8 follows. Choosing $\delta_0 = \delta_0(n, K)$ so that the above conditions for δ_0 and r_0 are satisfied, we obtain Corollary 1.8. □

6. Appendix: Curvature and injectivity radius

The main goal of this appendix is to prove Proposition 1.7. We begin with recalling some facts about Riemannian manifolds of bounded curvature and proving the estimate (1.1)

Let $M = (M, g)$ be a complete Riemannian manifold with $|\operatorname{Sec}_M| \le K$ where $K > 0$. For $p \in M$, consider the exponential map $\exp_p\colon T_pM \to M$. We restrict this map to the ball of radius $r < \pi/\sqrt{K}$ in T_pM centered at the origin. As a consequence of Rauch Comparison Theorem, $\exp_p$ is non-degenerate in this ball and we have the following estimates on its local bi-Lipschitz constants: for $y \in T_pM$ such that $|y| = r < \pi/\sqrt{K}$ and every $\xi \in T_pM \setminus \{0\}$,

$$\frac{\sin(\sqrt{K}r)}{\sqrt{K}r} \le \frac{|d_y \exp_p(\xi)|}{|\xi|} \le \frac{\sinh(\sqrt{K}r)}{\sqrt{K}r} \tag{6.1}$$

(see e.g. [61, Thm. 27 in Ch. 6] and [66, Thm. IV.2.5 and Remark IV.2.6]).

If $r \le \frac{\pi}{2\sqrt{K}}$ and $r \le \frac{1}{2}\operatorname{inj}_M(p)$ then the geodesic r-ball $B_r^M(p)$ is convex, i.e., minimizing geodesics with endpoints in this ball do not leave it (see e.g. [61, Thm. 29 in Ch. 6]). This makes the local bi-Lipschitz estimate (6.1) global:

$$\frac{\sin(\sqrt{K}r)}{\sqrt{K}r} \le \frac{d_M(\exp_p(y), \exp_p(z))}{|y - z|_{T_pM}} \le \frac{\sinh(\sqrt{K}r)}{\sqrt{K}r} \tag{6.2}$$

for all $y, z \in T_pM$ such that

$$\max\{|y|, |z|\} \le r \le \min\{\tfrac{\pi}{2\sqrt{K}}, \tfrac{1}{2}\operatorname{inj}_M(p)\}. \tag{6.3}$$

Since $\sin t \ge t - \frac{1}{6}t^3$ and $\sinh t \le t + \frac{1}{4}t^3$ for $0 \le t \le \pi/2$, (6.2) implies that

$$\big|d_M(\exp_x(y), \exp_p(z)) - |y - z|_{T_pM}\big| \le \tfrac{1}{2}Kr^3 \tag{6.4}$$

for all $y, z \in T_pM$ satisfying (6.3). This means that the restriction of $\exp_p$ to the r-ball in T_pM is a $(\frac{1}{2}Kr^3)$-isometry onto $B_r^M(p)$ and (1.1) follows.

In the sequel we will need Toponogov's Comparison Theorem (see e.g. [61, Thm. 79 in Ch. 11]), which can be formulated as follows. Let M^2_{-K} denote the rescaled hyperbolic plane of curvature $-K$. For real numbers $a, b > 0$ and $\alpha \in [0, \pi]$, denote by $\Upsilon_{-K}(a, b, \alpha)$ the length of the side x_1x_2 of a triangle $\triangle x_0x_1x_2$ in M^2_{-K} such that the sides x_0x_1 and x_0x_2 equal a and b and the angle at x_0 equals α. Note that $\Upsilon_{-K}(a, b, \alpha)$ is monotone in α: if $\alpha' > \alpha$, then $\Upsilon_{-K}(a, b, \alpha') > \Upsilon_{-K}(a, b, \alpha)$. Toponogov's Theorem asserts that, if γ_1 and γ_2 are minimizing geodesics in M connecting $p_0 \in M$ to p_1 and p_2, resp., with length$(\gamma_1) = a$, length$(\gamma_2) = b$ and $\angle(\gamma_1, \gamma_2) = \alpha$, then $d_M(p_1, p_2) \le \Upsilon_{-K}(a, b, \alpha)$.

The following lemma is the key part of the proof of Proposition 1.7.

Lemma 6.1. *There exists $\sigma_3 = \sigma_3(n) > 0$ such that the following holds. Let M and $\widetilde{M}$ be complete n-dimensional Riemannian manifolds with $|\operatorname{Sec}_M| \le K$ and $|\operatorname{Sec}_{\widetilde{M}}| \le K$, where $K > 0$, and*

$$0 < r \le \min\{\tfrac{\pi}{\sqrt{K}}, \operatorname{inj}_{\widetilde{M}}(\widetilde{x})\}.$$

Let $x \in M$, $\widetilde{x} \in \widetilde{M}$, and assume that

$$d_{GH}(B_r^M(x), B_r^{\widetilde{M}}(\widetilde{x})) < \delta \le \sigma_3 r.$$

Then $\operatorname{inj}_M(x) > r - 20\delta$.

Proof. We fix a metric on the disjoint union of the balls $B_r^M(x)$ and $B_r^{\widetilde{M}}(\widetilde{x})$ realizing the GH distance between them. We say that points $y \in B_r^M(x)$ and $\widetilde{y} \in B_r^{\widetilde{M}}(\widetilde{x})$ are *GH approximations* of each other if the distance between them in the metric on the union is less than δ. By the definition of the GH distance, every point in one ball has at least one GH approximation in the other ball. Since we are working with pointed GH distance, the centers x and $\widetilde{x}$ are GH approximations of each other.

The statement of the lemma is scale invariant so we may assume that $r = \pi$ and hence $K \le 1$. Let $r_0 = \operatorname{inj}_M(x)$ and suppose that

$$20\delta \le r_0 \le r - 20\delta. \tag{6.5}$$

Since $r_0 < r \le \frac{\pi}{\sqrt{K}}$, Klingenberg's Lemma (see e.g. [61, Lemma 16 in Ch. 5]) implies that there exists a geodesic loop γ of length $2r_0$ in M starting and ending at x. Let y be the midpoint of this loop and γ_1, γ_2 the two halves of γ between x and y. Note that γ_1 and γ_2 are minimizing geodesics and $d_M(x, y) = r_0$.

Let $\widetilde{y} \in B_r^{\widetilde{M}}(\widetilde{x})$ be a GH approximation of y. Then

$$d_{\widetilde{M}}(\widetilde{x}, \widetilde{y}) < d_M(x, y) + 2\delta < r - 18\delta.$$

Since $\operatorname{inj}_{\widetilde{M}}(\widetilde{x}) > r$, it follows that there is a point $\widetilde{z} \in B_r^{\widetilde{M}}(\widetilde{x})$ such that $d_{\widetilde{M}}(\widetilde{y}, \widetilde{z}) = 18\delta$ and $\widetilde{y}$ belongs to the minimizing geodesic from $\widetilde{x}$ to $\widetilde{z}$. Let $z \in B_r^M(x)$ be a GH approximation of z and let $a = d_M(y, z)$. Since the triangle inequality in $\widetilde{M}$ turns to equality for $\widetilde{x}, \widetilde{y}, \widetilde{z}$ and f is a δ-isometry, we have

$$r_0 + a = d_M(x, y) + d_M(y, z) < d_M(x, z) + 6\delta,$$

or, equivalently

$$d_M(x, z) > r_0 + a - 6\delta. \tag{6.6}$$

Also note that

$$|a - 18\delta| = |d_M(y, z) - d_{\widetilde{M}}(\widetilde{y}, \widetilde{z})| < 2\delta. \tag{6.7}$$

Let γ_3 be a minimizing geodesic between y and z. Consider the angles $\angle(\gamma_3, \gamma_1)$ and $\angle(\gamma_3, \gamma_2)$ at y. Their sum equals π, hence at least one of them is no greater than $\frac{\pi}{2}$. Assuming w.l.o.g. that $\angle(\gamma_3, \gamma_1) \le \frac{\pi}{2}$, we apply Toponogov's comparison to the hinge of γ_3 and γ_1. This yields

$$d_M(x, z) \le \curlyvee_{-K}(a, r_0, \tfrac{\pi}{2}) = \curlyvee_{-1}(a, r_0, \tfrac{\pi}{2}) \tag{6.8}$$

(recall that $K = 1$). Let us show that

$$\curlyvee_{-1}(a, r_0, \tfrac{\pi}{2}) < r_0 + \tfrac{1}{2}a \tag{6.9}$$

provided that δ is sufficiently small. Since $a \le r_0$ by (6.5) and (6.7), we have

$$\curlyvee_{-1}(a, r_0, \tfrac{\pi}{2}) \le \curlyvee_{-1}(a, a, \tfrac{\pi}{2}) + r_0 - a \tag{6.10}$$

by the triangle inequality in the hyperbolic plane. By rescaling,

$$\curlyvee_{-1}(a, a, \tfrac{\pi}{2}) = a \cdot \curlyvee_{-a^2}(1, 1, \tfrac{\pi}{2}) \sim a\sqrt{2}, \qquad a \to 0.$$

The asymptotic equality here follows from the fact that the rescaled hyperbolic plane converges to $\mathbb{R}^2$ as the curvature goes to 0. Since $\sqrt{2} < \frac{3}{2}$, it follows that $\curlyvee_{-1}(a, a, \frac{\pi}{2}) < \frac{3}{2}a$ if a is sufficiently small. This and (6.10) implies (6.9).

Now (6.9) and (6.8) imply that

$$d_M(x, z) < r_0 + \tfrac{1}{2}a = r_0 + a - \tfrac{1}{2}a < r_0 + a - 8\delta$$

where the last inequality follows from (6.7). This contradicts (6.6), therefore the assumption (6.5) was false.

Thus we have either $r_0 > r - 20\delta$ or $r_0 < 20\delta$. In the former case the assertion of the proposition holds, so it remains to rule out the case when $r_0 < 20\delta$.

Suppose that the proposition is false. Then there exists a sequence $\delta_i \to 0$ and complete Riemannian manifolds M_i, $\widetilde{M}_i$ with points $x_i \in M_i$ and $\widetilde{x}_i \in \widetilde{M}_i$ satisfying the assumptions of the proposition with $r = \pi$, $K = 1$, $\delta = \delta_i$ and such that $\mathrm{inj}_{M_i}(x_i) < 20\delta_i$. Due to uniformly bounded curvature, the sequences $\{(M_i, x_i)\}$ and $\{(\widetilde{M}_i, \widetilde{x}_i)\}$ are pre-compact in the pointed GH topology, see e.g. [61, Ch. 10, Corollary 31(2)]. Passing to a subsequence if necessary we may assume that (M_i, x_i) and $(\widetilde{M}_i, \widetilde{x}_i)$ converge to pointed metric spaces (X, x) and $(\widetilde{X}, \widetilde{x})$. The limit spaces X and $\widetilde{X}$ are Alexandrov spaces of curvature $\ge -K$. See [20, Ch. 10] or [21] for basics of Alexandrov space geometry. We are going to use the fact that Alexandrov spaces with curvature bounded below are dimensionally homogeneous. Furthermore an Alexandrov space X contains an open dense subset which is a Lipschitz manifold whose dimension equals the Hausdorff dimension of X.

Since $d_{GH}(B_1^{M_i}(x_i), B_1^{\widetilde{M}_i}(\widetilde{x}_i)) < \delta_i \to 0$, the balls $B_1^X(x)$ and $B_1^{\widetilde{X}}(\widetilde{x})$ in the limit spaces are isometric. Since $\mathrm{inj}_{M_i}(x_i) \to 0$, the sequence $\{M_i\}$ collapses, therefore the dimension of X_0 is strictly less than n. This follows from e.g. [41, Theorem 0.9], see also [24] and [43, Ch. 8]. On the other hand, since $\mathrm{inj}_{\widetilde{M}_i}(\widetilde{x}_i)$ is bounded away from zero, the limit space $\widetilde{X}$ is an n-dimensional manifold, see e.g. [43, §8D]. Due to dimensional homogeneity of Alexandrov spaces, it follows that $\dim_H(B_1^{\widetilde{X}}(\widetilde{x})) = n$ and $\dim_H(B_1^X(x)) < n$. Hence these balls are not isometric, a contradiction. This finishes the proof of Lemma 6.1. □

Proof of Proposition 1.7. 1. Define $C = \max\{20, \sigma_3^{-1}\}$ where σ_3 is the constant from Lemma 6.1. Let $d = d_{GH}(B_\rho^M(x), B_\rho^{\widetilde{M}}(\widetilde{x}))$. If $d < \sigma_3\rho$ then (1.11) follows from Lemma 6.1 by setting $r = \rho$. Otherwise (1.11) holds for the trivial reason that its right-hand side is nonpositive. This proves claim (i).

2. Let $\rho = \min\{\mathrm{inj}_{\widetilde{M}}(\widetilde{x}), \frac{\pi}{\sqrt{K}}\}$. Since M and $\widetilde{M}$ are $(1 + \delta r^{-1}, \delta)$-quasi-isometric and $r \le \rho$, (1.7) implies that for every $x \in M$ there exists $\widetilde{x} \in \widetilde{M}$ such that

$$d_{GH}(B_\rho^M(x), B_\rho^{\widetilde{M}}(\widetilde{x})) \le C\delta r^{-1}\rho.$$

Hence by the first part of the proposition we have

$$\mathrm{inj}_M(x) \geq \rho - C\delta r^{-1}\rho = (1 - C\delta r^{-1})\rho.$$

Since x is an arbitrary point of M, (1.13) follows. □

Acknowledgements. The authors express their gratitude to the Mittag-Leffler Institute, the Institut Henri Poincare, and the Fields Institute, where parts of this work have been done.

Ch.F. was partly supported AFOSR, grant DMS-1265524, and NSF, grant FA9550-12-1-0425. S.I. was partly supported RFBR, grant 14-01-00062, Y.K. was partly supported by EPSRC and the AXA professorship, and M.L. was supported by Academy of Finland, grants 273979 and 284715.

References

[1] M. Anderson, *Convergence and rigidity of manifolds under Ricci curvature bounds*, Invent. Math. **102** (1990), 429–445.

[2] M. Anderson, A. Katsuda, Y. Kurylev, M. Lassas, M. Taylor, *Boundary regularity for the Ricci equation, geometric convergence, and Gel'fand's inverse boundary problem*, Invent. Math. **158** (2004), 261–321.

[3] M. Belkin, P. Niyogi, *Laplacian eigenmaps and spectral techniques for embedding and clustering*, Adv. in Neural Inform. Process. Systems, **14** (2001), 586–691.

[4] M. Belkin, P. Niyogi, *Semi-Supervised Learning on Riemannian Manifolds*, Machine Learning, **56** (2004), 209–239.

[5] M. Belkin, P. Niyogi, *Convergence of Laplacian eigenmaps*, Adv. in Neural Inform. Process. Systems **19** (2007), 129–136.

[6] E. Beretta, M. de Hoop, L. Qiu, *Lipschitz Stability of an Inverse Boundary Value Problem for a Schrödinger-Type Equation* SIAM J. Math. Anal. **45** (2012), 679-699.

[7] E. Bierstone, P. Milman, W. Paulucki, *Differentiable functions defined on closed sets. A problem of Whitney*, Invent. Math., **151** (2003), 329–352.

[8] J. Boissonnat, L. Guibas, S. Oudot, *Manifold reconstruction in arbitrary dimensions using witness complexes*, Discrete & Computational Geometry 42 (2009), 37–70.

[9] L. Borcea, V. Druskin, F. Guevara Vasquez, *Electrical impedance tomography with resistor networks* Inverse Problems **24** (2008), 035013.

[10] L. Borcea, V. Druskin, L. Knizhnerman, *On the continuum limit of a discrete inverse spectral problem on optimal finite difference grids*, Comm. Pure Appl. Math. **58** (2005), 1231-1279.

[11] M. Brand, *Charting a manifold, NIPS 15* (2002), 985–992.

[12] S. Bromberg, *An extension in the class C1,* Bol. Soc. Mat. Mex. II, Ser. **27**, (1982), 35–44.

[13] Y. Brudnyi, *On an extension theorem,* Funk. Anal. i Prilzhen. 4 (1970), 97–98; English transl. in Func. Anal. Appl. **4** (1970), 252–253.

[14] Y. Brudnyi, P. Shvartsman, *The traces of differentiable functions to closed subsets of* $\mathbb{R}^n$, in Function Spaces (1989), Teubner-Texte Math. **120**, 206–210.

[15] Y. Brudnyi, P. Shvartsman, *A linear extension operator for a space of smooth functions defined on closed subsets of* $\mathbb{R}^n$, Dokl. Akad. Nauk SSSR **280** (1985), 268–270. English transl. in Soviet Math. Dokl. **31**, No. 1 (1985), 48–51.

[16] Y. Brudnyi, P. Shvartsman, *Generalizations of Whitney's extension theorem,* Int. Math. Research Notices **3** (1994), 129–139.

[17] Y. Brudnyi, P. Shvartsman, *The traces of differentiable functions to closed subsets of* $\mathbb{R}^n$, Dokl. Akad. Nauk SSSR **289** (1985), 268–270.

[18] Y. Brudnyi, P. Shvartsman, *The Whitney problem of existence of a linear extension operator*, J. Geom. Anal. **7**(1997), 515–574.

[19] Y. Brudnyi, P. Shvartsman, *Whitney's extension problem for multivariate* $C^{1,\omega}$ *functions,* Trans. Amer. Math. Soc. **353** No. 6 (2001), 2487–2512.

[20] D. Burago, Yu. Burago, S. Ivanov, *A course in metric geometry*, Graduate Studies in Mathematics, **33**, Amer. Math. Soc., (2001).

[21] Yu. Burago, M. Gromov, G. Perelman, *A. Aleksandrov spaces with curvatures bounded below*, Russian Math. Surveys **47** (1992), 1–58.

[22] D. Burago, S. Ivanov, Y. Kurylev, *A graph discretisation of the Laplace-Beltrami operator*, J. Spectr. Theory **4** (2014), 675–714.

[23] J. Cheeger, *Finiteness theorems for Riemannian manifolds.* Am. J. Math. **92** (1970), 61–75.

[24] J. Cheeger, K. Fukaya, M. Gromov, *Nilpotent structures and invariant metrics on collapsed manifolds*, J. Amer. Math. Soc. **5** 2 (1992), 327–372.
[25] S. Cheng, T. Dey, E. Ramos, *Manifold reconstruction from point samples, SODA* (2005), 1018–1027.
[26] D. Chigirev and W. Bialek, *Optimal Manifold Representation of Data: An Information Theoretic Approach*. In: Advances in Neural Information Processing Systems **16,** Ed. S. Thrun et al, The MIT press, 2004, pp. 164–168.
[27] H. Engl, M. Hanke, A. Neubauer, *Regularization of inverse problems,* Springer, 1996.
[28] Ch. Fefferman, *A sharp form of Whitney's extension theorem*, Ann. of Math. **161** (2005), 509–577.
[29] Ch. Fefferman, *Whitney's extension problem for* C^m, Ann. of Math. **164** (2006), 313–359.
[30] Ch. Fefferman, C^m*-extension by linear operators*, Ann. of Math. **166** (2007), 779–835.
[31] Ch. Fefferman, *Whitney's extension problem in certain function spaces,* (preprint).
[32] Ch. Fefferman, *A generalized sharp Whitney theorem for jets, Rev. Mat. Iberoam.* **21,** No. 2 (2005), 577–688.
[33] Ch. Fefferman, Extension of $C^{m,\omega}$ *smooth functions by linear operators,* Rev. Mat. Iberoam. **25,** No. 1 (2009), 1–48.
[34] Ch. Fefferman, S. Mitter, H. Narayanan, *Testing the manifold hypothesis*, arXiv: 1310.0425v2.
[35] Ch. Fefferman, B. Klartag, *Fitting* C^m*-smooth function to data I*, Ann. of Math, **169** (2009), 315–346.
[36] Ch. Fefferman, B. Klartag, *Fitting* C^m*-smooth function to data II*, Rev. Mat. Iberoam. **25** (2009), 49–273.
[37] R. Coifman, et al. *Geometric diffusions as a tool for harmonic analysis and structure definition of data* Part II: Multiscale methods. Proc. of Nat. Acad. Sci. **102** (2005), 7432–7438.
[38] T. Cox, M. and Cox *Multidimensional Scaling.* Chapman & Hall, London, (1994).
[39] S. Dasgupta, Y. Freund, *Random projection trees and low dimensional manifolds.* In Proc. the 40th ACM symposium on Theory of computing (2008), STOC '08 537–546.
[40] D. Donoho, D. Grimes, *Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data,* Proceedings of the National Academy of Sciences, **100**, 5591–5596.
[41] K. Fukaya, *A boundary of the set of the Riemannian manifolds with bounded curvatures and diameters* J. Differential Geom. **28** (1988), No. 1, 1–21.
[42] G. Glaeser, *Etudes de quelques algebres Tayloriennes*, J. d'Analyse **6** (1958), 1–124.
[43] M. Gromov with appendices by M. Katz, P. Pansu, and S. Semmes, *Metric Structures for Riemannian and Non-Riemanian Spaces.* Birkhauser (1999).
[44] H. Hotelling, Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* **24** (1933), 417–441, 498–520.
[45] E. Iversen, M. Tygel, B. Ursin, and M. V. de Hoop, *Kinematic time migration and demigration of re ections in pre-stack seismic data*, Geophys. J. Int. **189** (2012), 1635–1666.
[46] P. Jones, M., Maggioni, R. Schul, *Universal local parametrizations via heat kernels and eigenfunctions of the laplacian,* Ann. Acad. Scient. Fen. **35** (2010), 1–44.
[47] A. Katchalov, Y. Kurylev, M. Lassas, N. Mandache. *Equivalence of time-domain inverse problems and boundary spectral problem.* Inverse problems **20** (2004), 419–436.
[48] A. Katchalov, Y. Kurylev, M. Lassas: *Inverse Boundary Spectral Problems*, Monographs and Surveys in Pure and Applied Mathematics **123**, CRC-press, 2001, xi+290 pp.
[49] A. Katsuda, Y. Kurylev, M. Lassas, *Stability and Reconstruction in Gel'fand Inverse Boundary Spectral Problem*, in: New analytic and geometric methods in inverse problems. (Ed. K. Bingham, Y. Kurylev, and E. Somersalo), 309–320, Springer-Verlag, 2003.
[50] M. Lassas, G. Uhlmann, *Determining Riemannian manifold from boundary measurements,* Ann. Sci. École Norm. Sup. **34** (2001), 771–787.
[51] J. Lee, G. Uhlmann, *Determining anisotropic real-analytic conductivities by boundary measurements,* Comm. Pure Appl. Math. **42** (1989), 1097–1112.
[52] L. Ma, M. Crawford, J. W. Tian, *Generalised supervised local tangent space alignment for hyperspectral image classification*, Electronics Letters **46** (2010), 497.
[53] J. Mueller, S. Siltanen, *Linear and nonlinear inverse problems with practical applications.* SIAM, Philadelphia, 2012. xiv+351 pp.
[54] J. Nash, C^1*-isometric imbeddings*, Ann. of Math. **60** (1954), 383–396.
[55] J. Nash, *The imbedding problem for Riemannian manifolds*, Ann. of Math. **63** (1956), 20–63.
[56] L. Oksanen, *Solving an inverse problem for the wave equation by using a minimization algorithm and time-reversed measurements.* Inv. Prob. and Imag. **5** (2011), 731–744.
[57] G. Paternain, M. Salo, G. Uhlmann *Tensor Tomography on Simple Surfaces,* Inventiones Math. **193** (2013), 229–247.

[58] L. Pestov, G. Uhlmann, *Two Dimensional Compact Simple Riemannian manifolds are Boundary Distance Rigid,* Ann. of Math. **161** (2005), 1089–1106.

[59] K. Pearson, *On lines and planes of closest fit to systems of points in space,* Philosophical Magazine **2** (1901), 559–572.

[60] S. Peters, *Cheeger's finiteness theorem for diffeomorphism classes of Riemannian manifolds*, J. Reine Angew. Math. **349** (1984), 77–82.

[61] P. Petersen, *Riemannian geometry.* 2nd Ed. Springer, (2006), xvi+401 pp.

[62] D. O'Regan, Y. Cho, Y. Chen, *Topological Degree Theory and Applications*, CRC, (2006).

[63] G. Rosman, M. M. Bronstein, A. M. Bronstein, R. Kimmel, *Nonlinear Dimensionality Reduction by Topologically Constrained Isometric Embedding*, International Journal of Computer Vision, **89** (2010), 56–68.

[64] S. Roweis, L. Saul, *Nonlinear dimensionality reduction by locally linear embedding*, Science, **290** (2000), 2323–326.

[65] S. Roweis, L. Saul, G. Hinton, *Global coordination of local linear models,* Advances in Neural Information Processing Systems **14** (2001) 889–896.

[66] T. Sakai, *Riemannian geometry.* AMS, (1996), xiv+358 pp.

[67] J. Shawe-Taylor, N. Christianini, *Kernel Methods for Pattern Analysis,* Cambridge University Press, (2004).

[68] K. Shiohama, *An introduction to the geometry of Alexandrov spaces.* Lecture Notes Series, 8. Seoul National University, Seoul, (1993), ii+78 pp.

[69] P. Shvartsman, *Lipschitz selections of multivalued mappings and traces of the Zygmund class of functions to an arbitrary compact,* Dokl. Acad. Nauk SSSR **276** (1984), 559–562; English transl. in Soviet Math. Dokl. **29** (1984), 565–568.

[70] P. Shvartsman, *On traces of functions of Zygmund classes,* Sibirskyi Mathem. J. **28** (1987), 203–215; English transl. in Siberian Math. J. **28** (1987), 853–863.

[71] P. Shvartsman, *Lipschitz selections of set-valued functions and Helly's theorem,* J. Geom. Anal. **12** (2002), 289–324.

[72] J. Sylvester, G. Uhlmann, *A global uniqueness theorem for an inverse boundary value problem*, Ann. of Math. **125** (1987), 153–169.

[73] J. Tenenbaum, V. de Silva, J. Langford, *A global geometric framework for nonlinear dimensionality reduction*, Science, **290** 5500 (2000), 2319–2323.

[74] G. Uhlmann, A. Vasy, *The Inverse Problem for the Local Geodesic X-ray Transform*, Advances in Math., to appear (2015).

[75] K. Weinberger, L. Saul, *Unsupervised learning of image manifolds by semidefinite programming,* Int. J. Comput. Vision **70**, 1 (2006), 77–90.

[76] H. Whitney, *Analytic extensions of differentiable functions defined on closed sets*, Trans. Amer. Math. Soc., **36** (1934), 63–89.

[77] H. Whitney, *Differentiable functions defined in closed sets I,* Trans. Amer. Math. Soc. **36** (1934), 369–389.

[78] H. Whitney, *Functions differentiable on the boundaries of regions,* Ann. of Math. **35** (1934), 482–485.

[79] H. Whitney, *Differentiable manifolds*, Ann. of Math. **37** (1936), 645–680.

[80] H. Whitney, J. Eells, D. Toledo, eds., *The collected papers of Hassler Whitney.* Volumes I-II., Contemporary Mathematicians, Birkhäuser, (1992).

[81] Z. Zhang and H. Zha, *Principal manifolds and nonlinear dimension reduction via local tangent space alignment*, SIAM J. Sci. Computing, **26** (2005), 313–338.

[82] N. Zobin, *Whitney's problem on extendability of functions and an intrinsic metric*, Advances in Math. **133** (1998), 96–132.

[83] N. Zobin, *Extension of smooth functions from finitely connected planar domains*, J. Geom. Anal. **9** (1999), 489–509.

Charles Fefferman, Princeton University, Mathematics Department, Fine Hall, Washington Road, Princeton NJ, 08544-1000, USA.

Sergei Ivanov, Steklov Institute of Mathematics, Russian Academy of Sciences, 27 Fontanka, 191023 St.Petersburg, Russia.

Yaroslav Kurylev, University College London, Department of Mathematics, Gower Street WC1E 6BT, London, UK.

Matti Lassas, University of Helsinki, Department of Mathematics and Statistics, P.O. Box 68, 00014, Helsinki, Finland.

Hariharan Narayanan, University of Washington, Department of Mathematics, Box 354350, Seattle, WA 98195-4350, USA.