RECONSTRUCTING FUNCTIONS FROM RANDOM SAMPLES

Steve Ferry

Department of Mathematics
Rutgers University
Piscataway, NJ 08854, USA

Konstantin Mischaikow

Department of Mathematics and BioMaPS Institute
Rutgers University
Piscataway, NJ 08854, USA

Vidit Nanda

Department of Mathematics
University of Pennsylvania
Philadelphia, PA 19104, USA

(Communicated by Marian Mrozek)

Abstract. From a sufficiently large point sample lying on a compact Riemannian submanifold of Euclidean space, one can construct a simplicial complex which is homotopy-equivalent to that manifold with high confidence. We describe a corresponding result for a Lipschitz-continuous function between two such manifolds. That is, we outline the construction of a simplicial map which recovers the induced maps on homotopy and homology groups with high confidence using only finite sampled data from the domain and range, as well as knowledge of the image of every point sampled from the domain. We provide explicit bounds on the size of the point samples required for such reconstruction in terms of intrinsic properties of the domain, the co-domain and the function. This reconstruction is robust to certain types of bounded sampling and evaluation noise.

1. **Introduction.** The use of algebraic topological methods for the analysis of nonlinear data has become a subject of considerable interest with a wide variety of promising applications [3, 5, 7, 10]. With modern technology, large and high-dimensional datasets are easily collected. However, in many cases the data are generated by a nonlinear system with many fewer degrees of freedom than the ambient dimension, and thus one may expect that the data actually lie on a much lower dimensional manifold. In this setting, it seems reasonable that the geometry generated by the data may provide insight concerning the original system. Given that such data are typically finite and noisy, relatively crude invariants – such as homology or homotopy groups – appear to be natural tools for capturing aspects of their underlying topological structure. The same heuristic arguments extend to

transformations from one dataset to another. In such analysis, one is interested not only in the geometry associated with the data, but also in the action of an unknown nonlinear process on the data. This action may be partially characterized by the maps which it induces on homology and homotopy groups.

Given these arguments, an obvious but mostly unresolved mathematical issue is to quantify the extent to which we can extract correct topological features from noisy data. An important first step in this direction is the following result due to Partha Niyogi, Steve Smale, and Shmuel Weinberger [13]. Let $\mathbb{R}^n$ denote Euclidean n-dimensional space with the standard metric, and consider a compact k-dimensional Riemannian submanifold $\mathcal{X} \subset \mathbb{R}^n$. The *condition number* $1/\tau_{\mathcal{X}}$ of $\mathcal{X}$ is defined as follows: $\tau_{\mathcal{X}}$ is the largest positive real number such that for any $r \in (0, \tau_{\mathcal{X}})$, the normal bundle of radius r about $\mathcal{X}$ can be embedded in $\mathbb{R}^n$.

Theorem 1.1. ([13, Thm 3.1]) *Let $\mathcal{X}$ be a compact k-dimensional Riemannian submanifold of $\mathbb{R}^n$ with condition number $1/\tau_{\mathcal{X}}$. Given*

1. *some probability parameter $\delta \in (0, 1]$,*
2. *a radius $\epsilon < \tau_{\mathcal{X}}/2$, and*
3. *a finite set $X \subset \mathcal{X}$ of independent and identically distributed (i.i.d.) uniformly sampled points,*

let $\mathcal{U}(X)$ denote the union of n-dimensional open ϵ-balls centered at the points in X. If the sample size $\#X$ is larger than a bounding value $\beta_{\mathcal{X}}(\epsilon, \delta)$ (see Definition 3.1), then $\mathcal{U}_\epsilon(X)$ is homotopy-equivalent to $\mathcal{X}$ with probability exceeding $(1 - \delta)$.

Thus, the union of balls of a suitably chosen radius around a sufficiently large point sample suffices to recover the homotopy type of that manifold with high confidence. From a computational perspective, recall that the *nerve* of a cover [12, 15] is the abstract simplicial complex where each d-dimensional simplex corresponds to an intersection of $d + 1$ sets of that cover. If we let $N(X)$ denote the nerve corresponding to the cover of $\mathcal{U}(X)$ by its constituent open balls, then one obtains an isomorphism $H_*(\mathcal{X}) \simeq H_*^\Delta(N(X))$ between the singular homology of $\mathcal{X}$ and the simplicial homology of $N(X)$ by using the *nerve lemma*. Thus, it is possible to successfully compute the homology of an unknown manifold $\mathcal{X}$ with high probability from a sufficiently dense point sample $X \subset \mathcal{X}$.

An obvious next step is to obtain bounds on the probability of reconstructing, up to homotopy, a continuous function between Riemannian manifolds from images of dense samples. This is the focus of our work with the main result as follows.

Theorem 1.2. *Let $\mathcal{X} \subset \mathbb{R}^n$ and $\mathcal{Y} \subset \mathbb{R}^m$ be compact Riemannian submanifolds with condition numbers $1/\tau_{\mathcal{X}}$ and $1/\tau_{\mathcal{Y}}$ respectively and let $f : \mathcal{X} \to \mathcal{Y}$ be a Lipschitz continuous function with Lipschitz constant bounded above by some $\kappa \geq 0$. Given*

1. *probability parameters $\delta_{\mathcal{X}}, \delta_{\mathcal{Y}} \in (0, 1]$,*
2. *radii $\epsilon_{\mathcal{X}} < \tau_{\mathcal{X}}/2$ and $\epsilon_{\mathcal{Y}} < \tau_{\mathcal{Y}}/2$ satisfying $4\kappa \cdot \epsilon_{\mathcal{X}} < \epsilon_{\mathcal{Y}}$, and*
3. *finite sets $X \subset \mathcal{X}$ and $Y \subset \mathcal{Y}$ of independent and identically distributed (i.i.d.) uniformly sampled points,*

let $N(X)$ and $N(Y)$ be nerves of the covers generated by open balls of radius $\epsilon_{\mathcal{X}}$ and $\epsilon_{\mathcal{Y}}$ around X and Y respectively. If $\#X > \beta_{\mathcal{X}}(\epsilon_{\mathcal{X}}, \delta_{\mathcal{X}})$ and $\#Y > \beta_{\mathcal{Y}}(\epsilon_{\mathcal{Y}}, \delta_{\mathcal{Y}})$, then there exists a simplicial map $\phi : N(X) \to N(Y)$ which

1. *recovers the homotopy class of f with probability exceeding $(1 - \delta_{\mathcal{X}})(1 - \delta_{\mathcal{Y}})$, and*

2. *can be explicitly constructed using only* X, Y, $\epsilon_{\mathcal{X}}$, $\epsilon_{\mathcal{Y}}$, κ *and the restriction* $f|_X$.

In particular, the morphisms induced by ϕ on the simplicial homology as well as homotopy groups of the nerves faithfully capture their singular counterparts induced by f with high confidence. Note from Theorem 1.1 that the upper bound on the probability of failing to produce ϕ is no larger than the corresponding bound on failing to reconstruct both $\mathcal{X}$ and $\mathcal{Y}$ from the sample sets X and Y.

The third hypothesis in Theorem 1.1 which requires the sampled points to lie on the underlying manifold is too strong from the perspective of practical sampling considerations. A more reasonable hypothesis is that the data is *noisy* and lies near – rather than on – the respective underlying manifolds. This situation is also considered in [13]. Using their framework, we show that Theorem 1.2 can be extended to this more general setting (see Theorem 5.6).

The rest of the paper is organized as follows. In Section 2 we mention relevant definitions and tools from combinatorial algebraic topology. Section 3 describes results from [13] that are used in our work. The simplicial reconstruction of functions and its verification are presented in Section 4. Section 5 demonstrates the robustness of this reconstruction by providing a version of Theorem 1.2 which holds in the case of bounded sampling and evaluation noise. Our argument relies on a controlled version of the nerve lemma whose proof is described in Section 6.

2. **Carriers and nerves.** For the sake of completeness and to introduce relevant notation, we review some classical results from the theory of simplicial complexes. A much more complete treatment is available in standard texts, for instance [12, 15].

Let $\mathbb{U}$ be any finite set whose elements we call *vertices*. A *simplicial complex* K with vertex set $\mathbb{U}$ is a collection of nonempty subsets of $\mathbb{U}$ – called *simplices* – which contains all the vertices is closed under inclusion. More precisely, a collection K of subsets of $\mathbb{U}$ is a simplicial complex if

- for each $u \in \mathbb{U}$, we have $\{u\} \in K$, and
- if $\sigma \in K$ and $\tau \subset \sigma$, then $\tau \in K$.

The *dimension* of each simplex $\sigma \in K$ is the natural number given by $\dim \sigma = \#\sigma - 1$ where $\#$ indicates cardinality. A *subcomplex* of K is a sub-collection of simplices which forms a simplicial complex in its own right. The notation K^d is used to indicate all simplices of dimension d in K. Without loss of generality, we may identify K^0 with $\mathbb{U}$ by associating each 0-dimensional simplex with the unique element of $\mathbb{U}$ which it contains. Thus, each simplex is uniquely determined by its constituent vertices. Given simplicial complexes K and L, a *simplicial map* $\psi : K \to L$ associates to each vertex $v \in K^0$ a vertex $\psi(u) \in L^0$ so that for each simplex $\sigma \in K$ the image $\psi(\sigma)$ is a simplex in L.

2.1. **Carriers.** The *geometric realization* $|K|$ of a simplicial complex K is defined (in [15, Ch. 3.1], for instance) as the space of all maps $\alpha : K^0 \to [0, 1]$ called *barycentric functions* such that for each $\alpha \in |K|$,

1. there exists $\sigma \in K$ with $\{u \in K^0 \mid \alpha(u) \neq 0\} = \sigma$, and
2. the sum $\sum_{u \in K^0} \alpha(u)$ equals 1.

The realization of a simplex $\sigma \in K$ is defined as the closed subset $|\sigma| \subset |K|$ consisting of all barycentric functions $\alpha \in |K|$ such that $\alpha(u) = 0$ whenever $u \notin \sigma$. Observe that if $\sigma \subset \tau$ then $|\sigma| \subset |\tau|$; if $\sigma \in K$ then $|\sigma|$ is contractible in $|K|$; and $|K| = \bigcup_{\sigma \in K} |\sigma|$.

A simplicial map $\psi : K \to L$ induces a continuous function $|\psi| : |K| \to |L|$ between geometric realizations defined as follows. For any $\alpha \in |K|$, the action of the barycentric function $|\psi|(\alpha) \in |L|$ on a vertex $v \in L^0$ is given by

$$|\psi|(\alpha)(v) = \sum_{\psi(u)=v} \alpha(u).$$

It is readily seen from this definition that $|\psi|(|\sigma|) \subset |\psi(\sigma)|$ for each simplicial map $\psi : K \to L$ and each $\sigma \in K$.

Let $\mathcal{U}$ be any topological space and K a simplicial complex. A *contractible carrier* $C : K \Longrightarrow \mathcal{U}$ from K to $\mathcal{U}$ assigns to each simplex $\sigma \in K$ a contractible subset $C(\sigma)$ of $\mathcal{U}$ so that $C(\sigma) \subset C(\tau)$ whenever $\sigma \subset \tau$. A function $h : |K| \to \mathcal{U}$ is *carried* by C if $g(|\sigma|) \subset C(\sigma)$ for each simplex $\sigma \in K$. The following result is an extremely useful tool in combinatorial algebraic topology. We refer the reader to [1] and the references therein for details.

Lemma 2.1 (Carrier Lemma). *Let K be a simplicial complex, $\mathcal{U}$ a topological space and $C : K \Longrightarrow \mathcal{U}$ a contractible carrier. Then, there exists a continuous function from $|K|$ to $\mathcal{U}$ carried by C. Moreover, if two continuous functions $h, h' : |K| \to \mathcal{U}$ are carried by C, then*

1. *they are homotopic, i.e., $h \sim h'$, and*
2. *a homotopy $\Theta : |K| \times [0,1] \to S$ may be chosen so that for each $t \in [0,1]$ the section $\Theta(*, t) : |K| \to \mathcal{U}$ is also carried by C.*

2.2. **Nerves.** Let $\mathcal{U}$ be a topological space equipped with a finite cover $\mathbb{U}$ consisting of subsets of M. The *nerve* of $\mathbb{U}$ is the simplicial complex $N(\mathbb{U})$ with vertex set $\mathbb{U}$ where each subcollection $\sigma \subset \mathbb{U}$ constitutes a simplex if and only if the intersection $\bigcap_{u \in \sigma} u$ is a non-empty subset of $\mathcal{U}$. We call this intersection the *support* of σ and denote it by $\lfloor \sigma \rfloor \subset \mathcal{U}$. On the other hand, we also make use of the union $\lceil \sigma \rceil = \bigcup_{u \in \sigma} u \subset \mathcal{U}$. When considering an entire subcomplex $K \subset N(\mathbb{U})$ rather than a single simplex, one defines $\lceil K \rceil$ to be the union of supports of all simplices in K. It is easy to see[1] that $\lceil K \rceil = \bigcup_{u \in K^0} u$, and that $\lceil N(\mathbb{U}) \rceil = \mathcal{U}$.

A cover $\mathbb{U}$ of a topological space $\mathcal{U}$ is called *contractible* if the support $\lfloor \sigma \rfloor$ of each $\sigma \in N(\mathbb{U})$ is a contractible subset of $\mathcal{U}$. For instance, if $\mathcal{U}$ lies in a topological vector space and if each $u \in \mathbb{U}$ is convex, then all non-empty intersections are automatically convex and hence contractible. One reason to consider contractible nerves is the following classical result (see [2] or [11, Thm. 15.21]).

Lemma 2.2 (Nerve Lemma). *Let $\mathcal{U}$ be a paracompact topological space equipped with an open cover $\mathbb{U}$. If $\mathbb{U}$ is contractible, then the geometric realization $|N(\mathbb{U})|$ of its nerve is homotopy-equivalent to $\mathcal{U}$.*

Since we always restrict our attention to finite covers by open balls in Euclidean space, we may obtain the following *controlled* version of the nerve lemma by strengthening its hypotheses.

Lemma 2.3. *Let $\mathbb{U}$ be a finite collection of open balls in Euclidean space and let $\mathcal{U}$ be their union. Then,*

1. *$|N(\mathbb{U})|$ is homotopy-equivalent to $\mathcal{U}$, and*
2. *a homotopy-equivalence $\zeta : |N(\mathbb{U})| \to \mathcal{U}$ may be chosen so that $\zeta(|\sigma|) \subset \lceil \sigma \rceil$ for each simplex $\sigma \in N(\mathbb{U})$.*

---

[1]Note that the supports of 0-simplices are maximal among all supports in the partial order induced by inclusion.

Note that the first conclusion of Lemma 2.3 follows easily from the traditional nerve lemma; it is the second assertion which plays a fundamental role in the proofs of our main results. A detailed verification of the controlled nerve lemma is presented in Section 6.

3. **Recovering manifolds from samples.** Our main result focuses on recovering – up to homotopy type – a Lipschitz continuous function between finitely sampled unknown compact Riemannian submanifolds of Euclidean space. In order to accomplish this, we first use the finite sampled data to construct a simplicial complex homotopically faithful to the underlying manifold. In this section, we briefly survey the process from [13] which constructs such a simplicial complex.

Throughout this section, let $\mathcal{X} \subset \mathbb{R}^n$ be a compact Riemannian submanifold with condition number $1/\tau_{\mathcal{X}}$.

Definition 3.1. The *bounding function* $\beta_{\mathcal{X}} : \mathbb{R}^+ \times (0,1] \to \mathbb{R}$ is given by
$$\beta_{\mathcal{X}}(\epsilon, \delta) = \beta_1 \left[\log(\beta_2) + \log(1/\delta) \right], \tag{1}$$
where
$$\beta_1 = \frac{\mathrm{vol}(\mathcal{X})}{\cos^k\left(\arcsin(\epsilon/8\tau_{\mathcal{X}})\right) \cdot \mathrm{vol}(\mathcal{B}^k_{\epsilon/4})},$$
$$\beta_2 = \frac{\mathrm{vol}(\mathcal{X})}{\cos^k\left(\arcsin(\epsilon/16\tau_{\mathcal{X}})\right) \cdot \mathrm{vol}(\mathcal{B}^k_{\epsilon/8})},$$
and $\mathrm{vol}(\mathcal{B}^k_\epsilon)$ denotes the volume of the standard k-dimensional Euclidean ball of radius ϵ.

Let $\mathcal{B}_r(x)$ denote the n-dimensional Euclidean open ball of radius r centered at $x \in \mathbb{R}^n$. For any subset P of $\mathbb{R}^n$ and $\alpha > 0$, we denote by $\mathbb{U}_\alpha(P)$ the set of open balls $\{\mathcal{B}_\alpha(p) \mid p \in P\}$ and let $\mathcal{U}_\alpha(P)$ be their union. We say that P is α-*dense* in the manifold $\mathcal{X}$ if we have the inclusion $\mathcal{X} \subset \mathcal{U}_\alpha(P)$. The following proposition enables one to recover the homotopy type of $\mathcal{X}$ from a finite set X which is sufficiently dense in $\mathcal{X}$ relative to $\tau_{\mathcal{X}}$.

Proposition 3.2. ([13, Prop 3.1]) *Assume* $\epsilon \in \left(0, \sqrt{3/5}\, \tau_{\mathcal{X}}\right)$ *and that a finite set* $X \subset \mathbb{R}^n$ *is* $\epsilon/2$-*dense in* $\mathcal{X}$*. Then, the canonical projection map* $\pi_{\mathcal{X}} : \mathcal{U}_\epsilon(X) \to \mathcal{X}$ *defined by*
$$\pi_{\mathcal{X}}(w) = \arg \min_{x \in \mathcal{X}} \|w - x\|_{\mathbb{R}^n} \tag{2}$$
is a strong deformation-retraction.

As a corollary to this proposition one obtains a string of isomorphisms on homology:
$$\mathrm{H}_*(\mathcal{X}) \simeq \mathrm{H}_* \left(\mathcal{U}_\epsilon(X) \right) \simeq \mathrm{H}_*(|N_\epsilon(X)|) \simeq \mathrm{H}^\Delta_*(N_\epsilon(X)).$$
The first isomorphism comes from the fact that deformation retractions preserve homotopy type and homology is a homotopy-invariant. The second isomorphism results from applying the nerve lemma: since $\mathbb{U}_\epsilon$ is a convex cover of $\mathcal{U}_\epsilon(X)$ for each ϵ, the associated nerve is contractible. The last isomorphism is simply the equivalence of singular and simplicial homology. We remark here that the nerve of a cover by balls [6] and the homology groups of finite simplicial complexes [4, 9] are eminently computable. Thus, one can actually obtain a finite representation of the homology of $\mathcal{X}$ up to isomorphism from a sufficiently dense point sample.

Adopting the terminology of the proposition, we observe an important property of the projection map $\pi_{\mathcal{X}} : \mathcal{U}_\epsilon(X) \to \mathcal{X}$ which holds even when ϵ is allowed to range over the larger interval $(0, \tau_{\mathcal{X}})$. Note that for each $w \in \mathcal{U}_\epsilon(X)$ we have some $\xi \in X$ with $\|\xi - w\|_{\mathbb{R}^n} < \epsilon$. On the other hand, the distance between w and $\pi_{\mathcal{X}}(w)$ is at most ϵ as well, since $\pi_{\mathcal{X}}(w)$ is the nearest point of the manifold $\mathcal{X}$ to w. By the triangle inequality, one has the following estimate for each $\xi \in X$ and $w \in \mathcal{B}_\epsilon(\xi)$ whenever $\epsilon_{\mathcal{X}} \in (0, \tau_{\mathcal{X}})$:

$$\|\pi_{\mathcal{X}}(w) - \xi\|_{\mathbb{R}^n} < 2\epsilon_{\mathcal{X}}. \tag{3}$$

This is not the best possible estimate one could obtain, but it suffices for our purposes here.

The following proposition assumes that X is obtained by uniform i.i.d. sampling on $\mathcal{X}$ and provides a lower bound on the sample size $\#X$ which guarantees – with high confidence – the $\epsilon/2$-density needed by the previous proposition.

Proposition 3.3. ([13, Prop 3.2]) *Choose $\epsilon \in (0, \tau_{\mathcal{X}}/2)$ and the probability parameter $\delta \in (0, 1]$. Assume that X is obtained by i.i.d. uniform samplings from $\mathcal{X}$. If $\#X > \beta_{\mathcal{X}}(\epsilon, \delta)$, then X is $\epsilon/2$-dense in $\mathcal{X}$ with probability exceeding $(1 - \delta)$.*

Propositions 3.2 and 3.3 lead directly to Theorem 1.1, which is the main result of [13].

4. **Recovering functions from samples.** In this section we provide a proof of our main result, Theorem 1.2. The hypotheses of this theorem consist of a variety of assumptions and a-priori choices of parameters. To clarify their respective roles we present them via the following exhaustive list. The notation below will remain fixed throughout this section.

Cnd: Assume that $\mathcal{X} \subset \mathbb{R}^n$ and $\mathcal{Y} \subset \mathbb{R}^m$ are compact Riemannian submanifolds with condition numbers $1/\tau_{\mathcal{X}}$ and $1/\tau_{\mathcal{Y}}$, respectively.

Prb: Choose probability parameters $\delta_{\mathcal{X}}, \delta_{\mathcal{Y}} \in (0, 1]$.

Lip: Assume $f : \mathcal{X} \to \mathcal{Y}$ is a Lipschitz continuous function with Lipschitz constant less than $\kappa \geq 0$. More precisely, we require $\|f(x) - f(x')\|_{\mathbb{R}^m} \leq \kappa\|x - x'\|_{\mathbb{R}^n}$ for any pair of points $x, x' \in \mathcal{X}$.

Rad: Choose the radii $\epsilon_{\mathcal{X}} \in (0, \tau_{\mathcal{X}}/2)$ and $\epsilon_{\mathcal{Y}} \in (0, \tau_{\mathcal{Y}}/2)$ so that $4\kappa\epsilon_{\mathcal{X}} < \epsilon_{\mathcal{Y}}$.

Smp: Assume knowledge of the finite sets $X \subset \mathcal{X}$ and $Y \subset \mathcal{Y}$ obtained by i.i.d. uniform sampling from $\mathcal{X}$ and $\mathcal{Y}$ respectively. Furthermore, require $\#X > \beta_{\mathcal{X}}(\epsilon_{\mathcal{X}}, \delta_{\mathcal{X}})$ and $\#Y > \beta_{\mathcal{Y}}(\epsilon_{\mathcal{Y}}, \delta_{\mathcal{Y}})$.

Img: Assume knowledge of the restriction $f|_X : X \to \mathcal{Y} \subset \mathbb{R}^m$ of f to the point sample X.
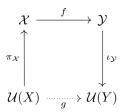
It is important to note that neither **Smp** nor **Img** imply that sampled points map to sampled points, so in general $f(X) \not\subset Y$. Since **Rad** fixes choices of $\epsilon_{\mathcal{X}}$ and $\epsilon_{\mathcal{Y}}$, we simplify our notation by declaring that $N(X) := N_{\epsilon_{\mathcal{X}}}(X)$ and $N(Y) := N_{\epsilon_{\mathcal{Y}}}(Y)$. Similarly, we denote the unions of balls $\mathcal{U}_{\epsilon_{\mathcal{X}}}(X)$ and $\mathcal{U}_{\epsilon_{\mathcal{Y}}}(Y)$ by $\mathcal{U}(X)$ and $\mathcal{U}(Y)$. Finally, define $\rho \in \mathbb{R}$ by

$$\rho = \epsilon_{\mathcal{Y}} - 2\kappa\epsilon_{\mathcal{X}}. \tag{4}$$

Note that by **Rad**, we have $\rho > \epsilon_{\mathcal{Y}}/2 > 0$.

By **Cnd**, **Prb**, **Rad** and **Smp**, Theorem 1.1 establishes that the vertical maps in the diagram below induce isomorphisms on homotopy with probability exceeding

$(1 - \delta_{\mathcal{X}})(1 - \delta_{\mathcal{Y}})$:

$$
\begin{array}{ccc}
\mathcal{X} & \xrightarrow{\ f\ } & \mathcal{Y} \\
{\scriptstyle \pi_{\mathcal{X}}}\big\uparrow & & \big\downarrow{\scriptstyle \iota_{\mathcal{Y}}} \\
\mathcal{U}(X) & \dashrightarrow[g] & \mathcal{U}(Y)
\end{array}
$$

Here, $\pi_{\mathcal{X}}$ is the canonical projection map (2) and $\iota_{\mathcal{Y}}$ is the inclusion of $\mathcal{Y}$ into the union of balls $\mathcal{U}(Y)$. Let $g : \mathcal{U}(X) \to \mathcal{U}(Y)$ be the composition $\iota_{\mathcal{Y}} \circ f \circ \pi_{\mathcal{X}}$. With probability exceeding $(1 - \delta_{\mathcal{X}})(1 - \delta_{\mathcal{Y}})$, the map induced by g on homotopy is naturally related by isomorphisms to the corresponding map induced by f.

Definition 4.1. A *simplicial reconstruction* of f is defined to be any simplicial map $\phi : N(X) \to N(Y)$ so that $g(\lceil \sigma \rceil) \subset \lceil \phi(\sigma) \rceil$ for each $\sigma \in N(X)$.

We know from Lemma 2.3 that it is possible to find continuous maps $\zeta_Z : |N(Z)| \to \mathcal{U}(Z)$ for $Z \in \{X, Y\}$ which induce isomorphisms on homotopy and satisfy $\zeta_Z(|\sigma|) \subset \lceil \sigma \rceil$ for each simplex $\sigma \in N(Z)$. For any choice of such maps, the next proposition establishes that the following diagram commutes up to homotopy whenever ϕ is a simplicial reconstruction of f:

$$
\begin{array}{ccc}
\mathcal{U}(X) & \xrightarrow{\ g\ } & \mathcal{U}(Y) \\
{\scriptstyle \zeta_X}\big\uparrow & & \big\uparrow{\scriptstyle \zeta_Y} \\
|N(X)| & \dashrightarrow[|\phi|] & |N(Y)|
\end{array}
$$

Proposition 4.2. *If ϕ is a simplicial reconstruction of f, then $\zeta_Y \circ |\phi|$ and $g \circ \zeta_X$ share a contractible carrier and hence are homotopic.*

Proof. For each $\sigma \in N(X)$, define $F(\sigma) = \lceil \phi(\sigma) \rceil$. Note that $\sigma \subset \sigma'$ implies $F(\sigma) \subset F(\sigma')$ since the latter is the union over a superset. The F-image of each $\sigma \in N(X)$ is contractible in $\mathcal{U}(Y)$, being a union of convex sets (in our case, balls in Euclidean space) with a non-empty intersection. Thus, $F : N(X) \Longrightarrow \mathcal{U}(Y)$ is a contractible carrier. Given any $\sigma \in N(X)$, we have

$$
\begin{aligned}
\zeta_Y \circ |\phi|(|\sigma|) &\subset \zeta_Y(|\phi(\sigma)|), \text{ since } \phi \text{ is a simplicial map,} \\
&\subset \lceil \phi(\sigma) \rceil, \text{ by assumption on } \zeta_Y, \\
&= F(\sigma).
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
g \circ \zeta_X(|\sigma|) &\subset g\left(\lceil \sigma \rceil\right), \text{ by assumption on } \zeta_X, \\
&\subset \lceil \phi(\sigma) \rceil, \text{ since } \phi \text{ is a simplicial reconstruction of } f, \\
&= F(\sigma).
\end{aligned}
$$

Thus, both $\zeta_Y \circ |\phi|$ and $g \circ \zeta_X$ are carried by F. $\qquad\square$

With the goal of building a simplicial reconstruction in mind, we consider the correspondence $\Delta : X \to 2^Y$ defined on each $\xi \in X$ by

$$
\Delta(\xi) = \{\eta \in Y \text{ so that } \|\eta - f(\xi)\|_{\mathbb{R}^m} < \rho\}. \tag{5}
$$

Proposition 4.3. *With probability exceeding* $(1-\delta_{\mathcal{Y}})$*, the following holds. For each* $\xi \in X$*, the set* $\Delta(\xi) \subset Y$ *is non-empty.*

Proof. Since **Cnd**, **Prb**, **Rad** and **Smp** satisfy the hypotheses of Proposition 3.3 for $\mathcal{Y}$, we see that Y is $\epsilon_{\mathcal{Y}}/2$-dense in $\mathcal{Y}$ with probability exceeding $(1 - \delta_{\mathcal{Y}})$. This density suffices to guarantee a non-empty $\Delta(\xi)$ for each $\xi \in X$ in the following way. For any $\xi \in X$ there exists some $\eta \in Y$ with $\|f(\xi) - \eta)\|_{\mathbb{R}^m} < \epsilon_{\mathcal{Y}}/2$. Since $\epsilon_{\mathcal{Y}}/2 < \rho$ as a consequence of **Rad** and (4), we have $\eta \in \Delta(\xi)$. □

Let $\mathbf{v}_X : N(X) \to 2^X$ denote the map taking each simplex $\sigma \in N(X)$ to its vertex set $\mathbf{v}_X(\sigma) \subset X$ which corresponds to (the centers of) those balls whose non-empty intersection determines the support of σ. Note that $\sigma \subset \sigma'$ in $N(X)$ if and only if we have the inclusion $\mathbf{v}_X(\sigma) \subset \mathbf{v}_X(\sigma')$. Define $\mathbf{v}_Y : N(Y) \to 2^Y$ similarly. For each $\sigma \in N(X)$, define $\Delta_\sigma \subset Y$ by

$$\Delta_\sigma = \bigcup_{\xi \in \mathbf{v}_X(\sigma)} \Delta(\xi). \tag{6}$$

Proposition 4.4. *With probability exceeding* $(1 - \delta_{\mathcal{X}})$*, the following holds. For each* $\sigma \in N(X)$*, any non-empty subset of* Δ_σ *determines a simplex in* $N(Y)$*.*

Proof. By **Cnd**, **Prb**, **Rad** and **Smp**, Proposition 3.3 holds for $\mathcal{X}$. Therefore, with probability exceeding $(1 - \delta_{\mathcal{X}})$ we are guaranteed that X is sufficiently dense in $\mathcal{X}$ so that the canonical projection map $\pi_{\mathcal{X}}$ from (2) induces homotopy-equivalence. Assuming this density, pick $w \in \lfloor \sigma \rfloor$ and recall that $g(w) = \iota_{\mathcal{Y}} \circ f \circ \pi_{\mathcal{X}}(w)$ by definition. Setting $x = \pi_{\mathcal{X}}(w)$, we note from (3) that $\|x - \xi\|_{\mathbb{R}^n} < 2\epsilon_{\mathcal{X}}$ for each $\xi \in \mathbf{v}_X(\sigma)$. By **Lip**, for each such ξ we have:

$$\|f(x) - f(\xi)\|_{\mathbb{R}^m} < 2\kappa\epsilon_{\mathcal{X}}. \tag{7}$$

Given $\eta \in \Delta_\sigma$, by (6) there is some $\xi_* \in \mathbf{v}_X(\sigma)$ so that $\eta \in \Delta(\xi_*)$, whence $\|\eta - f(\xi_*)\|_{\mathbb{R}^m} < \rho$ by (5). Since (7) implies $\|f(x) - f(\xi_*)\|_{\mathbb{R}^n} < 2\kappa\epsilon_{\mathcal{X}}$, the triangle inequality yields

$$\|f(x) - \eta\|_{\mathbb{R}^m} < \rho + 2\kappa\epsilon_{\mathcal{X}} < \epsilon_{\mathcal{Y}}.$$

Since the point $f(x)$ lies in the intersection $\bigcap_{\eta \in \Delta_\sigma} \mathcal{B}_{\epsilon_{\mathcal{Y}}}(\eta)$, this intersection is non-empty and must determine a simplex of $N(Y)$. Clearly, any subset of Δ_σ determines a face of this simplex, and hence constitutes a simplex in its own right. □

Proposition 4.3 guarantees with probability exceeding $(1 - \delta_{\mathcal{X}})$ that $\Delta(\xi)$ is nonempty for each $\xi \in X$ and so we may choose a *selector* function $h : X \to Y$ so that $h(\xi) \in \Delta(\xi)$ for each such ξ. By definition of Δ and (4), we have

$$g(\mathcal{B}_{\epsilon_{\mathcal{X}}}(\xi)) \subset \mathcal{B}_{\epsilon_{\mathcal{Y}}}(h(\xi)) \text{ for each } \xi \in X. \tag{8}$$

Proposition 4.4 guarantees with probability exceeding $(1-\delta_{\mathcal{Y}})$ that for each $\sigma \in N(X)$, the collection $\{h(\xi) \mid \xi \in \mathbf{v}_X(\sigma)\}$ determines a simplex of $N(Y)$. Therefore, with probability exceeding $(1-\delta_{\mathcal{X}})(1-\delta_{\mathcal{Y}})$, there exists a map $h : X \to Y$ of points which induces a simplicial map $\phi_h : N(X) \to N(Y)$. The following proposition demonstrates that the homotopy type of the induced simplicial map is independent of the choice of h.

Proposition 4.5. *Given any pair* $h, h' : X \to Y$ *of selectors, the maps* $|\phi_h|$ *and* $|\phi_{h'}|$ *from* $|N(X)|$ *to* $|N(Y)|$ *are homotopic.*

Proof. For each $\sigma \in N(X)$, we know that $\tau_\sigma := \mathbf{v}_Y^{-1}(\Delta_\sigma)$ is a simplex of $N(Y)$ by Proposition 4.4. Moreover, if $\sigma \subset \sigma'$ then $\Delta_\sigma \subset \Delta_{\sigma'}$ by (6) and therefore $\tau_\sigma \subset \tau_{\sigma'}$.. For any $\sigma \in N(X)$, note that since $\phi_h(\sigma)$ and $\phi_{h'}(\sigma)$ are faces of τ_σ, both $|\phi_h(\sigma)|$ and $|\phi_{h'}(\sigma)|$ are subsets of $|\tau_\sigma|$. Now $|\tau_\sigma|$, being the realization of a single simplex, is contractible. Therefore, $|\phi_h|$ and $|\phi_{h'}|$ share that contractible carrier which associates each $\sigma \in N(X)$ to $|\tau_\sigma| \subset |N(Y)|$. $\qquad\square$

For any selector $h : X \to Y$, the induced map $\phi_h : N(X) \to N(Y)$ is a simplicial reconstruction of f. To see this, note that for any $\sigma \in N(X)$ we can apply (8) to each element of $\mathbf{v}_X(\sigma)$ in order to conclude $g(\lceil\sigma\rceil) \subset \lceil\phi_h(\sigma)\rceil$. Combining this with Proposition 4.2 concludes the proof of Theorem 1.2.

5. **Robustness to bounded noise.** As is indicated in the Introduction, we would like to extend the results of the previous section to the case where data samples are *noisy*. That is, we assume that the sampled points lie close to, rather than on, the underlying manifolds. Such sampling discrepancies – aside from being ubiquitous in experimental data – cascade into imprecise knowledge of the images under an unknown function, particularly if the evaluation of that function is also subject to some inherent measurement errors. This generalization requires a framework to describe the noise, and so we adopt the model of [13, Sec. 7]. For any subset P of Euclidean space $\mathbb{R}^n$ and any $\alpha > 0$, define the *tubular neighborhood of radius α around P* as follows:

$$\mathrm{Tub}_\alpha(P) = \{x \in \mathbb{R}^n \mid \text{ there is some } p \in P \text{ with } \|p - x\|_{\mathbb{R}^n} < \alpha\}.$$

Definition 5.1. Given a subset $P \subset \mathbb{R}^n$ and some $r > 0$, a probability measure μ on $\mathbb{R}^n$ is called *r-conditioned about P* if

1. the support of μ is contained in $\mathrm{Tub}_r(P)$, and
2. for each $s \in (0, r)$, there exists some constant Ω_s so that

$$\inf_{p \in P} \mu(\mathcal{B}_s(p)) = \Omega_s > 0$$

where $\mathcal{B}_s(p)$ denotes the n-dimensional open ball of radius s about p.

We write $\Omega(\mu)$ to denote the constant $\Omega_{r/2} > 0$.

As in the noiseless case, the following fundamental results concerning sampling of manifolds are reproduced from [13]. Let $\mathcal{X}$ be a compact Riemannian submanifold of $\mathbb{R}^n$ with condition number $1/\tau_\mathcal{X}$. For $r > 0$ define the functions

$$\Gamma_\mathcal{X}^\pm(r) = \frac{(\tau_\mathcal{X} + r) \pm \sqrt{\tau_\mathcal{X}^2 + r^2 - 6\tau_\mathcal{X}r}}{2}$$

and note that $0 < \Gamma_\mathcal{X}^- < \Gamma_\mathcal{X}^+$ when the quantity under the square root is strictly positive. It is straightforward to check that this positivity holds for $r < (3-\sqrt{8})\tau_\mathcal{X}$. Pick such an r and assume that X is a finite set lying in $\mathrm{Tub}_r(\mathcal{X}) \subset \mathbb{R}^n$. For each $\alpha > 0$, let $N_\alpha(X)$ be the nerve generated by open balls of radius α about the points in X and let $\mathcal{U}_\alpha(X)$ be their union. The following result is the noisy analogue of Proposition 3.2.

Proposition 5.2. ([13, Prop 7.1]) *Assume that X is r-dense in $\mathcal{X}$ for some $0 < r < (3-\sqrt{8})\tau_\mathcal{X}$ and choose a radius ϵ satisfying $\Gamma_\mathcal{X}^-(r) < \epsilon < \Gamma_\mathcal{X}^+(r)$. Then, the canonical projection map $\pi_\mathcal{X} : \mathcal{U}_\epsilon(X) \to \mathcal{X}$ as defined in (2) is a strong deformation retraction.*

Recall that given $\nu > 0$, the ν-*covering number of* $\mathcal{X}$ – denoted $\Lambda_\nu(\mathcal{X})$ – is defined to be the minimum possible $q \in \mathbb{N}$ satisfying the following property: there exists some finite set $S \subset \mathcal{X}$ of cardinality q such that the collection $\{\mathcal{B}_\alpha(s) \mid s \in S\}$ of n-dimensional open balls covers $\mathcal{X}$. Given an r-conditioned probability measure μ about $\mathcal{X}$ and a probability parameter $\delta \in [0, 1)$, define the new bounding function $\gamma_{\mathcal{X}}$ as follows:

$$\gamma_{\mathcal{X}}(\mu, \delta) = \frac{1}{\Omega(\mu)} \left(\log(\Lambda_{r/2}(\mathcal{X})) + \log(1/\delta) \right). \tag{9}$$

The next result replaces Proposition 3.3 in the setting of conditioned noise.

Proposition 5.3. ([13, Prop 7.2]) *Assume that the real numbers* r *and* ϵ *satisfy* $0 < r < (3 - \sqrt{8})\tau_{\mathcal{X}}$ *and* $\Gamma_{\mathcal{X}}^-(r) < \epsilon < \Gamma_{\mathcal{X}}^+(r)$. *Let* μ *be any* r-*conditioned probability measure about* $\mathcal{X}$ *and assume that a finite set* X *is drawn from* $\mathbb{R}^n$ *in i.i.d. fashion with respect to* μ. *Given a parameter* $\delta \in (0, 1]$, *if* $\#X > \gamma_{\mathcal{X}}(\mu, \delta)$ *then* X *is* r-*dense in* $\mathcal{X}$ *with probability exceeding* $(1 - \delta)$.

Combining the preceding propositions yields the main result of [13] as adapted for conditioned noise.

Theorem 5.4. ([13, Thm 7.1]) *Let* $\mathcal{X} \subset \mathbb{R}^n$ *be a compact Riemannian submanifold with condition number* $1/\tau_{\mathcal{X}}$. *Fix* $r \in (0, (3 - \sqrt{8})\tau_{\mathcal{X}})$ *and choose a radius* ϵ *satisfying* $\Gamma_{\mathcal{X}}^-(r) < \epsilon < \Gamma_{\mathcal{X}}^+(r)$. *Assume that* μ *is an* r-*conditioned probability measure about* $\mathcal{X}$ *and that* $X \subset \mathbb{R}^n$ *is a finite set obtained by* μ-*i.i.d. sampling. If* $\#X > \gamma_{\mathcal{X}}(\mu, \delta)$ *for some* $\delta \in (0, 1]$, *then* $\mathcal{U}_\epsilon(X)$ *strong deformation retracts onto* $\mathcal{X}$ *with probability exceeding* $(1 - \delta)$.

The introduction of sampling noise requires the following modifications to our assumptions and choices.

Lip': Assume that $f : \mathcal{X} \to \mathcal{Y}$ is a Lipschitz-continuous function whose Lipschitz constant is bounded above by some $\kappa \geq 0$ satisfying $4\kappa \cdot \tau_{\mathcal{X}} < (\sqrt{2} - 1)\tau_{\mathcal{Y}}$.

Nse: Choose positive noise bounds $r_{\mathcal{Z}} < \alpha \cdot \tau_{\mathcal{Z}}$ where $\alpha = (3 - \sqrt{8})$ and $\mathcal{Z} \in \{\mathcal{X}, \mathcal{Y}\}$. Assume that $\mu_{\mathcal{Z}}$ is a $r_{\mathcal{Z}}$-conditioned probability measure about $\mathcal{Z}$.

Rad': Choose radii $\epsilon_{\mathcal{Z}}$ satisfying $\Gamma_{\mathcal{Z}}^-(r_{\mathcal{Z}}) < \epsilon_{\mathcal{Z}} < \Gamma_{\mathcal{Z}}^+(r_{\mathcal{Z}})$ for $\mathcal{Z} \in \{\mathcal{X}, \mathcal{Y}\}$ so that

$$4\kappa \cdot (\epsilon_{\mathcal{X}} + r_{\mathcal{X}}) < (\epsilon_{\mathcal{Y}} - r_{\mathcal{Y}}). \tag{10}$$

Smp': Assume knowledge of the finite sets $X' \subset \mathbb{R}^n$ and $Y' \subset \mathbb{R}^m$ obtained by i.i.d. $\mu_{\mathcal{X}}$ and $\mu_{\mathcal{Y}}$ sampling respectively. We require $\#X' > \gamma_{\mathcal{X}}(\mu_{\mathcal{X}}, \delta_{\mathcal{X}})$ and $\#Y' > \gamma_{\mathcal{Y}}(\mu_{\mathcal{Y}}, \delta_{\mathcal{Y}})$.

Img': Assume knowledge of $f' : X' \to \mathbb{R}^m$ so that $\|f'(\xi) - f \circ \pi_{\mathcal{X}}(\xi)\| \leq d$ for each $\xi \in X'$, where $\pi_{\mathcal{X}}$ is the canonical projection map from (2) and $d > 0$ satisfies the following bound:

$$d < \frac{(\epsilon_{\mathcal{Y}} - r_{\mathcal{Y}}) - 2\kappa \cdot (\epsilon_{\mathcal{X}} + r_{\mathcal{X}})}{2}. \tag{11}$$

The assumptions **Cnd** and **Prb** of Section 4 remain unchanged. As usual, we simplify notation by dropping the fixed quantities $\epsilon_{\mathcal{X}}$ and $\epsilon_{\mathcal{Y}}$ from various subscripts. Thus, the nerve $N_{\epsilon_{\mathcal{X}}}(X')$ and the union $\mathcal{U}_{\epsilon_{\mathcal{X}}}(X')$ are denoted by $N(X')$ and $\mathcal{U}(X')$ respectively, and similar simplifications are made for the Y' analogues. Note that in the assumption **Img'** we allow *evaluation noise*. That is, we only assume knowledge of the true image $f \circ \pi_{\mathcal{X}}(\xi) \in \mathcal{Y}$ of each $\xi \in X'$ up to a distance of d.

The inequality (10) is a constraint that involves the Lipschitz constant of f, the models for conditioned noise, and the radii for the nerves. It guarantees that the

restriction (11) is always positive. The following result provides conditions on the manifolds, the noise models and the function under which (10) can be satisfied.

Proposition 5.5. *If $4\kappa \cdot \tau_{\mathcal{X}} < (\sqrt{2} - 1)\tau_{\mathcal{Y}}$, then there exist valid choices of $\epsilon_{\mathcal{X}}$ and $\epsilon_{\mathcal{Y}}$ which satisfy (10).*

Proof. First, we check that $(\epsilon_{\mathcal{X}} + r_{\mathcal{X}}) < \tau_{\mathcal{X}}$ on the domain $0 < r_{\mathcal{X}} < (3 - \sqrt{8})\tau_{\mathcal{X}}$ imposed by **Nse**. Recall that $\epsilon_{\mathcal{X}} < \Gamma_{\mathcal{X}}^{+}(r_{\mathcal{X}})$ by **Rad'**, and consider the following function

$$2(\Gamma_{\mathcal{X}}^{+}(r_{\mathcal{X}}) + r_{\mathcal{X}}) = (\tau_{\mathcal{X}} + 3r_{\mathcal{X}}) + \sqrt{\tau_{\mathcal{X}}^2 + r_{\mathcal{X}}^2 - 6\tau_{\mathcal{X}}r_{\mathcal{X}}}$$

This function has no local maximum in its domain and attains a maximum value of $2\tau_{\mathcal{X}}$ at the left endpoint, so $(\epsilon_{\mathcal{X}} + r_{\mathcal{X}}) < \tau_{\mathcal{X}}$ as desired. Since (10) imposes a lower bound of $4\kappa \cdot (\epsilon_{\mathcal{X}} + r_{\mathcal{X}}) + r_{\mathcal{Y}}$ on $\epsilon_{\mathcal{Y}}$, it suffices to show that the over-estimate $4\kappa \cdot \tau_{\mathcal{X}} + r_{\mathcal{Y}}$ of this lower bound is smaller than the upper bound $\Gamma_{\mathcal{Y}}^{+}(r_{\mathcal{Y}})$ imposed on $\epsilon_{\mathcal{Y}}$ by **Rad'**. Equivalently, we must show that $\Gamma_{\mathcal{Y}}^{+}(r_{\mathcal{Y}}) - r_{\mathcal{Y}} > 4\kappa \cdot \tau_{\mathcal{X}}$. Observe that the function

$$\Gamma_{\mathcal{Y}}^{+}(r_{\mathcal{Y}}) - r_{\mathcal{Y}} = \frac{(\tau_{\mathcal{Y}} - r_{\mathcal{Y}}) + \sqrt{\tau_{\mathcal{Y}}^2 + r_{\mathcal{Y}}^2 - 6\tau_{\mathcal{Y}}r_{\mathcal{Y}}}}{2}$$

has no local minima on the domain $(0, (3 - \sqrt{8})\tau_{\mathcal{Y}})$ imposed by **Nse** and attains a minimum value of $(\sqrt{2} - 1)\tau_{\mathcal{Y}}$ at the right endpoint. Thus, it is possible to satisfy (10) if $4\kappa \cdot \tau_{\mathcal{X}} < (\sqrt{2} - 1)\tau_{\mathcal{Y}}$. $\square$

The main result of this section is the following theorem.

Theorem 5.6. *Assume* **Cnd**, **Lip'**, **Nse**, **Rad'**, **Prb**, **Smp'**, *and* **Img'**. *Then, with probability exceeding $(1 - \delta_{\mathcal{X}})(1 - \delta_{\mathcal{Y}})$ there exists a simplicial map $\phi : N(X') \to N(Y')$ which*

1. *is a simplicial reconstruction of f, and*
2. *can be explicitly constructed using only X', Y', $\epsilon_{\mathcal{X}}$, $\epsilon_{\mathcal{Y}}$, κ and f'.*

For the most part, the proof of this theorem is analogous to that of Theorem 1.2. Aside from Proposition 5.5, the only major modification is that the domain's radius $\epsilon_{\mathcal{X}}$ is augmented by the noise bound $r_{\mathcal{X}}$ whereas the range's radius $\epsilon_{\mathcal{Y}}$ is diminished by the corresponding bound $r_{\mathcal{Y}}$. The following quantity plays the role of ρ from (4).

$$\rho' = r_{\mathcal{Y}} + d. \tag{12}$$

Define $\Delta' : X' \to 2^{Y'}$ as follows: for each $\xi \in X'$,

$$\Delta'(\xi) = \{\eta \in Y' \text{ so that } \|f'(\xi) - \eta\|_{\mathbb{R}^m} < \rho'\}. \tag{13}$$

The noisy analogue of Proposition 4.3 is as follows.

Proposition 5.7. *With probability exceeding $(1 - \delta_{\mathcal{Y}})$, the following holds. For each $\xi \in X'$, the set $\Delta'(\xi) \subset Y$ is non-empty.*

Proof. Since **Cnd**, **Prb**, **Nse**, **Rad'** and **Smp'** satisfy the hypotheses of Proposition 5.3 for $\mathcal{Y}$ with probability measure $\mu_{\mathcal{Y}}$, the sampled set Y' is $r_{\mathcal{Y}}$-dense in $\mathcal{Y}$ with probability exceeding $(1 - \delta_{\mathcal{Y}})$. Assume that this density holds. Now, $\|f'(\xi) - f \circ \pi_{\mathcal{X}}(\xi)\|_{\mathbb{R}^m} < d$ by **Img'** and there exists some $\eta \in Y$ with $\|f \circ \pi_{\mathcal{X}}(\xi) - \eta\|_{\mathbb{R}^m} < r_{\mathcal{Y}}$ by the assumed density of Y'. By the triangle inequality, $\|f'(x) - \eta\| < r_{\mathcal{Y}} + d$, and hence $\eta \in \Delta'(\xi)$. $\square$

As before, we define maps $\mathbf{v}_Z : N(Z) \to 2^Z$ for $Z \in \{X', Y'\}$ taking each simplex to its vertex set. For each $\sigma \in N(X')$, define

$$\Delta'_\sigma = \bigcup_{\xi \in \mathbf{v}_{X'}(\sigma)} \Delta'(\xi). \tag{14}$$

Proposition 5.8. *With probability exceeding $(1 - \delta_{\mathcal{X}})$, the following is true. For each $\sigma \in N(X')$, any non-empty subset of Δ'_σ determines a simplex in $N(Y')$.*

Proof. By **Cnd**, **Prb**, **Rad'**, **Nse** and **Smp'**, Proposition 5.3 holds for $\mathcal{X}$ and the probability measure $\mu_{\mathcal{X}}$. Therefore, with probability exceeding $(1 - \delta_{\mathcal{X}})$ we are guaranteed that X' is $r_{\mathcal{X}}$-dense in $\mathcal{X}$ via Proposition 5.3 and hence that $\pi_{\mathcal{X}}$ is a strong deformation retraction via Proposition 5.2. Assuming this density, note that the distance from any $\xi \in X'$ to its nearest neighbor $\pi_{\mathcal{X}}(\xi)$ in $\mathcal{X}$ is at most $r_{\mathcal{X}}$, since $\mu_{\mathcal{X}}$ is $r_{\mathcal{X}}$-conditioned about $\mathcal{X}$ by **Nse**. Similarly, given any $w \in \mathcal{U}(X')$, we have $\|w - \pi_{\mathcal{X}}(w)\|_{\mathbb{R}^n} < \epsilon_{\mathcal{X}} + r_{\mathcal{X}}$. For any $w \in \lfloor \sigma \rfloor$ and $\xi \in \mathbf{v}_{X'}(\sigma)$, we have $\|w - \xi\|_{\mathbb{R}^n} < \epsilon_{\mathcal{X}}$ by definition, so we may make the following estimate:

$$\begin{aligned} \|\pi_{\mathcal{X}}(w) - \pi_{\mathcal{X}}(\xi)\|_{\mathbb{R}^n} &\leq \|\pi_{\mathcal{X}}(w) - w\|_{\mathbb{R}^n} + \|w - \xi\|_{\mathbb{R}^n} + \|\xi - \pi_{\mathcal{X}}(\xi)\|_{\mathbb{R}^n} \\ &< (\epsilon_{\mathcal{X}} + r_{\mathcal{X}}) + \epsilon_{\mathcal{X}} + r_{\mathcal{X}} \\ &= 2(\epsilon_{\mathcal{X}} + r_{\mathcal{X}}). \end{aligned}$$

Write $x = \pi_{\mathcal{X}}(w)$, and observe by **Lip'** that

$$\|f(x) - f \circ \pi_{\mathcal{X}}(\xi)\|_{\mathbb{R}^m} < 2\kappa(\epsilon_{\mathcal{X}} + r_{\mathcal{X}}).$$

By (11) we know $\|f'(\xi) - f \circ \pi_{\mathcal{X}}(\xi)\|_{\mathbb{R}^m} < d$, so we obtain

$$\|f(x) - f'(\xi)\|_{\mathbb{R}^m} < 2\kappa(\epsilon_{\mathcal{X}} + r_{\mathcal{X}}) + d. \tag{15}$$

Pick any $\eta \in \Delta_\sigma$, so by (14) there is some $\xi_* \in \mathbf{v}_{X'}(\sigma)$ with $\eta \in \Delta'(\xi_*)$, whence $\|\eta - f'(\xi_*)\|_{\mathbb{R}^m} < \rho'$ by (13). Since (15) implies $\|f(x) - f'(\xi_*)\|_{\mathbb{R}^m} < 2\kappa(\epsilon_{\mathcal{X}} + r_{\mathcal{X}}) + d$, the triangle inequality followed by (12) yields

$$\|f(x) - \eta\|_{\mathbb{R}^m} < 2\kappa(\epsilon_{\mathcal{X}} + r_{\mathcal{X}}) + d + \rho' = 2\kappa(\epsilon_{\mathcal{X}} + r_{\mathcal{X}}) + r_{\mathcal{Y}} + 2d.$$

Using (11), we finally obtain $\|f(x) - \eta\|_{\mathbb{R}^m} < \epsilon_{\mathcal{Y}}$. Thus, the intersection $\bigcap_{\eta \in \Delta_\sigma} \mathcal{B}_{\epsilon_{\mathcal{Y}}}(\eta)$ is non-empty: it contains $f(x)$, and therefore determines a simplex of $N(Y')$. Any subset of Δ'_σ determines a face of this simplex and must also be a simplex in its own right. $\square$

The proof of Theorem 5.6 concludes with the observation that Propositions 5.7 and 5.8 permit – with probability larger than $(1 - \delta_{\mathcal{X}})(1 - \delta_{\mathcal{Y}})$ – the construction of a selector $h' : X' \to Y'$ for Δ' which satisfies $h'(\xi) \in \Delta'(\xi)$. It is easy to check that $\iota_{\mathcal{Y}} \circ f \circ \pi_{\mathcal{X}}(\mathcal{B}_{\epsilon_{\mathcal{X}}}(\xi)) \subset \mathcal{B}_{\epsilon_{\mathcal{Y}}}(h'(\xi))$ for any such h', so the induced simplicial map $\phi_{h'} : N(X') \to N(Y')$ is a simplicial reconstruction of f as desired.

6. **A proof of the controlled nerve lemma.** In this section we prove Lemma 2.3, which was used to control the maps $\zeta_Z : |N(Z)| \to \mathcal{U}(Z)$ for $Z \in \{X, Y\}$ in the proof of Proposition 4.2.

6.1. **General Considerations.** The *mapping cylinder* $M(\gamma)$ of a continuous function $\gamma : \mathcal{U} \to \mathcal{V}$ between topological spaces is given by the quotient

$$M(\gamma) = \frac{(\mathcal{U} \times [0,1]) \coprod \mathcal{V}}{(x,1) \sim \gamma(x)}.$$

In other words, the mapping cylinder of γ consists of the product of its domain $\mathcal{U}$ with $[0,1]$ glued onto the co-domain $\mathcal{V}$ via the identification of each $(x,1)$ with the image $\gamma(x)$. It is easy to see that there is a strong deformation retraction from $M(\gamma)$ to the codomain $\mathcal{V}$ obtained by sliding each $(x,t) \in \mathcal{U} \times [0,1]$ to the endpoint $\gamma(x)$. If γ is a homotopy-equivalence, then $M(\gamma)$ also admits a strong deformation retraction to the domain $\mathcal{U} \simeq \mathcal{U} \times \{0\} \subset M(\gamma)$.

Definition 6.1. A metrizable topological space $\mathcal{V}$ is an *absolute neighborhood retract* (henceforth abbreviated **ANR**) if for each triple $(\mathcal{U}, \mathcal{A}, \psi)$ of metrizable space $\mathcal{U}$, closed subset $\mathcal{A} \subset \mathcal{U}$ and continuous function $\psi : \mathcal{A} \to \mathcal{V}$, there exists a pair $(\mathcal{A}', \psi')$ consisting of an open set $\mathcal{A}'$ with $\mathcal{A} \subset \mathcal{A}' \subset \mathcal{U}$ and a map $\psi' : \mathcal{A}' \to \mathcal{V}$ whose restriction $\psi'|_{\mathcal{A}}$ to $\mathcal{A}$ equals ψ.

Throughout the sequel, every **ANR** is assumed to be *finite-dimensional* in the sense that it admits a continuous embedding into $\mathbb{R}^m$ for some $m \geq 0$. We will highlight all those basic facts about finite-dimensinal **ANR**s which are relevant to our proofs and refer the reader to [8, Ch. 11] for details and proofs. Recall that a pair $(\mathcal{U}, \mathcal{A})$ of topological spaces satisfies the *homotopy extension property* if for any triple $(\mathcal{V}, \omega_t, \gamma)$ of topological space $\mathcal{V}$, homotopy $\omega_t : \mathcal{A} \to \mathcal{Y}$ and function $\gamma : \mathcal{U} \to \mathcal{V}$ with $\gamma|_{\mathcal{A}} \equiv \omega_0$, there exist continuous functions $\omega_t' : \mathcal{U} \to \mathcal{V}$ so that $\omega_0' \equiv \gamma$ on $\mathcal{U}$ and $\omega_t'|_{\mathcal{A}} \equiv \omega_t$ for each $t \in [0,1]$.

We make use of the following facts in our next lemma:

1. the mapping cylinder of a continuous function between **ANR**s is also an **ANR**, and
2. any pair $(\mathcal{U}, \mathcal{A})$ of compact **ANR**s with $\mathcal{A}$ is closed in $\mathcal{U}$ satisfies the homotopy extension property.

Lemma 6.2. *Let $\gamma : (\mathcal{U}, \mathcal{A}) \to (\mathcal{V}, \mathcal{B})$ be a map of **ANR** pairs with $\mathcal{A}$ and $\mathcal{B}$ closed in $\mathcal{U}$ and $\mathcal{V}$, such that $\gamma : \mathcal{U} \to \mathcal{V}$ and $\gamma|_{\mathcal{A}} : \mathcal{A} \to \mathcal{B}$ are homotopy-equivalences. Then, γ is a homotopy-equivalence of pairs.*

Proof. Let $\omega_t : M(\gamma|_{\mathcal{A}}) \to M(\gamma|_{\mathcal{A}})$ be a strong deformation retraction from $M(\gamma|_{\mathcal{A}})$ to $\mathcal{A}$. By the homotopy extension property, we can construct a map $\omega_t' : M(\gamma) \to M(\gamma)$ extending ω_t so that $\omega_0'|_{\mathcal{U}}$ is the identity on $\mathcal{U}$. Let $\nu_t : M(\gamma) \to M(\gamma)$ be a strong deformation retraction from $M(\gamma)$ to $\mathcal{U}$. Define

$$\theta_t = \begin{cases} \omega_{2t}' & 0 \leq t \leq 1/2 \\ \nu_{2t-1} \circ \omega_1' & 1/2 \leq t \leq 1 \end{cases}.$$

This is a strong deformation retraction of pairs from $(M(\gamma), M(\gamma|_{\mathcal{A}}))$ to $(\mathcal{U}, \mathcal{A})$. Therefore, $(\mathcal{U}, \mathcal{A})$ is homotopy-equivalent as a pair to $(M(\gamma), M(\gamma|_{\mathcal{A}}))$, which in turn is homotopy-equivalent as a pair to $(\mathcal{V}, \mathcal{B})$. $\square$

As a corollary of the proof, we see that any strong deformation retraction of $M(\gamma|_{\mathcal{A}})$ to $\mathcal{A}$ can be extended to a strong deformation retraction of $M(\gamma)$ to $\mathcal{U}$ with a new parameterization. The given strong deformation retraction over $\mathcal{A}$ takes place in the first half of the interval [0,1] and remains stationary thereafter.

Proposition 6.3. *Let $\mathcal{L}$ be a topological space, N a finite simplicial complex and $\rho : \mathcal{L} \to |N|$ a continuous surjection. If $\mathcal{L}^\sigma := \rho^{-1}(|\sigma|)$ is a contractible* **ANR** *for each simplex $\sigma \in N$, then there is a strong deformation retraction from $M(\rho)$ to $\mathcal{L}$ whose restriction to the mapping cylinder of $\rho|_{\mathcal{L}^\sigma} : \mathcal{L}^\sigma \to |\sigma|$ is a strong deformation retraction onto $\mathcal{L}^\sigma$ for each $\sigma \in N$.*

Proof. We proceed by induction on skeleta of N using Lemma 6.2. Pick any 0-dimensional simplex $\sigma \in N^0$. Now, $\rho|_{\mathcal{U}^\sigma} : \mathcal{U}^\sigma \to |\sigma|$ is a homotopy-equivalence by assumption, and therefore $\mathcal{U}^\sigma$ is a strong deformation retract of the mapping cylinder $M(\rho|_{\mathcal{U}^\sigma})$. The disjoint union of such strong deformation retractions as σ ranges over N^0 yields a strong deformation retraction over all N^0, thus establishing the base case of our induction. Assume the inductive hypothesis for the $(d-1)$-skeleton of N for some $d \geq 1$. For each $\sigma \in N^d$, we have a strong deformation retraction from the mapping cylinder $M(\rho|_{\mathcal{L}^\sigma})$ to $\mathcal{L}^\sigma$ since the restriction of ρ to $\mathcal{L}^\sigma$ is a homotopy-equivalence. On the other hand, the inductive hypothesis guarantees a strong deformation retraction from the mapping cylinder $M(\rho|_{\partial\sigma})$ to $\partial\sigma$, where $\partial\sigma$ denotes the boundary of σ in N. Lemma 6.2 now gives us a strong deformation retraction from $M(\rho|_{\mathcal{L}^\sigma})$ to $\mathcal{L}^\sigma$ extending the strong deformation retraction over $\partial\sigma$. Piecing these extensions together for all $\sigma \in N^d$ gives the desired result. $\square$

The following corollary is immediate; it plays a central role in our proof of the controlled nerve lemma.

Corollary 6.4. *Let $\mathcal{L}$ be a topological space, N a finite simplicial complex and $\rho : \mathcal{L} \to |N|$ a continuous surjection. If $\rho^{-1}(|\sigma|)$ is a contractible* **ANR** *for each $\sigma \in N$, then ρ admits a homotopy-inverse $\iota : |N| \to \mathcal{L}$ so that $\iota(|\sigma|) \subset \rho^{-1}(|\sigma|)$ for each $\sigma \in N$.*

6.2. **Specific Considerations.** Let $\mathbb{U}$ be a finite collection of open balls in Euclidean space $\mathbb{R}^n$ whose union is denoted by $\mathcal{U}$. Let $N(\mathbb{U})$ be the nerve of the cover of $\mathcal{U}$ by the balls in $\mathbb{U}$. We remark that a finite union of balls in $\mathbb{R}^n$ is an **ANR**, and so is the geometric realization of a finite simplicial complex. Note also that **ANR**s are closed under the operation of taking finite products.

The following consequence of the main theorem from [14] will be used frequently in order to establish homotopy-equivalences.

Theorem 6.5 (The Vietoris-Smale Theorem). *Let $\gamma : \mathcal{U} \to \mathcal{V}$ be a proper surjective continuous map of connected, locally contractible separable metric spaces. If the fiber $\gamma^{-1}(v) \subset \mathcal{U}$ is contractible for each $v \in \mathcal{V}$, then the map induced by γ on the homotopy groups of $\mathcal{U}$ and $\mathcal{V}$ is an isomorphism.*

In order to use this theorem, one must construct proper functions[2] and hence deal with closed rather than open balls.

Remark 6.6. We may modify $\mathbb{U}$ into a finite collection of closed balls without altering the homotopy type of $N(\mathbb{U})$. For each top-dimensional simplex $\sigma \in N(\mathbb{U})$, fix a witness point $x_\sigma \in \lfloor \sigma \rfloor$ and note by openness that there exists some $d_\sigma > 0$ so that the ball of radius d_σ around x_σ lies entirely in $\lfloor \sigma \rfloor$. Now, set
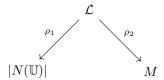
$$d_{\mathbb{U}} = \min_\sigma d_\sigma,$$

where σ ranges over all top-dimensional simplices in $N(\mathbb{U})$. Pick any $d \in (0, d_{\mathbb{U}})$ and replace each open ball $u \in \mathbb{U}$ by a closed ball u' with the same center and with radius

---

[2]Recall that $\gamma : \mathcal{U} \to \mathcal{V}$ is *proper* if $\gamma^{-1}(K) \subset \mathcal{V}$ is compact whenever $K \subset \mathcal{U}$ is compact.

shrunk by d. This modification preserves the non-emptiness of all intersections. The contractibility of the new closed intersections follows immediately from the fact that they are convex sets.

In light of the preceding remark, we assume that $\mathbb{U}$ is a finite collection of closed balls in $\mathbb{R}^n$ and hence that the union $\mathcal{U}$ is also closed. Our goal is to establish the existence of a homotopy-equivalence $\zeta : N(\mathbb{U}) \to \mathcal{U}$ so that $\zeta(|\sigma|)$ is contained in the union of balls $\lceil \sigma \rceil$. To this end, consider the subset $\mathcal{L}$ of the product $|N(\mathbb{U})| \times \mathcal{U}$ defined as follows. We include (α, x) in $\mathcal{L}$ if and only if there exists some simplex $\sigma \in N(\mathbb{U})$ with $\alpha \in |\sigma|$ and $x \in \lfloor \sigma \rfloor$.

We now have surjective, continuous and proper projection maps from $\mathcal{L}$ onto each of the constituent factors of the product:

$$
\begin{array}{ccc}
 & \mathcal{L} & \\
 {\scriptstyle \rho_1} \swarrow & & \searrow {\scriptstyle \rho_2} \\
|N(\mathbb{U})| & & M
\end{array}
$$

Proposition 6.7. *The maps ρ_1 and ρ_2 are homotopy-equivalences.*

Proof. Since all spaces in sight certainly satisfy the requirements of the Vietoris-Smale theorem, it suffices to check that the fibers of ρ_1 and ρ_2 are contractible. First consider the fiber $\rho_1^{-1}(\alpha)$ over some barycentric function $\alpha \in |N(\mathbb{U})|$. By finiteness of the cover $\mathbb{U}$, there is a unique simplex $\sigma_\alpha \in N(\mathcal{U})$ of minimal dimension so that $\alpha \in |\sigma_\alpha|$. By definition, we have

$$
\begin{aligned}
\rho_1^{-1}(\alpha) &= \{(\alpha, x) \mid x \in \lfloor \sigma \rfloor \text{ for some } \sigma \supset \sigma_\alpha\} \\
&= \{\alpha\} \times \bigcup_{\sigma \supset \sigma_\alpha} \lfloor \sigma \rfloor \\
&= \{\alpha\} \times \lfloor \sigma_\alpha \rfloor .
\end{aligned}
$$

The second factor $\lfloor \sigma_\alpha \rfloor$ is contractible, being an intersection of convex sets – and therefore, so is the fiber $\rho_1^{-1}(\alpha)$. Applying the Vietoris-Smale theorem concludes this half of the proof.

On the other hand, pick any $x \in \mathcal{U}$ and consider the fiber $\rho_2^{-1}(x) \subset \mathcal{L}$. Since the cover $\mathbb{U}$ is finite by assumption, there is a unique simplex $\sigma_x \in N(\mathbb{U})$ of maximal dimension satisfying $x \in \lfloor \sigma_x \rfloor$. By definition, we have

$$
\begin{aligned}
\rho_2^{-1}(x) &= \{(\alpha, x) \in |\sigma| \times \{x\} \mid \sigma \in N(\mathbb{U}) \text{ satisfies } x \in \lfloor \sigma \rfloor\} \\
&= |\sigma_x| \times \{x\}
\end{aligned}
$$

Since the realization of any simplex is contractible, the fiber $\rho_2^{-1}(x)$ is the product of contractible sets and hence is contractible itself. By the Vietoris-Smale theorem, ρ_2 induces an isomorphism of homotopy groups. $\qquad \square$

In order to apply Corollary 6.4, we re-examine the fibers of ρ_1. For each $\sigma \in N(\mathbb{U})$ one obtains the following:

$$
\begin{aligned}
\rho_1^{-1}(|\sigma|) &= \{(\alpha, x) \mid \exists \sigma' \subset \sigma \text{ with } \alpha \in |\sigma'| \text{ and } x \in \lfloor \sigma' \rfloor\} \\
&= |\sigma| \times \bigcup_{\sigma' \subset \sigma} \lfloor \sigma' \rfloor \\
&= |\sigma| \times \lceil \sigma \rceil .
\end{aligned}
$$

The second factor $\lceil\sigma\rceil$ is contractible, being a union of convex sets (i.e., balls in Euclidean space) with a non-empty intersection. Now, $|\sigma| \times \lceil\sigma\rceil$ is a product of contractible **ANR**s, and hence a contractible **ANR**. Therefore, ρ_1 satisfies the hypotheses of Corollary 6.4 and hence admits a homotopy-inverse $\iota : |N(\mathbb{U})| \to \mathcal{L}$ so that

$$\iota(|\sigma|) \subset \rho_1^{-1}(|\sigma|) = |\sigma| \times \lceil\sigma\rceil.$$

Letting $\zeta : |N(\mathbb{U})| \to \mathcal{U}$ be the composition $\rho_2 \circ \iota$ of homotopy-equivalences gives the desired inclusion $\zeta(|\sigma|) \subset \lceil\sigma\rceil$ and concludes the proof of Lemma 2.3.

REFERENCES

[1] A. Bjorner, Nerves, fibers and homotopy groups, *Journal of Combinatorial Theory, Series A*, **102** (2003), 88–93.

[2] K. Borsuk, On the imbedding of systems of compacta in simplicial complexes, *Fundamenta Mathematicae*, **35** (1948), 217–234.

[3] G. Carlsson, Topology and data, *Bulletin of the American Mathematical Society*, **46** (2009), 255–308.

[4] J.-G. Dumas, F. Heckenbach, B. D. Saunders and V. Welker, Computing simplicial homology based on efficient Smith normal form algorithms, *Proceedings of Algebra, Geometry and Software Systems*, (2003), 177–206.

[5] H. Edelsbrunner and J. L. Harer, *Computational Topology – an Introduction*, American Mathematical Society, Providence, RI, 2010.

[6] K. Fischer, B. Gaertner and M. Kutz, Fast-smallest-enclosing-ball computation in high dimensions, *Proceedings of the 11th Annual European Symposium on Algorithms (ESA)*, **2832** (2003), 630–641.

[7] R. Ghrist, Three examples of applied and computational homology, *Nieuw Archief voor Wiskunde*, **9** (2008), 122–125.

[8] A. Granas and J. Dugundji, *Fixed Point Theory*, Springer-Verlag, New York, 2003.

[9] S. Harker, K. Mischaikow, M. Mrozek and V. Nanda, Discrete Morse theoretic algorithms for computing homology of complexes and maps, *Foundations of Computational Mathematics*, **14** (2014), 151–184.

[10] T. Kaczynski, K. Mischaikow and M. Mrozek, *Computational Homology*, Springer-Verlag, New York, 2004.

[11] D. Kozlov, *Combinatorial Algebraic Topology*, Springer, 2008.

[12] J. R. Munkres, *Elements of Algebraic Topology*, Addison-Wesley, Menlo Park, 1984.

[13] P. Niyogi, S. Smale and S. Weinberger, Finding the homology of submanifolds with high confidence from random samples, *Discrete and Computational Geometry*, **39** (2008), 419–441.

[14] S. Smale, A Vietoris mapping theorem for homotopy, *Proceedings of the American mathematical society*, **8** (1957), 604–610.

[15] E. H. Spanier, *Algebraic Topology*, Springer-Verlag, New York, 1981. Corrected reprint.

E-mail address: sferry@math.rutgers.edu
E-mail address: mischaik@math.rutgers.edu
E-mail address: vnanda@sas.upenn.edu