HAL open science

Adversarial Manifold Estimation

Eddie Aamari, Alexander Knop

To cite this version:

Eddie Aamari, Alexander Knop. Adversarial Manifold Estimation. Foundations of Computational Mathematics, 2022. hal-02995288v2

HAL Id: hal-02995288

https://hal.science/hal-02995288v2

Submitted on 12 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Adversarial Manifold Estimation

Eddie Aamari
aamari@lpsm.paris
LPSM, CNRS
Université Paris Cité, Sorbonne Université
Paris, France

Alexander Knop
aknop@ucsd.edu
Department of Mathematics
University of California, San Diego
La Jolla, CA, USA

Abstract

This paper studies the statistical query (SQ) complexity of estimating d-dimensional submanifolds in $\mathbb{R}^n$. We propose a purely geometric algorithm called Manifold Propagation, that reduces the problem to three natural geometric routines: projection, tangent space estimation, and point detection. We then provide constructions of these geometric routines in the SQ framework. Given an adversarial STAT(τ) oracle and a target Hausdorff distance precision $\varepsilon = \Omega(\tau^{2/(d+1)})$, the resulting SQ manifold reconstruction algorithm has query complexity $\tilde{O}(n\varepsilon^{-d/2})$, which is proved to be nearly optimal. In the process, we establish low-rank matrix completion results for SQ's and lower bounds for randomized SQ estimators in general metric spaces.

Contents

1 Introduction **3**
1.1 Context . . . 3
1.2 Contribution . . . 5
1.2.1 Main Results . . . 5
1.2.2 Manifold Propagation Algorithm . . . 7
1.2.3 Statistical Query Algorithms for the Routines . . . 8
1.2.4 Lower Bound Techniques . . . 10

2 Preliminaries **11**
2.1 Statistical Query Framework . . . 11
2.2 Manifold Regularity and Distributional Assumptions . . . 14
2.2.1 General Notation and Differential Geometry . . . 14
2.2.2 Geometric and Statistical Models . . . 15
2.2.3 On Some Implicit Bounds on the Model Parameters . . . 17
2.3 Manifold Reconstruction from Point Clouds . . . 18

3 Manifold Propagation Algorithm **19**

4 Geometric Routines with Statistical Queries **22**
4.1 Projection . . . 23
4.2 Tangent Space . . . 23
4.3 Seed Point . . . 24

5 Manifold Estimation with Statistical Queries **25**
5.1 Fixed Point Model . . . 25
5.2 Bounding Ball Model . . . 26

6 Further Directions **28**

Acknowledgements **28**

A Proofs of the Properties of Manifold Propagation **28**

B Preliminary Geometric Results **33**
B.1 Local Mass of Balls Estimates . . . 33
B.2 Euclidean Packing and Covering Estimates . . . 33
B.3 Global Volume Estimates . . . 35

C Projection Routine **36**
C.1 Bias of the Local Conditional Mean for Projection . . . 36
C.2 Metric Projection with Statistical Queries . . . 39

D Tangent Space Estimation Routine **40**
D.1 Bias of Local Principal Component Analysis . . . 40
D.2 Matrix Decomposition and Principal Angles . . . 42
D.3 Low-rank Matrix Recovery . . . 43
D.3.1 Restricted Isometry Property and Low-Rank Matrix Recovery . . . 43
D.3.2 Building a Good Matrix Sensing Operator . . . 44
D.3.3 Mean Matrix Completion with Statistical Queries . . . 45
D.4 Tangent Space Estimation with Statistical Queries . . . 46

E Seed Point Detection **47**
E.1 Detecting a Raw Initial Point . . . 47
E.2 Refined Point Detection . . . 49

F Proof for the Main Statistical Query Manifold Estimators **50**

G Statistical Query Lower Bounds in Metric Spaces **53**
G.1 Information-Theoretic Lower Bound for Randomized SQ Algorithms . . . 53
G.2 Computational Lower Bound . . . 54
G.2.1 Probabilistic Covering and Packing Number . . . 55
G.2.2 Proof of the Computational Lower Bounds for Randomized SQ Algorithms 56

H Lower Bounds for Manifold Models **56**
H.1 Diffeomorphisms and Geometric Model Stability . . . 56
H.2 Building a Large-Volume Submanifold with Small Euclidean Diameter . . . 58
H.2.1 The Statement . . . 58
H.2.2 Widget Gluing: From Paths on the Discrete Grid to Manifolds . . . 59
H.2.3 Existence of Long Paths on the Grid . . . 64
H.3 Informational Lower Bounds: Hypotheses for Le Cam's Lemma . . . 65
H.3.1 Construction of the Hypotheses . . . 65
H.3.2 Proof of the Informational Lower Bounds for Manifold Estimation . . . 68
H.4 Computational Lower Bounds: Packing Number of Manifold Classes . . . 69
H.4.1 Global Ambient Packings . . . 69
H.4.2 Local Intrinsic Packings . . . 70
H.4.3 Proof of the Computational Lower Bounds for Manifold Estimation . . . 74

1 Introduction

In the realm of massive data acquisition, the curse of dimensionality phenomena led to major developments of computationally efficient statistical and machine learning techniques. Central to them are topological data analysis and geometric methods, which have recently garnered a lot of attention and proved fruitful in both theoretical and applied areas [Was18]. These realms refer to a collection of statistical methods that find intrinsic structure in data. In short, this field is based upon the idea that data described with a huge number of features n may be subject to redundancies and correlations, so that they include only $d \ll n$ intrinsic and informative degrees of freedom. This low-dimensional paradigm naturally leads to the problem of recovering this intrinsic structure, for data visualization or to mitigate the curse of dimensionality. This problem is usually referred to as *support estimation* [CF97] or *dimensionality reduction* [LV07].

Linear dimensionality reduction techniques, such as Principal Component Analysis and LASSO-types methods [HTF09], assume that the data of interest lie on a low-dimensional linear subspace. This assumption appears to be often too strong in practice, so that one may use problem-specific featurization techniques, or other non-linear methods. On the other hand, non-linear dimensionality reduction techniques such as Isomap [Ten97], Local Linear Embedding [RS00] and Maximum Variance Unfolding [ACP13], work under the relaxed assumption that the data of interest lie on an embedded d-dimensional manifold of $\mathbb{R}^n$ with $d \ll n$, hence allowing natively for non-linearities.

1.1 Context

Geometric Inference from Samples. The classical statistical framework, based on data points, is usually referred to as *PAC-learning* [Val84] or *sample framework*. In this setting, the learner is given a set $\{x_1, \dots, x_s\}$ of s samples drawn, most commonly independently, from an unknown distribution D. From these samples, the learner then aims at estimating a parameter of interest $\theta(D)$, which in our context will naturally be taken as the support $\theta(D) = \mathrm{Supp}(D) \subseteq \mathbb{R}^n$. As described above, for this problem to make sense, D shall be assumed to be supported on (or near) a low-dimensional structure, i.e. a d-dimensional submanifold with $d \ll n$. Here, the precision is usually measured with the Hausdorff distance, a L^∞-type distance between compact subsets of $\mathbb{R}^n$.

The Hausdorff estimation of manifolds in the noiseless sample framework is now well understood. The first minimax manifold estimation results in the sample framework are due to [GPPVW12b, GPPVW12a]. At the core of their work is the reach, a scale parameter that quantitatively encodes $\mathcal{C}^2$ regularity of the unknown manifold $M = \mathrm{Supp}(D) \subseteq \mathbb{R}^n$, and that allows measuring the typical scale at which M looks like $\mathbb{R}^d$ [AKC+19] (see Definition 2.3). Under reach assumptions, the estimator of [GPPVW12a] achieves a worst-case average precision at most $O((\log s/s)^{2/d})$, but with a computationally intractable method. This rate was later shown to be log-tight optimal by [KZ15]. Given a target precision $\varepsilon > 0$, these results hence reformulate to yield sample complexity of order $s = O(\varepsilon^{-d/2} \log(1/\varepsilon))$. Later, [AL18] gave a constructive estimator combining local PCA and the computational geometry algorithm from [BG14], which outputs a triangulation of the sample points in polynomial time, and linear time in the ambient dimension n [BG14]. More recently, [Div21a] proposed a computationally tractable minimax adaptive method that automatically selects the intrinsic dimension and reach. Let us also mention that by using local polynomials, faster sample rates can also be achieved when the manifolds are smoother than $\mathcal{C}^2$ [AL19].

Manifold estimation in the presence of noise is by far less understood. The only known optimal statistical method able to handle samples corrupted with additive noise [GPPVW12b] is intractable. Currently, the best algorithmically tractable estimators in this context require either the noise level to vanish fast enough as the sample size grows [PS22], or very specific distributional assumptions on the noise: either Gaussian [FILN19, DW21] or ambient uni-

form [AS21]. To date, the only computationally tractable sample method that truly is robust to some type of noise is due to [AL18], in which the authors consider the so-called *clutter noise* model introduced by [GPPVW12a]. In this model, the samples are generated by a mixture of a distribution D on M and a *uniform* distribution in the ambient space, with respective coefficients $\beta \in (0, 1]$ and $(1 - \beta)$. That is, the s-sample consists of unlabelled points in $\mathbb{R}^n$, with approximately βs informative points on M and $(1 - \beta)s$ non-informative ambient clutter points. In [GPPVW12a, KZ15], the optimal sample complexity was shown to be $s = O(\beta^{-1}\varepsilon^{-d/2}\log(1/\varepsilon))$, but this rate was obtained with an intractable method. On the other hand, [AL18] proposed to label the non-informative data points, which allows to apply a clutter-free estimation procedure to the remaining decluttered points. This results in a computationally tractable minimax optimal method, with an additional computational cost due to the decluttering. However, the success of this extra step relies heavily on the assumption that the ambient clutter is uniform.

Overall, the existing reasonable manifold estimation methods are heavily attached to the *individual* data points and do not tolerate much noise, as the change of a single point may have the method fail completely. Let us mention from now that in sharp contrast, the statistical query framework considered in this work is inherently robust to clutter noise without an artificial decluttering, no matter the clutter noise distribution (see Remark 2.2).

Private Learning. Beyond the classical sample complexity, the modern practice of statistics raised concerns leading to quantitative and qualitative estimation constraints [Wai14]. For instance, in many applications of learning methods, the studied data is contributed by individuals, and features represent their (possibly) private characteristics such as gender, race, or health history. Hence, it is essential not to reveal too much information about any particular individual. The seminal paper [KLN+11] on this topic introduces the notion of private learning, a learning framework inspired by differentially private algorithms [DMNS06]. Given samples $\{x_1, \ldots, x_s\}$, this framework imposes privacy to a learner by requiring it not to be significantly affected if a particular sample x_i is replaced with an arbitrary x_i'.

In contrast to precision, which is analyzed with respect to a model, the level of differential privacy is a worst-case notion. Hence, when analyzing the privacy guarantees of a learner, no assumption should be made on the underlying generative model. Indeed, such a generative assumption could fall apart in the presence of background knowledge that the adversary might have: conditioned on this background knowledge, the model may change drastically.

There are two main types of differentially private algorithms. *Global differential privacy* assumes that there is a trusted entity (i.e. a central data aggregator) that can give private answers to database queries [KLN+11]. This approach was used by LinkedIn to share advertisements data [RSP+20], by Uber's system for internal analytics [JNS18], and is about to be implemented by the U.S. Census Bureau for publishing data [Jar19].

In contrast, *local differential privacy*, as defined by [EGS03, KLN+11], even further restricts the learners. It requires that even if an adversary has access to the personal revealed data of individuals, this adversary will still be unable to learn too much about the user's actual personal data. The simplest example of a locally private learning protocol was originally introduced to encourage truthfulness in surveys [War65]. In local differential privacy, a trusted entity is not necessarily present, and each individual protects their own privacy, for instance, by adding noise to their data separately.

Statistical Queries. Instead of sample complexity, this paper considers the notion of statistical query complexity, which was proved to be equivalent to the locally private learning complexity up to a polynomial factor [KLN+11], and that naturally enforces robustness to clutter noise (see Remark 2.2).

First introduced by Kearns [Kea98], the statistical query (SQ) framework is a restriction of

PAC-learning, where the learner is only allowed to obtain approximate averages of the unknown distribution D via an adversarial oracle, but cannot see any sample. That is, given a tolerance parameter $\tau > 0$, a STAT(τ) oracle for the distribution D accepts functions $r : \mathbb{R}^n \to [-1, 1]$ as queries from the learner, and can answer *any* value $\mathrm{a} \in \mathbb{R}$ such that $|\mathbb{E}_{x\sim D}[r(x)] - \mathrm{a}| \leq \tau$. Informally, the fact that the oracle is adversarial is the counterpart to the fact that differential privacy is a worst-case notion. We emphasize that in the statistical query framework, estimators (or learners) are only given access to such an oracle, and not to the data themselves. Limiting the learner's accessible information to adversarially perturbed averages both restricts the range of the usable algorithms, and effectively forces them to be robust and efficient.

Kearns showed that any statistical query learner can be transformed into a classical PAC-learning algorithm with robustness to random classification noise [Kea98]. Conversely, many commonly used PAC-learning algorithms have statistical query implementations [Kea98, Byl94, DV08]. Though, Kearns also showed that there are information-theoretic obstacles that are specific to the statistical query framework; e.g. parity functions require an exponential number of queries [Kea98]. In other words, PAC-learning is strictly stronger than SQ-learning.

We have already mentioned the connection between statistical queries and private learning. On top of this, the simplicity of the SQ framework allowed its application in several other fields, such as (theoretical and practical) learning algorithms for distributed data systems. Indeed, a problem has an efficient SQ algorithm if and only if it has an efficient distributed learner [BD98, SVW16]. Another incentive to study statistical queries arises from quantum computations: general quantum PAC learners can perform complex entangling measurements which do not seem realizable for near-term quantum computers. To overcome this issue, Arunachalam, Grilo, and Yuen [AGY20] introduced the notion of quantum statistical query learning, for which practical implementations would only require to measure a single quantum state at a time.

Overall, certainly the most interesting property of statistical query algorithms is the possibility of proving unconditional lower bounds on the complexity of statistical problems. Considering the number of learning algorithms that are implementable in the statistical query framework, these lower bounds provide strong evidence of hardness of these problems. Moreover, for many learning problems, the known unconditional lower bounds for the statistical query framework closely match the known computational complexity upper bounds. For instance, [BFJ+94] proved that SQ algorithms require a quasi-polynomial number of queries to learn disjunctive normal forms (DNF), which matches the running time upper bound by Verbeurgt [Ver90]. Similar results were proved by [FGR+17] for the planted clique problem, and by [DKS17] for high-dimensional Gaussian mixtures learning. Finally, some problem-specific statistical query lower bounds directly imply lower bounds against general convex relaxations of Boolean constraint satisfaction problems [FPV18, FGV21], lower bounds on approximation of Boolean functions by polynomials [DFT+15], and lower bounds on dimension complexity of Boolean function classes [She08, FGV21].

1.2 Contribution

This paper is the long and complete version of [AK21]. We establish nearly matching upper and lower bounds on the statistical query complexity of manifold learning. As a corollary, it provides an efficient and natural noise-tolerant sample manifold estimation technique; as another side-product, it also provides, to the best of our knowledge, the first private manifold estimation method. In some regimes of the parameters, it also exhibits another example of a natural statistical problem with different sample and statistical query complexities.

1.2.1 Main Results

Upper Bounds. The main contribution of this paper is the construction of a low-complexity *deterministic* SQ algorithm that uniformly estimates the compact connected d-dimensional $\mathcal{C}^2$-

manifolds $M \subseteq \mathbb{R}^n$ with reach $\mathrm{rch}_M \geq \mathrm{rch}_{\min} > 0$ (i.e. curvature roughly bounded by $1/\mathrm{rch}_{\min}$), from distributions D with support $\mathrm{Supp}(D) = M$ that have a Lipschitz density function bounded below by $f_{\min} > 0$ on M (i.e. volume of M roughly bounded by $1/f_{\min}$). See Definition 2.5 for a formal definition. The estimation error is measured in Hausdorff distance, which plays the role of a sup-norm between compact subsets of $\mathbb{R}^n$.

In Proposition 2.1, we prove that without any prior information about the location of the manifolds, SQ algorithms cannot estimate them, even with an unlimited number of queries. It is worth noting that this phenomenon is specific to the SQ framework and does not occur in the sample framework. We consider two ways to "localize" the model. Namely, we either assume: that the manifold contains the origin (fixed point model), or that the manifold lies within the ball of radius $R > 0$ centered at the origin (bounding ball model).

[Fixed point model] Theorem 5.1 presents a *deterministic* algorithm which, given the information that $0 \in M$, achieves precision ε using

$$q = O\left(\frac{n\,\mathrm{polylog}(n)}{f_{\min}}\left(\frac{1}{\mathrm{rch}_{\min}\varepsilon}\right)^{d/2}\right)$$

queries to a $\mathrm{STAT}(\tau)$ oracle, provided that

$$\varepsilon = \Omega\left(\mathrm{rch}_{\min}\left(\frac{\tau}{f_{\min}\mathrm{rch}_{\min}^d}\right)^{2/(d+1)}\right), \text{ and } \tau = O(f_{\min}\mathrm{rch}_{\min}^d).$$

[Bounding ball model] Theorem 5.4 shows that the same estimation problem can still be solved using $O(n\log(R/\varepsilon))$ extra queries to $\mathrm{STAT}(\tau)$ if M is only assumed to be contained in the ball $\mathrm{B}(0, R)$. That is, it shows that manifold estimation with precision ε in the bounding ball model can be done deterministically with

$$q = O\left(n\log\left(\frac{R}{\varepsilon}\right) + \frac{n\,\mathrm{polylog}(n)}{f_{\min}}\left(\frac{1}{\mathrm{rch}_{\min}\varepsilon}\right)^{d/2}\right)$$

queries to a $\mathrm{STAT}(\tau)$ oracle, under similar conditions as above.

Notice the limited quasi-linear dependency on the ambient dimension n in these bounds. Actually, in the fixed point model, the given query complexity corresponds to the sample complexity up to the $n\,\mathrm{polylog}(n)$ factor [KZ15, Div21a]. This apparent discrepancy can be explained by the fact that a single sample of $M \subseteq \mathbb{R}^n$ corresponds to n coordinates, while statistical queries are forced to be real-valued. More interestingly, the extra cost $O(n\log(R/\varepsilon))$ in the bounding ball model is specific to the statistical query framework and does not appear in the sample framework [KZ15], although this term would dominate only in the regime where R is exponentially bigger than $\mathrm{rch}_{\min}$.

The insights underlying these upper bounds are described in Sections 1.2.2 and 1.2.3, and the formal statements in Sections 3 to 5.

Differentially Private Manifold Estimation. As a direct corollary (see [KLN+11, Theorem 5.7]), these SQ upper bounds transform into private learning upper bounds. They yield, to the best of our knowledge, the first private learning algorithms for manifold estimation. More precisely, we proved that for all $\varepsilon = O(\mathrm{rch}_{\min})$, there is a local δ-differentially[1] private algorithm estimating the d-dimensional $\mathcal{C}^2$-manifolds M with precision ε that requires no more than

$$s_{\delta\text{-private}}(\varepsilon) = \tilde{O}\left(\frac{n}{(f_{\min}\mathrm{rch}_{\min}^d)^3\delta^2}\left(\frac{\mathrm{rch}_{\min}}{\varepsilon}\right)^{(3d+2)/2}\right)$$

[1] As the present paper uses ε for precision, we use δ as the privacy parameter, contrary to the standard notation.

samples in the fixed point model, where $\tilde{O}$ hides the logarithmic terms of the complexity. See [KLN+11] for more formal and thorough developments on differential privacy.

Lower Bounds. Complementing these upper bounds on the statistical query complexity of manifold estimation, we prove a *computational* and an *informational* lower bound, that both nearly match. To examine whether or not randomness may facilitate manifold learning, the below lower bounds apply to *randomized* SQ algorithms, which are allowed to use randomness and to fail with probability at most $\alpha \in [0, 1)$. Recall that the above upper bounds stand for *deterministic* ($\alpha = 0$) SQ algorithms.

First, we prove the following computational lower bounds.

[Fixed point model] Theorem 5.3 asserts that any *randomized* SQ algorithm estimating M with precision ε and probability of error at most α in the fixed point model requires at least

$$q = \Omega\left(\frac{\frac{n}{f_{\min}}\left(\frac{1}{\mathrm{rch}_{\min}\varepsilon}\right)^{d/2} + \log(1-\alpha)}{\log\left(1+\frac{1}{\tau}\right)}\right)$$

queries to a STAT(τ) oracle.

[Bounding ball model] Theorem 5.6 states that any *randomized* SQ algorithm estimating M with precision ε and probability of error at most α in the bounding ball model requires at least

$$q = \Omega\left(\frac{n\log\left(\frac{R}{\varepsilon}\right) + \frac{n}{f_{\min}}\left(\frac{1}{\mathrm{rch}_{\min}\varepsilon}\right)^{d/2} + \log(1-\alpha)}{\log\left(1+\frac{1}{\tau}\right)}\right)$$

queries to a STAT(τ) oracle.

In words, this proves that for any fixed probability of error $\alpha < 1$, the above manifold estimators are optimal up to a polylog$(n, 1/\tau)$ factor. Hence, randomized algorithms are not significantly more powerful than deterministic ones in these models.

Second, we establish informational lower bounds (Theorems 5.2 and 5.5) that advocate for the necessity of the requirements on ε and τ made in the upper bounds. More precisely, they assert that if we either have $\varepsilon = o\left(\mathrm{rch}_{\min}\left(\frac{\tau}{f_{\min}\mathrm{rch}_{\min}^d}\right)^{2/d}\right)$, or $\tau = \Omega(f_{\min}\mathrm{rch}_{\min}^d)$ and $\varepsilon = o(\mathrm{rch}_{\min})$, then no SQ algorithm (even randomized) can estimate manifolds in these models with precision ε, regardless of its number of queries. Said otherwise, the adversarial tolerance parameter τ prevents the learner to have arbitrary precision ε , with quantitative lower bound

$$\varepsilon = \Omega\left(\mathrm{rch}_{\min}\min\left\{1, \left(\frac{\tau}{f_{\min}\mathrm{rch}_{\min}^d}\right)^{2/d}\right\}\right)$$

no matter the computational power of the statistician.

The high level exposition of these lower bounds stands in Section 1.2.4, and all the necessary details and formal statements in Section 5.

1.2.2 Manifold Propagation Algorithm

The core component of the upper bounds (Theorems 5.1 and 5.4) is a purely geometric algorithm, which we call `Manifold Propagation`, parametrized by an initialization method $\hat{x}_0$ and two routines $\hat{T}(\cdot)$ and $\hat{\pi}(\cdot)$ related to the manifold M:

(Seed point) This initialization method finds a point $\hat{x}_0$ that is η-close to M, for some $\eta \geq 0$.

(Tangent space) Given a point x_0 that is η-close to M, this routine finds a linear subspace $\hat{T}(x_0)$ that is $(\sin\theta)$-close to the tangent space $T_{\pi_M(x_0)}M$ at its projection $\pi_M(x_0)$ (i.e. the closest point on M), for some $\theta \geq 0$.

(Projection) Given a point x_0 that is Λ-close to M, this routine finds a point $\hat{\pi}(x_0)$ that is η-close to its projection $\pi_M(x_0)$, where $\Lambda \geq \eta$.

Then, given a tuning scale parameter $\Delta > 0$, `Manifold Propagation` iteratively explores the connected manifold M starting from the seed point, and constructs a $\Omega(\Delta)$-sparse and $O(\Delta)$-dense point cloud $\mathcal{O}$ of points close to M (see Theorem 3.1). This algorithm is reminiscent of breadth-first search and can be roughly described as follows:

1. Start with the seed point $\hat{x}_0$ in the vicinity of the manifold and initialize a queue of points to $\mathcal{Q} = \{\hat{x}_0\}$, and the final output point cloud to $\mathcal{O} = \emptyset$.
2. Pick a point $x_0 \in \mathcal{Q}$, remove x_0 from $\mathcal{Q}$ and add it to $\mathcal{O}$. Compute the approximate tangent space $\hat{T}(x_0)$ of M at x_0.
3. Consider a covering $y_1, \ldots, y_k$ of a sphere of radius Δ in $\hat{T}(x_0)$. To avoid backtracking, remove all of these points that are too close to (the already-visited) points from $\mathcal{Q} \cup \mathcal{O}$. To account for the linear approximation made and the past estimation errors, "project" the remaining points y_i's on M with $\hat{\pi}(\cdot)$ and add them to $\mathcal{Q}$.
4. If $\mathcal{Q}$ is empty, terminate and output $\mathcal{O}$. Otherwise go to Step 2.

Note the importance of the proximity check of Step 3, without which the algorithm would not terminate, even with infinitely precise routines.

Then, given such a point cloud $\mathcal{O}$ that forms a $O(\Delta)$-dense sample of M, existing algorithms from computational geometry (Theorem 2.1) allow to reconstruct a manifold with precision $O(\Delta^2/\mathrm{rch}_{\min})$. This quadratic gain is made possible by the $\mathcal{C}^2$-smoothness of M [BG14, AL18]. Hence, running `Manifold Propagation` with step size $\Delta = O(\sqrt{\mathrm{rch}_M \varepsilon})$ and applying Theorem 2.1 yields a dynamic method to estimate a manifold M with reach $\mathrm{rch}_{\min} > 0$. Namely, to estimate M with precision ε in Hausdorff distance, it shows that it is sufficient to design routines for M that have precision $\eta = O(\varepsilon)$ for the seed point, $\sin\theta = O(\sqrt{\varepsilon/\mathrm{rch}_{\min}})$ for the tangent spaces, and $\eta = O(\varepsilon)$ for the projection.

To our knowledge, this provides the first computational geometric result involving the three routines above only. It also does a single call to $\hat{x}_0$ for initialization, and provably no more than $O(\mathcal{H}^d(M)/\Delta^d) = O(\mathcal{H}^d(M)/(\mathrm{rch}_{\min}\varepsilon)^{d/2})$ calls to the routines $\hat{\pi}(\cdot)$ and $\hat{T}(\cdot)$, where $\mathcal{H}^d(M)$ stands for the surface area of M. In particular, this number of calls is blind to the ambient dimension. Overall, `Manifold Propagation` manages to have this possible ambient dependency supported by $\hat{x}_0$, $\hat{T}(\cdot)$ and $\hat{\pi}(\cdot)$ only. See Section 3 for the formal statements and further discussion.

1.2.3 Statistical Query Algorithms for the Routines

In order to plug `Manifold Propagation` in the SQ framework, we then provide SQ implementations of its geometric routines.

Projection Routine. As mentioned above, the projection routine should allow to find a point $\hat{\pi}(x_0)$ that is η-close to $\pi_M(x_0)$, provided that x_0 is Λ-close to M. To implement this using a $\mathrm{STAT}(\tau)$ oracle, we first note that the conditional expectation of D in the neighborhood of x_0 has small bias for estimating $\pi_M(x_0)$. That is, $\|\pi_M(x_0) - \mathbb{E}_{x\sim D}[x \,|\mathrm{B}(x_0,h)]\|$ is small for a properly tuned bandwidth h (see Lemma C.2). Hence, it is enough to estimate

$$m_D(x_0,h) = \mathop{\mathbb{E}}_{x\sim D}[x\,|\mathrm{B}(x_0,h)] = x_0 + h\frac{\mathbb{E}_{x\sim D}\left[\frac{(x-x_0)}{h}\cdot \mathbb{1}_{\|x-x_0\|\leq h}\right]}{D(\mathrm{B}(x_0,h))},$$

where $D(\mathrm{B}(x_0, h))$ stands for the mass of the ball $\mathrm{B}(x_0, h)$ for distribution D. Written as a ratio of two means, one easily sees how to estimate $m_D(x_0, h)$ in $\mathrm{STAT}(\tau)$, as we now explain. The denominator only requires one query $r = \mathbb{1}_{\mathrm{B}(x_0,h)}$ to the oracle. As about the numerator, which is a n-dimensional mean vector, the naive approach that would query each coordinate of its integrand separately would end up with the dimension-dependent precision $\sqrt{n}\tau$ in Euclidean norm. Instead, by using tight frames, an algorithm of Feldman, Guzmán, and Vempala [FGV21] allows to get precision $O(\tau)$ in only $2n$ queries.
At the end of the day, we achieve precision $\eta = O\left(\Lambda^2/\mathrm{rch}_{\min}\right)$ with $O(n)$ queries to $\mathrm{STAT}(\tau)$, provided that: (see Theorem 4.1)

$$\Lambda = \Omega\left(\mathrm{rch}_{\min}\left(\frac{\tau}{f_{\min}\mathrm{rch}_{\min}^d}\right)^{1/(d+1)}\right).$$

Tangent Space Routine. Here, the tangent space routine should allow to estimate the tangent space $T_{\pi_M(x_0)}M$ of M at $\pi_M(x_0)$, provided that x_0 is η-close to M. Local Principal Component Analysis proved fruitful in the sample framework [AL18]. Inspired by it, we notice that the local covariance matrix of D at x_0

$$\Sigma_D(x_0, h) = \mathop{\mathbb{E}}_{x\sim D}\left[\frac{(x-x_0)(x-x_0)^\top}{h^2}\mathbb{1}_{\|x-x_0\|\le h}\right]$$

allows to approximate $T_{\pi_M(x_0)}M$. That is, $\Sigma_D(x_0, h)$ is almost rank-d and its first d principal components span a d-plane close to $T_{\pi_M(x_0)}M$, for a properly tuned bandwidth h (see Lemma D.1). Next, aiming at estimating $\Sigma_D(x_0, h) \in \mathbb{R}^{n\times n}$ in $\mathrm{STAT}(\tau)$, we note that seeing it as a mean vector of $\mathbb{R}^{n^2}$ and using tight frames [FGV21] directly would cost $O(n^2)$ queries for precision $O(\tau)$, but would not exploit the low-rank (hence redundant) structure of $\Sigma_D(x_0, h)$. Instead, we use matrix compression arguments [FCRP08] to present a new general SQ algorithm estimating low-rank mean matrices (Lemma D.4). This allows to mitigate the query complexity from $O(n^2)$ to $O(dn\log^6(n))$ while keeping precision $O(\tau)$ in Frobenius norm.
At the end of the day, coming back to our initial problem of tangent space estimation in $\mathrm{STAT}(\tau)$, we achieve precision $\sin\theta = O\left(\sqrt{\eta/\mathrm{rch}_{\min}}\right)$ with $O(dn\,\mathrm{polylog}(n))$ queries to $\mathrm{STAT}(\tau)$, provided that: (see Theorem 4.2)

$$\eta = \Omega\left(\mathrm{rch}_{\min}\left(\frac{\tau}{f_{\min}\mathrm{rch}_{\min}^d}\right)^{2/(d+1)}\right).$$

Seed Point Detection. Finally, the seed point $\hat{x}_0$ should be η-close to M. In the fixed point model, this method is trivial since $0 \in M$ by assumption. In the bounding ball model, where it is only assumed that $M \subseteq \mathrm{B}(0, R)$, we proceed in two stages:

- First, we use a divide and conquer strategy over $\mathrm{B}(0, R)$ (Theorem E.1). The algorithm (`SQ Ambient Binary Search`) recurses down over a discretization of $\mathrm{B}(0, R)$ with unions of small balls, maintained to intersect $M = \mathrm{Supp}(D)$ by querying their indicator functions, i.e. by checking that they have non-zero mass for D. It stops when there is only one ball left and outputs its center $\hat{x}_0^{raw}$. Unfortunately, the output point $\hat{x}_0^{raw}$ of this simple strategy might only be $O(\mathrm{rch}_{\min}\left(\tau/(f_{\min}\mathrm{rch}_{\min}^d)\right)^{1/d})$-close to M, since this procedure does not use the $\mathcal{C}^2$-smoothness of M.

- Starting from $\hat{x}_0^{raw}$, the algorithm applies iteratively the projection routine $\hat{\pi}(\cdot)$ described above. Since $\hat{x}_0^{raw}$ is already close to M, the procedure is guaranteed to improve precision quadratically at each step, and to output a point $\hat{x}_0$ that is η-close to M after a logarithmic number of iterations.

At the end of the day, we achieve precision η with $O(n\log(R/\eta))$ queries to $\mathrm{STAT}(\tau)$, provided that: (see Theorem 4.3)

$$\eta = \Omega\left(\mathrm{rch}_{\min}\left(\frac{\tau}{f_{\min}\mathrm{rch}_{\min}^d}\right)^{2/(d+1)}\right).$$

1.2.4 Lower Bound Techniques

The standard SQ lower bound techniques, such as those involving the so-called statistical dimension [Fel17], appear to be ill-suited to our context. In fact, the informational bounds on the statistical dimension naturally involve Kullback-Leibler or chi-squared divergences [Fel17, DKS17], which are non-informative in non-dominated statistical models such as manifolds ones. Indeed, two low-dimensional submanifolds $M_0, M_1 \subseteq \mathbb{R}^n$ that are not equal would differ in a non-trivial area, yielding distributions are not absolutely continuous with respect to one another. This results in infinite Kullback-Leibler and chi-squared divergences, and hence non-informative lower bounds.

To overcome this issue, we present a two-step technique — involving a computational and an informational lower bound — that does not involve these quantities. The method applies in general metric spaces (see Appendix G), although we shall limit its exposition and application to the case of manifolds with Hausdorff distance in this introduction.

Computational Lower Bounds. We aim at deriving a lower bound the number $q(\varepsilon)$ of queries necessary to achieve precision ε. For this, we observe that since a SQ algorithm should cope with *any* adversarial oracle, it has to cope with the oracle that answers roundings $\mathrm{a} = \tau\lfloor \mathbb{E}_{x\sim D}[r(x)]/\tau\rfloor$ of the true queried mean to the nearest integer multiple of τ in $[-1,1]$. As this oracle only generates $(1+1/\tau)$ different answers, any SQ manifold estimation algorithm that makes only q queries to this discrete oracle produces at most $\mathcal{N} \leq (1+1/\tau)^q$ possible outputs $\hat{M}$. Hence, if this estimator has precision ε, these outputs should form an ε-covering of the manifold class of interest $\mathcal{M}$. Hence, deriving a lower bound on $q = q(\varepsilon) \geq \log\mathcal{N}(\varepsilon)/\log(1+1/\tau)$ boils down to deriving a lower bound on the ε-covering number of $\mathcal{M}$ for the Hausdorff distance, or equivalently, by duality, on its ε-packing number. This argument also extends to randomized SQ algorithms (see Appendix G.2).

We then explicitly build ε-packings of the manifold classes associated to the models, with a general combinatorial scheme (see Proposition H.5) based on a single initial manifold M_0. The construction bumps M_0 at many different locations, with bumps of height $\Omega(\varepsilon)$ scattered in all the available $(n-d)$ codimensions of space, as shown in Figure 1. Note that the $\mathcal{C}^2$-like

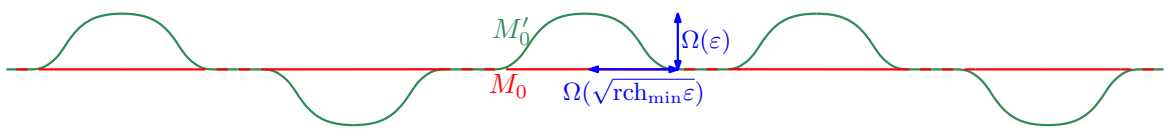


Figure 1: Construction of an ε-packing of the manifold class by local bumping. Here, in dimension $d = 1$ and codimension $n - d = 1$, each bump has the two options "upwards" and "downwards" within each of the $N \gg 1$ locations, yielding 2^N ε-separated manifolds.

assumption $\mathrm{rch}_M \geq \mathrm{rch}_{\min}$ forces to spread each of these bumps on a ball of radius $\Omega(\sqrt{\mathrm{rch}_{\min}\varepsilon})$. Intuitively, in this construction, the larger the surface area $\mathcal{H}^d(M_0)$ of the base manifold M_0, the more room to include many bumps on it, and hence the stronger the lower bound. Hence, in the bounding ball model, we exhibit manifolds that can have large volume, while remaining in $\mathrm{B}(0,R)$ and with reach $\mathrm{rch}_M \geq \mathrm{rch}_{\min}$. This is done by gluing next to each other linkable *widgets* along a long path in the cubic grid in $\mathrm{B}(0,R)$ (see Appendix H.2). Overall, this construction

allows to get the correct dependency in $1/f_{\min}$ — which plays the role of a maximal volume, see Section 2.2.3 — in the bounds.

[Fixed point model] If $0 \in M_0$, the above construction is possible while preserving the point $0 \in \mathbb{R}^n$ within all the bumped manifolds, yielding the lower bound (Theorem 5.3).

[Bounding ball model] As in this model, no point is fixed, we build another type of ε-packing by translating a base manifold $M_0 \subseteq \mathrm{B}(0, R/2)$ in the ambient space by all the possible vectors of an ε-packing of the ambient ball $\mathrm{B}(0, R/2)$, which has size $\Omega((R/\varepsilon)^n)$. This yields the first term of the lower bound, while the second term follows as described above, by locally bumping a manifold $M_0 \subseteq \mathrm{B}(0, R)$ (Theorem 5.6).

Informational Lower Bounds. In addition, forgetting about the number of queries SQ algorithms may do, they have a limited precision ε given tolerance τ. Hence, aiming at lower bounding this best precision $\varepsilon(\tau)$ achievable in $\mathrm{STAT}(\tau)$, we notice that two distributions that are closer than $\tau/2$ in total variation distance allow an adversarial oracle to swap their respective answers, and hence make them — and their supports — indistinguishable using SQ's. This idea is at the core of standard lower bounds in the sample framework [Yu97], and is formalized in the so-called *Le Cam's lemma* for SQ's (Theorem G.1).

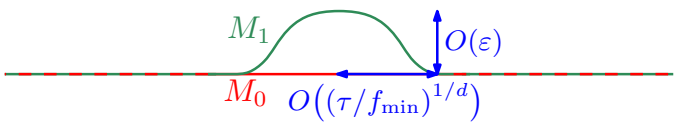


Figure 2: Indistinguishable manifolds for the informational lower bound. The measure on which they differ being of order $O(\tau)$, an adversarial $\mathrm{STAT}(\tau)$-oracle may fool the learner and force them to make an error proportional to the Hausdorff distance between them.

To build such indistinguishable manifolds, we locally bump a base manifold M_0 at a single location. As M_0 supports a d-dimensional measure with density lower bounded by $f_{\min}$, the largest possible width of such a bump is of order $\delta = \Omega((\tau/f_{\min})^{1/d})$, since the d-volume of this area multiplied by $f_{\min}$ (i.e. total variation) gets of order $\Omega(\tau)$. Similarly as above, given the width δ of this bump, its largest possible height is $\varepsilon = \Omega\left(\delta^2/\mathrm{rch}_{\min}\right) = \Omega\left(\mathrm{rch}_{\min}\left(\frac{\tau}{f_{\min}\mathrm{rch}_{\min}^d}\right)^{2/d}\right)$, which provides the ε-separated manifolds indistinguishable in $\mathrm{STAT}(\tau)$. This yields the announced informational lower bounds (Theorems 5.2 and 5.5), after picking manifolds M_0 in the fixed point and bounding model respectively that have volume of order $1/f_{\min}$, and the uniform distributions on them.

2 Preliminaries

2.1 Statistical Query Framework

To begin with the formal presentation of this work, let us define the statistical query (SQ) framework used throughout the paper. In the SQ framework, the algorithm (or learner) is allowed to access to the unknown underlying distribution D over $\mathbb{R}^n$ via an adversarial oracle O who knows it perfectly. The learner also has access to some prior information on D via the knowledge of a *model* $\mathcal{D}$, i.e. a set of probability distributions over $\mathbb{R}^n$ assumed to contain D. For a measurable function $r : \mathbb{R}^n \to [-1, 1]$, called *query*, the oracle answers the mean value $\mathbb{E}_{x\sim D}[r(x)]$ of r with respect to D, up to some adversarial error τ known to both parties.

Let $\mathfrak{F}$ denote the set of Borel-measurable functions from $\mathbb{R}^n$ to $[-1, 1]$. An *oracle* provides answers $\mathrm{a} : \mathfrak{F} \to \mathbb{R}$. Given a query $r \in \mathfrak{F}$ and a tolerance parameter $\tau \geq 0$, we say that O is a *valid*

STAT(τ) oracle for the distribution D over $\mathbb{R}^n$ if its answers are such that $|\mathrm{a}(r) - \mathbb{E}_{x\sim D}[r(x)]| \leq \tau$. Let us insist on the fact that the oracle is adversarial, meaning that it can answer *any* such values. Its adversarial strategy can also adapt to the previous queries made by the learner. See formal Definition 2.1).

We now describe the estimation framework using SQ's. Given a metric space (Θ, ρ), a target precision $\varepsilon > 0$ and a parameter of interest $\theta : \mathcal{D} \to \Theta$, the learner aims at estimating $\theta(D)$ with precision ε for the metric ρ with a minimum number of queries $r : \mathbb{R}^n \to [-1, 1]$, uniformly over the model $\mathcal{D}$. The present framework is a particular case of the *search problems* considered in [Fel17], where a metric on Θ is not necessarily available.

Remark 2.1. Manifold estimation will naturally bring us to consider the support $\theta(D) = \mathrm{Supp}(D) \subseteq \mathbb{R}^n$ as the parameter of interest, and the Hausdorff distance $\rho = \mathrm{d_H}$. However, we present the broader setting of a general metric space (Θ, ρ) of estimation, to also cover the intermediate results required by the SQ versions of the routines of `Manifold Propagation` (see Section 4). Namely, it will involve the Euclidean space $(\mathbb{R}^n, \|\cdot\|)$ for point estimation, and the matrix spaces $(\mathbb{R}^{n\times n}, \|\cdot\|_{\mathrm{F}})$ and $(\mathbb{R}^{n\times n}, \|\cdot\|_{\mathrm{op}})$.

This paper considers interactive SQ algorithms, meaning that the learner is allowed to interact with the oracle dynamically and does not have to send all their queries at once before any answer. That is, query functions are allowed to depend arbitrarily on the previous answers given by the oracle. More formally, we give the following Definition 2.1.

Definition 2.1 (Deterministic Statistical Query Estimation Framework)**.**

- A *statistical query algorithm making q queries* is a tuple $\mathtt{A} = (r_1, \ldots, r_q, \hat{\theta})$, where

$$r_1 \in \mathfrak{F},\ r_2 : \mathbb{R} \to \mathfrak{F},\ \cdots,\ r_q : \mathbb{R}^{q-1} \to \mathfrak{F},\ \text{and } \hat{\theta} : \mathbb{R}^q \to \Theta.$$

- Let $\mathrm{a}_1 : \mathfrak{F} \to \mathbb{R}$, $\mathrm{a}_2 : \mathfrak{F}^2 \to \mathbb{R}$, $\ldots$, $\mathrm{a}_q : \mathfrak{F}^q \to \mathbb{R}$. We say that $\mathsf{O} = (\mathrm{a}_1, \ldots, \mathrm{a}_q)$ is a STAT(τ) *oracle for SQ algorithms making q queries on the distribution D*, if for all $r_1 \in \mathfrak{F}$, $r_2 : \mathbb{R} \to \mathfrak{F}$, $\ldots$, $r_q : \mathbb{R}^{q-1} \to \mathfrak{F}$,

$$\begin{aligned} \big|\mathrm{a}_1(r_1) - \mathop{\mathbb{E}}_{x\sim D}\big[r_1(x)\big]\big| &\leq \tau, \\ \big|\mathrm{a}_2(r_1, r_2) - \mathop{\mathbb{E}}_{x\sim D}\big[r_2(\mathrm{a}_1(r_1))(x)\big]\big| &\leq \tau, \\ &\vdots \\ \big|\mathrm{a}_q(r_1, \ldots, r_q) - \mathop{\mathbb{E}}_{x\sim D}\big[r_q(\mathrm{a}_1(r_1), \ldots, \mathrm{a}_{q-1}(r_1))(x)\big]\big| &\leq \tau. \end{aligned}$$

- The *output* of $\mathtt{A} = (r_1, \ldots, r_q, \hat{\theta})$ when it interacts with the oracle $\mathsf{O} = (\mathrm{a}_1, \ldots, \mathrm{a}_q)$ is defined by

$$\hat{\theta}(r_1, \ldots, r_q; \mathsf{O}) = \hat{\theta}(\mathrm{a}_1(r_1), \ldots, \mathrm{a}_q(r_1, \ldots, r_q)).$$

- Given a model $\mathcal{D}$ over $\mathbb{R}^n$ (i.e. a set of probability distributions), we say that a SQ algorithm $\mathtt{A}$ is a STAT(τ) *estimator with precision ε for the statistical estimation problem $\theta : \mathcal{D} \to \Theta$* if for all $D \in \mathcal{D}$ and all valid STAT(τ) oracle O for D,

$$\rho\big(\theta(D), \hat{\theta}(r_1, \ldots, r_q; \mathsf{O})\big) \leq \varepsilon.$$

Beyond deterministic algorithms, one may allow the learner to access randomness, and to fail at estimating the parameter of interest with some controlled probability $\alpha < 1$ [Kea98, Fel17]. This gives rise to the following Definition 2.2.

Definition 2.2 (Randomized SQ Estimation Framework)**.**

- A *randomized SQ algorithm* $\mathcal{A}$ is a distribution over SQ algorithms.
- Given a model $\mathcal{D}$ over $\mathbb{R}^n$, we say that a randomized SQ algorithm $\mathcal{A}$ is a STAT(τ) *algorithm with precision* ε *and probability of failure (or error)* α *over* $\mathcal{D}$, if for all distribution $D \in \mathcal{D}$ and all valid STAT(τ) oracle O for D,

$$\Pr_{\mathtt{A}=(r_1,\ldots,r_q,\hat{\theta})\sim\mathcal{A}} \left[\rho\big(\theta(D), \hat{\theta}(r_1,\ldots,r_q;\mathsf{O})\big) \leq \varepsilon\right] \geq 1-\alpha.$$

Naturally, any SQ algorithm making at most q queries can be emulated into a PAC algorithm by considering empirical averages. Indeed, given independent and identically distributed data $\{X_1,\ldots,X_s\}$ with common distribution D, Hoeffding's inequality yields that the oracle which answers $\mathrm{a} = \frac{1}{s}\sum_{i=1}^{s} r(X_i)$ to any query $r : \mathbb{R}^n \to [-1,1]$ satisfies $|\mathrm{a} - \mathbb{E}_P[r]| \leq \tau$ with probability $\geq 1 - 2e^{-s\tau^2/2}$. Therefore, if we can estimate $\theta : \mathcal{D} \to \Theta$ "efficiently" in STAT(τ), so do we in the PAC (sample) framework, with sample complexity $s \leq \frac{q\sqrt{\log(q/\alpha)}}{\tau^2}$. .

A priori, randomized algorithms may require significantly less queries than deterministic ones to achieve an estimation task [Fel17]. However, we will show that this phenomenon does not occur for manifold estimation, as soon as the probability of error α is not considerably close to 1. For this, we will exhibit upper bounds using *deterministic* algorithms, and matching lower bounds on *randomized* algorithms. See Section 5 for the precise statements.

Remark 2.2 (About Noise)**.** The statistical models considered in this work (Definition 2.5) are noise-free, in the sense that the STAT(τ) oracle — although adversarial — has access to the exact underlying distribution D. Beyond such an exact model, a noise model that is particularly popular in the manifold estimation literature is the so-called *clutter noise* model [AL18, GPPVW12a]. Given a nuisance parameter $\beta \in (0,1]$ and a fixed noise distribution Q_0 over $\mathbb{R}^n$ — usually the ambient uniform distribution over a compact set of $\mathbb{R}^n$ —, the associated clutter noise model is the set of mixtures

$$\mathcal{D}^{(\text{clutter})}_{\beta,Q_0} = \{\beta D + (1-\beta)Q_0, D \in \mathcal{D}\}.$$

In particular, $\mathcal{D}^{(\text{clutter})}_{\beta=1,Q_0}$ coincides with $\mathcal{D}$. For $\beta < 1$, in the independent and identically distributed (i.i.d.) sampling framework, it yields samples with a proportion of approximately β informative points and $(1-\beta)$ of non-informative clutter points.

As mentioned in the introduction, this type of noise model gave rise to subtle iterative decluttering procedures that rely heavily on the properties of Q_0 (i.e. being ambient uniform) [AL18]. This noise-specificity is also a limitation of the (intractable) estimator of [GPPVW12a], which would also fail with clutter distributions Q_0 other than uniform. In contrast, in the statistical query framework, if β and Q_0 are known, then estimation techniques need not be much more elaborate for the case $\beta < 1$ than for $\beta = 1$. Indeed, the statistical query complexity of an estimation problem in STAT(τ) over $\mathcal{D}^{(\text{clutter})}_{\beta,Q_0}$ coincides with its counterpart in STAT(τ/β) over $\mathcal{D}$.

The correspondence is explicit: algorithms designed for $\beta = 1$ naturally generalize for $\beta < 1$ and vice-versa. To see this, let $r : \mathbb{R}^n \to [-1,1]$ be a query to a STAT(τ) oracle with true distribution $D^{(\text{clutter})} = \beta D + (1-\beta)Q_0$. Say that the learner gets answer $\mathrm{a} \in \mathbb{R}$, then the function returning $\mathrm{a}' = (\mathrm{a} - (1-\beta)\,\mathbb{E}_{Q_0}[r])/\beta$, which can be computed by the learner who knows Q_0 and β, clearly simulates a valid STAT(τ/β) oracle to the query for the distribution D. For the same reason, conversely, any STAT(τ/β)-algorithm over $\mathcal{D}$ yields a STAT(τ)-algorithm in $\mathcal{D}^{(\text{clutter})}_{\beta,Q_0}$.

This shows that the statistical query complexity in STAT(τ) over $\mathcal{D}$ coincides with its counterpart in STAT($\beta\tau$) over $\mathcal{D}^{(\text{clutter})}_{\beta,Q_0}$ for any fixed $0 < \beta \leq 1$ and clutter distribution Q_0.

Conversely, any SQ algorithm in the clutter-free model can easily be made robust to clutter noise, as soon as the clutter distribution Q_0 and noise level $0 < \beta \leq 1$ are known to the learner.

As a first illustration of a non-trivial SQ estimation problem, let us describe that of the mean $\mathbb{E}_{x\sim D}[F(x)]$ of a bounded vector-valued function $F : \mathbb{R}^n \to \mathbb{R}^k$, where $\|F(x)\| \leq 1$, see [FGV21]. Here and below, $\|\cdot\|$ stands for the Euclidean norm. This example will be central in the construction of our SQ geometric routines (Theorems 4.1 to 4.3), and hence for the final SQ manifold estimation algorithms (Theorems 5.1 and 5.4).

One query to a STAT(τ) oracle allows to compute the mean value of a function ranging in $[-1, 1]$ with precision τ. Hence, the k coordinate functions $r_i(x) = \langle e_i, F(x)\rangle \in [-1, 1]$ are valid queries, and allow to estimate each coordinate of $\mathbb{E}_{x\sim D}[F(x)]$ with precision τ. This naive strategy results in a deterministic SQ algorithm making k queries to STAT(τ) and precision τ for the sup-norm, but only $\sqrt{k}\tau$ for the Euclidean norm. The following Lemma 2.1 shows that the learner may ask $2k$ queries to a STAT(τ) oracle, while still preserving a precision of order τ for the Euclidean norm. The strategy consists in querying F in a suitable frame of $\mathbb{R}^k$ [FGV21, Theorem 3.9], i.e. a redundant family of vectors of $\mathbb{R}^k$ which avoids the extra $\sqrt{k}$ factor of the non-redundant coordinate-wise queries.

Lemma 2.1. *Let D be a Borel probability distribution on $\mathbb{R}^n$, and $F : \mathbb{R}^n \to \mathbb{R}^k$ be such that $\|F(x)\| \leq 1$ for all $x \in \mathbb{R}^n$.*

There exists a deterministic statistical query algorithm making $2k$ queries to a STAT(τ) oracle, and that estimates $\mathbb{E}_{x\sim D}[F(x)]$ with precision $C\tau$ for the Euclidean norm, where $C > 0$ is a universal constant.

Proof of Lemma 2.1. Let us denote by D' the pushforward distribution of D by F. As for all measurable function $r : \mathbb{R}^k \to [-1, 1]$,

$$\mathop{\mathbb{E}}_{x'\sim D'}[r(x')] = \mathop{\mathbb{E}}_{x\sim D}[r(F(x))],$$

any valid STAT(τ) oracle for D simulates a valid STAT(τ) oracle for D'. Hence, applying [FGV21, Theorem 3.9] to D', we get the desired result. □

2.2 Manifold Regularity and Distributional Assumptions

2.2.1 General Notation and Differential Geometry

From now on, $n \geq 2$ is referred to as the ambient dimension and $\mathbb{R}^n$ is endowed with the Euclidean inner product $\langle\cdot,\cdot\rangle$ and the associated norm $\|\cdot\|$. The closed Euclidean ball of center x and radius r is denoted by $\mathrm{B}(x, r)$. The volume of the d-dimensional unit ball $\mathrm{B}_d(0, 1)$ is denoted by ω_d, and that of the d-dimensional unit sphere $\mathcal{S}^d(0, 1) \subseteq \mathbb{R}^{d+1}$ by σ_d.

We will consider compact connected submanifolds M of $\mathbb{R}^n$, without boundary, and with dimension $d < n$ [dC92]. Given a point $p \in M$, the tangent space of M at p, denoted by T_pM, is the d-dimensional linear subspace of $\mathbb{R}^n$ spanned by the velocity vectors at p of $\mathcal{C}^1$ curves of M. The Grassmannian $\mathbb{G}^{n,d}$ is the set of all the d-dimensional linear subspaces of $\mathbb{R}^n$, so that $T_pM \in \mathbb{G}^{n,d}$ for all $p \in M$. In addition to the Euclidean structure induced by $\mathbb{R}^n$ on $M \subseteq \mathbb{R}^n$, we also endow M with its intrinsic geodesic distance d_M, with $\mathrm{B}_M(p, s)$ denoting the closed geodesic ball of center $p \in M$ and of radius s. More precisely, given a $\mathcal{C}^1$ curve $c : [a, b] \to M$, the length of c is defined as $\mathrm{Length}(c) = \int_a^b \|c'(t)\|\, dt$. Given $p, q \in M$, there always exists a path $\gamma_{p\to q}$ of minimal length joining p and q [dC92]. Such a curve $\gamma_{p\to q}$ is called geodesic, and the geodesic distance between p and q is given by $\mathrm{d}_M(p, q) = \mathrm{Length}(\gamma_{p\to q})$ [BBI01, Chapter 2]. In particular, (M, d_M) is a length space [BBI01, Remark 5.1.7]. A geodesic γ such that $\|\gamma'(t)\| = 1$ for all t is called arc-length parametrized. Unless stated otherwise, we always assume that geodesics are parametrized by arc-length. For all $p \in M$ and all unit vectors $v \in T_pM$, we denote by $\gamma_{p,v}$

the unique arc-length parametrized geodesic of M such that $\gamma_{p,v}(0) = p$ and $\gamma'_{p,v}(0) = v$ [dC92]; the exponential map is defined as $\exp_p^M(vt) = \gamma_{p,v}(t)$. Note that from the compactness of M, $\exp_p^M : T_pM \to M$ is defined globally on T_pM [BBI01, Theorem 2.5.28].

2.2.2 Geometric and Statistical Models

Let us detail the geometric assumptions we will make throughout. Besides the differential structure given by low-dimensional submanifolds, the core regularity assumption of this work will be encoded by the *reach*, a central quantity in the statistical analysis of geometric structures (see [AKC+19] and references therein), and that we now describe.

To this aim, let us define the *medial axis* $\mathrm{Med}(K)$ of a closed subset $K \subseteq \mathbb{R}^n$ as the set of ambient points that have at least two nearest neighbors on K. Namely, if we let $\mathrm{d}(z, K) = \inf_{p\in K} \|p - z\|$ denote the distance function to K,

$$\mathrm{Med}(K) = \{z \in \mathbb{R}^n \mid \exists p \neq q \in K, \|p - z\| = \|q - z\| = \mathrm{d}(z, K)\}.$$

By definition of the medial axis, the *metric projection onto* K, given by

$$\pi_K(z) = \operatorname*{argmin}_{p\in K} \|p - z\|$$

is well defined exactly on $\mathbb{R}^n \setminus \mathrm{Med}(K)$. The reach of K is then defined as the minimal distance from K to $\mathrm{Med}(K)$.

Definition 2.3 ([Fed59, Theorem 4.18])**.** The *reach* of a closed subset $K \subseteq \mathbb{R}^n$ is defined by

$$\mathrm{rch}_K = \inf_{z\in \mathrm{Med}(K)} \mathrm{d}(z, K).$$

Furthermore, if $K = M \subseteq \mathbb{R}^n$ is a $\mathcal{C}^2$-submanifold, then its reach can be written as

$$\mathrm{rch}_M = \inf_{p\neq q\in M} \frac{\|q - p\|^2}{\mathrm{d}(q - p, T_pM)}.$$

The second formulation of Definition 2.3 assesses how a large positive reach testifies of a quantitative uniform regularity of $M \subseteq \mathbb{R}^n$. Indeed, the submanifold M being $\mathcal{C}^2$-smooth essentially means that locally, M deviates at most quadratically from its tangent spaces. Adding the condition $\mathrm{rch}_M \geq \mathrm{rch}_{\min} > 0$ yields the quantitative bound $\mathrm{d}(q-p, T_pM) \leq \|q - p\|^2 / (2\mathrm{rch}_{\min})$ for all $p, q \in M$. In particular, this condition bounds curvature and intrinsic metric properties (see Lemma 2.2). We shall refer the interested reader to [AKC+19] for further discussions on the reach.

Definition 2.4. We let $\mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$ denote the class of compact connected d-dimensional $\mathcal{C}^2$-submanifolds M of $\mathbb{R}^n$, without boundary, that have reach bounded below by $\mathrm{rch}_M \geq \mathrm{rch}_{\min}$.

Among the key properties shared by submanifolds M with reach bounded below $\mathrm{rch}_M \geq \mathrm{rch}_{\min}$ are a quantitative equivalence between the Euclidean and geodesic distances, and the fact that their curvature is uniformly bounded by $1/\mathrm{rch}_{\min}$.

Lemma 2.2. *Let $M \in \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$ and $p, q \in M$. If $\|q - p\| < 2\mathrm{rch}_{\min}$, then*

$$\|q - p\| \leq \mathrm{d}_M(p, q) \leq 2\mathrm{rch}_{\min} \arcsin\left(\frac{\|q - p\|}{2\mathrm{rch}_{\min}}\right).$$

In particular, for all $r < 2\mathrm{rch}_{\min}$,

$$\begin{aligned}\mathrm{B}\left(p, r\left(1-(r/\mathrm{rch}_{\min})^2/24\right)\right) \cap M &\subseteq \mathrm{B}_M(p,r) \\ &\subseteq \mathrm{B}(p,r) \cap M \\ &\subseteq \mathrm{B}_M\left(p, r\left(1+(r/\mathrm{rch}_{\min})^2/4\right)\right).\end{aligned}$$

Furthermore, if $\gamma : [a,b] \to M$ *is an arc-length parametrized geodesic, then for all* $t \in [a,b]$, $\|\gamma''(t)\| \leq 1/\mathrm{rch}_{\min}$.

Proof of Lemma 2.2. We clearly have $\|q-p\| \leq \mathrm{d}_M(p,q)$, and the upper bound comes from [BLW19, Lemma 3]. The ball inclusions then follow from the elementary bounds $\sin s \geq s(1-s^2/6)$ for $s \geq 0$, and $\arcsin u \leq u(1+u^2)$ for $0 \leq u \leq 1$. The last claim is a rephrasing of [NSW08, Proposition 6.1]. □

These estimates will be used to compare, in a quantitative way, the (curved) geometry of M with that of the (flat) Euclidean d-dimensional space. Finally, we present the following uniform estimate on the massivity of submanifolds $M \in \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$, which we will use below to show that `Manifold Propagation` terminates. For $\delta > 0$, the δ-packing number $\mathrm{pk}_M(\delta)$ of $M \subseteq \mathbb{R}^n$ is the maximal cardinal k of a set of points $\{p_i\}_{1 \leq i \leq k} \subseteq M$ such that $\mathrm{B}(p_i,\delta) \cap \mathrm{B}(p_j,\delta) = \emptyset$ for all $i \neq j$ (i.e. $\|p_i - p_i\| > 2\delta$) (see Appendix B.2 for more details).

Lemma 2.3. *Let* $M \in \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$. *Then for all* $\delta \leq \mathrm{rch}_{\min}/8$,

$$\mathrm{pk}_M(\delta) \leq \frac{\mathcal{H}^d(M)}{\omega_d(\delta/4)^d},$$

where $\mathcal{H}^d(M)$ *denotes the surface area of* M.

Proof of Lemma 2.3. Follows from Proposition B.2 and Proposition B.3. □

Based on the geometric model above (Definition 2.4), we now describe the statistical model (i.e. set of probability distributions) of this work. Every $M \in \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$ inherits a non-trivial finite measure induced by the d-dimensional Hausdorff measure $\mathcal{H}^d$ on $\mathbb{R}^n \supseteq M$, defined by $\mathrm{vol}_M = \mathbb{1}_M \mathcal{H}^d$, and called the volume measure of M. Note that with this normalization, $\mathrm{vol}_M(M) = \mathcal{H}^d(M)$ corresponds to the d-dimensional surface area of M, and $\mathrm{vol}_M/\mathcal{H}^d(M)$ corresponds to the uniform probability distribution on M.

Definition 2.5. We let $\mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ denote the set of Borel probability distributions D on $\mathbb{R}^n$ with $M = \mathrm{Supp}(D) \in \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$ and a density f with respect to vol_M such that:

- f is bounded away from zero and infinity: $0 < f_{\min} \leq f(x) \leq f_{\max} < \infty$ for all $x \in M$.
- f is L-Lipschitz over M: $|f(x) - f(y)| \leq L\,\|x-y\|$ for all $x, y \in M$.

In this model, as will be clear below, the extra degree of freedom allowed by the density f being non-constant will contribute in the final estimation rate and query complexity, especially through the lower bound $f_{\min}$. On the geometric side, $f_{\min}^{-1}$ and $f_{\max}^{-1}$ impose quantitative restrictions on the volume $\mathcal{H}^d(M)$ of M (see Section 2.2.3).

Since $\mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ is invariant by translations in $\mathbb{R}^n$, this model actually provides insufficient prior information to derive any uniform SQ complexity bound over it. This contrasts sharply with the sample framework [AL18, GPPVW12a], where the sample points provide automatic location information and yields finite sample complexity over $\mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$.

Proposition 2.1. *Assume that $\sigma_d f_{\min} \mathrm{rch}_{\min}^d \leq 1$. Then for all $\varepsilon > 0$, manifold estimation over $\mathcal{D}_{\mathrm{rch}_{\min}}^{n,d}(f_{\min}, f_{\max}, L)$ with precision ε has infinite randomized statistical query complexity.*

The assumption that $\sigma_d f_{\min} \mathrm{rch}_{\min}^d \leq 1$ is made to preclude degeneracy of the model. It can be shown to be necessary (see Section 2.2.3 below for a more detailed discussion). The proof of Proposition 2.1 relies on the fact that the supports $\mathrm{Supp}(D)$ of distributions $D \in \mathcal{D}_{\mathrm{rch}_{\min}}^{n,d}(f_{\min}, f_{\max}, L)$ form an unbounded class for the Hausdorff distance. It is therefore natural to add an extra location constraint to the model. We study two different such constraints. The first one fixes membership of a distinguished point to M, which we take to be the origin $0 \in \mathbb{R}^n$ without loss of generality. The second one bounds the problem in an ambient ball of radius $R > 0$, which we take to be centered at the origin $\mathrm{B}(0, R)$ without loss of generality.

Definition 2.6. Completing the framework of Definition 2.5, we consider the two following models.

- Fixed point model:
  - $\{0\} \sqcup \mathcal{M}_{\mathrm{rch}_{\min}}^{n,d}$ denotes the set of manifolds $M \in \mathcal{M}_{\mathrm{rch}_{\min}}^{n,d}$ such that $0 \in M$;
  - The model $\{0\} \sqcup \mathcal{D}_{\mathrm{rch}_{\min}}^{n,d}(f_{\min}, f_{\max}, L)$ stands for the set of distributions $D \in \mathcal{D}_{\mathrm{rch}_{\min}}^{n,d}(f_{\min}, f_{\max}, L)$ with support such that $0 \in M = \mathrm{Supp}(D)$.
- Bounding ball model: given $R > 0$,
  - $\mathrm{B}(0, R) \sqcap \mathcal{M}_{\mathrm{rch}_{\min}}^{n,d}$ denotes the set of manifolds $M \in \mathcal{M}_{\mathrm{rch}_{\min}}^{n,d}$ such that $M \subseteq \mathrm{B}(0, R)$;
  - The model $\mathrm{B}(0, R) \sqcap \mathcal{D}_{\mathrm{rch}_{\min}}^{n,d}(f_{\min}, f_{\max}, L)$ stands for the set of distributions $D \in \mathcal{D}_{\mathrm{rch}_{\min}}^{n,d}(f_{\min}, f_{\max}, L)$ with support such that $\mathrm{Supp}(D) = M \subseteq \mathrm{B}(0, R)$.

Let us now discuss some features imposed by the above models.

2.2.3 On Some Implicit Bounds on the Model Parameters

Although not explicit in Definition 2.6, parameters of the models are not arbitrary. That is, $\mathcal{D}_{\mathrm{rch}_{\min}}^{n,d}(f_{\min}, f_{\max}, L)$ might be degenerate or even empty in some regimes of parameters, making the manifold estimation problem vacuous. The reason for this resides in implicit volume bounds imposed by the reach. Indeed, if $D \in \mathcal{D}_{\mathrm{rch}_{\min}}^{n,d}(f_{\min}, f_{\max}, L)$ has support M, then since D is a probability distribution,

$$f_{\min} \mathcal{H}^d(M) \leq 1 = \int_M f \mathrm{d}\mathcal{H}^d \leq f_{\max} \mathcal{H}^d(M).$$

As a result, the volume estimates of Proposition B.5 yield

$$f_{\min} \leq \frac{1}{\mathcal{H}^d(M)} \leq \frac{1}{\sigma_d \mathrm{rch}_{\min}^d} \leq \frac{1}{\omega_d \mathrm{rch}_{\min}^d}.$$

If furthermore, $D \in \mathrm{B}(0, R) \sqcap \mathcal{D}_{\mathrm{rch}_{\min}}^{n,d}(f_{\min}, f_{\max}, L)$ (i.e. $M \subseteq \mathrm{B}(0, R)$), then

$$f_{\max} \geq \frac{1}{\mathcal{H}^d(M)} \geq \frac{1}{\left(\frac{18R}{\mathrm{rch}_{\min}}\right)^n \omega_d \left(\frac{\mathrm{rch}_{\min}}{2}\right)^d}.$$

Note that Proposition B.5 also yields that $R \geq \mathrm{rch}_{\min}/\sqrt{2}$. Consequently, to ensure non-vacuity of the models, and without loss of generality, it is natural to take the following setup. Here, $C_\square$ stands for a constant depending only on $\square$.

- When working over $\{0\} \sqcup \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$, we will always assume that $f_{\min} \leq f_{\max}$, $R \geq C\mathrm{rch}_{\min}$, and
$$\omega_d f_{\min} \mathrm{rch}^d_{\min} \leq C_d^{-1},$$
for some large enough constant $C_d > 0$.

- When working over $\mathrm{B}(0,R) \sqcap \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$, we will always assume that $f_{\min} \leq f_{\max}$, $R \geq C\mathrm{rch}_{\min}$,
$$\omega_d f_{\min} \mathrm{rch}^d_{\min} \leq C_d^{-1} \quad \text{and} \quad \omega_d f_{\max} \mathrm{rch}^d_{\min} \geq C_{n,d} \left(\frac{\mathrm{rch}_{\min}}{R} \right)^n,$$
for some large enough constants $C, C_d, C_{n,d} > 0$.

See Appendix B for a more thorough exposition of the technical properties of the models $\{0\} \sqcup \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ and $\mathrm{B}(0,R) \sqcap \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$.

2.3 Manifold Reconstruction from Point Clouds

Following the recent line of research on manifold estimation [GPPVW12b, GPPVW12a, AL18, AL19, Div21a], we will measure the accuracy of estimators $\hat{M}$ of manifolds M via the so-called *Hausdorff distance*, which plays the role of an L^∞-distance between compact subsets of $\mathbb{R}^n$. To this aim, we will need the following piece of notation. For $K \subseteq \mathbb{R}^n$ and $r \geq 0$, we let K^r denote the *r-offset of K*:
$$K^r := \{z \in \mathbb{R}^n, \mathrm{d}(z,K) \leq r\}, \tag{1}$$
where we recall that $\mathrm{d}(z,K) = \inf_{p \in K} \|p - z\|$ is the function distance to K.

Definition 2.7 (Hausdorff Distance [BBI01, Section 7.3.1])**.** Given two compact subsets $K, K' \subseteq \mathbb{R}^n$, the *Hausdorff distance* between them is
$$\begin{aligned} \mathrm{d_H}(K,K') &= \sup_{x \in \mathbb{R}^n} |\mathrm{d}(x,K) - \mathrm{d}(x,K')| \\ &= \inf \left\{ r > 0, K \subseteq (K')^r \text{ and } K' \subseteq K^r \right\}. \end{aligned}$$

Manifold reconstruction from point clouds has been extensively studied in the area of computational geometry [Dey07, BG14]. In this field, the learner is given a sample of M, usually seen as deterministic, and the goal is to build efficiently a reliable triangulation $\hat{M}$ of M, either topologically, geometrically, or both. Such a construction actually is always possible, provided that the point cloud is sufficiently close and dense in M, and that the learner is provided with tangent space estimates at these points. This is formalized in the following Theorem 2.1, where $\|\cdot\|_{\mathrm{op}}$ stands for the operator norm over the set of matrices.

Theorem 2.1 (Adapted from [AL18, Theorem 4.4])**.** *There exists $\lambda_d > 0$ such that for all $\varepsilon \leq \lambda_d \mathrm{rch}_{\min}$ and all $M \in \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$, the following holds.*

Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a finite point cloud and $\mathbb{T}_{\mathcal{X}} = \{T_x\}_{x \in \mathcal{X}} \subseteq \mathbb{G}^{n,d}$ be a family of d-dimensional linear subspaces of $\mathbb{R}^n$ such that

- $\max_{x \in \mathcal{X}} \mathrm{d}(x,M) \leq \eta$,
- $\max_{p \in M} \mathrm{d}(p,\mathcal{X}) \leq \Delta$,
- $\max_{x \in \mathcal{X}} \left\| \pi_{T_{\pi_M(x)}M} - \pi_{T_x} \right\|_{\mathrm{op}} \leq \sin\theta$.

If $\theta \leq \Delta/(1140\mathrm{rch}_{\min})$ and $\eta \leq \Delta^2/(1140\mathrm{rch}_{\min})$, then one can build a triangulation $\hat{M} = \hat{M}(\mathcal{X}, \mathbb{T}_{\mathcal{X}})$ with vertices in $\mathcal{X}$ such that

- $\mathrm{d_H}(M, \hat{M}) \leq C_d \Delta^2 / \mathrm{rch}_{\min}$,
- *M and $\hat{M}$ are ambient isotopic.*

Proof of Theorem 2.1. We apply [AL18, Theorem 4.4] on a sparsified subset $\mathcal{X}'$ of $\mathcal{X}$, which is a pruned version of $\mathcal{X}$ that is ε-sparse but still dense enough in M. This subsample $\mathcal{X}'$ can be built explicitly by using the so-called farthest point sampling algorithm to $\mathcal{X}$ [AL18, Section 3.3]. For this, initialize $\mathcal{X}'$ with $\mathcal{X}' = \{x_0\}$, where $x_0 \in \mathcal{X}$ is chosen arbitrarily. Then, while $\max_{x \in \mathcal{X}} \mathrm{d}(x, \mathcal{X}') > \Delta$, find the farthest point to $\mathcal{X}'$ in $\mathcal{X}$, and add it to $\mathcal{X}'$. That is, $\mathcal{X}' \leftarrow \mathcal{X}' \cup \{\mathrm{argmax}_{x \in \mathcal{X}} \mathrm{d}(x, \mathcal{X}')\}$ (and if the argmax is not a singleton, pick an arbitrary element of it). The output $\mathcal{X}' \subseteq \mathcal{X}$ of this algorithm clearly satisfies $\min_{x' \neq y' \in \mathcal{X}'} \|y' - x'\| \geq \Delta$, and furthermore,

$$\max_{p \in M} \mathrm{d}(p, \mathcal{X}') \leq \max_{p \in M} \mathrm{d}(p, \mathcal{X}) + \max_{x \in \mathcal{X}} \mathrm{d}(x, \mathcal{X}') \leq 2\Delta.$$

Therefore, [AL18, Theorem 4.4] applies to $\mathcal{X}'$ and $\mathbb{T}_{\mathcal{X}'}$, and $\hat{M}(\mathcal{X}', \mathbb{T}_{\mathcal{X}'})$ provides the announced triangulation. □

Although we will not emphasize on exact topology recovery in the present work, let us mention that the triangulation $\hat{M}$ actually exhibits the extra feature of sharing the same topology as M, i.e. M and $\hat{M}$ are isotopy equivalent. Let us also mention that the triangulation can be built in linear time in n, with an explicit polynomial time and space complexity [BG14, Section 4.6].

Said otherwise, Theorem 2.1 asserts that manifold reconstruction with precision ε can be achieved if a sample that is $(\sqrt{\mathrm{rch}_{\min}\varepsilon})$-dense and ε-close to M, together with associated estimated tangent spaces with precision $\sqrt{\varepsilon / \mathrm{rch}_{\min}}$, are available to the learner. As opposed to the sample framework, the statistical framework does not provide the learner with such data directly. In STAT(τ), our strategy will therefore be to build such a point cloud and tangent spaces iteratively from queries, using the following purely geometric `Manifold Propagation` algorithm.

3 Manifold Propagation Algorithm

We now present the `Manifold Propagation` algorithm and its properties, which works in a setting where only geometric routines are available to the learner. Although we will eventually apply this algorithm in the context of statistical queries (see Section 5), let us insist on the fact that the framework detailed in this Section 3 is purely geometric, and does not rely specifically on statistical queries.

As mentioned in the introduction, the idea is to explore the unknown manifold M via the access to only three complementary geometric routines. Roughly speaking, `Manifold Propagation` explores M in a greedy way, while building a point cloud with associated tangent spaces, by using:

- A seed point $\hat{x}_0 \in \mathbb{R}^n$, known to be close to M, and that allows to initialize the process.
- A tangent space routine $\hat{T} : \mathbb{R}^n \to \mathbb{G}^{n,d}$, that allows to make linear approximations of M nearby points, and hence to provide local candidate directions to explore next.
- A projection routine $\hat{\pi} : \mathbb{R}^n \to \mathbb{R}^n$, that compensates for the errors made by the previous steps, by approximately projecting points back to M.

To avoid redundancy, all this is done while checking that the new candidate points are not too close to some already-visited region of the space. More formally, the algorithm runs as described on page 20.

In spirit, `Manifold Propagation` is similar to the *marching cube algorithm* of [LC87] and the *tracing algorithm* of [BKW19], which use calls to an *intersection oracle* answering whether

Algorithm 1 `Manifold Propagation`

Require:
Seed point $\hat{x}_0 \in \mathbb{R}^n$
Tangent space routine $\hat{T} : \mathbb{R}^n \to \mathbb{G}^{n,d}$
Projection routine $\hat{\pi} : \mathbb{R}^n \to \mathbb{R}^n$
Tuning parameters $\Delta, \delta > 0$ (scales) and $0 < \alpha < \pi/2$ (angle)

1: Initialize $\mathcal{Q} \leftarrow \{\hat{x}_0\}$, $\mathcal{O} \leftarrow \emptyset$ and $\mathbb{T}_{\mathcal{O}} \leftarrow \emptyset$
2: **while** $\mathcal{Q} \neq \emptyset$ **do**
3: Pick $x \in \mathcal{Q}$
4: Set $T \leftarrow \hat{T}(x)$ and $\mathbb{T}_{\mathcal{O}} \leftarrow \mathbb{T}_{\mathcal{O}} \cup \{T\}$
5: Consider a maximal $(\sin\alpha)$-packing $v_1, \ldots, v_k$ of the sphere $\mathcal{S}_T^{d-1}(0,1) \subseteq T$
6: **for** $i \in \{1, \ldots, k\}$ **do**
7: **if** $\mathrm{d}\big(x + \Delta v_i, \mathcal{Q} \cup \mathcal{O}\big) \geq \delta$ **then**
8: $\mathcal{Q} \leftarrow \mathcal{Q} \cup \big\{\hat{\pi}(x + \Delta v_i)\big\}$
9: **end if**
10: **end for**
11: $\mathcal{Q} \leftarrow \mathcal{Q} \setminus \{x\}$ and $\mathcal{O} \leftarrow \mathcal{O} \cup \{x\}$
12: **end while**
13: **return** $\mathcal{O}$ and $\mathbb{T}_{\mathcal{O}}$

a candidate element of a partition of the ambient space intersects the manifold. However, the approaches of [LC87] and [BKW19] use static partitions of $\mathbb{R}^n$ (cubes and a Coxeter triangulation respectively), which translates into an exploration complexity of M — measured in the number of calls made to the oracles/routines — that strongly depends on the ambient dimension [BKW19, Theorem 24]. In contrast, `Manifold Propagation` builds a point cloud nearby M dynamically, which allows to adapt to its intrinsic low-dimensional geometry. This results in an exploration complexity that is completely oblivious to the ambient space. That is, the overall dependency in the ambient dimension is fully supported by the geometric routines themselves. This can be explained by the intermediate tangent space estimation routine, that allows the algorithm to only explore the d local (approximate) tangent directions of M only, while being oblivious to the $(n-d) \gg d$ non-informative codimensions. As a counterpart, `Manifold Propagation` needs to compensate for these local linear approximations which, although possibly negligible at each iteration, may cumulate into substantial deviations from the manifold after several steps. This possible global drift is taken care of via the projection routine, which somehow reboots the precision of the process when a point is added. To the best of our knowledge, `Manifold Propagation` is the first instance of an algorithm working only with the three geometric routines described above. We now state the main result presenting its properties.

Theorem 3.1 (Properties of `Manifold Propagation`)**.** *Let $M \in \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$, and assume that there exist $0 \leq \eta \leq \Lambda < \mathrm{rch}_{\min}$ and $0 \leq \theta < \pi/2$ such that:*

(i) $\mathrm{d}(\hat{x}_0, M) \leq \eta$;

(ii) For all $x \in \mathbb{R}^n$ such that $\mathrm{d}(x, M) \leq \eta$, $\left\|\pi_{T_{\pi_M(x)}M} - \pi_{\hat{T}(x)}\right\|_{\mathrm{op}} \leq \sin\theta$;

(iii) For all $x \in \mathbb{R}^n$ such that $\mathrm{d}(x, M) \leq \Lambda$, $\|\pi_M(x) - \hat{\pi}(x)\| \leq \eta$.

Assume furthermore that

$$64\eta \leq \Delta \leq \mathrm{rch}_{\min}/24, \;\; \max\{\sin\alpha, \sin\theta\} \leq 1/64,$$
$$5\Delta^2/(8\mathrm{rch}_{\min}) + \eta + \Delta\sin\theta \leq \Lambda, \text{ and } 3\Delta/10 \leq \delta \leq 7\Delta/10.$$

Then, `Manifold Propagation` *terminates, and the number N_{loop} of iterations performed in the* while *loop (Lines 2–12) satisfies*

1. $N_{\text{loop}} \leq \dfrac{\mathcal{H}^d(M)}{\omega_d(\delta/32)^d}$, *where* $\mathcal{H}^d(M)$ *denotes the surface area of* M.

Furthermore, it outputs a finite point cloud $\mathcal{O} \subseteq \mathbb{R}^n$ *that:*

2. *Is* η*-close to* M*:* $\max_{x\in\mathcal{O}} \mathrm{d}(x, M) \leq \eta$*;*

3. *Is a* $(\Delta+\eta)$*-covering of* M*:* $\max_{p\in M} \mathrm{d}(p, \mathcal{O}) \leq \Delta + \eta$*;*

together with a family $\mathbb{T}_{\mathcal{O}} = \left\{\hat{T}(x)\right\}_{x\in\mathcal{O}} \subseteq \mathbb{G}^{n,d}$ *of linear spaces that:*

4. θ*-approximate tangent spaces:* $\max_{x\in\mathcal{O}} \left\|\pi_{T_{\pi_M(x)}M} - \pi_{\hat{T}(x)}\right\|_{\text{op}} \leq \sin\theta$.

To get to Theorem 3.1, we will need the following series of lemmas, which are proved in Appendix A. The first statement asserts that the point clouds $\mathcal{Q}$ and $\mathcal{O}$ that the algorithm builds remain η-close to M at all times. The reason for this resides in the fact that this property holds for the seed point $\hat{x}_0$ by assumption, and that the projection routine $\hat{\pi}$ maintains this η-closeness when points are added to $\mathcal{Q}$, and hence to $\mathcal{Q}\cup\mathcal{O}$.

Lemma 3.1. *Let* $M \in \mathcal{M}^{n,d}_{\text{rch}_{\min}}$*, and assume that* $\eta < \text{rch}_{\min}$*,* $\Delta \leq \text{rch}_{\min}/4$ *and* $\frac{5}{8}\frac{\Delta^2}{\text{rch}_{\min}} + \eta + \Delta\sin\theta \leq \Lambda$*. Then when running* `Manifold Propagation`*, the following inequality is maintained:*

$$\max_{x\in\mathcal{Q}\cup\mathcal{O}} \mathrm{d}(x, M) \leq \eta.$$

The second statement ensures that points in $\mathcal{Q}\cup\mathcal{O}$ remain far away from each other, so that they always form a packing with fixed radius. This property, maintained by the proximity test at Line 7 of `Manifold Propagation`, is the key ingredient for the termination of the algorithm and its complexity.

Lemma 3.2. *Let* $M \in \mathcal{M}^{n,d}_{\text{rch}_{\min}}$*, and assume that* $\eta < \text{rch}_{\min}$*,* $\Delta \leq \text{rch}_{\min}/4$ *and* $\frac{5}{8}\frac{\Delta^2}{\text{rch}_{\min}} + \eta + \Delta\sin\theta \leq \Lambda$*. Then when running* `Manifold Propagation`*, the following inequality is maintained:*

$$\min_{\substack{x,y\in\mathcal{Q}\cup\mathcal{O}\\ x\neq y}} \|x-y\| \geq \delta - \frac{5}{8}\frac{\Delta^2}{\text{rch}_{\min}} - 2\eta - \Delta\sin\theta.$$

The third and last statement roughly asserts that if `Manifold Propagation` terminates, then all the Δ-neighborhoods of M have been visited by the output $\mathcal{O}$, i.e. that the greedy tangential exploration strategy is somehow exhaustive at scale Δ.

Lemma 3.3. *Let* $M \in \mathcal{M}^{n,d}_{\text{rch}_{\min}}$*, and assume that* $\Delta \leq \text{rch}_{\min}/24$*,* $\eta < \Delta/64$*, and* $\max\{\sin\alpha, \sin\theta\} \leq 1/64$*. Assume furthermore that,* $\frac{5}{8}\frac{\Delta^2}{\text{rch}_{\min}} + \eta + \Delta\sin\theta \leq \Lambda$ *and* $\delta \leq 7\Delta/10$*. If* `Manifold Propagation` *terminates, then its output* $\mathcal{O}$ *satisfies*

$$\max_{p\in M}\min_{x\in\mathcal{O}} \mathrm{d}_M(p, \pi_M(x)) \leq \Delta,$$

where $\mathrm{d}_M(\cdot,\cdot)$ *is the geodesic distance of* M.

We are now in position to prove Theorem 3.1.

Proof of Theorem 3.1. 1. By construction of `Manifold Propagation`, any visit of the *while* loop (Lines 2–12) finishes with the addition of a point to $\mathcal{O}$. Since $\mathcal{O} = \emptyset$ at initialization, the

number of already performed loops is maintained to satisfy $N_{\text{loop}} = |\mathcal{O}|$ when the algorithm runs. Furthermore, by Lemma 3.1 and Lemma 3.2, we have at all times

$$\begin{aligned}
&\min_{\substack{x,y\in\mathcal{O}\\ x\neq y}} \|\pi_M(x)-\pi_M(y)\| \\
&\qquad \geq \min_{\substack{x,y\in\mathcal{Q}\cup\mathcal{O}\\ x\neq y}} \|\pi_M(x)-\pi_M(y)\| \\
&\qquad \geq \min_{\substack{x,y\in\mathcal{Q}\cup\mathcal{O}\\ x\neq y}} \big(\|x-y\| - \|x-\pi_M(x)\| - \|y-\pi_M(y)\|\big) \\
&\qquad \geq \left(\delta - \frac{5}{8}\frac{\Delta^2}{\operatorname{rch}_{\min}} - 2\eta - \Delta\sin\theta\right) - 2\eta \\
&\qquad \geq \frac{173}{960}\Delta \geq \frac{173}{960}\frac{10}{7}\delta > \frac{\delta}{4} > 0.
\end{aligned}$$

This shows that $\pi_M : \mathcal{O} \to \pi_M(\mathcal{O})$ is one-to-one, and that the set $\pi_M(\mathcal{O}) \subseteq M$ is a $(\delta/8)$-packing of M. As a consequence, we have at all times

$$N_{\text{loop}} = |\mathcal{O}| = |\pi_M(\mathcal{O})| \leq \operatorname{pk}_M(\delta/8) \leq \frac{\mathcal{H}^d(M)}{\omega_d(\delta/32)^d},$$

where the last inequality follows from Lemma 2.3.
As $N_{\text{loop}} < \infty$, this first item also shows that `Manifold Propagation` terminates.

2. This statement follows directly from Lemma 3.1.

3. We have already shown that `Manifold Propagation` terminates. Therefore, Lemma 3.3 applies, and combining it with Item 2, we get

$$\begin{aligned}
\max_{p\in M} \mathrm{d}(p,\mathcal{O}) &= \max_{p\in M}\min_{x\in\mathcal{O}} \|p-x\| \\
&\leq \max_{p\in M}\min_{x\in\mathcal{O}} \|p-\pi_M(x)\| + \max_{x\in\mathcal{O}} \|x-\pi_M(x)\| \\
&\leq \max_{p\in M}\min_{x\in\mathcal{O}} \mathrm{d}_M(p,\pi_M(x)) + \eta \\
&\leq \Delta + \eta.
\end{aligned}$$

4. Follows straightforwardly from Item 2 above, and the assumption that $\left\|\pi_{T_{\pi_M(x)}M} - \pi_{\hat{T}(x)}\right\|_{\text{op}} \leq \sin\theta$ for all $x \in \mathbb{R}^n$ such that $\mathrm{d}(x,M) \leq \eta$.

□

4 Geometric Routines with Statistical Queries

Coming back to manifold estimation with SQ's, we notice that by combining together:
(Exploration) the greedy point cloud construction of Theorem 3.1 using geometric routines only,
(Reconstruction) the point cloud-based reconstruction method of Theorem 2.1,
we have reduced the problem to constructing SQ algorithms emulating these routines with a STAT(τ) oracle. We now present constructions of SQ algorithms for the projection routine $\hat{\pi}(\cdot)$ (Section 4.1), the tangent space estimation routine $\hat{T}(\cdot)$ (Section 4.2), and the seed point detection $\hat{x}_0$ (Section 4.3).

4.1 Projection

Given a point $x_0 \in \mathbb{R}^n$ nearby $M = \mathrm{Supp}(D)$, we aim at estimating its metric projection $\pi_M(x_0)$ onto M with statistical queries to $\mathrm{STAT}(\tau)$. As mentioned earlier, the reasoning we adopt is as follows:

- For a properly chosen bandwidth $h > 0$, the local conditional mean

$$m_D(x_0, h) = \mathop{\mathbb{E}}_{x\sim D}\left[x \,|\mathrm{B}(x_0, h)\right] = x_0 + h\frac{\mathbb{E}_{x\sim D}\left[\frac{(x-x_0)}{h}\mathbb{1}_{\|x-x_0\|\le h}\right]}{D(\mathrm{B}(x_0, h))}$$

  of D around x_0 has small bias for estimating $\pi_M(x_0)$ (Lemma C.2).

- As $m_D(x_0, h) \in \mathbb{R}^n$ writes as a functional of the two means $D(\mathrm{B}(x_0, h)) = \mathbb{E}_{x\sim D}[\mathbb{1}_{\|x-x_0\|\le h}] \in \mathbb{R}$, and $\mathbb{E}_{x\sim D}\left[\frac{(x-x_0)}{h}\mathbb{1}_{\|x-x_0\|\le h}\right] \in \mathbb{R}^n$, it can be estimated using $2n+1$ queries (Lemma 2.1).

The proof of these results are to be found in Appendix C. Combined together, we then prove the correctness of the SQ projection estimation procedure (Theorem 4.1) in Appendix C.2.

Theorem 4.1 (SQ Projection Estimation)**.** *Let $D \in \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ have support $M = \mathrm{Supp}(D)$. Assume that*

$$\frac{\tau}{\omega_d f_{\min}\mathrm{rch}^d_{\min}} \le c^{d(d+1)}\Gamma^d \text{ and } \Lambda \le \frac{\mathrm{rch}_{\min}}{16},$$

for some small enough absolute constant $c > 0$, where $\Gamma = \Gamma_{f_{\min}, f_{\max}, L} = \frac{f_{\min}}{f_{\max}+L\mathrm{rch}_{\min}}$.

Then for all $x_0 \in \mathbb{R}^n$ such that $\mathrm{d}(x_0, M) \le \Lambda$, there exists a SQ algorithm making $2n+1$ queries to $\mathrm{STAT}(\tau)$, that outputs a point $\hat{\pi}(x_0) \in \mathbb{R}^n$ estimating $\pi_M(x_0)$ with precision

$$\|\hat{\pi}(x_0) - \pi_M(x_0)\| \le \eta = \frac{C^d}{\Gamma}\max\left\{\frac{\Lambda^2}{\mathrm{rch}_{\min}}, \Gamma^{\frac{2}{d+1}}\mathrm{rch}_{\min}\left(\frac{\tau}{\omega_d f_{\min}\mathrm{rch}^d_{\min}}\right)^{\frac{2}{d+1}}\right\},$$

where $C > 0$ is an absolute constant.

4.2 Tangent Space

Given a point $x_0 \in \mathbb{R}^n$ nearby $M = \mathrm{Supp}(D)$, we aim at estimating the tangent space $T_{\pi_M(x_0)}M$ with statistical queries to $\mathrm{STAT}(\tau)$. The strategy we propose is based on local Principal Components Analysis, combined with low-rank matrix recovery with SQ's. This local PCA approach is similar to that of [ACLZ17, AL18]. As described above, the reasoning is as follows:

- For a properly chosen bandwidth $h > 0$, the local (rescaled) covariance matrix

$$\Sigma_D(x_0, h) = \mathop{\mathbb{E}}_{x\sim D}\left[\frac{(x-x_0)(x-x_0)^\top}{h^2}\mathbb{1}_{\|x-x_0\|\le h}\right] \in \mathbb{R}^{n\times n}$$

  of D around x_0 is nearly rank-d, and its first d components span a d-plane close to $T_{\pi_M(x_0)}M \in \mathbb{G}^{n,d}$ (Lemma D.1).

- Principal components being stable to perturbations (Lemma D.2), estimating $\Sigma_D(x_0, h) \in \mathbb{R}^{n\times n}$ is sufficient to estimate $T_{\pi_M(x_0)}M \in \mathbb{G}^{n,d}$.

- Estimating $\Sigma_D(x_0, h) \in \mathbb{R}^{n\times n} = \mathbb{R}^{n^2}$ using $O(n^2)$ queries (Lemma 2.1) is too costly and would be redundant since $\Sigma_D(x_0, h)$ is nearly rank $d \ll n$. Instead, we use matrix compression arguments (Theorem D.1) and an explicit construction of a matrix sensing operator (Lemma D.3) to derive a general mean low-rank matrix SQ algorithm (Lemma D.4). This result roughly asserts that a mean matrix $\Sigma = \mathbb{E}_{x\sim D}[F(x)] \in \mathbb{R}^{n\times n}$ of a bounded function $F : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ that has nearly rank d can be estimated with precision $C\tau$ using $nd\,\mathrm{polylog}(n)$ queries to $\mathrm{STAT}(\tau)$.

The proof of these results are to be found in Appendix D. All combined together, we then prove the correctness of the SQ tangent space estimation procedure (Theorem 4.2) in Appendix D.4.

Theorem 4.2 (SQ Tangent Space Estimation)**.** *Let $D \in \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ have support $M = \mathrm{Supp}(D)$. Assume that*

$$\frac{\tau}{\omega_d f_{\max} \mathrm{rch}^d_{\min}} \leq \left(\frac{1}{8\sqrt{d}}\right)^{d+1} \quad \textit{and } \eta \leq \frac{\mathrm{rch}_{\min}}{64d}.$$

Then for all $x_0 \in \mathbb{R}^n$ such that $\mathrm{d}(x_0, M) \leq \eta$, there exists a deterministic SQ algorithm making at most $Cdn\log^6(n)$ queries to $\mathrm{STAT}(\tau)$, and that outputs a d-plane $\hat{T}(x_0) \in \mathbb{G}^{n,d}$ estimating $T_{\pi_M(x_0)}M$ with precision

$$\begin{aligned}\left\|\pi_{\hat{T}(x_0)} - \pi_{T_{\pi_M(x_0)}M}\right\|_{\mathrm{op}} &\leq \sin\theta \\ &= \tilde{C}^d \frac{f_{\max}}{f_{\min}} \max\left\{\sqrt{\frac{\eta}{\mathrm{rch}_{\min}}}, \left(\frac{\tau}{\omega_d f_{\max}\mathrm{rch}^d_{\min}}\right)^{\frac{1}{d+1}}\right\},\end{aligned}$$

where $\tilde{C} > 0$ is an absolute constant.

4.3 Seed Point

Given a ball of radius $R > 0$ guaranteed to encompass $M = \mathrm{Supp}(D) \subseteq \mathrm{B}(0, R)$, and a target precision $\eta > 0$, we aim at finding a point that is η-close to M with statistical queries to $\mathrm{STAT}(\tau)$. The strategy we propose is as follows:

- Starting from $\mathrm{B}(0, R)$, we use a divide and conquer strategy (Theorem E.1). The algorithm (`SQ Ambient Binary Search`) queries indicator functions of an interactively chosen union of balls (i.e. the queried balls depend on the previous answers of the oracle), stops when there is only one ball left and outputs its center $\hat{x}_0^{raw}$. This method only uses estimates on the local mass of balls for D (Lemma B.1), and forgets about the differential structure and $\mathcal{C}^2$-smoothness of M. Hence, although efficient, it only obtains a precision $O(\max\{\eta, \tau^{1/d}\})$, that can be much larger than the prescribed one $O(\max\{\eta, \tau^{2/(d+1)}\})$.

- Starting from $\hat{x}_0^{raw}$, we then refine this detected point by iterating the SQ projection routine $\hat{\pi}(\cdot)$ (Theorem 4.1), which does use extensively the $\mathcal{C}^2$-smoothness of M. As $\hat{x}_0^{raw}$ is close to M, this procedure is guaranteed to enhance precision quadratically at each step, and is hence satisfactory (i.e. has precision η) after a logarithmic number of iterations.

The proof of these results are to be found in Appendix E.

Theorem 4.3 (SQ Point Detection)**.** *Let $D \in \mathrm{B}(0, R) \sqcap \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ have support $M = \mathrm{Supp}(D) \subseteq \mathrm{B}(0, R)$. Assume that*

$$\frac{\tau}{\omega_d f_{\min}\mathrm{rch}^d_{\min}} \leq c\Gamma^d \min\left\{c^d, (n\log(R/(\Gamma\mathrm{rch}_{\min}))^{-1/2}\right\}^d, \quad \textit{and } \eta \leq \frac{\mathrm{rch}_{\min}}{16}$$

for some small enough $c > 0$, where $\Gamma = \Gamma_{f_{\min}, f_{\max}, L} = \frac{f_{\min}}{f_{\max} + L\mathrm{rch}_{\min}}$,

Then there exists a deterministic SQ algorithm making at most $6n\log(6R/\eta)$ queries to $\mathrm{STAT}(\tau)$, and that outputs a point $\hat{x}_0 \in \mathrm{B}(0, R)$ such that

$$\mathrm{d}(\hat{x}_0, M) \leq \max\left\{\eta, C^d \Gamma^{\frac{2}{d+1}-1} \mathrm{rch}_{\min} \left(\frac{\tau}{\omega_d f_{\min} \mathrm{rch}_{\min}^d}\right)^{\frac{2}{d+1}}\right\},$$

where $C > 0$ is the absolute constant of Theorem 4.1.

5 Manifold Estimation with Statistical Queries

We are now in position to state the main results of this work, namely bounds on the statistical query complexity of manifold estimation in $\mathrm{STAT}(\tau)$. We split the results into the two studied models $\{0\} \sqcup \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ and $\mathrm{B}(0, R) \sqcap \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$. For each model, we get an upper bound by combining the results of Sections 3 and 4. It is followed by an informational and a computational lower bound, coming from the general lower bound techniques of Appendix G and the constructions of Appendix H.

5.1 Fixed Point Model

In $\{0\} \sqcup \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$, the origin $0 \in \mathbb{R}^n$ is known to belong to M. The SQ algorithm we propose consists in running `Manifold Propagation` with seed point $\hat{x}_0 = 0$ and the SQ projection and tangent space routines of Theorems 4.1 and 4.2. This leads to the following upper bound. Let us mention that one could easily extend this result and relax the assumption that $0 \in M$ to $\mathrm{d}(0, M)$ being small enough.

Theorem 5.1. *Let $D \in \{0\} \sqcup \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ have support $M = \mathrm{Supp}(D)$. Writing $\Gamma = \Gamma_{f_{\min}, f_{\max}, L} = \frac{f_{\min}}{f_{\max} + L\mathrm{rch}_{\min}}$, let us assume that*

$$\frac{\tau}{f_{\min} \mathrm{rch}_{\min}^d} \leq c^d \Gamma^{7(d+1)/2} \text{ and } \varepsilon \leq \tilde{c}^d \Gamma^3 \mathrm{rch}_{\min},$$

for some small enough absolute constants $c, \tilde{c} > 0$. Then there exists a deterministic SQ algorithm making at most

$$q \leq n \log^6 n \frac{C_d}{f_{\min} \mathrm{rch}_{\min}^d} \left(\frac{\mathrm{rch}_{\min}}{\varepsilon}\right)^{d/2}$$

queries to $\mathrm{STAT}(\tau)$, and that outputs a finite triangulation $\hat{M} \subseteq \mathbb{R}^n$ that has the same topology as M, and such that

$$\mathrm{d}_{\mathrm{H}}(M, \hat{M}) \leq \max\left\{\varepsilon, \frac{\tilde{C}^d}{\Gamma^3} \mathrm{rch}_{\min} \left(\frac{\tau}{f_{\min} \mathrm{rch}_{\min}^d}\right)^{2/(d+1)}\right\},$$

where $C_d > 0$ depends on d and $\tilde{C} > 0$ is an absolute constant.

The algorithm of Theorem 5.1 has a statistical query complexity comparable to the optimal sample complexity $s = O(\varepsilon^{-2/d}\log(1/\varepsilon))$ over the model $\mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ [GPPVW12a, KZ15, AL18], and can provably achieve precision $O(\tau^{2/(d+1)})$. Furthermore, the assumptions made as well as the final precision are completely insensitive to n. The ambient dimension n only appears as a quasi-linear factor in the query complexity. This contrasts with the sample complexity which does not depends on n. However, notice that a single sample nearby $M \subseteq \mathbb{R}^n$ consists of n coordinates, while statistical queries are forced to be real-valued (one dimensional) pieces of information, which explains this apparent discrepancy.

Discussing its optimality, one may first wonder if the assumption made on τ is necessary, and whether the precision barrier of order $O(\tau^{2/(d+1)})$ is improvable in $\mathrm{STAT}(\tau)$. The following statement answers to these questions, regardless the statistical query complexity.

Theorem 5.2. *Let $\alpha < 1/2$ be a probability of error. Assume that $f_{\min} \leq f_{\max}/4$ and*

$$2^{d+1}\sigma_d f_{\min}\mathrm{rch}_{\min}^d \leq 1.$$

Then no randomized SQ algorithm can estimate $M = \mathrm{Supp}(D)$ over $\{0\} \sqcup \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ with precision

$$\varepsilon < \frac{\mathrm{rch}_{\min}}{2^{21}} \min\left\{\frac{1}{2^{20}d^2}, \left(\frac{\tau}{\omega_d f_{\min}\mathrm{rch}_{\min}^d}\right)^{2/d}\right\}$$

and probability $1-\alpha$, no matter its number of queries.

This result justifies why the quantity $\tau/(\omega_d f_{\min}\mathrm{rch}_{\min}^d)$ is required to be small enough in Theorem 5.1: this actually is necessary so as to reach a precision of order at least $O(\mathrm{rch}_{\min})$. Second, this informational lower bound shows that the learner cannot hope to achieve precision better than

$$\varepsilon = \Omega\big(\mathrm{rch}_{\min}\left(\tau/(\omega_d f_{\min}\mathrm{rch}_{\min}^d)\right)^{2/d}\big),$$

even with the most costly randomized SQ algorithms. In fact, the precision $O(\tau^{2/(d+1)})$ of Theorem 5.1 is nearly optimal. Here, the assumptions made on $f_{\min}, f_{\max}$ and $\mathrm{rch}_{\min}$ are also necessary to ensure non-degeneracy of the model $\{0\} \sqcup \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$, as mentioned in Section 2.2.3. Beyond the above informational considerations, we turn to the computational one, i.e. to the minimal number of queries to $\mathrm{STAT}(\tau)$ that a learner must make to achieve precision ε.

Theorem 5.3. *Let $\alpha < 1$ be a probability of error, and $\varepsilon \leq \mathrm{rch}_{\min}/(2^{34}d^2)$. Assume that $f_{\min} \leq f_{\max}/4$ and*

$$2^{d+1}\sigma_d f_{\min}\mathrm{rch}_{\min}^d \leq 1.$$

Then any randomized SQ algorithm estimating $M = \mathrm{Supp}(D)$ over $\{0\} \sqcup \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ with precision ε and with probability of success at least $1-\alpha$ must make at least

$$q \geq \left(n\frac{1}{\omega_d f_{\min}\mathrm{rch}_{\min}^d}\left(\frac{\mathrm{rch}_{\min}}{2^{21}\varepsilon}\right)^{d/2} + \log(1-\alpha)\right) \Big/ \log(1+1/\tau)$$

queries to $\mathrm{STAT}(\tau)$.

For deterministic SQ algorithms ($\alpha = 0$), the statistical query complexity of the algorithm of Theorem 5.1 is therefore optimal up to $\mathrm{polylog}(n, 1/\tau)$ factors. It even performs nearly optimally within all the possible randomized algorithms, provided that their probability of error α is not too close to 1, which would allow for a naive random pick among an ε-covering of the space $\big(\{0\} \sqcup \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}, \mathrm{d_H}\big)$ (with zero query to $\mathrm{STAT}(\tau)$) to be a valid algorithm.

5.2 Bounding Ball Model

In $\mathrm{B}(0,R) \sqcap \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$, no distinguished point of $\mathbb{R}^n$ is known to belong to M, but a location area $\mathrm{B}(0,R)$ containing M is available to the learner. Hence, the strategy of the previous section cannot initialize directly. However, Theorem 4.3 allows to find a seed point $\hat{x}_0$ close to M using a limited number of queries to $\mathrm{STAT}(\tau)$. Starting from $\hat{x}_0$ and, as above, running `Manifold Propagation` with the SQ projection and tangent space routines of Theorems 4.1 and 4.2 leads to the following upper bound.

Theorem 5.4. *Let $D \in \mathrm{B}(0,R) \sqcap \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ have support $M = \mathrm{Supp}(D)$. Writing $\Gamma = \Gamma_{f_{\min},f_{\max},L} = \frac{f_{\min}}{f_{\max}+L\mathrm{rch}_{\min}}$, let us assume that*

$$\frac{\tau}{f_{\min}\mathrm{rch}^d_{\min}} \leq \min\left\{c^d\Gamma^{7(d+1)/2}, \Gamma^d\left(n\log\left(R/(\Gamma\mathrm{rch}_{\min})\right)\right)^{-d/2}\right\}$$

and

$$\varepsilon \leq \tilde{c}^d\Gamma^3\mathrm{rch}_{\min},$$

for some small enough absolute constants $c, \tilde{c} > 0$. Then there exists a deterministic SQ algorithm making at most

$$q \leq Cn\log\left(\frac{R}{\varepsilon}\right) + n\log^6 n\frac{C_d}{f_{\min}\mathrm{rch}^d_{\min}}\left(\frac{\mathrm{rch}_{\min}}{\varepsilon}\right)^{d/2}$$

queries to $\mathrm{STAT}(\tau)$, and that outputs a finite triangulation $\hat{M} \subseteq \mathbb{R}^n$ that has the same topology as M, and such that

$$\mathrm{d_H}(M, \hat{M}) \leq \max\left\{\varepsilon, \frac{\tilde{C}^d}{\Gamma^3}\mathrm{rch}_{\min}\left(\frac{\tau}{f_{\min}\mathrm{rch}^d_{\min}}\right)^{2/(d+1)}\right\},$$

where $C_d > 0$ depends on d and $\tilde{C} > 0$ is an absolute constant.

Compared to Theorem 5.1, observe the extra $O(n\log(R/\varepsilon))$ queries made in Theorem 5.4, which come from the seed point search performed at initialization. Passed this difference, the two results are tightly similar. In the same fashion as above, we first discuss the necessity of the assumptions made on τ and the precision threshold $O(\tau^{2/(d+1)})$.

Theorem 5.5. *Let $\alpha < 1/2$ be a probability of error. Assume that $\mathrm{rch}_{\min} \leq R/144$ and $f_{\min} \leq f_{\max}/96$, and*

$$\min_{1\leq k\leq n}\left(\frac{192\mathrm{rch}_{\min}\sqrt{k}}{R}\right)^k \leq 36 \times 8^d\sigma_{d-1}f_{\min}\mathrm{rch}^d_{\min} \leq 1.$$

Then no randomized SQ algorithm can estimate $M = \mathrm{Supp}(D)$ over $\mathrm{B}(0,R)\sqcap\mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ with precision

$$\varepsilon < \frac{\mathrm{rch}_{\min}}{2^{31}}\min\left\{\frac{1}{2^{10}d^2}, \left(\frac{\tau}{\omega_d f_{\min}\mathrm{rch}^d_{\min}}\right)^{2/d}\right\}$$

and probability $1-\alpha$, no matter its number of queries.

As above, we emphasize the fact that the assumptions made on $f_{\min}$, $f_{\max}$, $\mathrm{rch}_{\min}$ and R are necessary to guarantee the non-degeneracy of the model $\mathrm{B}(0,R) \sqcap \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$, and hence a non-trivial estimation problem (see Section 2.2.3). As for the fixed point model, we hence see that the assumptions made on τ and the precision ε cannot be omitted. Let us though notice the slightly more stringent assumption made on τ that depends on n in the upper bound (Theorem 5.4) but not in the lower bound (Theorem 5.5). This dependency originates from the seed point detection method that we developed (Theorem 4.3) and we do not claim it to be optimal. As about the computational lower bound for this model, we state the following result.

Theorem 5.6. *Let $\alpha < 1$ be a probability of error, and $\varepsilon \leq \mathrm{rch}_{\min}/(2^{34}d^2)$. Assume that $\mathrm{rch}_{\min} \leq R/144$, $f_{\min} \leq f_{\max}/96$, and*

$$\min_{1\leq k\leq n}\left(\frac{192\mathrm{rch}_{\min}\sqrt{k}}{R}\right)^k \leq 36 \times 8^d\sigma_{d-1}f_{\min}\mathrm{rch}^d_{\min} \leq 1.$$

Then any randomized SQ algorithm estimating $M = \mathrm{Supp}(D)$ *over* $\mathrm{B}(0,R) \sqcap \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ *with precision* ε *and with probability of success at least* $1-\alpha$ *must make at least*

$$q \geq \frac{n \max\left\{\log\left(\frac{R}{4\varepsilon}\right), \frac{1}{\omega_d f_{\min} \mathrm{rch}^d_{\min}} \left(\frac{\mathrm{rch}_{\min}}{2^{31}\varepsilon}\right)^{d/2}\right\} + \log(1-\alpha)}{\log(1+1/\tau)}$$

queries to $\mathrm{STAT}(\tau)$.

As a result, the extra $O(n \log(R/\varepsilon))$ queries of Theorem 5.4 are necessary in the SQ framework. This contrasts sharply with the sample model, where no prior location information is necessary appears in the sample complexity [GPPVW12a, KZ15, AL18]. Roughly speaking, this is explained by the fact that a single sample (the first, say) does provide location information for free, while in the SQ framework, the learner is left with the whole ball $\mathrm{B}(0,R)$ to explore at initialization. However, as mentioned above, note that the term $\Omega(n \log(R/\varepsilon))$ attributable to this initialization step would only dominate in the regime where R is exponentially bigger than $\mathrm{rch}_{\min}$.

6 Further Directions

As mentioned above, a byproduct of these results is that manifold estimation is possible in a locally private way. However, the transformation used to pass from statistical query learning to locally private learning has a polynomial blowup [KLN+11]. Hence, the derived locally private upper bound may not be optimal, so that a close study of the private framework directly is still necessary. Coming back to SQ's, the derived bounds on the best achievable precision ε in $\mathrm{STAT}(\tau)$ do not match, as they are respectively of the form $\varepsilon = O\left(\mathrm{rch}_{\min}\left(\frac{\tau}{f_{\min}\mathrm{rch}^d_{\min}}\right)^{2/(d+1)}\right)$ for the upper bounds, and $\varepsilon = \Omega\left(\mathrm{rch}_{\min}\left(\frac{\tau}{f_{\min}\mathrm{rch}^d_{\min}}\right)^{2/d}\right)$ for the lower bounds, so that this gap remains to be breached.

The case of smoother $\mathcal{C}^k$ manifolds ($k > 2$) would be of fundamental interest, as their estimation is at the basis of plugin methods of higher order. This includes, for instance, density estimation [BHHS22], or distribution estimation in Wasserstein distance [Div21b]. For $k > 2$, local linear approximations are not optimal [AL19], and local polynomials of higher order — that were shown to be optimal in the sample framework — might adapt to statistical queries.

On a more technical side, note that we have assumed throughout that the density f is L-Lipschitz and satisfies $f_{\min} \leq f \leq f_{\max}$, although the lower bounds do not let L and $f_{\max}$ appear, but only $f_{\min}$. While the Lipschitz assumption could be dropped for the tangent space routine, it actually is crucial in the proposed projection routine to bound the bias term. Hence, it remains unclear to us how to design an efficient projection routine without this assumption, as well as how to carry the whole analysis with $f_{\max} = \infty$.

Acknowledgements

This work was partially funded by CNRS PEPS JCJC. The authors would like to thank warmly the *Department of Mathematics of UC San Diego*, the *Steklov Institute*, and all the members of the *Laboratoire de Probabilités, Statistiques et Modélisation* for their support and insightful comments.

A Proofs of the Properties of Manifold Propagation

When running `Manifold Propagation`, linear approximations of the manifold are done via its (approximate) tangent spaces. A key point in the proof of its correctness is the (quantitative)

validity of this approximation, which is ensured by the reach assumption $\mathrm{rch}_M \geq \mathrm{rch}_{\min}$, which bounds curvature. Recall from (1) that $M^r = \{z \in \mathbb{R}^n, \mathrm{d}(z, M) \leq r\}$ stands for the r-offset of M.

Lemma A.1. *Let $M \in \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$, and $x \in M^\eta$ with $\eta < \mathrm{rch}_{\min}$. Take $T \in \mathbb{G}^{n,d}$ such that $\left\|\pi_{T_{\pi_M(x)}M} - \pi_T\right\|_{\mathrm{op}} \leq \sin\theta$. Then for all $\Delta \leq \mathrm{rch}_{\min}/4$, and all unit vector $v \in T$,*

$$\mathrm{d}\big(x + \Delta v, M\big) \leq \frac{5}{8}\frac{\Delta^2}{\mathrm{rch}_{\min}} + \eta + \Delta \sin\theta.$$

Proof of Lemma A.1. By assumption on T, there exists a unit vector $v' \in T_{\pi_M(x)}M$ such that $\|v - v'\| \leq \sin\theta$. Hence, since $\mathrm{d}(\cdot, M)$ is 1-Lipschitz, we have

$$\begin{aligned}
\mathrm{d}\big(x + \Delta v, M\big) &\leq \mathrm{d}\big(x + \Delta v', M\big) + \Delta \sin\theta \\
&\leq \mathrm{d}\big(\pi_M(x) + \Delta v', M\big) + \eta + \Delta \sin\theta \\
&\leq \left\|\pi_M(x) + \Delta v' - \exp^M_{\pi_M(x)}(\Delta v')\right\| + \eta + \Delta \sin\theta \\
&\leq \frac{5}{8}\frac{\Delta^2}{\mathrm{rch}_{\min}} + \eta + \Delta \sin\theta,
\end{aligned}$$

where the last inequality follows from [AL19, Lemma 1]. □

We are now in position to prove Lemma 3.1, that guarantees that `Manifold Propagation` builds point clouds that do not deviate from M.

Proof of Lemma 3.1. The points added to $\mathcal{O}$ are all first added to $\mathcal{Q}$: therefore, it is sufficient to check that all the points x added to $\mathcal{Q}$ satisfy $\mathrm{d}(x, M) \leq \eta$. To see this, proceed by induction:

- As $\mathcal{Q}$ is initialized to $\{\hat{x}_0\}$ with $\mathrm{d}(\hat{x}_0, M) \leq \eta$, the inequality holds true at Line 1, before the first loop.
- If $\bar{x} \neq \hat{x}_0$ was added to $\mathcal{Q}$, it can be written as $\bar{x} = \hat{\pi}(x_0 + \Delta v_i)$, for some point $x_0 \in \mathcal{Q}$ and a unit vector $v_i \in \hat{T}(x_0)$. By induction, we have $\mathrm{d}(x_0, M) \leq \eta$. But since $\hat{T}(\cdot)$ is assumed to have precision $\sin\theta$ over M^η, we hence obtain that $\left\|\pi_{T_{\pi_M(x_0)}M} - \pi_{\hat{T}(x_0)}\right\|_{\mathrm{op}} \leq \sin\theta$. As a result, from Lemma A.1,

  $$\mathrm{d}(x_0 + \Delta v_i, M) \leq \frac{5}{8}\frac{\Delta^2}{\mathrm{rch}_{\min}} + \eta + \Delta \sin\theta \leq \Lambda,$$

  and therefore

  $$\mathrm{d}(\bar{x}, M) \leq \|\bar{x} - \pi_M(x_0 + \Delta v_i)\| = \|\hat{\pi}(x_0 + \Delta v_i) - \pi_M(x_0 + \Delta v_i)\| \leq \eta$$

  since $\hat{\pi}(\cdot)$ is assumed to have precision η over M^Λ.

This concludes the induction and hence the proof. □

Next we show Lemma 3.2, asserting that the radius of sparsity of the point clouds built by `Manifold Propagation` is maintained at all times.

Proof of Lemma 3.2. At initialization of `Manifold Propagation`, $\mathcal{Q} \cup \mathcal{O} = \{\hat{x}_0\}$, so that the inequality trivially holds at Line 1. Then, if a point $\bar{x}$ is added to $\mathcal{Q}$ at Line 8, it means that it

can be written as $\bar{x} = \hat{\pi}(x_0 + \Delta v_{i_0})$, with $\mathrm{d}(x_0 + \Delta v_{i_0}, \mathcal{Q} \cup \mathcal{O}) \geq \delta$. Consequently, by induction, we have

$$\begin{aligned}
\min_{\substack{x,y \in \mathcal{Q} \cup \mathcal{O} \cup \{\bar{x}\} \\ x \neq y}} \|x - y\| &= \min\Big\{ \min_{\substack{x,y \in \mathcal{Q} \cup \mathcal{O} \\ x \neq y}} \|x - y\|, \mathrm{d}(\bar{x}, \mathcal{Q} \cup \mathcal{O}) \Big\} \\
&\geq \min\Big\{ \delta - \frac{5}{8} \frac{\Delta^2}{\mathrm{rch}_{\min}} - 2\eta - \Delta \sin\theta, \\
&\qquad \mathrm{d}(x_0 + \Delta v_{i_0}, \mathcal{Q} \cup \mathcal{O}) - \|(x_0 + \Delta v_{i_0}) - \bar{x}\| \Big\} \\
&\geq \min\left\{ \delta - \frac{5}{8} \frac{\Delta^2}{\mathrm{rch}_{\min}} - 2\eta - \Delta \sin\theta, \delta - \|(x_0 + \Delta v_{i_0}) - \bar{x}\| \right\}.
\end{aligned}$$

In addition, Lemma A.1 and Lemma 3.1 combined yield

$$\begin{aligned}
\|(x_0 + \Delta v_{i_0}) - \bar{x}\| &= \|(x_0 + \Delta v_{i_0}) - \hat{\pi}(x_0 + \Delta v_{i_0})\| \\
&\leq \|\hat{\pi}(x_0 + \Delta v_{i_0}) - \pi_M(x_0 + \Delta v_{i_0})\| \\
&\quad + \|\pi_M(x_0 + \Delta v_{i_0}) - (x_0 + \Delta v_{i_0})\| \\
&\leq \eta + \left(\frac{5}{8} \frac{\Delta^2}{\mathrm{rch}_{\min}} + \eta + \Delta \sin\theta \right).
\end{aligned}$$

As a result, after the update $\mathcal{Q} \leftarrow \mathcal{Q} \cup \{\bar{x}\}$, the announced inequality still holds. Finally, we notice that Line 11, which swaps a point from $\mathcal{Q}$ to $\mathcal{O}$, leaves $\mathcal{Q} \cup \mathcal{O}$ unchanged. By induction, this concludes the proof. □

Finally we prove Lemma 3.3, that states that if `Manifold Propagation` terminates, it outputs a point cloud dense enough nearby M.

Proof of Lemma 3.3. Assume for contradiction that there exists $p_0 \in M$ such that for all $x \in \mathcal{O}$, $\mathrm{d}_M\big(p_0, \pi_M(x)\big) > \Delta$. Let $x_0 \in \mathcal{O}$ (which is not empty since $\hat{x}_0 \in \mathcal{O}$) be such that

$$\mathrm{d}_M\big(p_0, \pi_M(x_0)\big) = \min_{x \in \mathcal{O}} \mathrm{d}_M(p_0, \pi_M(x)) := r_0 > \Delta,$$

and write $y_0 := \pi_M(x_0)$. Let $\gamma := \gamma_{y_0 \to p_0} : [0, r_0] \to M$ denote an arc-length parametrized geodesic joining y_0 and p_0. Finally, set $q_0 := \gamma(\Delta) \in M$ and $v_0 := \gamma'(0) \in T_{y_0}M$.

Consider the sets $\mathcal{Q}$ and $\mathcal{O}$ of `Manifold Propagation` right after x_0 was removed from $\mathcal{Q}$ and added to $\mathcal{O}$ (Line 11). By construction, all the elements $v_1, \ldots, v_k$ of a maximal $(\sin\alpha)$-packing of $\mathcal{S}^{d-1}_{\hat{T}(x_0)}$ were tested to enter $\mathcal{Q}$ (Loop from Line 6 to Line 10). Because the packing is maximal, it is also a $(2\sin\alpha)$-covering of $\mathcal{S}^{d-1}_{\hat{T}(x_0)}$ (see the proof of Proposition B.2). As a result, by assumption on the precision of $\hat{T}(x_0)$, there exists v_{i_0} in this packing such that $\|v_0 - v_{i_0}\| \leq 2\sin\alpha + \sin\theta$.

As γ is a distance-minimizing path on M from y_0 to p_0, so it is along its two sub-paths with endpoint q_0, as otherwise, one could build a strictly shorter path between y_0 and p_0. In particular, since $\Delta < r_0 = \mathrm{d}_M(y_0, p_0)$, we have $\mathrm{d}_M(y_0, q_0) = \mathrm{d}_M(y_0, \gamma(\Delta)) = \Delta$ and $\mathrm{d}_M(p_0, q_0) = \mathrm{d}_M(p_0, \gamma(\Delta)) = r_0 - \Delta$. As a result,

$$\begin{aligned}
\mathrm{d}_M\left(p_0, \pi_M\left(\hat{\pi}\left(x_0 + \Delta v_{i_0}\right)\right)\right) &\leq \mathrm{d}_M\left(p_0, q_0\right) + \mathrm{d}_M\left(q_0, \pi_M\left(\hat{\pi}\left(x_0 + \Delta v_{i_0}\right)\right)\right) \\
&= r_0 - \Delta + \mathrm{d}_M\left(q_0, \pi_M\left(\hat{\pi}\left(x_0 + \Delta v_{i_0}\right)\right)\right). \qquad (2)
\end{aligned}$$

But from Lemma 2.2, we get

$$\begin{aligned}\mathrm{d}_M\left(q_0, \pi_M\left(\hat{\pi}\left(x_0+\Delta v_{i_0}\right)\right)\right) &\leq 2\mathrm{rch}_{\min} \arcsin\left(\frac{\left\|q_0-\pi_M\left(\hat{\pi}\left(x_0+\Delta v_{i_0}\right)\right)\right\|}{2\mathrm{rch}_{\min}}\right) \\ &\leq \frac{\left\|q_0-\pi_M\left(\hat{\pi}\left(x_0+\Delta v_{i_0}\right)\right)\right\|}{\sqrt{1-\left(\frac{\left\|q_0-\pi_M\left(\hat{\pi}\left(x_0+\Delta v_{i_0}\right)\right)\right\|}{2\mathrm{rch}_{\min}}\right)^2}}, \end{aligned} \tag{3}$$

and furthermore,

$$\begin{aligned}\left\|q_0-\pi_M\left(\hat{\pi}\left(x_0+\Delta v_{i_0}\right)\right)\right\| \leq &\left\|q_0-\left(y_0+\Delta v_0\right)\right\| + \left\|\left(y_0+\Delta v_0\right)-\left(x_0+\Delta v_{i_0}\right)\right\| \\ &+ \left\|\left(x_0+\Delta v_{i_0}\right)-\hat{\pi}\left(x_0+\Delta v_{i_0}\right)\right\| \\ &+ \left\|\hat{\pi}\left(x_0+\Delta v_{i_0}\right)-\pi_M\left(\hat{\pi}\left(x_0+\Delta v_{i_0}\right)\right)\right\|. \end{aligned} \tag{4}$$

We now bound the right hand side of Equation (4) term by term. The first term is bounded by

$$\left\|q_0-\left(y_0+\Delta v_0\right)\right\| = \left\|\gamma(\Delta)-\left(\gamma(0)+\Delta\gamma'(0)\right)\right\| \leq \frac{\Delta^2}{2\mathrm{rch}_{\min}},$$

where the inequality follows from a Taylor expansion and Lemma 2.2. For the second term, write

$$\left\|\left(y_0+\Delta v_0\right)-\left(x_0+\Delta v_{i_0}\right)\right\| \leq \left\|y_0-x_0\right\| + \Delta\left\|v_0-v_{i_0}\right\| \leq \eta + \Delta(2\sin\alpha+\sin\theta).$$

For the third term, we combine Lemma A.1 and Lemma 3.1 to get

$$\begin{aligned}\left\|\left(x_0+\Delta v_{i_0}\right)-\hat{\pi}\left(x_0+\Delta v_{i_0}\right)\right\| &\leq \mathrm{d}\left(x_0+\Delta v_{i_0}, M\right) \\ &\quad + \left\|\pi_M(x_0+\Delta v_{i_0})-\hat{\pi}\left(x_0+\Delta v_{i_0}\right)\right\| \\ &\leq \frac{5}{8}\frac{\Delta^2}{\mathrm{rch}_{\min}} + 2\eta + \Delta\sin\theta, \end{aligned}$$

and for the fourth term, applying again Lemma A.1 and Lemma 3.1 yields

$$\begin{aligned}\left\|\hat{\pi}\left(x_0+\Delta v_{i_0}\right)-\pi_M\left(\hat{\pi}\left(x_0+\Delta v_{i_0}\right)\right)\right\| &= \mathrm{d}\big(\hat{\pi}\left(x_0+\Delta v_{i_0}\right), M\big) \\ &\leq \eta. \end{aligned}$$

Plugging these four bounds in Equation (4), we have shown that

$$\left\|q_0-\pi_M\left(\hat{\pi}\left(x_0+\Delta v_{i_0}\right)\right)\right\| \leq \frac{9\Delta^2}{8\mathrm{rch}_{\min}} + 4\eta + 2\Delta(\sin\alpha+\sin\theta). \tag{5}$$

Combining Equation (5), Equation (3), and the assumptions on the parameters $\Delta, \eta, \theta, \alpha$ hence yields

$$\mathrm{d}_M\left(y_0, \pi_M\left(\hat{\pi}\left(x_0+\Delta v_{i_0}\right)\right)\right) \leq 2\left\|q_0-\pi_M\left(\hat{\pi}\left(x_0+\Delta v_{i_0}\right)\right)\right\| \leq \Delta/2,$$

so that Equation (2) gives

$$\begin{aligned}\mathrm{d}_M\left(p_0, \pi_M\left(\hat{\pi}\left(x_0+\Delta v_{i_0}\right)\right)\right) &\leq r_0 - \Delta + \Delta/2 \\ &< r_0 \\ &= \mathrm{d}_M(p_0, \pi_M(x_0)) = \min_{x\in\mathcal{O}} \mathrm{d}_M(p_0, \pi_M(x)). \end{aligned}$$

In particular, $\hat{\pi}\left(x_0+\Delta v_{i_0}\right)$ was not added to $\mathcal{Q}$ in the Loop of Lines 6 to 10 investigating the neighbors of x_0 (i.e. when x_0 was picked Line 3). Since $\mathcal{Q}\cup\mathcal{O}$ is an increasing sequence of

sets as `Manifold Propagation` runs and that $\mathcal{Q} = \emptyset$ when it terminates, this means that there exists x_1 in the final output $\mathcal{O}$ such that $\|x_0 + \Delta v_{i_0} - x_1\| \leq \delta$.

The existence of this particular point x_1 in $\mathcal{O}$ which is δ-close to $x_0 + \Delta v_{i_0}$ will lead us to a contradiction: we will show that $\pi_M(x_1)$ will be closer to p_0 than $\pi_M(x_0)$ is in geodesic distance. To get there, we first notice that any such $x_1 \in \mathcal{O}$ would satisfy $\mathrm{d}(x_1, M) \leq \eta$ from Lemma 3.1, so that

$$\begin{aligned}
\|\pi_M(x_0 + \Delta v_{i_0}) - \pi_M(x_1)\| &\leq \|\pi_M(x_0 + \Delta v_{i_0}) - (x_0 + \Delta v_{i_0})\| \\
&\quad + \|(x_0 + \Delta v_{i_0}) - x_1\| + \|x_1 - \pi_M(x_1)\| \\
&\leq \delta + \frac{5}{8}\frac{\Delta^2}{\mathrm{rch}_{\min}} + 2\eta + \Delta \sin\theta \\
&\leq \delta + \frac{17}{192}\Delta \leq \frac{25}{192}\mathrm{rch}_{\min},
\end{aligned}$$

where the last-but-one line follows from Lemma A.1, and the last one from the assumptions on the parameters Δ, η, θ and δ. As a result, from Lemma 2.2,

$$\begin{aligned}
\mathrm{d}_M(\pi_M(x_0 + \Delta v_{i_0}), \pi_M(x_1)) &\leq \frac{\|\pi_M(x_0 + \Delta v_{i_0}) - \pi_M(x_1)\|}{\sqrt{1 - \left(\frac{25}{2\times 192}\right)^2}} \\
&\leq \left(1 + \frac{3}{10000}\right)\left(\delta + \frac{17}{192}\Delta\right). \qquad (6)
\end{aligned}$$

Furthermore, using a similar decomposition as for Equation (5), we have

$$\begin{aligned}
\|q_0 - \pi_M(x_0 + \Delta v_{i_0})\| &\leq \|q_0 - (y_0 + \Delta v_0)\| + \|(y_0 + \Delta v_0) - (x_0 + \Delta v_{i_0})\| \\
&\quad + \|(x_0 + \Delta v_{i_0}) - \pi_M(x_0 + \Delta v_{i_0})\| \\
&\leq \frac{\Delta^2}{2\mathrm{rch}_{\min}} + (\eta + \Delta(2\sin\alpha + \sin\theta)) \\
&\quad + \left(\frac{5}{8}\frac{\Delta^2}{\mathrm{rch}_{\min}} + \eta + \Delta\sin\theta\right) \\
&\leq \frac{11}{64}\Delta \leq \frac{11}{1536}\mathrm{rch}_{\min},
\end{aligned}$$

from which we finally get

$$\begin{aligned}
\mathrm{d}_M(q_0, \pi_M(x_0 + \Delta v_{i_0})) &\leq \frac{\|q_0 - \pi_M(x_0 + \Delta v_{i_0})\|}{\sqrt{1 - \left(\frac{11}{2\times 1536}\right)^2}} \\
&\leq \frac{3}{16}\Delta. \qquad (7)
\end{aligned}$$

This takes us to the desired contradiction, since:

- on one hand, $x_1 \in \mathcal{O}$ forces to have

$$\mathrm{d}_M(p_0, \pi_M(x_1)) \geq r_0 = \min_{x \in \mathcal{O}} \mathrm{d}_M(p_0, \pi_M(x)) = \mathrm{d}_M(p_0, \pi_M(x_0));$$

- on the other hand, Equation (6) and Equation (7) combined yield

$$\begin{aligned}
\mathrm{d}_M(p_0, \pi_M(x_1)) &\leq \mathrm{d}_M(p_0, q_0) + \mathrm{d}_M(q_0, \pi_M(x_0 + \Delta v_{i_0})) \\
&\quad + \mathrm{d}_M(\pi_M(x_0 + \Delta v_{i_0}), \pi_M(x_1)) \\
&\leq r_0 - \Delta + \frac{3}{16}\Delta + \left(1 + \frac{3}{10000}\right)\left(\delta + \frac{17}{192}\Delta\right) \\
&< r_0,
\end{aligned}$$

  where we used that $\delta \leq 7\Delta/10$.

As a result, we have proved that

$$\max_{p\in M}\min_{x\in\mathcal{O}} \mathrm{d}_M(p,\pi_M(x)) \le \Delta,$$

which is the announced result. □

B Preliminary Geometric Results

B.1 Local Mass of Balls Estimates

To prove the properties of the statistical query routines, we will need the following two geometric results about manifolds with bounded reach. In what follows, $t_+ := \max\{0,t\}$ stands for the positive part of $t \in \mathbb{R}$.

Proposition B.1 ([AL18, Proposition 8.2])**.** *Let $M \in \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$, $x \in \mathbb{R}^n$ such that $\mathrm{d}(x,M) \le \mathrm{rch}_{\min}/8$, and $h \le \mathrm{rch}_{\min}/8$. Then,*

$$\mathrm{B}\left(\pi_M(x), r_h^-\right) \cap M \subseteq \mathrm{B}(x,h) \cap M \subseteq \mathrm{B}\left(\pi_M(x), r_h^+\right) \cap M,$$

where $r_h = (h^2 - \mathrm{d}(x,M)^2)_+^{1/2}$, $(r_h^-)^2 = \left(1 - \frac{\mathrm{d}(x,M)}{\mathrm{rch}_{\min}}\right) r_h^2$, and $(r_h^+)^2 = \left(1 + \frac{2\mathrm{d}(x,M)}{\mathrm{rch}_{\min}}\right) r_h^2$.

As a result, one may show that any ball has large mass with respect to a measure $D \in \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$.

Lemma B.1. *Let $D \in \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ have support $M = \mathrm{Supp}(D)$.*

- *For all $p \in M$ and $h \le \mathrm{rch}_{\min}/4$,*

$$a_d f_{\min} h^d \le D\big(\mathrm{B}(p,h)\big) \le A_d f_{\max} h^d,$$

  where $a_d = 2^{-d}\omega_d$ and $A_d = 2^d \omega_d$.

- *For all $x_0 \in \mathbb{R}^n$ and $h \le \mathrm{rch}_{\min}/8$,*

$$a'_d f_{\min} (h^2 - \mathrm{d}(x_0,M)^2)_+^{d/2} \le D\big(\mathrm{B}(x_0,h)\big) \le A'_d f_{\max} (h^2 - \mathrm{d}(x_0,M)^2)_+^{d/2},$$

  where $a'_d = (7/8)^{d/2} a_d$ and $A'_d = (5/4)^{d/2} A_d$.

Proof of Lemma B.1. The first statement is a direct consequence of [AL18, Propositions 8.6 & 8.7]. The second one follows by combining the previous point with Proposition B.1. □

B.2 Euclidean Packing and Covering Estimates

For sake of completeness, we include in this section some standard packing and covering bounds that are used in our analysis. We recall the following definitions.

A r-covering of $K \subseteq \mathbb{R}^n$ is a subset $\mathcal{X} = \{x_1,\dots,x_k\} \subseteq K$ such that for all $x \in K$, $\mathrm{d}(x,\mathcal{X}) \le r$. A r-packing of K is a subset $\mathcal{Y} = \{y_1,\dots,y_k\} \subseteq K$ such that for all $y, y' \in \mathcal{Y}$, $\mathrm{B}(y,r) \cap \mathrm{B}(y',r) = \emptyset$ (or equivalently $\|y' - y\| > 2r$).

Definition B.1 (Covering and Packing numbers)**.** For $K \subseteq \mathbb{R}^n$ and $r > 0$, the covering number $\mathrm{cv}_K(r)$ of K is the minimum number of balls of radius r that are necessary to cover K:

$$\mathrm{cv}_K(r) = \min\left\{k > 0 \mid \text{there exists a } r\text{-covering of cardinality } k\right\}.$$

The packing number $\mathrm{pk}_K(r)$ of K is the maximum number of disjoint balls of radius r that can be packed in K:

$$\mathrm{pk}_K(r) = \max\left\{k > 0 \mid \text{there exists a } r\text{-packing of cardinality } k\right\}.$$

Packing and covering numbers are tightly related, as shown by the following well-known statement.

Proposition B.2. *For all subset $K \subseteq \mathbb{R}^n$ and $r > 0$,*

$$\mathrm{pk}_K(2r) \leq \mathrm{cv}_K(2r) \leq \mathrm{pk}_K(r).$$

Proof of Proposition B.2. For the left-hand side inequality, notice that if K is covered by a family of balls of radius $2r$, each of these balls contains at most one point of a maximal $2r$-packing. Conversely, the right-hand side inequality follows from the fact that a maximal r-packing is always a $2r$-covering. Indeed, if it was not the case one could add a point $x_0 \in K$ that is $2r$-away from all of the r-packing elements, which would contradict the maximality of this packing. □

We then bound the packing and covering numbers of the submanifolds with reach bounded below. Note that these bounds depend only on the intrinsic dimension and volumes, but not on the ambient dimension.

Proposition B.3. *For all $M \in \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$ and $r \leq \mathrm{rch}_{\min}/8$,*

$$\mathrm{pk}_M(r) \geq \frac{\mathcal{H}^d(M)}{\omega_d(4r)^d},$$

and

$$\mathrm{cv}_M(r) \leq \frac{\mathcal{H}^d(M)}{\omega_d(r/4)^d}.$$

Proof of Proposition B.3. First, we have $\mathrm{pk}_M(r) \geq \mathrm{cv}_M(2r)$ from Proposition B.2. In addition, if $\{p_i\}_{1\leq i\leq N} \subseteq M$ is a minimal $(2r)$-covering of M, then by considering the uniform distribution $D_M = \mathbb{1}_M \mathcal{H}^d/\mathcal{H}^d(M)$ over M, using a union bound and applying Lemma B.1, we get

$$1 = D_M\left(\cup_{i=1}^N \mathrm{B}(p_i, 2r)\right) \leq \sum_{i=1}^N D_M(\mathrm{B}(p_i, 2r)) \leq N 2^d \omega_d (2r)^d/\mathcal{H}^d(M).$$

As a result, $\mathrm{pk}_M(r) \geq \mathrm{cv}_M(2r) = N \geq \frac{\mathcal{H}^d(M)}{\omega_d(4r)^d}$.

For the second bound, use again Proposition B.2 to get $\mathrm{cv}_M(r) \leq \mathrm{pk}_M(r/2)$. Now, by definition, a maximal $(r/2)$-packing $\{q_j\}_{1\leq j\leq N'} \subseteq M$ of M provides us with a family of disjoint balls of radii $r/2$. Hence, from Lemma B.1, we get

$$1 \geq D_M\left(\cup_{i=j}^{N'} \mathrm{B}(q_j, r/2)\right) = \sum_{j=1}^{N'} D_M(\mathrm{B}(q_j, r/2)) \geq N' 2^{-d} \omega_d (r/2)^d/\mathcal{H}^d(M),$$

so that $\mathrm{cv}_M(r) \leq \mathrm{pk}_M(r/2) = N' \leq \frac{\mathcal{H}^d(M)}{\omega_d(r/4)^d}$. □

Bounds on the same discretization-related quantities computed on the Euclidean n-balls and k-spheres will also be useful.

Proposition B.4.

- *For all $r > 0$,*

$$\mathrm{pk}_{\mathrm{B}(0,R)}(r) \geq \left(\frac{R}{2r}\right)^n \text{ and } \mathrm{cv}_{\mathrm{B}(0,R)}(r) \leq \left(1 + \frac{2R}{r}\right)^n.$$

- *For all integer $1 \leq k < n$ and $r \leq 1/8$,*

$$\mathrm{pk}_{\mathcal{S}^k(0,1)}(r) \geq 2\left(\frac{1}{4r}\right)^k.$$

Proof of Proposition B.4. • From Proposition B.2, we have $\mathrm{pk}_{\mathrm{B}(0,R)}(r) \geq \mathrm{cv}_{\mathrm{B}(0,R)}(2r)$. Furthermore, if $\cup_{i=1}^{N} \mathrm{B}(x_i, 2r) \supseteq \mathrm{B}(0,R)$ is a minimal $2r$-covering of $\mathrm{B}(0,R)$, then by a union bound, $\omega_n R^n = \mathcal{H}^n(\mathrm{B}(0,R)) \leq N\omega_n(2r)^n$, so that $\mathrm{pk}_{\mathrm{B}(0,R)}(r) \geq \mathrm{cv}_{\mathrm{B}(0,R)}(2r) = N \geq (R/(2r))^n$.

For the second bound, we use again Proposition B.2 to get $\mathrm{cv}_{\mathrm{B}(0,R)}(r) \leq \mathrm{pk}_{\mathrm{B}(0,R)}(r/2)$, and we notice that any maximal $(r/2)$-packing of $\mathrm{B}(0,R)$ with cardinality N' provides us with a family of disjoint balls of radii $r/2$, all contained in $\mathrm{B}(0,R)^{r/2} = \mathrm{B}(0, R + r/2)$. A union bound hence yields $\omega_n(R + r/2)^n = \mathcal{H}^n(\mathrm{B}(0, R + r/2)) \geq N'\mathcal{H}^n(\mathrm{B}(0, r/2)) = N'\omega_n(r/2)^n$, yielding $\mathrm{cv}_{\mathrm{B}(0,R)}(r) \leq \mathrm{pk}_{\mathrm{B}(0,R)}(r/2) = N' \leq (1 + 2R/r)^n$.

- Notice that $\mathcal{S}^k(0,1) \subseteq \mathbb{R}^n$ is a compact k-dimensional submanifold without boundary, reach $\mathrm{rch}_{\mathcal{S}^k(0,1)} = 1$, and volume $\mathcal{H}^k(\mathcal{S}^k(0,1)) = \sigma_k$. Applying Proposition B.3 together with elementary calculations hence yield

$$\begin{aligned}
\mathrm{pk}_{\mathcal{S}^k(0,1)}(r) &\geq \frac{\sigma_k}{\omega_k}\left(\frac{1}{4r}\right)^k \\
&= \left(\frac{2\pi^{(k+1)/2}}{\Gamma\left(\frac{k+1}{2}\right)}\right)\left(\frac{\pi^{k/2}}{\Gamma\left(\frac{k}{2}+1\right)}\right)^{-1}\left(\frac{1}{4r}\right)^k \\
&= 2\sqrt{\pi}\frac{\Gamma\left(\frac{k}{2}+1\right)}{\Gamma\left(\frac{k+1}{2}\right)}\left(\frac{1}{4r}\right)^k \\
&\geq 2\left(\frac{1}{4r}\right)^k.
\end{aligned}$$

□

B.3 Global Volume Estimates

The following bounds on the volume and diameter of low-dimensional submanifolds of $\mathbb{R}^n$ with positive reach are at the core of Section 2.2.3. They exhibit some implicit constraints on the parameters for the statistical models not to be degenerate.

Proposition B.5. *For all $M \in \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$,*

$$\mathcal{H}^d(M) \geq \sigma_d \mathrm{rch}^d_{\min},$$

with equality if and only if M is a d-dimensional sphere of radius $\mathrm{rch}_{\min}$. Furthermore, if $M \subseteq \mathrm{B}(0,R)$ then $\mathrm{rch}_{\min} \leq \sqrt{2}R$ and

$$\mathcal{H}^d(M) \leq \left(\frac{18R}{\mathrm{rch}_{\min}}\right)^n \omega_d \left(\frac{\mathrm{rch}_{\min}}{2}\right)^d.$$

Proof of Proposition B.5. For the first bound, note that the operator norm of the second fundamental form of M is everywhere bounded above by $1/\mathrm{rch}_{\min}$ [NSW08, Proposition 6.1], so that [Alm86, (3)] applies and yields the result.

For the next statement, note that [Hat02, Theorem 3.26] ensures that M is not homotopy equivalent to a point. As a result, [AKC+19, Lemma A.3] applies and yields

$$\begin{aligned}
\mathrm{rch}_{\min} &\leq \mathrm{rch}_M \\
&\leq \mathrm{diam}(M)/\sqrt{2} \\
&\leq \mathrm{diam}(\mathrm{B}(0,R))/\sqrt{2} \\
&= \sqrt{2}R.
\end{aligned}$$

For the last bound, consider a $(\mathrm{rch}_{\min}/8)$-covering $\{z_i\}_{1\le i\le N}$ of $\mathrm{B}(0,R)$, which can be chosen so that $N \le \left(1+\frac{2R}{\mathrm{rch}_{\min}/8}\right)^n \le \left(\frac{18R}{\mathrm{rch}_{\min}}\right)^n$ from Proposition B.4. Applying Lemma B.1 with $h=\mathrm{rch}_{\min}/8$, we obtain

$$\begin{aligned}\mathcal{H}^d(M\cap \mathrm{B}(z_i,\mathrm{rch}_{\min}/8)) &\le (5/4)^{d/2}\times 2^d\omega_d((\mathrm{rch}_{\min}/8)^2-\mathrm{d}(z_i,M)^2)_+^{d/2}\\ &\le \omega_d\left(\frac{\mathrm{rch}_{\min}}{2}\right)^d,\end{aligned}$$

for all $i\in\{1,\dots,N\}$. A union bound then yields

$$\begin{aligned}\mathcal{H}^d(M) &= \mathcal{H}^d\left(\cup_{i=1}^N M\cap \mathrm{B}(z_i,\mathrm{rch}_{\min}/8)\right)\\ &\le N\omega_d\left(\frac{\mathrm{rch}_{\min}}{2}\right)^d\\ &\le \left(\frac{18R}{\mathrm{rch}_{\min}}\right)^n\omega_d\left(\frac{\mathrm{rch}_{\min}}{2}\right)^d,\end{aligned}$$

which concludes the proof. □

C Projection Routine

We now build the SQ projection routine $\hat{\pi}:\mathbb{R}^n\to\mathbb{R}^n$ (Theorem 4.1), which is used repeatedly in the SQ emulation of `Manifold Propagation` (Theorems 5.1 and 5.4). Recall that given a point $x_0\in\mathbb{R}^n$ nearby $M=\mathrm{Supp}(D)$, we aim at estimating its metric projection $\pi_M(x_0)$ onto M with statistical queries to $\mathrm{STAT}(\tau)$. We follow the strategy of proof described in Section 4.1.

C.1 Bias of the Local Conditional Mean for Projection

In what follows, we will write

$$m_D(x_0,h) = \mathop{\mathbb{E}}_{x\sim D}\left[x\,|\mathrm{B}(x_0,h)\right] = \frac{\mathbb{E}_{x\sim D}\left[x\mathbb{1}_{\|x-x_0\|\le h}\right]}{D(\mathrm{B}(x_0,h))} \tag{8}$$

for the local conditional mean of D given $\mathrm{B}(x_0,h)$. In order to study the bias of $m_D(x_0,h)$ with respect to $\pi_M(x_0)$, it will be convenient to express it (up to approximation) with intrinsic geodesic balls $\mathrm{B}_M(\cdot,\cdot)$ instead of the extrinsic Euclidean balls $\mathrm{B}(\cdot,\cdot)$ that appears in its definition (Equation (8)). This change of metric is stated in the following result.

Lemma C.1. *Let $D\in\mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min},f_{\max},L)$ have support $M=\mathrm{Supp}(D)$, and $p\in M$. Recall that $\omega_d=\mathcal{H}^d\left(\mathrm{B}_d(0,1)\right)$ denotes the volume of the d-dimensional unit Euclidean ball. Then for all $r\le \mathrm{rch}_{\min}/4$,*

$$\left\|\mathop{\mathbb{E}}_{x\sim D}\left[x\mathbb{1}_{\mathrm{B}_M(p,r)}(x)\right]-D\left(\mathrm{B}_M(p,r)\right)p\right\|\le C^d\omega_d\left(\frac{f_{\max}}{\mathrm{rch}_{\min}}+L\right)r^{d+2},$$

and for $r\le\bar{r}\le\mathrm{rch}_{\min}/4$,

$$D\left(\mathrm{B}_M(p,\bar{r})\setminus\mathrm{B}_M(p,r)\right)\le (C')^d\omega_d f_{\max}\bar{r}^{d-1}(\bar{r}-r),$$

where $C,C'>0$ are absolute constants.

Proof of Lemma C.1. First apply the area formula [Fed69, Section 3.2.5] to write the mean of any measurable function G defined on M as

$$\mathop{\mathbb{E}}_{x\sim D}\left[G(x)\mathbb{1}_{\mathrm{B}_M(p,r)}(x)\right]=\int_0^r\int_{\mathcal{S}^{d-1}}J(t,v)f\left(\exp_p^M(tv)\right)G\left(\exp_p^M(tv)\right)dvdt,$$

where $J(t,v)$ is the Jacobian of the volume form of M expressed in polar coordinates around p for $0 \le t \le r \le \mathrm{rch}_{\min}/4$ and unit $v \in T_pM$. That is, $J(t,v) = t^{d-1}\sqrt{\det\left(A_{t,v}^\top A_{t,v}\right)}$ where $A_{t,v} = \mathrm{d}_{tv}\exp_p^M$. But from [AKC+19, Proposition A.1 (iv)], for all $w \in T_pM$, we have

$$\left(1 - \frac{t^2}{6\mathrm{rch}_{\min}^2}\right)\|w\| \le \|A_{t,v}w\| \le \left(1 + \frac{t^2}{\mathrm{rch}_{\min}^2}\right)\|w\|.$$

As a consequence,

$$\left(1 - \frac{t^2}{6\mathrm{rch}_{\min}^2}\right)^d \le \sqrt{\det\left(A_{t,v}^\top A_{t,v}\right)} \le \left(1 + \frac{t^2}{\mathrm{rch}_{\min}^2}\right)^d$$

and in particular,

$$R_J(t,v) := \left|J(t,v) - t^{d-1}\right| \le C^d t^{d-1}\left(\frac{t}{\mathrm{rch}_{\min}}\right)^2,$$

where $C > 0$ is an absolute constant. Also, by assumption on the model, f is L-Lipschitz, so

$$\begin{aligned}
|R_f(t,v)| := \left|f\left(\exp_p^M(tv)\right) - f(p)\right| &= \left|f\left(\exp_p^M(tv)\right) - f(\exp_p^M(0))\right| \\
&\le L\left\|\exp_p^M(tv) - \exp_p^M(0)\right\| \\
&\le L d_M(\exp_p^M(0), \exp_p^M(tv)) \\
&= Lt.
\end{aligned}$$

Finally, from [AL19, Lemma 1], we have

$$\|R_{\exp}(t,v)\| := \left\|\exp_p^M(tv) - (p + tv)\right\| \le 5t^2/(8\mathrm{rch}_{\min}).$$

Putting everything together, we can now prove the first bound by writing

$$\begin{aligned}
&\left\|\mathop{\mathbb{E}}_{x\sim D}\left[x\mathbb{1}_{\mathrm{B}_M(p,r)}(x)\right] - D\left(\mathrm{B}_M(p,r)\right)p\right\| \\
&= \left\|\int_0^r \int_{\mathcal{S}^{d-1}} J(t,v) f\left(\exp_p^M(tv)\right)\left(\exp_p^M(tv) - p\right) dv dt\right\| \\
&= \left\|\int_0^r \int_{\mathcal{S}^{d-1}} \left(t^{d-1} + R_J(t,v)\right)\left(f(p) + R_f(t,v)\right)\left(tv + R_{\exp}(t,v)\right) dv dt\right\| \\
&\le \tilde{C}^d \omega_d \left(\frac{f_{\max}}{\mathrm{rch}_{\min}} + L\right) r^{d+2},
\end{aligned}$$

where the last inequality used the fact that $\int_0^r \int_{\mathcal{S}^{d-1}} t^d f(p) v dv dt = 0$. Similarly, to derive the second bound, we write

$$\begin{aligned}
D\left(\mathrm{B}_M(p,\bar{r}) \setminus \mathrm{B}_M(p,r)\right) &= \int_r^{\bar{r}} \int_{\mathcal{S}^{d-1}} J(t,v) f\left(\exp_p^M(tv)\right) dv dt \\
&\le \sigma_{d-1} f_{\max} \int_r^{\bar{r}} t^{d-1}\left(1 + C^d\left(t/\mathrm{rch}_{\min}\right)^2\right) dt \\
&\le (C')^d \sigma_{d-1} f_{\max} \int_r^{\bar{r}} t^{d-1} dt \\
&\le (C'')^d \omega_d f_{\max} \bar{r}^{d-1}(\bar{r} - r),
\end{aligned}$$

which concludes the proof. □

We are now in position to bound the bias of $m_D(x_0, h)$.

Lemma C.2. *Let $D \in \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ have support $M = \mathrm{Supp}(D)$, and $x_0 \in \mathbb{R}^n$ be such that $\mathrm{d}(x_0, M) < h \leq \mathrm{rch}_{\min}/8$. Then,*

$$\|\pi_M(x_0) - m_D(x_0, h)\| \leq C^d \left(\frac{f_{\max} + L\mathrm{rch}_{\min}}{f_{\min}} \right) \frac{hr_h}{\mathrm{rch}_{\min}},$$

where $r_h = (h^2 - \mathrm{d}(x_0, M)^2)^{1/2}$ and $C > 0$ is an absolute constant.

Proof of Lemma C.2. For short, let us write $p_0 = \pi_M(x_0)$. All the expected values $\mathbb{E}$ are taken with respect to $x \sim D$. Before any calculation, we combine Proposition B.1 and Lemma 2.2 to assert that

$$\mathrm{B}_M\left(p_0, r_h^-\right) \subseteq \mathrm{B}(x_0, h) \cap M \subseteq \mathrm{B}_M\left(p_0, R_h^+\right), \tag{9}$$

where we wrote $(r_h^-)^2 = (1 - \mathrm{d}(x_0, M)/\mathrm{rch}_{\min})\, r_h^2$ and $R_h^+ = r_h^+ \left(1 + (r_h^+/\mathrm{rch}_{\min})^2\right)$, with $(r_h^+)^2 = (1 + 2\mathrm{d}(x_0, M)/\mathrm{rch}_{\min})\, r_h^2$. We note by now from the definition $0 < r_h^- \leq R_h^+ \leq \mathrm{rch}_{\min}/4$ since $\mathrm{d}(x_0, M) < h \leq \mathrm{rch}_{\min}/8$, and that

$$R_h^+ - r_h^- \leq \frac{C' r_h}{\mathrm{rch}_{\min}} \left(\mathrm{d}(x_0, M) + r_h^2/\mathrm{rch}_{\min}\right) \leq \frac{2C' h r_h}{\mathrm{rch}_{\min}}, \tag{10}$$

for some absolute constant $C' > 0$.

We can now proceed and derive the asserted bound. From triangle inequality,

$$\begin{aligned}
&\|m_D(x_0, h) - \pi_M(x_0)\| \\
&= \frac{\left\|\mathbb{E}_{x\sim D}\left[(x - p_0)\mathbb{1}_{\mathrm{B}(x_0,h)}(x)\right]\right\|}{D(\mathrm{B}(x_0, h))} \\
&\leq \left\| \frac{\mathbb{E}\left[(x - p_0)\mathbb{1}_{\mathrm{B}(x_0,h)}(x)\right]}{D(\mathrm{B}(x_0, h))} - \frac{\mathbb{E}\left[(x - p_0)\mathbb{1}_{\mathrm{B}_M(p_0,R_h^+)}(x)\right]}{D(\mathrm{B}_M(p_0, R_h^+))} \right\| \\
&\quad + \frac{\left\|\mathbb{E}\left[(x - p_0)\mathbb{1}_{\mathrm{B}_M(p_0,R_h^+)}(x)\right]\right\|}{D(\mathrm{B}_M(p_0, R_h^+))}.
\end{aligned}$$

Combining Equation (9), Lemma C.1, Proposition B.1 and Lemma B.1, the first term of the right hand side can be further upper bounded by

$$\begin{aligned}
&\left\| \frac{\mathbb{E}\left[(x - p_0)\mathbb{1}_{\mathrm{B}(x_0,h)}(x)\right]}{D(\mathrm{B}(x_0, h))} - \frac{\mathbb{E}\left[(x - p_0)\mathbb{1}_{\mathrm{B}_M(p_0,R_h^+)}(x)\right]}{D(\mathrm{B}_M(p_0, R_h^+))} \right\| \\
&\leq \frac{\left\|\mathbb{E}\left[(x - p_0)\mathbb{1}_{\mathrm{B}(x_0,h)}(x)\right]\right\|}{D(\mathrm{B}(x_0, h))D(\mathrm{B}_M(p_0, R_h^+))} \left|D(\mathrm{B}_M(p_0, R_h^+)) - D(\mathrm{B}(x_0, h))\right| \\
&\quad + \frac{\left\|\mathbb{E}\left[(x - p_0)\left(\mathbb{1}_{\mathrm{B}_M(p_0,R_h^+)}(x) - \mathbb{1}_{\mathrm{B}(x_0,h)}(x)\right)\right]\right\|}{D(\mathrm{B}_M(p_0, R_h^+))} \\
&\leq \frac{2R_h^+ D\left(\mathrm{B}_M(p_0, R_h^+) \setminus \mathrm{B}_M(p_0, r_h^-)\right)}{D(\mathrm{B}_M(p_0, R_h^+))} \\
&\leq \frac{(C'')^d \omega_d f_{\max} (R_h^+)^d (R_h^+ - r_h^-)}{c^d \omega_d f_{\min} (R_h^+)^d} \\
&\leq \tilde{C}^d \frac{f_{\max}}{f_{\min}} \frac{hr_h}{\mathrm{rch}_{\min}},
\end{aligned}$$

where the last bound uses (10). For the second term, we use Lemma C.1 and Lemma B.1 to derive

$$\frac{\left\|\mathbb{E}\left[(x-p_0)\mathbb{1}_{\mathrm{B}_M(p_0,R_h^+)}(x)\right]\right\|}{D(\mathrm{B}_M(p_0,R_h^+))} \leq \frac{(\tilde{C}')^d\omega_d\left(\frac{f_{\max}}{\mathrm{rch}_{\min}}+L\right)(R_h^+)^{d+2}}{c^d\omega_d f_{\min}(R_h^+)^d}$$
$$\leq (\tilde{C}'')^d\left(\frac{f_{\max}+L\mathrm{rch}_{\min}}{f_{\min}}\right)\frac{r_h^2}{\mathrm{rch}_{\min}}.$$

Since $r_h \leq h$, this concludes the proof by setting $C = \tilde{C} + \tilde{C}''$. □

C.2 Metric Projection with Statistical Queries

We finally prove the main announced statement of Appendix C.

Proof of Theorem 4.1. First note that under the assumptions of the theorem, $\mathrm{d}(x_0, M) \leq \Lambda \leq \mathrm{rch}_{\min}/8$. We hence let $h > 0$ be a bandwidth to be specified later, but taken such that $\mathrm{d}(x_0, M) < \sqrt{2}\Lambda \leq h \leq \mathrm{rch}_{\min}/8$.

Consider the map $F(x) = \frac{(x-x_0)}{h}\mathbb{1}_{\|x-x_0\|\leq h}$ for $x \in \mathbb{R}^n$. As $\|F(x)\| \leq 1$ for all $x \in \mathbb{R}^n$, Lemma 2.1 asserts that there exists a deterministic statistical query algorithm making $2n$ queries to $\mathrm{STAT}(\tau)$ and that outputs a vector $\hat{W} = \hat{V}/h \in \mathbb{R}^n$ such that $\left\|\mathbb{E}_{x\sim D}\left[F(x)\right] - \hat{V}/h\right\| \leq C\tau$. Furthermore, with the single query $r = \mathbb{1}_{\mathrm{B}(x_0,h)}$ to $\mathrm{STAT}(\tau)$, we obtain $\hat{a} \in \mathbb{R}$ such that $|D(\mathrm{B}(x_0,h) - \hat{a}| \leq \tau$. Let us set $\hat{\pi}(x_0) := x_0 + \hat{V}/\hat{a}$ and prove that it satisfies the claimed bound. For this, use $|V/a - \hat{V}/\hat{a}| \leq |a-\hat{a}|V/(a\hat{a}) + |V-\hat{V}|/\hat{a}$ to write

$$\begin{aligned}
&\|m_D(x_0,h) - \hat{\pi}(x_0)\| \\
&= \left\|\frac{\mathbb{E}_{x\sim D}\left[(x-x_0)\mathbb{1}_{\|x-x_0\|\leq h}\right]}{D(\mathrm{B}(x_0,h))} - \frac{\hat{V}}{\hat{a}}\right\| \\
&\leq \frac{|D(\mathrm{B}(x_0,h)) - \hat{a}|\left\|\mathbb{E}_{x\sim D}\left[(x-x_0)\mathbb{1}_{\|x-x_0\|\leq h}\right]\right\|}{D(\mathrm{B}(x_0,h))\hat{a}} + \frac{\left\|\mathbb{E}_{x\sim D}\left[(x-x_0)\mathbb{1}_{\|x-x_0\|\leq h}\right] - \hat{V}\right\|}{\hat{a}} \\
&\leq \frac{|D(\mathrm{B}(x_0,h)) - \hat{a}|\,h + \left\|\mathbb{E}_{x\sim D}\left[(x-x_0)\mathbb{1}_{\|x-x_0\|\leq h}\right] - \hat{V}\right\|}{D(\mathrm{B}(x_0,h)) - |D(\mathrm{B}(x_0,h) - \hat{a}|} \\
&\leq \frac{(C+1)\tau h}{D(\mathrm{B}(x_0,h)) - \tau} \\
&\leq \frac{(C+1)\tau h}{\tilde{c}^d\omega_d f_{\min}(h/\sqrt{2})^d - \tau},
\end{aligned}$$

where the last inequality comes from Lemma B.1, and $r_h = (h^2 - \mathrm{d}(x_0,M)^2)^{d/2} \geq h/\sqrt{2}$ since $h \geq \sqrt{2}\Lambda$. If in addition, one assumes that $\tilde{c}^d\omega_d f_{\min}(h/\sqrt{2})^d \geq 2\tau$, we obtain the lower bound $\tilde{c}^d\omega_d f_{\min}(h/\sqrt{2})^d - \tau \geq \tilde{c}^d\omega_d f_{\min}(h/\sqrt{2})^d/2$, so that the previous bound further simplifies to

$$\|m_D(x_0,h) - \hat{\pi}(x_0)\| \leq \frac{(C')^d}{\omega_d f_{\min}}\tau h^{1-d}.$$

On the other hand, Lemma C.2 yields that the bias term is not bigger than

$$\|\pi_M(x_0) - m_D(x_0,h)\| \leq \tilde{C}^d\left(\frac{f_{\max}+L\mathrm{rch}_{\min}}{f_{\min}}\right)\frac{hr_h}{\mathrm{rch}_{\min}},$$

with $r_h \leq h$. As a result,

$$\begin{aligned}\|\pi_M(x_0) - \hat{\pi}(x_0)\| &\leq \|\pi_M(x_0) - m_D(x_0, h)\| + \|m_D(x_0, h) - \hat{\pi}(x_0)\| \\ &\leq \frac{(C' \vee \tilde{C})^d}{f_{\min}} \left((f_{\max} + L\mathrm{rch}_{\min}) \frac{h}{\mathrm{rch}_{\min}} + \frac{\tau}{\omega_d h^d} \right) h.\end{aligned}$$

Taking bandwidth

$$\begin{aligned}h &= \max\left\{ 2\Lambda, \left(\frac{\mathrm{rch}_{\min}}{f_{\max} + L\mathrm{rch}_{\min}} \right)^{\frac{1}{d+1}} \left(\frac{\tau}{\omega_d} \right)^{\frac{1}{d+1}} \right\} \\ &= \max\left\{ 2\Lambda, \mathrm{rch}_{\min} \left(\frac{f_{\min}}{f_{\max} + L\mathrm{rch}_{\min}} \right)^{\frac{1}{d+1}} \left(\frac{\tau}{\omega_d f_{\min} \mathrm{rch}_{\min}^d} \right)^{\frac{1}{d+1}} \right\},\end{aligned}$$

we have by assumption on the parameters of the model that $\mathrm{rch}_{\min}/8 \geq h \geq 2\Lambda \geq \sqrt{2}\Lambda$, and that $\tilde{c}^d \omega_d f_{\min} (h/\sqrt{2})^d \geq 2\tau$ as soon as $c > 0$ is small enough. Finally, plugging the value of h in the above bound and recalling that $\Gamma = \frac{f_{\min}}{f_{\max} + L\mathrm{rch}_{\min}}$ yields

$$\|\pi_M(x_0) - \hat{\pi}(x_0)\| \leq \frac{\tilde{C}^d}{\Gamma} \max\left\{ \frac{\Lambda^2}{\mathrm{rch}_{\min}}, \Gamma^{\frac{2}{d+1}} \mathrm{rch}_{\min} \left(\frac{\tau}{\omega_d f_{\min} \mathrm{rch}_{\min}^d} \right)^{\frac{2}{d+1}} \right\},$$

which concludes the proof. □

D Tangent Space Estimation Routine

We now build the SQ tangent space routine $\hat{T} : \mathbb{R}^n \to \mathbb{G}^{n,d}$ (Theorem 4.2), which is used repeatedly in the SQ emulation of `Manifold Propagation` (Theorems 5.1 and 5.4). Recall that given a point $x_0 \in \mathbb{R}^n$ nearby $M = \mathrm{Supp}(D)$, we aim at estimating the tangent space $T_{\pi_M(x_0)}M$ with statistical queries to $\mathrm{STAT}(\tau)$. We follow the strategy of proof described in Section 4.2.

To fix notation from now on, we let $\langle A, B \rangle = \mathrm{tr}(A^* B)$ stand for the Euclidean inner product between $A, B \in \mathbb{R}^{k \times k}$. We also write $\|\Sigma\|_{\mathrm{F}} = \sqrt{\langle \Sigma, \Sigma \rangle}$ for the Frobenius norm, $\|\Sigma\|_{\mathrm{op}} = \max_{\|v\| \leq 1} \|\Sigma v\|$ for the operator norm, and $\|\Sigma\|_* = \max_{\|X\|_{\mathrm{op}} \leq 1} \langle \Sigma, X \rangle$ for the nuclear norm. In what follows, for a symmetric matrix $A \in \mathbb{R}^{n \times n}$, we let $\mu_i(A)$ denote its i-th largest singular value.

D.1 Bias of Local Principal Component Analysis

In what follows, we will write

$$\Sigma_D(x_0, h) = \mathop{\mathbb{E}}_{x \sim D} \left[\frac{(x - x_0)(x - x_0)^\top}{h^2} \mathbb{1}_{\|x - x_0\| \leq h} \right] \tag{11}$$

for the re-scaled local covariance-like matrix of D at $x_0 \in \mathbb{R}^n$ with bandwidth $h > 0$. Notice that for simplicity, this local covariance-like matrix is computed with centering at the current point x_0, and not at the local conditional mean $\mathbb{E}_{x \sim D}[x | \|x - x_0\| \leq h]$. This choice simplifies our analysis and will not impact the subsequent estimation rates. Let us first decompose this matrix and exhibit its link with the target tangent space $T_{\pi_M(x_0)}M \in \mathbb{G}^{n,d}$.

Lemma D.1. *Let $D \in \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ have support $M = \mathrm{Supp}(D)$, $x_0 \in \mathbb{R}^n$ and $h > 0$. If $\mathrm{d}(x_0, M) \leq \eta \leq h/\sqrt{2}$ and $h \leq \mathrm{rch}_{\min}/(8\sqrt{d})$, then there exists a symmetric matrix $\Sigma_0 \in \mathbb{R}^{n \times n}$ with $\mathrm{Im}(\Sigma_0) = T_{\pi_M(x_0)}M$ such that*

$$\Sigma_D(x_0, h) = \Sigma_0 + R,$$

with $\mu_d(\Sigma_0) \geq \omega_d f_{\min}(ch)^d$ *and* $\|R\|_* \leq \omega_d f_{\max}(Ch)^d \left(\frac{\eta}{h} + \frac{h}{\mathrm{rch}_{\min}}\right)$*, where* $c, C > 0$ *are absolute constants.*

Proof of Lemma D.1. This proof roughly follows the ideas of [AL18, Section E.1], with a different center point in the covariance matrix (x_0 itself instead of the local mean around x_0) and finer (nuclear norm) estimates on residual terms. For brevity, we let $p_0 = \pi_M(x_0)$. We first note that the integrand defining $h^2\Sigma_D(x_0, h)$ decomposes as

$$\begin{aligned}(x - x_0)(x - x_0)^\top &= (x - p_0)(x - p_0)^\top + (x_0 - p_0)(x_0 - p_0)^\top \\ &\quad + (x - p_0)(x_0 - p_0)^\top + (x_0 - p_0)(x - p_0)^\top,\end{aligned} \tag{12}$$

for all $x \in \mathrm{B}(x_0, h) \cap M$. After integrating them with respect to $x \sim D$, we bound the last two terms, by writing

$$\begin{aligned}\left\| \mathop{\mathbb{E}}_{x\sim D}\left[(x - p_0)(x_0 - p_0)^\top \mathbb{1}_{\|x-x_0\|\leq h}\right]\right\|_* &= \left\| \mathop{\mathbb{E}}_{x\sim D}\left[(x_0 - p_0)(x - p_0)^\top \mathbb{1}_{\|x-x_0\|\leq h}\right]\right\|_* \\ &\leq \mathop{\mathbb{E}}_{x\sim D}\left[\left\|(x_0 - p_0)(x - p_0)^\top\right\|_* \mathbb{1}_{\|x-x_0\|\leq h}\right] \\ &= \mathop{\mathbb{E}}_{x\sim D}\left[\|x_0 - p_0\| \, \|x - p_0\| \, \mathbb{1}_{\|x-x_0\|\leq h}\right] \\ &\leq \eta h D(\mathrm{B}(x_0, h)) \\ &\leq C^d \omega_d f_{\max} \eta h^{d+1},\end{aligned}$$

where the last inequality uses Lemma B.1. Similarly, for the second term of Equation (12), we have

$$\begin{aligned}\left\| \mathop{\mathbb{E}}_{x\sim D}\left[(x_0 - p_0)(x_0 - p_0)^\top \mathbb{1}_{\|x-x_0\|\leq h}\right]\right\|_* &= \|x_0 - p_0\|^2 \, D(\mathrm{B}(x_0, h)) \\ &\leq C^d \omega_d f_{\max} \eta^2 h^d.\end{aligned}$$

Given $v \in \mathbb{R}^n$, write $v_{/\!/} = \pi_{T_{p_0}M}(v)$ and $v_\perp = v - v_{/\!/} = \pi_{T_{p_0}M^\perp}(v)$. We now focus on the first term of Equation (12), which we further decompose as

$$\begin{aligned}(x - p_0)(x - p_0)^\top &= (x - p_0)_{/\!/}(x - p_0)_{/\!/}^\top + (x - p_0)_\perp (x - p_0)_\perp^\top \\ &\quad + (x - p_0)_\perp (x - p_0)_{/\!/}^\top + (x - p_0)_{/\!/}(x - p_0)_\perp^\top,\end{aligned} \tag{13}$$

for all $x \in \mathrm{B}(x_0, h) \cap M$. Note that for those points $x \in \mathrm{B}(x_0, h) \cap M$, we have $\left\|(x - p_0)_{/\!/}\right\| \leq \|x - p_0\| \leq 2h$, and from [Fed59, Theorem 4.18], $\|(x - p_0)_\perp\| \leq \|x - p_0\|^2 / (2\mathrm{rch}_{\min}) \leq 4h^2/(2\mathrm{rch}_{\min})$. Hence, for the last two terms of Equation (13),

$$\begin{aligned}\left\| \mathop{\mathbb{E}}_{x\sim D}\left[(x - p_0)_\perp (x - p_0)_{/\!/}^\top \mathbb{1}_{\|x-x_0\|\leq h}\right]\right\|_* &= \left\| \mathop{\mathbb{E}}_{x\sim D}\left[(x - p_0)_{/\!/}(x - p_0)_\perp^\top \mathbb{1}_{\|x-x_0\|\leq h}\right]\right\|_* \\ &\leq \mathop{\mathbb{E}}_{x\sim D}\left[\left\|(x - p_0)_{/\!/}\right\| \, \|(x - p_0)_\perp\| \, \mathbb{1}_{\|x-x_0\|\leq h}\right] \\ &\leq C^d \omega_d f_{\max} h^{d+3} / \mathrm{rch}_{\min},\end{aligned}$$

where we used Lemma B.1 again. Dealing now with the second term of Equation (13),

$$\begin{aligned}\left\| \mathop{\mathbb{E}}_{x\sim D}\left[(x - p_0)_\perp (x - p_0)_\perp^\top \mathbb{1}_{\|x-x_0\|\leq h}\right]\right\|_* &\leq \mathop{\mathbb{E}}_{x\sim D}\left[\|(x - p_0)_\perp\| \, \|(x - p_0)_\perp\| \, \mathbb{1}_{\|x-x_0\|\leq h}\right] \\ &\leq C^d \omega_d f_{\max} h^{d+4} / (4\mathrm{rch}_{\min}^2).\end{aligned}$$

Finally, let us write

$$\Sigma_0 = \mathop{\mathbb{E}}_{x\sim D}\left[\frac{(x-p_0)_{/\!/}(x-p_0)_{/\!/}^\top}{h^2}\mathbb{1}_{\|x-x_0\|\le h}\right].$$

The matrix Σ_0 is symmetric and clearly has image $\mathrm{Im}(\Sigma_0) \subseteq T_{p_0}M$. Furthermore, since $\mathrm{d}(x_0, M) \le \eta \le h/\sqrt{2}$ and $h \le \mathrm{rch}_{\min}/8$, Proposition B.1 and Lemma 2.2 yield that $M \cap \mathrm{B}(x_0, h) \supseteq M \cap \mathrm{B}\big(p_0, \sqrt{7}h/4\big) \supseteq \mathrm{B}_M(p_0, h/2)$. Hence, for all $u \in T_{p_0}M$,

$$\begin{aligned}
h^2\langle \Sigma_0 u, u\rangle &= \mathop{\mathbb{E}}_{x\sim D}\left[\big\langle (x-p_0)_{/\!/}, u\big\rangle^2 \mathbb{1}_{\|x-x_0\|\le h}\right] \\
&= \mathop{\mathbb{E}}_{x\sim D}\left[\langle x-p_0, u\rangle^2 \mathbb{1}_{\|x-x_0\|\le h}\right] \\
&\ge f_{\min}\int_{\mathrm{B}_M(p_0,h/2)} \langle x-p_0, u\rangle^2 \,\mathrm{d}\mathcal{H}^d(x) \\
&= f_{\min}\int_{\mathrm{B}_d(0,h/2)} \big\langle \exp_{p_0}^M(v)-p_0, u\big\rangle^2 \left|\det\left(\mathrm{d}_v \exp_{p_0}^M\right)\right| \mathrm{d}v,
\end{aligned}$$

where $\mathcal{H}^d$ is the d-dimensional Hausdorff measure on $\mathbb{R}^n$, and $\exp_{p_0}^M : T_{p_0}M \to M$ is the exponential map of M at p_0. But [AL18, Proposition 8.7] states that there exists $c > 0$ such that for all $v \in \mathrm{B}_d(0, \mathrm{rch}_{\min}/4)$, $\left|\det\left(\mathrm{d}_v \exp_{p_0}^M\right)\right| \ge c^d$, and [AL19, Lemma 1] yields the bound $\left\|\exp_{p_0}^M(v) - (p_0+v)\right\| \le 5\|v\|^2/(8\mathrm{rch}_{\min})$. As a result, using the fact that $(a-b)^2 \ge a^2/2 - 3b^2$ for all $a, b \in \mathbb{R}$, we have

$$\begin{aligned}
h^2\langle \Sigma_0 u, u\rangle &\ge c^d f_{\min}\int_{\mathrm{B}_d(0,h/2)} \left(\langle v, u\rangle - \big\langle \exp_{p_0}^M(v) - (p_0+v), u\big\rangle\right)^2 \mathrm{d}v \\
&\ge c^d f_{\min}\int_{\mathrm{B}_d(0,h/2)} \langle v, u\rangle^2/2 - 3\big\langle \exp_{p_0}^M(v) - (p_0+v), u\big\rangle^2 \mathrm{d}v \\
&\ge c^d f_{\min}\int_{\mathrm{B}_d(0,h/2)} \langle v, u\rangle^2/2 - 3\|u\|^2\left(5\|v\|^2/(8\mathrm{rch}_{\min})\right)^2 \mathrm{d}v \\
&= c^d f_{\min}\sigma_{d-1}\left(\frac{1}{2d(d+2)} - \frac{3(5/8)^2}{d+4}\left(\frac{h}{2\mathrm{rch}_{\min}}\right)^2\right)\left(\frac{h}{2}\right)^{d+2}\|u\|^2 \\
&\ge (c')^d \omega_d f_{\min} h^{d+2}\|u\|^2,
\end{aligned}$$

as soon as $h \le \mathrm{rch}_{\min}/\sqrt{d}$. In particular, the last bound shows that the image of Σ_0 is exactly $T_{p_0}M$, and that $\mu_d(\Sigma_0) \ge \omega_d f_{\min}(c'h)^d$. Summing up the above, we have shown that

$$\Sigma_D(x_0, h) = \Sigma_0 + R,$$

where Σ_0 is symmetric, $\mathrm{Im}(\Sigma_0) = T_{\pi_M(x_0)}M$, $\mu_d(\Sigma_0) \ge \omega_d f_{\min}(c'h)^d$, and

$$\begin{aligned}
\|R\|_* &\le \omega_d f_{\max}(C'h)^d\left(\frac{\eta}{h} + \frac{\eta^2}{h^2} + \frac{h}{\mathrm{rch}_{\min}} + \frac{h^2}{\mathrm{rch}_{\min}^2}\right) \\
&\le \omega_d f_{\max}(C''h)^d\left(\frac{\eta}{h} + \frac{h}{\mathrm{rch}_{\min}}\right),
\end{aligned}$$

which is the announced result. □

D.2 Matrix Decomposition and Principal Angles

The following lemma ensures that the principal components of a matrix A are stable to perturbations, provided that A has a large-enough spectral gap. For a symmetric matrix $A \in \mathbb{R}^{n\times n}$, recall that $\mu_i(A)$ denotes its i-th largest singular value.

Lemma D.2 (Davis-Kahan)**.** *Let $\hat{A}, A \in \mathbb{R}^{n\times n}$ be symmetric matrices such that* $\operatorname{rank}(A) = d$*. If $\hat{T} \in \mathbb{G}^{n,d}$ denotes the linear space spanned by the first d eigenvectors of $\hat{A}$, and $T = \operatorname{Im}(A) \in \mathbb{G}^{n,d}$, then*

$$\angle(T, \hat{T}) := \left\|\pi_{\hat{T}} - \pi_T\right\|_{\mathrm{op}} \leq \frac{2\left\|\hat{A} - A\right\|_{\mathrm{F}}}{\mu_d(A)}.$$

Proof of Lemma D.2. It is a direct application of [YWS15, Theorem 2] with $r = 1$ and $s = d$. □

D.3 Low-rank Matrix Recovery

Proceeding further in the strategy described in Section 4.2, we now explain how to estimate the local covariance matrix $\Sigma_D(x_0, h) \in \mathbb{R}^{n\times n}$ (Equation (11)) in $\mathrm{STAT}(\tau)$.

Because $\Sigma_D(x_0, h) \in \mathbb{R}^{n\times n} = \mathbb{R}^{n^2}$ can be seen as a mean vector with respect to the unknown distribution D, $2n^2$ queries to $\mathrm{STAT}(\tau)$ would yield error $O(\tau)$ from Lemma 2.1. However, this would not use the low-rank structure of $\Sigma_D(x_0, h)$, i.e. some redundancy of its entries. To mitigate the query complexity of this estimation problem, we will use compressed sensing techniques [FCRP08]. Mimicking the vector case (Lemma 2.1), we put our problem in the broader context of the estimation of $\Sigma = \mathbb{E}_{x\sim D}[F(x)] \in \mathbb{R}^{k\times k}$ in $\mathrm{STAT}(\tau)$, where $F : \mathbb{R}^n \to \mathbb{R}^{k\times k}$ and Σ are approximately low rank (see Lemma D.4).

D.3.1 Restricted Isometry Property and Low-Rank Matrix Recovery

Let us first present some fundamental results coming of matrix recovery. Following [FCRP08, Section II], assume that we observe $y \in \mathbb{R}^q$ such that

$$y = \mathcal{L}(\Sigma) + z, \tag{14}$$

where $\Sigma \in \mathbb{R}^{k\times k}$ is the matrix of interest, $\mathcal{L} : \mathbb{R}^{k\times k} \to \mathbb{R}^q$ is a linear map seen as a sampling operator, and $z \in \mathbb{R}^q$ encodes noise and has small Euclidean norm $\|z\| \leq \xi$.

In general, when $q < k^2$, $\mathcal{L}$ has non-empty kernel, and hence one has no hope to recover Σ only from y, even with no noise. However, if Σ is (close to being) low-rank and that $\mathcal{L}$ does not shrink low-rank matrices too much, $\mathcal{L}(\Sigma)$ may not actually censor information on Σ, while compressing the dimension from k^2 to q. A way to formalize this idea states as follows.

Definition D.1 (Restricted Isometry Property)**.** Let $\mathcal{L} : \mathbb{R}^{k\times k} \to \mathbb{R}^q$ be a linear map, and $d \leq k$. We say that $\mathcal{L}$ satisfies the d-restricted isometry property with constant $\delta > 0$ if for all matrix $X \in \mathbb{R}^{k\times k}$ of rank at most d,

$$(1 - \delta)\|X\|_{\mathrm{F}} \leq \|\mathcal{L}(X)\| \leq (1 + \delta)\|X\|_{\mathrm{F}}.$$

We let $\delta_d(\mathcal{L})$ denote the smallest such δ.

To recover Σ only from the knowledge of y, consider the convex optimization problem (see [FCRP08]) over $X \in \mathbb{R}^{k\times k}$:

$$\begin{array}{ll} \text{minimize} & \|X\|_* \\ \text{subject to} & \|y - \mathcal{L}(X)\| \leq \xi. \end{array} \tag{15}$$

Let Σ_{opt} denote the solution of Equation (15). To give insights, the nuclear norm is seen here as a convex relaxation of the rank function [FCRP08], so that Equation (15) is expected to capture a low-rank matrix close to Σ. If $\mathcal{L}$ satisfies the restricted isometry property, the next result states that (15) does indeed capture such a low-rank matrix. In what follows, we let $\Sigma^{(d)} \in \mathbb{R}^{k\times k}$ denote the matrix closest to Σ among all the matrices of rank d, where closeness is indifferently measured in nuclear, Frobenius, or operator norm. That is, $\Sigma^{(d)}$ is the truncated singular value decomposition of Σ.

Theorem D.1 ([FCRP08, Theorem 4])**.** *Assume that $\delta_{5d} < 1/10$. Then the solution Σ_{opt} of Equation (15) satisfies*

$$\|\Sigma_{opt} - \Sigma\|_{\mathrm{F}} \leq C_0 \frac{\left\|\Sigma - \Sigma^{(d)}\right\|_*}{\sqrt{d}} + C_1 \xi,$$

where $C_0, C_1 > 0$ are universal constants.

D.3.2 Building a Good Matrix Sensing Operator

We now detail a standard way to build a sampling operator $\mathcal{L}$ that satisfies the restricted isometry property (Definition D.1), thereby allowing to recover low-rank matrices from a few measurements (Theorem D.1). For purely technical reasons, we shall present a construction over the complex linear space $\mathbb{C}^{k\times k}$. This will eventually enable us to recover results over $\mathbb{R}^{k\times k}$ via the isometry $\mathbb{R}^{k\times k} \hookrightarrow \mathbb{C}^{k\times k}$.

First, we note that given an orthonormal $\mathbb{C}$-basis $\mathbb{W} = (W_1, \ldots, W_{k^2})$ of $\mathbb{C}^{k\times k}$ for the Hermitian inner product $\langle A, B\rangle = \mathrm{tr}(A^*B)$, we can build a sampling operator $\mathcal{L}_{\mathbb{S}} : \mathbb{C}^{k\times k} \to \mathbb{C}^q$ by projecting orthogonally onto the space spanned by only q (randomly) pre-selected $\mathbb{S} \subseteq \mathbb{W}$ elements of the basis.

When $k = 2^\ell$, an orthonormal basis of $\mathbb{C}^{k\times k}$ of particular interest is the so-called *Pauli* basis [Liu11]. Its construction goes as follows:

- For $k = 2$ ($\ell = 1$), it is defined by $W_i^{(1)} = \sigma_i/\sqrt{2}$, where

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \sigma_4 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

  Note that the σ_i's have two eigenvalues, both belonging to $\{-1, 1\}$, so that they are both Hermitian and unitary. In particular, $\left\|W_i^{(1)}\right\|_{\mathrm{op}} = 1/\sqrt{2}$ and $\left\|W_i^{(1)}\right\|_{\mathrm{F}} = 1$ for all $i \in \{1, \ldots, 4\}$. One easily checks that $\left(W_i^{(1)}\right)_{1\leq i\leq 4}$ is an orthonormal basis of $\mathbb{C}^{2\times 2}$.

- For $k = 2^\ell$ ($\ell \geq 2$), the Pauli basis $\left(W_i^{(\ell)}\right)_{1\leq i\leq 2^\ell}$ is composed of matrices acting on the tensor space $\left(\mathbb{C}^2\right)^{\otimes \ell} \simeq \mathbb{C}^{2^\ell}$, and defined as the family of all the possible ℓ-fold tensor products of elements of $\left(W_i^{(1)}\right)_{1\leq i\leq 4}$. As tensor products preserve orthogonality, we get that $\left(W_i^{(\ell)}\right)_{1\leq i\leq 2^\ell}$ is an orthonormal basis of $\mathbb{C}^{2^\ell\times 2^\ell}$. Furthermore, as $\|W \otimes W'\|_{\mathrm{op}} = \|W\|_{\mathrm{op}} \|W'\|_{\mathrm{op}}$, we get that for all $i \in \left\{1, \ldots, 2^\ell\right\}$,

$$\left\|W_i^{(k)}\right\|_{\mathrm{op}} = \left(\frac{1}{\sqrt{2}}\right)^\ell = \frac{1}{\sqrt{k}}. \tag{16}$$

  Since $\|W\|_{\mathrm{F}} \leq \sqrt{k}\,\|W\|_{\mathrm{op}}$, the value $1/\sqrt{k}$ actually is the smallest possible common operator norm of an orthonormal basis of $\mathbb{C}^{k\times k}$. As will be clear in the proof of Lemma D.3, this last property — called incoherence in the matrix completion literature [Liu11] — is key to design a good sampling operator.

Still considering the case $k = 2^\ell$, we let $\mathcal{L}_{\mathrm{Pauli}} : \mathbb{C}^{k\times k} \to \mathbb{C}^q$ denote the random sampling operator defined by

$$\mathcal{L}_{\mathrm{Pauli}}(X) = \left(\frac{k}{\sqrt{q}} \left\langle W_{I_i}^{(\ell)}, X\right\rangle\right)_{1\leq i\leq q}, \tag{17}$$

where $(I_i)_{1\le i\le q}$ is an i.i.d. sequence with uniform distribution over $\{1,\dots,k^2\}$. Up to the factor $k/\sqrt{q}$, $\mathcal{L}_{\text{Pauli}}$ is the orthogonal projector onto the space spanned by $(W^{(\ell)}_{I_1},\dots,W^{(\ell)}_{I_q})$. This normalisation $k/\sqrt{q}$ is chosen so that for all $X\in\mathbb{C}^{k\times k}$,

$$\begin{aligned}\mathbb{E}\big[\|\mathcal{L}_{\text{Pauli}}(X)\|^2\big] &= \frac{k^2}{q}\sum_{i=1}^{k^2} q\Pr\left(I_1=i\right)\left|\left\langle W_i^{(\ell)},X\right\rangle\right|^2\\ &= \sum_{i=1}^{k^2}\left|\left\langle W_i^{(\ell)},X\right\rangle\right|^2 = \|X\|_{\text{F}}^2\,.\end{aligned}$$

That is, roughly speaking, $\mathcal{L}_{\text{Pauli}}$ satisfies the restricted isometry property (RIP, Definition D.1) on average. Actually, as soon as q is large enough compared to d, the result below states that $\mathcal{L}_{\text{Pauli}}$ does fulfill a restricted isometry property with high probability.

Lemma D.3. *Assume that $k=2^\ell$, and fix $0<\alpha\le 1$. There exist universal constants $c_0,c_1>0$ such that if $q\ge c_0 kd\log^6(k)\log(c_1/\alpha)$, then with probability at least $1-\alpha$, the following holds.*

For all $X\in\mathbb{R}^{k\times k}$ such that $\|X\|_*\le\sqrt{5d}\,\|X\|_{\text{F}}$,

$$|\|\mathcal{L}_{\text{Pauli}}(X)\|-\|X\|_{\text{F}}|\le\frac{\|X\|_{\text{F}}}{20}.$$

In particular, on the same event of probability at least $1-\alpha$, $\delta_{5d}\left(\mathcal{L}_{\text{Pauli}}\right)<1/10$.

Proof of Lemma D.3. The Pauli basis is an orthonormal basis of $\mathbb{C}^{k\times k}$, and from Equation (16), its elements all have operator norm smaller than $1/\sqrt{k}$. Hence, applying [Liu11, Theorem 2.1] with $K=\sqrt{k}\max_{1\le i\le k}\left\|W_i^{(\ell)}\right\|_{\text{op}}=1$, $r=5d$, $C=c_0\log(c_1/\alpha)$, and $\delta=1/20$ yields the first bound. The second one follows by recalling that any rank-r matrix $X\in\mathbb{R}^{k\times k}$ satisfies $\|X\|_*\le\sqrt{r}\,\|X\|_{\text{F}}$. □

D.3.3 Mean Matrix Completion with Statistical Queries

The low-rank matrix recovery of Appendices D.3.1 and D.3.2 combined with mean vector estimation in STAT(τ) for the Euclidean norm (see Lemma 2.1) lead to the following result.

Lemma D.4. *For all $\alpha\in(0,1]$, there exists a family of statistical query algorithms indexed by maps $F:\mathbb{R}^n\to\mathbb{R}^{k\times k}$ such that the following holds on an event of probability at least $1-\alpha$ (uniformly over F).*

Let D be a Borel probability distribution over $\mathbb{R}^n$, and $F:\mathbb{R}^n\to\mathbb{R}^{k\times k}$ be a map such that for all $x\in\mathbb{R}^n$, $\|F(x)\|_{\text{F}}\le 1$ and $\|F(x)\|_*\le\sqrt{5d}\,\|F(x)\|_{\text{F}}$. Write $\Sigma=\mathbb{E}_{x\sim D}\left[F(x)\right]$, and $\Sigma^{(d)}$ for the matrix closest to Σ among all the matrices of rank $d\le k$. Assume that $\Sigma\in\Xi$, where $\Xi\subseteq\mathbb{R}^{k\times k}$ is a known linear subspace of $\mathbb{R}^{k\times k}$.

Then, there exists a statistical query algorithm making at most $c_0 dk\log^6(k)\log(c_1/\alpha)$ queries to STAT(τ)*, and that outputs a matrix $\hat\Sigma\in\Xi$ that satisfies*

$$\left\|\hat\Sigma-\Sigma\right\|_{\text{F}}\le C_0\frac{\left\|\Sigma-\Sigma^{(d)}\right\|_*}{\sqrt{d}}+C_1\tau$$

on the event of probability at least $1-\alpha$ described above, where $C_0,C_1>0$ are universal constants.

Proof of Lemma D.4. Without loss of generality, we can assume that $k=2^\ell$. Indeed, one can always embed $\mathbb{R}^{k\times k}$ isometrically into $\mathbb{R}^{2^\ell\times 2^\ell}$, with $2^\ell=2^{\lceil\log_2(k)\rceil}\le 2k$, via the linear map

$$\mathbb{R}^{k\times k}\ni A\longmapsto\tilde A=\left(\begin{array}{c|c}A&0\\\hline 0&0\end{array}\right)\in\mathbb{R}^{2^\ell\times 2^\ell},$$

which preserves both the rank, the Frobenius and nuclear norms.

Let $q \geq 1$ be a fixed integer to be specified later, and $(I_i)_{1\leq i\leq q}$ be and i.i.d. sequence with uniform distribution over $\{1, \ldots, k^2\}$, and for $X \in \mathbb{R}^{k\times k}$, write

$$\mathcal{L}_{\text{Pauli}}(X) = \left(\frac{k}{\sqrt{q}}\left\langle W_{I_i}^{(\ell)}, X\right\rangle\right)_{1\leq i\leq q} \in \mathbb{C}^q = \mathbb{R}^{2q}, \tag{18}$$

as in Equation (17). For $x \in \mathbb{R}^n$, write $G(x) = \mathcal{L}_{\text{Pauli}}(F(x))/2 \in \mathbb{R}^{2q}$. From Lemma D.3, with probability at least $1-\alpha$ (over the randomness of $(I_i)_{1\leq i\leq q}$),

$$\|G(x)\| = \|\mathcal{L}_{\text{Pauli}}(F(x))\|/2 \leq (1+1/20)\|F(x)\|_{\text{F}}/2 \leq \|F(x)\|_{\text{F}} \leq 1$$

holds simultaneously for all the described $F : \mathbb{R}^{k\times k} \to \mathbb{R}^{2q}$. Hence, on this event of probability at least $1-\alpha$, Lemma 2.1 applies to G and provides a deterministic statistical query algorithm making $4q$ queries to $\text{STAT}(\tau)$, and that outputs a vector $y \in \mathbb{R}^{2q}$ such that

$$\left\|y - \mathop{\mathbb{E}}_{x\sim D}[G(x)]\right\| \leq C\tau,$$

where $C > 0$ is a universal constant. But on the other hand, by linearity,

$$\mathop{\mathbb{E}}_{x\sim D}[2G(x)] = \mathop{\mathbb{E}}_{x\sim D}[\mathcal{L}_{\text{Pauli}}(F(x))] = \mathcal{L}_{\text{Pauli}}\left(\mathop{\mathbb{E}}_{x\sim D}[F(x)]\right) = \mathcal{L}_{\text{Pauli}}(\Sigma),$$

where all the expected values are taken with respect D, conditionally on $(I_i)_{1\leq i\leq q}$. Hence, as soon as $q \geq c_0 dk\log^6(k)\log(c_1/\alpha)$, Theorem D.1 and Lemma D.3 combined together yields the following: on the same event of probability at least $1-\alpha$ as before, the solution Σ_{opt} to the convex optimization problem over $X \in \mathbb{R}^{k\times k}$ given by

$$\begin{array}{ll} \text{minimize} & \|X\|_* \\ \text{subject to} & \|2y - \mathcal{L}_{\text{Pauli}}(X)\| \leq 2C\tau, \end{array}$$

satisfies

$$\|\Sigma_{\text{opt}} - \Sigma\|_{\text{F}} \leq C_0\frac{\left\|\Sigma - \Sigma^{(d)}\right\|_*}{\sqrt{d}} + C_1(2C\tau).$$

Hence, the projected solution $\hat{\Sigma} = \pi_{\Xi}(\Sigma_{\text{opt}})$ onto $\Xi \subseteq \mathbb{R}^{k\times k}$ belongs to Ξ and satisfies

$$\begin{aligned} \left\|\hat{\Sigma} - \Sigma\right\|_{\text{F}} = \|\pi_{\Xi}(\Sigma_{\text{opt}} - \Sigma)\|_{\text{F}} &\leq \|\Sigma_{\text{opt}} - \Sigma\|_{\text{F}} \\ &\leq C_0\frac{\left\|\Sigma - \Sigma^{(d)}\right\|_*}{\sqrt{d}} + C_1'\tau, \end{aligned}$$

which concludes the proof. □

D.4 Tangent Space Estimation with Statistical Queries

We finally prove the main announced statement of Appendix D.

Proof of Theorem 4.2. Let $h > 0$ be a bandwidth to be specified later, such that $\eta \leq h/\sqrt{2}$ and $h \leq \text{rch}_{\min}/(8\sqrt{d})$. First note that $\Sigma_D(x_0, h) = \mathbb{E}_{x\sim D}[F(x)]$, where the function $F(x) = (x-x_0)(x-x_0)^\top/h^2\mathbb{1}_{\|x-x_0\|\leq h}$ is defined for all $x \in \mathbb{R}^n$, and is such that $\|F(x)\|_{\text{F}} \leq 1$ and $\text{rank}(F(x)) \leq 1$. In particular, $\|F(x)\|_* = \|F(x)\|_{\text{F}} \leq \sqrt{5d}\|F(x)\|_{\text{F}}$ for all $x \in \mathbb{R}^n$. Furthermore, $\Sigma_D(x_0, h)$ belongs to the linear space Ξ of symmetric matrices. Working on the event on which Lemma D.4 holds (with $\alpha = 1/2$, say), yields the existence of a deterministic SQ algorithm

making at most $c_0 dn \log^6(n) \log(2c_1)$ queries to STAT(τ), and that outputs a symmetric matrix $\hat{\Sigma}$ that satisfies

$$\left\|\hat{\Sigma} - \Sigma_D(x_0, h)\right\|_{\mathrm{F}} \leq C_0 \frac{\left\|\Sigma_D(x_0, h) - \Sigma_D^{(d)}(x_0, h)\right\|_*}{\sqrt{d}} + C_1 \tau,$$

with probability at least $1 - \alpha$. On the other hand, from Lemma D.1, provided that $\sqrt{2}\eta \leq h \leq \mathrm{rch}_{\min}/(8\sqrt{d})$, one can write

$$\Sigma_D(x_0, h) = \Sigma_0 + R,$$

where the symmetric matrix Σ_0 satisfies $\mathrm{Im}(\Sigma_0) = T_{\pi_M(x_0)}M$, $\mu_d(\Sigma_0) \geq \omega_d f_{\min}(ch)^d$ and $\|R\|_{\mathrm{F}} \leq \|R\|_* \leq \omega_d f_{\max}(Ch)^d \left(\frac{\eta}{h} + \frac{h}{\mathrm{rch}_{\min}}\right)$. As $\mathrm{rank}(\Sigma_0) = d$, we have in particular that,

$$\left\|\Sigma_D(x_0, h) - \Sigma_D^{(d)}(x_0, h)\right\|_* \leq \|\Sigma_D(x_0, h) - \Sigma_0\|_* = \|R\|_* .$$

Therefore, taking $\hat{T}(x_0)$ as the linear space spanned by the first d eigenvectors of $\hat{\Sigma}$, Lemma D.2 yields

$$\begin{aligned}
\angle\big(T_{\pi_M(x_0)}M, \hat{T}(x_0)\big) &= \left\|\pi_{\hat{T}(x_0)} - \pi_{T_{\pi_M(x_0)}M}\right\|_{\mathrm{op}} \\
&\leq \frac{2\left\|\hat{\Sigma} - \Sigma_0\right\|_{\mathrm{F}}}{\mu_d(\Sigma_0)} \\
&\leq 2\frac{\left\|\hat{\Sigma} - \Sigma_D(x_0, h)\right\|_{\mathrm{F}} + \|\Sigma_D(x_0, h) - \Sigma_0\|_{\mathrm{F}}}{\mu_d(\Sigma_0)} \\
&\leq \frac{2}{\omega_d f_{\min}(ch)^d}\left(C_0 \frac{\|R\|_*}{\sqrt{d}} + C_1\tau + \|R\|_{\mathrm{F}}\right) \\
&\leq \frac{C'^d}{\omega_d f_{\min}}\left(\omega_d f_{\max}\left\{\frac{\eta}{h} + \frac{h}{\mathrm{rch}_{\min}}\right\} + \frac{\tau}{h^d}\right).
\end{aligned}$$

We conclude by setting $h = \mathrm{rch}_{\min}\left\{\sqrt{\frac{\eta}{\mathrm{rch}_{\min}}} \vee \left(\frac{\tau}{\omega_d f_{\max}\mathrm{rch}_{\min}^d}\right)^{1/(d+1)}\right\}$ in this last bound. This value for h does satisfy $\sqrt{2}\eta \leq h \leq \mathrm{rch}_{\min}/(8\sqrt{d})$ since $\eta \leq \mathrm{rch}_{\min}/(64d)$ and $\frac{\tau}{\omega_d f_{\max}\mathrm{rch}_{\min}^d} \leq \left(\frac{1}{8\sqrt{d}}\right)^{d+1}$, so that the whole analysis applies, and yields the announced result. □

E Seed Point Detection

We now build the SQ point detection algorithm $\hat{x}_0 \in \mathbb{R}^n$ (Theorem 4.3), which is used to initialize in the SQ emulation of `Manifold Propagation` yielding the SQ reconstruction algorithm in the model $\mathrm{B}(0, R) \sqcap \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ where no seed point is available (Definition 2.6).

Recall that given a ball of radius $R > 0$ guaranteed to encompass $M = \mathrm{Supp}(D) \subseteq \mathrm{B}(0, R)$, and a target precision $\eta > 0$, we aim at finding a point that is η-close to M with statistical queries to STAT(τ). We follow the strategy of proof described in Section 4.3.

E.1 Detecting a Raw Initial Point

Starting from the whole ball $\mathrm{B}(0, R)$, the following result allows us to find a point nearby M using a binary search, with best precision of order $\Omega(\tau^{1/d})$. Let us note that it does not explicitly rely on any differential property of M, but only the behavior of the mass of balls for D (Lemma B.1).

Theorem E.1. *Let $D \in \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ have support $M = \mathrm{Supp}(D) \subseteq \mathrm{B}(0,R)$. Let $\Lambda_0 \leq \mathrm{rch}_{\min}/8$ be fixed, and assume that $\frac{\Lambda_0}{\sqrt{\log(6R/\Lambda_0)}} \geq 21\mathrm{rch}_{\min}\sqrt{n}\left(\frac{\tau}{\omega_d f_{\min}\mathrm{rch}^d_{\min}}\right)^{1/d}$.*

Then there exists a deterministic statistical query algorithm making at most $3n\log(6R/\Lambda_0)$ queries to $\mathrm{STAT}(\tau)$, and that outputs a point $\hat{x}_0^{raw} \in \mathrm{B}(0,R)$ such that

$$\mathrm{d}(\hat{x}_0^{raw}, M) \leq \Lambda_0.$$

Remark E.1. Recall from Section 2.2.3 that we always assume that $R \geq \mathrm{rch}_{\min}/\sqrt{2}$ to ensure that the model is nonempty. As a result $\log(6R/\Lambda_0) \geq 0$ for all $\Lambda_0 \leq \mathrm{rch}_{\min}/8$.

Proof of Theorem E.1. The idea is to use a divide and conquer strategy over a covering $\{x_i\}_{1\leq i\leq N}$ of $\mathrm{B}(0,R)$. The algorithm recurses over a subset of indices $\mathcal{I} \subseteq \{1,\dots,N\}$ that is maintained to fulfill $\cup_{i\in\mathcal{I}}\mathrm{B}(x_i,h) \cap M \neq \emptyset$ for some known $h > 0$. This property can be checked with the single query $r = \mathbb{1}_{\cup_{i\in\mathcal{I}}\mathrm{B}(x_i,h)}$ to $\mathrm{STAT}(\tau)$, provided that $D(\cup_{i\in\mathcal{I}}\mathrm{B}(x_i,h)) > \tau$. To ensure the later, the radius $h > 0$ is dynamically increased at each iteration. The algorithm stops when $\mathcal{I}$ is reduced to a singleton. More formally, we consider `SQ Ambient Binary Search`.

Algorithm 2 `SQ Ambient Binary Search`

Require:
Model parameters $d, \mathrm{rch}_{\min}, f_{\min}$
Precision $\Lambda_0 > 0$
Initialize value $h \leftarrow \Lambda_0/2$, and set $\Delta = 6\mathrm{rch}_{\min}\left(\frac{\tau}{\omega_d f_{\min}\mathrm{rch}^d_{\min}}\right)^{1/d}$
Consider a minimal $(\Lambda_0/2)$-covering $\{x_i\}_{1\leq i\leq N}$ of $\mathrm{B}(0,R)$, where $N = \mathrm{cv}_{\mathrm{B}(0,R)}(\Lambda_0/2)$
Initialize sets $\mathcal{I} \leftarrow \{1,\dots,N\}$, $\mathcal{L} \leftarrow \emptyset$ and $\mathcal{R} \leftarrow \emptyset$
while $|\mathcal{I}| > 1$ **do**
  Split $\mathcal{I} = \mathcal{L} \cup \mathcal{R}$ into two disjoint sets $\mathcal{L} \cap \mathcal{R} = \emptyset$ such that $||\mathcal{L}| - |\mathcal{R}|| \leq 1$
  Query $r = \mathbb{1}_{\cup_{i\in\mathcal{L}}\mathrm{B}\left(x_i,\sqrt{h^2+\Delta^2}\right)}$ to the $\mathrm{STAT}(\tau)$ oracle
  a $\leftarrow$ Value answered by the oracle
  if a $> \tau$ **then**
    $\mathcal{I} \leftarrow \mathcal{L}$
  else
    $\mathcal{I} \leftarrow \mathcal{R}$
  end if
  $h \leftarrow \sqrt{h^2+\Delta^2}$
end while
return The only element of $\hat{x}_0^{raw}$ of $\{x_i\}_{i\in\mathcal{I}}$

Because $|\mathcal{I}|$ is a decreasing sequence of integers, it is clear that `SQ Ambient Binary Search` terminates, and that $|\mathcal{I}_{final}| = 1$ so that the output $\hat{x}_0^{raw}$ is well defined. As each *while* loop does only one query to $\mathrm{STAT}(\tau)$, and that $N = \mathrm{cv}_{\mathrm{B}(0,R)}(\Lambda_0/2) \leq (6R/\Lambda_0)^n$ from Proposition B.4 and $\Lambda_0 \leq R$, it makes at most $\lfloor \log_2(N)+1 \rfloor \leq \lfloor n\log(6R/\Lambda_0)/\log(2)+1 \rfloor \leq 3n\log(6R/\Lambda_0)$ queries in total.

Let us now prove that the output $\hat{x}_0^{raw}$ satisfies $\mathrm{d}(\hat{x}_0^{raw}, M) \leq \Lambda_0$. For this, we show that when running `SQ Ambient Binary Search`, the inequality $\min_{i\in\mathcal{I}}\mathrm{d}(x_i,M) \leq h$ is maintained (recall that both $\mathcal{I}$ and h are dynamic), or equivalently that $\cup_{i\in\mathcal{I}}\mathrm{B}(x_i,h) \cap M \neq \emptyset$. At initialization, this is clear because $\mathcal{I} = \{1,\dots,N\}$, $h = \Lambda_0/2$, and $\{x_i\}_{1\leq i\leq N}$ is a $(\Lambda_0/2)$-covering of $\mathrm{B}(0,R) \supseteq M$. Then, proceeding by induction, assume that $\cup_{i\in\mathcal{I}}\mathrm{B}(x_i,h) \cap M \neq \emptyset$ when entering an iteration of the *while* loop. Let $i_0 \in \mathcal{I}$ be such that $\mathrm{d}(x_{i_0},M) \leq h$. From Lemma B.1,

provided that $\sqrt{h^2+\Delta^2} \leq \mathrm{rch}_{\min}/8$, we have

$$\begin{aligned} D\left(\cup_{i\in\mathcal{I}}\mathrm{B}\left(x_i, \sqrt{h^2+\Delta^2}\right)\right) &\geq D\left(\mathrm{B}\left(x_{i_0}, \sqrt{h^2+\Delta^2}\right)\right) \\ &\geq (\sqrt{7/24})^d \omega_d f_{\min}\left((h^2+\Delta^2) - \mathrm{d}(x_{i_0}, M)^2\right)^{d/2} \\ &\geq (\sqrt{7/24})^d \omega_d f_{\min}\Delta^d \\ &= (\sqrt{7/24})^d 6^d \tau \\ &> 2\tau. \end{aligned} \tag{19}$$

Hence, if we let a denote the answer of the oracle to the query $r = \mathbb{1}_{\cup_{i\in\mathcal{L}}\mathrm{B}\left(x_i,\sqrt{h^2+\Delta^2}\right)}$, we have:

- If $\mathrm{a} > \tau$, then

$$D\left(\cup_{i\in\mathcal{L}}\mathrm{B}\left(x_i, \sqrt{h^2+\Delta^2}\right)\right) \geq \mathrm{a} - \tau > 0,$$

  so that after the updates $\mathcal{I} \leftarrow \mathcal{L}$ and $h \leftarrow \sqrt{h^2+\Delta^2}$, we still have $\cup_{i\in\mathcal{I}}\mathrm{B}\left(x_i, h\right) \cap M \neq \emptyset$.

- Otherwise $a \leq \tau$, so that from Equation (19),

$$\begin{aligned} D\left(\cup_{i\in\mathcal{R}}\mathrm{B}\left(x_i, \sqrt{h^2+\Delta^2}\right)\right) &\geq D\left(\cup_{i\in\mathcal{I}}\mathrm{B}\left(x_i, \sqrt{h^2+\Delta^2}\right)\right) - D\left(\cup_{i\in\mathcal{L}}\mathrm{B}\left(x_i, \sqrt{h^2+\Delta^2}\right)\right) \\ &> 2\tau - (\mathrm{a}+\tau) \\ &\geq 0. \end{aligned}$$

  So as above, after the updates $\mathcal{I} \leftarrow \mathcal{R}$ and $h \leftarrow \sqrt{h^2+\Delta^2}$, we still have $\cup_{i\in\mathcal{I}}\mathrm{B}\left(x_i, h\right) \cap M \neq \emptyset$.

Consequently, when the algorithm terminates, we have

$$\begin{aligned} \mathrm{d}(\hat{x}_0^{raw}, M) &\leq h_{final} \\ &\leq \sqrt{\left(\frac{\Lambda_0}{2}\right)^2 + 3n\log(6R/\Lambda_0)\Delta^2} \\ &\leq \frac{\Lambda_0}{2} + \sqrt{3n\log(6R/\Lambda_0)}6\mathrm{rch}_{\min}\left(\frac{\tau}{\omega_d f_{\min}\mathrm{rch}_{\min}^d}\right)^{1/d} \\ &\leq \Lambda_0, \end{aligned}$$

since $\frac{\Lambda_0}{\sqrt{\log(6R/\Lambda_0)}} \geq 21\mathrm{rch}_{\min}\sqrt{n}\left(\frac{\tau}{\omega_d f_{\min}\mathrm{rch}_{\min}^d}\right)^{1/d}$. The above also shows that when running the algorithm we have $\sqrt{h^2+\Delta^2} \leq h_{final} \leq \Lambda_0 \leq \mathrm{rch}_{\min}/8$, which ensures that Equation (19) is valid throughout and concludes the proof. □

E.2 Refined Point Detection

We finally prove the main announced statement of Appendix E.

Proof of Theorem 4.3. The idea is to first detect a possibly coarse base point $\hat{x}_0^{raw}$ using a divide and conquer strategy in the ambient space (Theorem E.1), and then refine it by considering iterated projections of $\hat{x}_0^{raw}$ given by the local conditional mean (Theorem 4.1). More precisely, let $\hat{x}_0^{raw}$ be the output of the point detection SQ algorithm of Theorem E.1 applied with parameter

$$\Lambda_0 = \max\left\{\eta, \min\left\{\frac{1}{16}, \frac{\Gamma}{2C^d}\right\}\mathrm{rch}_{\min}\right\},$$

where $C^d, \Gamma > 0$ are the constants of Theorem 4.1. From the assumptions on the parameters (recall also that we necessarily have $R \geq \mathrm{rch}_{\min}/\sqrt{2}$, see Section 2.2.3), we have $\Lambda_0 \leq \mathrm{rch}_{\min}/8$ and

$$\frac{\Lambda_0}{\sqrt{\log(6R/\Lambda_0)}} \geq 21\mathrm{rch}_{\min}\sqrt{n}\left(\frac{\tau}{\omega_d f_{\min}\mathrm{rch}_{\min}^d}\right)^{1/d},$$

so that Theorem E.1 applies and guarantees that $\hat{x}_0^{raw}$ can be obtained with at most $3n\log(6R/\Lambda_0)$ queries to $\mathrm{STAT}(\tau)$ and satisfies $\mathrm{d}(\hat{x}_0^{raw}, M) \leq \Lambda_0$.

If $\Lambda = \eta$ — condition which can be checked by the learner since the parameters η, Γ, d and $\mathrm{rch}_{\min}$ are assumed to be known —, then $\hat{x}_0 := \hat{x}_0^{raw}$ clearly satisfies $\mathrm{d}(\hat{x}_0, M) = \mathrm{d}(\hat{x}_0^{raw}, M) \leq \eta$, and has required at most $3n\log(6R/\Lambda_0) = 3n\log(6R/\eta)$ queries to $\mathrm{STAT}(\tau)$. Otherwise, $\eta \leq \Lambda_0$, and we iterate the SQ approximate projections $\hat{\pi}(\cdot)$ given by Theorem 4.1. Namely, we let $\hat{y}_0 = \hat{x}_0^{raw}$ and for all integer $k \geq 1$, $\hat{y}_k = \hat{\pi}(\hat{y}_{k-1})$. In total, note that the computation of $\hat{y}_k$ requires at most $3n\log(6R/\eta) + k(2n+1) \leq 3n\big(\log(6R/\eta) + k\big)$ queries to $\mathrm{STAT}(\tau)$. Similarly as above, from the assumptions on the parameters, one easily shows by induction that since $\mathrm{d}(\hat{y}_0, M) \leq \Lambda_0 \leq \frac{\mathrm{rch}_{\min}}{16}$, Theorem 4.1 applies to each $\hat{y}_k$ and guarantees that

$$\begin{aligned}
\mathrm{d}(\hat{y}_k, M) &= \mathrm{d}(\hat{\pi}(\hat{y}_{k-1}), M) \\
&\leq \|\hat{\pi}(\hat{y}_{k-1}) - \pi_M(\hat{y}_{k-1}))\| \\
&\leq \max\left\{\frac{C^d \mathrm{d}(\hat{y}_{k-1}, M)^2}{\Gamma \mathrm{rch}_{\min}}, C^d \Gamma^{\frac{2}{d+1}-1}\mathrm{rch}_{\min}\left(\frac{\tau}{\omega_d f_{\min}\mathrm{rch}_{\min}^d}\right)^{\frac{2}{d+1}}\right\} \\
&\leq \max\left\{\frac{\mathrm{d}(\hat{y}_{k-1}, M)}{2}, C^d \Gamma^{\frac{2}{d+1}-1}\mathrm{rch}_{\min}\left(\frac{\tau}{\omega_d f_{\min}\mathrm{rch}_{\min}^d}\right)^{\frac{2}{d+1}}\right\} \\
&\leq \max\left\{\frac{\Lambda_0}{2^k}, C^d \Gamma^{\frac{2}{d+1}-1}\mathrm{rch}_{\min}\left(\frac{\tau}{\omega_d f_{\min}\mathrm{rch}_{\min}^d}\right)^{\frac{2}{d+1}}\right\}.
\end{aligned}$$

To conclude, fix $k_0 := \lceil \log_2(\Lambda_0/\eta) \rceil \leq \log(6\Lambda_0/\eta)$, and set $\hat{x}_0 := \hat{y}_{k_0}$. From the previous bound, we obtain that

$$\mathrm{d}(\hat{x}_0, M) \leq \max\left\{\eta, C^d \Gamma^{\frac{2}{d+1}-1}\mathrm{rch}_{\min}\left(\frac{\tau}{\omega_d f_{\min}\mathrm{rch}_{\min}^d}\right)^{\frac{2}{d+1}}\right\},$$

with $\hat{x}_0$ requiring at most $3n\big(\log(6R/\eta) + \log(6\Lambda_0/\eta)\big) \leq 6n\log(6R/\eta)$ queries to $\mathrm{STAT}(\tau)$ to be computed, which concludes the proof. □

F Proof for the Main Statistical Query Manifold Estimators

This section is devoted to the proof of the two SQ manifold estimation upper bounds: the first one in the fixed point model $\{0\} \sqcup \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ (Theorem 5.1), and the second one for the bounding ball model $\mathrm{B}(0,R) \sqcap \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ (Theorem 5.4).

Proof of Theorem 5.1. Let us write

$$\Delta := \mathrm{rch}_{\min}\max\left\{\sqrt{\varepsilon/(\mathrm{rch}_{\min}\bar{C}_d)}, \mathbf{C}^{\frac{3}{2}}\left(\tau/(\omega_d f_{\min}\mathrm{rch}_{\min}^d)\right)^{\frac{1}{d+1}}\right\},$$

for some large enough $\bar{C}_d > 0$ depending on d and $\mathbf{C}$ to be chosen later, and $\delta = \Delta/2$. We will run `Manifold Propagation` with scale parameters Δ, δ, angle $\sin\alpha = 1/64$, and initialization point $\hat{x}_0 = 0 \in M$, the SQ projection routine $\hat{\pi}(\cdot)$ of Theorem 4.1 and the SQ tangent space routine $\hat{T}(\cdot)$ of Theorem 4.2. If we prove that these routines are precise enough, then Theorem 3.1 will

assert that the output point cloud $\mathcal{O}$ and associated tangent space estimates $\mathbb{T}_{\mathcal{O}}$ of `Manifold Propagation` fulfill the assumptions of Theorem 2.1. This will hence allow to reconstruct M with a good triangulation, as claimed.

Note by now that at each iteration `Manifold Propagation`, exactly one call to each SQ routine $\hat{\pi}(\cdot)$ and $\hat{T}(\cdot)$ are made, yielding at most $(2n+1)+Cdn\log^6(n) \leq C'dn\log^6(n)$ statistical queries. But if Theorem 3.1 applies, we get that the number of iteration N_{loop} of `Manifold Propagation` satisfies

$$\begin{aligned} N_{\text{loop}} &\leq \frac{\mathcal{H}^d(M)}{\omega_d(\delta/32)^d} \\ &\leq \frac{\bar{C}'_d}{f_{\min}(\sqrt{\text{rch}_{\min}}\varepsilon)^d} \\ &= \frac{\bar{C}'_d}{f_{\min}\text{rch}_{\min}^d}\left(\frac{\text{rch}_{\min}}{\varepsilon}\right)^{d/2}, \end{aligned}$$

where the second inequality comes from the fact that $1 = \int_M f \mathrm{d}\mathcal{H}^d \geq f_{\min}\mathcal{H}^d(M)$. In total, the resulting SQ algorithm hence makes at most

$$\begin{aligned} q &\leq \left(C'dn\log^6(n)\right)\frac{\bar{C}'_d}{f_{\min}\text{rch}_{\min}^d}\left(\frac{\text{rch}_{\min}}{\varepsilon}\right)^{d/2} \\ &= n\log^6 n\frac{C_d}{f_{\min}\text{rch}_{\min}^d}\left(\frac{\text{rch}_{\min}}{\varepsilon}\right)^{d/2} \end{aligned}$$

queries to $\text{STAT}(\tau)$, which is the announced complexity. It only remains to verify that the SQ routines $\hat{\pi}(\cdot)$ and $\hat{T}(\cdot)$ are indeed precise enough so that Theorem 3.1 applies, and to bound the final precision given by the triangulation of Theorem 2.1.

To this aim, we notice that the assumption made on τ puts it in the regime of validity of Theorem 4.1 and Theorem 4.2. Let us write

$$\mathbf{C} := \max\left\{C^d\Gamma^{\frac{2}{d+1}-1}, \tilde{C}^d\frac{f_{\max}}{f_{\min}}\right\} \leq \frac{(\max\{C,\tilde{C}\})^d}{\Gamma},$$

where $C > 0$ is the constant of Theorem 4.1 and $\tilde{C} > 1$ that of Theorem 4.2. Note by now that since $f_{\max} \geq f_{\min}$, we have $\mathbf{C} \geq 1$. For short, we also let $\tilde{\tau} := \tau/(\omega_d f_{\min}\text{rch}_{\min}^d)$.

At initialization, and since $D \in \{0\} \sqcup \mathcal{D}^{n,d}_{\text{rch}_{\min}}(f_{\min}, f_{\max}, L)$, the seed point $\hat{x}_0 = 0$ belongs to M, meaning that

$$\mathrm{d}(\hat{x}_0, M) = 0 \leq \eta := \text{rch}_{\min}\max\left\{\frac{1}{\mathbf{C}^2}\left(\frac{\Delta}{\text{rch}_{\min}}\right)^2, \mathbf{C}\tilde{\tau}^{\frac{2}{d+1}}\right\}.$$

Note that from the assumptions on the parameters, $\eta \leq \text{rch}_{\min}/(64d)$. Hence, on the η-offset M^η of M, Theorem 4.2 asserts that $\hat{T}(\cdot)$ has precision

$$\begin{aligned} \sin\theta &\leq \max\left\{\tilde{C}^d\frac{f_{\max}}{f_{\min}}\frac{1}{\mathbf{C}}\frac{\Delta}{\text{rch}_{\min}}, \tilde{C}^d\frac{f_{\max}}{f_{\min}}\sqrt{\mathbf{C}}\tilde{\tau}^{\frac{1}{d+1}}\right\} \\ &\leq \max\left\{\frac{\Delta}{\text{rch}_{\min}}, \mathbf{C}^{\frac{3}{2}}\tilde{\tau}^{\frac{1}{d+1}}\right\} \end{aligned}$$

for estimating tangent spaces. As a result, we have

$$\begin{aligned}
\frac{5\Delta^2}{8\mathrm{rch}_{\min}} + \eta + \Delta \sin\theta &\leq 3\mathrm{rch}_{\min} \max\left\{ \left(\frac{\Delta}{\mathrm{rch}_{\min}}\right)^2, \frac{\eta}{\mathrm{rch}_{\min}}, \frac{\Delta}{\mathrm{rch}_{\min}} \sin\theta \right\} \\
&\leq 3\mathrm{rch}_{\min} \max\left\{ \left(\frac{\Delta}{\mathrm{rch}_{\min}}\right)^2, \frac{\eta}{\mathrm{rch}_{\min}}, \sin^2\theta \right\} \\
&\leq 3\mathrm{rch}_{\min} \max\left\{ \left(\frac{\Delta}{\mathrm{rch}_{\min}}\right)^2, \mathbf{C}^3 \tilde{\tau}^{\frac{2}{d+1}} \right\} \\
&:= \Lambda.
\end{aligned}$$

Using again the assumptions on the parameters, we have $\Lambda \leq \mathrm{rch}_{\min}/8$. Hence, applying Theorem 4.1 and elementary simplifications given by the assumptions on the parameters yield that, over the Λ-offset M^Λ of M, the projection $\hat{\pi}(\cdot)$ has precision at most

$$\begin{aligned}
\eta' &\leq \mathrm{rch}_{\min} \max\left\{ \left(\frac{9C^d}{\Gamma}\left(\frac{\Delta}{\mathrm{rch}_{\min}}\right)^2\right)\left(\frac{\Delta}{\mathrm{rch}_{\min}}\right)^2, \right. \\
&\qquad\qquad \left. \max\left\{ \frac{9C^d\mathbf{C}^6}{\Gamma}\tilde{\tau}^{2/(d+1)}, C^d\Gamma^{\frac{2}{d+1}-1} \right\} \tilde{\tau}^{2/(d+1)} \right\} \\
&= \mathrm{rch}_{\min} \max\left\{ \left(\frac{9C^d}{\Gamma}\left(\frac{\Delta}{\mathrm{rch}_{\min}}\right)^2\right)\left(\frac{\Delta}{\mathrm{rch}_{\min}}\right)^2, C^d\Gamma^{\frac{2}{d+1}-1}\tilde{\tau}^{2/(d+1)} \right\} \\
&\leq \mathrm{rch}_{\min} \max\left\{ \frac{1}{\mathbf{C}^2}\left(\frac{\Delta}{\mathrm{rch}_{\min}}\right)^2, \mathbf{C}\tilde{\tau}^{2/(d+1)} \right\} \\
&= \eta.
\end{aligned}$$

Additionally, one easily checks that $\Delta \leq \mathrm{rch}_{\min}/24$, $\eta \leq \Delta/24$ and $\max\{\sin\alpha, \sin\theta\} \leq 1/64$, so that Theorem 3.1 applies: `Manifold Propagation` terminates and outputs a finite point cloud $\mathcal{O}$ such that $\max_{x\in\mathcal{O}} \mathrm{d}(x, M) \leq \eta$ and $\max_{p\in M} \mathrm{d}(p, \mathcal{O}) \leq \Delta + \eta \leq 2\Delta$, together with tangent space estimates $\mathbb{T}_{\mathcal{O}}$ with error at most $\sin\theta$. Hence, applying Theorem 2.1 with parameters $\Delta' = 2\Delta$, η and $\sin\theta$ (for which one easily checks that they fulfill its assumptions), we get that the triangulation $\hat{M}$ of Theorem 2.1 computed over $\mathcal{O}$ and $\mathbb{T}_{\mathcal{O}}$ achieves precision

$$\mathrm{d_H}(M, \hat{M}) \leq \frac{C_d \Delta'^2}{\mathrm{rch}_{\min}} \leq \max\left\{ \varepsilon, \mathbf{C}^3 \mathrm{rch}_{\min} \tilde{\tau}^{\frac{2}{d+1}} \right\},$$

which yields the announced result since $\mathbf{C} \leq (C \vee \tilde{C})^d/\Gamma$. □

Proof of Theorem 5.4. The proof follows the same lines as that of Theorem 5.1, except for the seed point $\hat{x}_0$ which is not trivially available, and requires extra statistical queries. More precisely, we let $\hat{x}_0$ be the output point given by the SQ detection algorithm of Theorem 4.3 applied with precision parameter $\varepsilon/2$. This point requires no more than $6n\log(6R/\varepsilon)$ statistical queries to $\mathrm{STAT}(\tau)$. Furthermore, adopting the same notation as in the proof of Theorem 5.1 we have

$$\begin{aligned}
\mathrm{d}(\hat{x}_0, M) &\leq \max\left\{ \frac{\varepsilon}{2}, C^d\Gamma^{\frac{2}{d+1}-1}\mathrm{rch}_{\min}\left(\frac{\tau}{\omega_d f_{\min}\mathrm{rch}_{\min}^d}\right)^{\frac{2}{d+1}} \right\} \\
&\leq \mathrm{rch}_{\min} \max\left\{ \frac{1}{\mathbf{C}^2}\left(\frac{\Delta}{\mathrm{rch}_{\min}}\right)^2, \mathbf{C}\tilde{\tau}^{\frac{2}{d+1}} \right\},
\end{aligned}$$

so that the rest of the proof runs exactly as that of Theorem 5.1, and yields the result. □

G Statistical Query Lower Bounds in Metric Spaces

In spirit, the lower bound techniques developed below are similar to the *statistical dimension* of [Fel17], developed for general search problems. However, when working with manifold models, this tool appears difficult to handle, due to the singular nature of low-dimensional distributions, yielding non-dominated models. Indeed, if D_0 and D_1 are distributions that have supports being d-dimensional submanifolds $M_0, M_1 \subseteq \mathbb{R}^n$, and that $M_0 \neq M_1$, then D_0 and D_1 cannot be absolutely continuous with respect to one another. As a result, any lower bound technique involving Kullback-Leibler or chi-squared divergences becomes non-informative (see for instance [Fel17, DKS17]).

Instead, we present techniques that are well-suited for non-dominated models. They apply for SQ estimation in metric spaces (Θ, ρ) (see Section 2.1), as opposed to the (more general) setting of *search problems* of [Fel17]. We decompose these results into two different types of lower bounds:

- (Appendix G.1) The *information-theoretic* ones, yielding a maximal estimation precision $\varepsilon = \varepsilon(\tau)$ given a tolerance τ;
- (Appendix G.2) The *computational* ones, yielding a minimal number of queries $q = q(\varepsilon)$ to achieve a given precision ε.

G.1 Information-Theoretic Lower Bound for Randomized SQ Algorithms

The proofs of the informational lower bounds Theorems 5.2 and 5.5 are based on the following Theorem G.1, which is similar to so-called *Le Cam's Lemma* [Yu97]. To introduce this result we define the *total variation distance* between probability distributions.

Definition G.1 (Total Variation Distance)**.** Given two probability distributions D_0 and D_1 over $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, the *total variation distance* between them is defined by

$$\begin{aligned}\mathrm{TV}(D_0, D_1) &= \sup_{B \in \mathcal{B}(\mathbb{R}^n)} \big|D_0(B) - D_1(B)\big| \\ &= \sup_{\substack{r:\mathbb{R}^n \to [-1,1] \\ \text{measurable}}} \frac{1}{2}\big|\mathop{\mathbb{E}}_{D_0}[r] - \mathop{\mathbb{E}}_{D_1}[r]\big|.\end{aligned}$$

The second formula above for the total variation suggests how it can measure an impossibility of estimation with $\mathrm{STAT}(\tau)$ oracles: two distributions that are close in total variation distance provide a malicious oracle to make them — and their parameter of interest — indistinguishable using SQ's, . This lower bound insight is what underlies Le Cam's Lemma [Yu97] in the sample model, and it adapts easily to (randomized) SQ's in the following way.

Theorem G.1 (Le Cam's Lemma for Statistical Queries)**.** *Consider a model $\mathcal{D}$ and a parameter of interest $\theta : \mathcal{D} \to \Theta$ in the metric space (Θ, ρ). Assume that there exist* hypotheses $D_0, D_1 \in \mathcal{D}$*, such that*

$$\mathrm{TV}(D_0, D_1) \leq \tau/2 \text{ and } \delta < \rho\big(\theta(D_0), \theta(D_1)\big)/2.$$

If $\alpha < 1/2$, then no $\mathrm{STAT}(\tau)$ *randomized SQ algorithm can estimate θ with precision $\varepsilon \leq \delta$ and probability of success $1 - \alpha$ over $\mathcal{D}$ (no matter how many queries it does).*

Proof of Theorem G.1. We prove the contrapositive. For this purpose, assume that a randomized SQ algorithm $\mathtt{A} \sim \mathcal{A}$ estimates θ with precision $\varepsilon \leq \delta$ and probability at least $1 - \alpha$ over $\mathcal{D}$. We will show that $\alpha \geq 1/2$.

Consider the oracle which, given a query $r : \mathbb{R}^n \to [-1, 1]$ to the distribution $D \in \mathcal{D}$, returns the answer:

- $\mathrm{a} = \mathbb{E}_{D_0}[r]$ if $D = D_1$;
- $\mathrm{a} = \mathbb{E}_{D}[r]$ if $D \in \mathcal{D} \setminus \{D_1\}$.

As for all query $r : \mathbb{R}^n \to [-1, 1]$, $|\mathbb{E}_{D_0}[r] - \mathbb{E}_{D_1}[r]| \leq 2\,\mathrm{TV}(D_0, D_1) \leq \tau$, it is a valid $\mathrm{STAT}(\tau)$ oracle. Furthermore, notice that the answers of this oracle are the same for $D = D_0$ and $D = D_1$. Writing $\mathtt{A} = (r_1, \ldots, r_q, \hat{\theta}) \sim \mathcal{A}$, we denote these answers by $\mathrm{a}_1, \ldots, \mathrm{a}_q$. The a_i's are random variables, with randomness driven by the randomness of $\mathtt{A} \sim \mathcal{A}$. For $i \in \{0, 1\}$, let us consider the event

$$B_i = \left\{ \rho\left(\theta(D_i), \hat{\theta}(\mathrm{a}_1, \ldots, \mathrm{a}_q)\right) \leq \varepsilon \right\}.$$

The fact that $\mathcal{A}$ estimates θ with precision ε and probability at least $1 - \alpha$ over $\mathcal{D}$ translates into $\Pr_{\mathtt{A} \sim \mathcal{A}}(B_i) \geq 1 - \alpha$, for $i \in \{0, 1\}$. But since $\varepsilon \leq \delta < \rho(\theta(D_0), \theta(D_1))/2$, the events B_0 and B_1 are disjoint (i.e. $B_0 \cap B_1 = \emptyset$). As a result,

$$1 \geq \Pr_{\mathtt{A} \sim \mathcal{A}}(B_0 \cup B_1) = \Pr_{\mathtt{A} \sim \mathcal{A}}(B_0) + \Pr_{\mathtt{A} \sim \mathcal{A}}(B_1) \geq 2(1 - \alpha),$$

which yields $\alpha \geq 1/2$ and concludes the proof. □

G.2 Computational Lower Bound

This section is dedicated to prove the following Theorem G.2, that provides a computational lower bound for support estimation in Hausdorff distance. It involves the generalized notion of metric packing, which is defined right below.

Theorem G.2. *Given a model $\mathcal{D}$ over $\mathbb{R}^n$, any randomized SQ algorithm estimating $M = \mathrm{Supp}(D) \subseteq \mathbb{R}^n$ with precision ε for the Hausdorff distance, and with probability of success at least $1 - \alpha$, must make at least*

$$q \geq \frac{\log\big((1 - \alpha)\mathrm{pk}_{(\mathcal{M}, \mathrm{d_H})}(\varepsilon)\big)}{\log(1 + \lfloor 1/\tau \rfloor)}$$

queries to $\mathrm{STAT}(\tau)$, where $\mathcal{M} = \{\mathrm{Supp}(D), D \in \mathcal{D}\}$.

Similarly to Appendix G.1, we put Theorem G.2 in the broader context of SQ estimation in metric spaces (see Section 2.1), and state the more general Theorem G.3 below. To this aims, and similarly to the Euclidean case (Definition B.1), let us recall the definitions of metric packings and coverings. We let (Θ, ρ) be a metric space, $\mathcal{M} \subseteq \Theta$ a subset of Θ, and a radius $\varepsilon > 0$.

- An *ε-covering* of $\mathcal{M}$ is a subset $\{\theta_1, \ldots, \theta_k\} \subseteq \mathcal{M}$ such that for all $\theta \in \mathcal{M}$, we have $\min_{1 \leq i \leq k} \rho(\theta, \theta_i) \leq \varepsilon$. The covering number $\mathrm{cv}_{(\mathcal{M}, \rho)}(\varepsilon)$ of $\mathcal{M}$ at scale ε is the smallest cardinality k of such an ε-covering.
- An *ε-packing* of $\mathcal{M}$ is a subset $\{\theta_1, \ldots, \theta_k\} \subseteq \mathcal{M}$ such that for all $1 \leq i < j \leq k$, $\mathrm{B}_{(\Theta, \rho)}(\theta_i, \varepsilon) \cap \mathrm{B}_{(\Theta, \rho)}(\theta_j, \varepsilon) = \emptyset$ (or equivalently $\rho(\theta_i, \theta_j) > 2\varepsilon$), where $\mathrm{B}_{(\Theta, \rho)}(\theta, \varepsilon) = \{\theta' \in \Theta, \rho(\theta, \theta') \leq \varepsilon\}$ is the closed ball in (Θ, ρ). The covering number $\mathrm{pk}_{(\mathcal{M}, \rho)}(\varepsilon)$ of $\mathcal{M}$ at scale ε is the largest cardinality k of such an ε-packing.

Theorem G.3. *Given a model $\mathcal{D}$ and a parameter of interest $\theta : \mathcal{D} \to \Theta$ in the metric space (Θ, ρ), any randomized SQ algorithm estimating $\theta(D)$ over $\mathcal{D}$ with precision ε and probability of success at least $1 - \alpha$, must make at least*

$$q \geq \frac{\log\big((1 - \alpha)\mathrm{pk}_{(\theta(\mathcal{D}), \rho)}(\varepsilon)\big)}{\log(1 + \lfloor 1/\tau \rfloor)}$$

queries to $\mathrm{STAT}(\tau)$, where $\theta(\mathcal{D}) = \{\theta(D), D \in \mathcal{D}\}$.

Proof of Theorem G.2. Apply Theorem G.3 with parameter of interest $\theta(D) = \mathrm{Supp}(D)$ and distance $\rho = \mathrm{d_H}$. □

G.2.1 Probabilistic Covering and Packing Number

To prove Theorem G.3, we will use the following notion of probabilistic covering. Given a set S and an integer $k \geq 0$, we denote by $\binom{S}{\leq k}$ the set of all subsets of S of cardinality at most k.

Definition G.2. Let (Θ, ρ) be a metric space. We say that a probabilistic measure μ over $\binom{\Theta}{\leq d}$ is a probabilistic (ε, α)-covering of (Θ, ρ) by d points if for all $\theta \in \Theta$,

$$\Pr_{\mathbf{p} \sim \mu}\left[\theta \in \bigcup_{q \in \mathbf{p}} \mathrm{B}_{(\Theta,\rho)}(q, \varepsilon)\right] \geq 1 - \alpha.$$

We denote by $\mathrm{cv}_{(\Theta,\rho)}(\varepsilon, \alpha)$ the minimal d such that there is a probabilistic (ε, α)-covering of (Θ, ρ) with d points.

This clearly generalizes (deterministic) coverings, since $\mathrm{cv}_{(\Theta,\rho)}(\varepsilon, \alpha = 0)$ coincides with the standard covering number $\mathrm{cv}_{(\Theta,\rho)}(\varepsilon)$. However, this quantity might be involved to compute since it involves randomness. Before proving Theorem G.3, let us show how to lower bound $\mathrm{cv}_{(\Theta,\rho)}(\varepsilon, \alpha)$ in practice.

Theorem G.4. *Let (Θ, ρ) be a metric space. Assume that there is a probability measure ν on Θ such that for all $q_1, \ldots, q_\ell \in \Theta$,*

$$\nu\left(\bigcup_{i=1}^{\ell} \mathrm{B}_{(\Theta,\rho)}(q_i, \varepsilon)\right) < 1 - \alpha.$$

Then $\mathrm{cv}_{(\Theta,\rho)}(\varepsilon, \alpha) > \ell$.

Proof of Theorem G.4. Take any probability measure μ over $\binom{\Theta}{\leq \ell}$, and consider the map $f(\mathbf{p}, \theta) = \mathbb{1}_{\cup_{q \in \mathbf{p}} \mathrm{B}_{(\Theta,\rho)}(q,\varepsilon)}(\theta)$ for all $\mathbf{p} \in \binom{\Theta}{\leq \ell}$ and $\theta \in \Theta$. By assumption, for all fixed $\mathbf{p} \in \binom{\Theta}{\leq \ell}$,

$$1 - \alpha > \nu\left(\bigcup_{q \in \mathbf{p}} \mathrm{B}_{(\Theta,\rho)}(q, \varepsilon)\right) = \int_{\Theta} f(\mathbf{p}, \theta)\nu(d\theta);$$

hence, by integration with respect to $\mu(d\mathbf{p})$ and Fubini–Tonelli,

$$1 - \alpha > \int_{\binom{\Theta}{\leq \ell}} \left(\int_{\Theta} f(\mathbf{p}, \theta)\nu(d\theta)\right) \mu(d\mathbf{p}) = \int_{\Theta} \left(\int_{\binom{\Theta}{\leq \ell}} f(\mathbf{p}, \theta)\mu(d\mathbf{p})\right) \nu(d\theta).$$

As ν is a probability distribution, this yields the existence of a fixed $\theta = \theta_\mu \in \Theta$ such that

$$1 - \alpha > \int_{\binom{\Theta}{\leq \ell}} f(\mathbf{p}, \theta)\mu(d\mathbf{p}) = \Pr_{\mathbf{p} \sim \mu}\left[\theta \in \bigcup_{q \in \mathbf{p}} \mathrm{B}_{(\Theta,\rho)}(q, \varepsilon)\right].$$

In other words, we have shown that no probability distribution μ over $\binom{\Theta}{\leq \ell}$ can be an (ε, α)-covering of (Θ, ρ) (Definition G.2). Hence, $\mathrm{cv}_{(\Theta,\rho)}(\varepsilon, \alpha) > \ell$. □

As a byproduct of Theorem G.4, we can now show that probabilistic coverings are closely related to the usual notions of metric covering and packing numbers.

Theorem G.5. *Let (Θ, ρ) be a metric space, and $\alpha < 1$. Then,*

$$\mathrm{cv}_{(\Theta,\rho)}(\varepsilon) \geq \mathrm{cv}_{(\Theta,\rho)}(\varepsilon, \alpha) \geq (1 - \alpha)\mathrm{pk}_{(\Theta,\rho)}(\varepsilon).$$

Proof of Theorem G.5. If any of the three terms is infinite, then all the terms involved clearly are infinite, so that the announced bounds hold. Otherwise, any given ε-covering of (Θ, ρ) is also a (ε, α)-covering (where we identify a finite set to the uniform measure on it), which gives the left-hand bound. For the right-hand bound, write $k = \mathrm{pk}_{(\Theta,\rho)}(\varepsilon) < \infty$, and let $\{\theta_1, \ldots, \theta_k\}$ be an ε-packing of (Θ, ρ). That is, for all $i \neq j$, $\rho(\theta_i, \theta_j) > 2\varepsilon$.

Take ν to be the uniform probability distribution over this packing, that is set $\nu(S) = |\{\theta_1, \ldots, \theta_k\} \cap S|/k$ for all $S \subseteq \Theta$. Note that since $\{\theta_1, \ldots, \theta_k\}$ is an ε-packing, we have $\nu\big(\mathrm{B}_{(\Theta,\rho)}(\theta, \varepsilon)\big) \leq 1/k$ for all $\theta \in \Theta$, and as a result,

$$\nu\left(\bigcup_{i=1}^{\ell} \mathrm{B}_{(\Theta,\rho)}(\theta_i, \varepsilon)\right) \leq \frac{\ell}{k}$$

for all $\theta_1, \ldots, \theta_\ell \in \Theta$.

Taking $\ell = \lceil (1-\alpha)k \rceil - 1$, Theorem G.4 implies that $\mathrm{cv}_{(\Theta,\rho)}(\varepsilon, \alpha) > \lceil (1-\alpha)k \rceil - 1$, and hence

$$\mathrm{cv}_{(\Theta,\rho)}(\varepsilon, \alpha) \geq \lceil (1-\alpha)k \rceil \geq (1-\alpha)k = (1-\alpha)\mathrm{pk}_{(\Theta,\rho)}(\varepsilon).$$

□

G.2.2 Proof of the Computational Lower Bounds for Randomized SQ Algorithms

We are now in position to prove the lower bounds on (randomized) SQ algorithms in general metric spaces.

Proof of Theorem G.3. For all $i \in \{0, \ldots, \lceil 1/\tau \rceil\}$, write $L_i = \min\{-1 + (2i+1)\tau, 1\}$. The L_i's form a τ-cover of $[-1, 1]$, meaning that for all $t \in [-1, 1]$, there is a least one $0 \leq i \leq \lfloor 1/\tau \rfloor$ with $|L_i - t| \leq \tau$. Hence we can define $f : [-1, 1] \to [-1, 1]$ by $f(t) = L_{i_0}$, where L_{i_0} is smallest L_i such that $|L_i - t| \leq \tau$. Note that f takes only $\lfloor 1/\tau \rfloor + 1$ different values, and that $|f(t) - t| \leq \tau$ for all $t \in [-1, 1]$.

Let us now consider the oracle O which, given a query $r : \mathbb{R}^n \to [-1, 1]$ to the distribution D, returns the answer $\mathrm{a}_D(r) = f(\mathbb{E}_D[r])$. Roughly speaking, the oracle discretizes the segment $[-1, 1]$ into $\lfloor 1/\tau \rfloor + 1$ points and returns the projection of the correct mean value $\mathbb{E}_D[r]$ onto this discretization. Clearly, O is a valid $\mathrm{STAT}(\tau)$ oracle since $|f(t) - t| \leq \tau$ for all $t \in [-1, 1]$.

Let $\mathcal{A}$ be a randomized SQ algorithm estimating θ over $\mathcal{D}$, and $\mathtt{A} = (r_1, \ldots, r_q, \hat{\theta}) \sim \mathcal{A}$. Let us write $d = (\lfloor 1/\tau \rfloor + 1)^q$, and consider the random subset of Θ given by

$$C(\mathtt{A}) = \left\{\hat{\theta}\Big(\mathrm{a}_D(r_1), \ldots, \mathrm{a}_D(r_q)\Big)\right\}_{D \in \mathcal{D}}.$$

Note that by construction of the oracle O, $C(\mathtt{A}) \in \binom{\mathcal{D}}{\leq d}$. Let us consider the probability distribution μ over $\binom{\mathcal{D}}{\leq d}$ such that the measure of a set S is equal to $\Pr_{\mathtt{A} \sim \mathcal{A}}[C(\mathtt{A}) \in S]$.

It is clear that if a deterministic algorithm $\mathtt{A}_0$ estimates $\theta(D)$ with precision ε using the oracle O, then $\theta(D) \in \cup_{t \in C(\mathtt{A}_0)} \mathrm{B}_{(\Theta,\rho)}(t, \varepsilon)$. As $\mathcal{A}$ estimates θ with precision ε and probability at least $1 - \alpha$ over $\mathcal{D}$, this means that μ is a probabilistic (ε, α)-covering of $\theta(\mathcal{D})$ with $(\lfloor 1/\tau \rfloor + 1)^q$ points (Definition G.2). As a result, by definition of $\mathrm{cv}_{(\theta(\mathcal{D}),\rho)}(\varepsilon, \alpha)$, we have $(\lfloor 1/\tau \rfloor + 1)^q \geq \mathrm{cv}_{(\theta(\mathcal{D}),\rho)}(\varepsilon, \alpha)$. Finally, from Theorem G.5 we have $\mathrm{cv}_{(\theta(\mathcal{D}),\rho)}(\varepsilon, \alpha) \geq (1-\alpha)\mathrm{pk}_{(\theta(\mathcal{D}),\rho)}(\varepsilon)$, which gives the announced result. □

H Lower Bounds for Manifold Models

H.1 Diffeomorphisms and Geometric Model Stability

The following result will allow us to build different elements of $\mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$ in a simple way, by considering diffeomorphic smooth perturbations of a base manifold M_0. Here and below, I_n is

the identity map of $\mathbb{R}^n$. Given a regular map $\Phi : \mathbb{R}^n \to \mathbb{R}^n$, $d_x\Phi$ and $d_x^2\Phi$ stand for its first and second order differentials at $x \in \mathbb{R}^n$.

Proposition H.1. *Let $M_0 \in \mathcal{M}^{n,d}_{2\mathrm{rch}_{\min}}$ and $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ be a proper $\mathcal{C}^2$ map, i.e. $\lim_{\|x\|\to\infty} \|\Phi(x)\| = \infty$. If $\sup_{x\in\mathbb{R}^n} \|I_n - d_x\Phi\|_{\mathrm{op}} \leq 1/(10d)$ and $\sup_{x\in\mathbb{R}^n} \left\|d_x^2\Phi\right\|_{\mathrm{op}} \leq 1/\left(4\mathrm{rch}_{\min}\right)$, then Φ is a global diffeomorphism, and $\Phi(M_0) \in \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$. Furthermore, $1/2 \leq \mathcal{H}^d(\Phi(M_0))/\mathcal{H}^d(M_0) \leq 2$.*

Proof of Proposition H.1. As $\sup_x \|d_x\Phi - I_n\|_{\mathrm{op}} < 1$, $d_x\Phi$ is invertiblefor all $x \in \mathbb{R}^n$. Hence, the inverse function theorem yields that Φ is everywhere a local diffeomorphism. As, $\lim_{\|x\|\to\infty} \|\Phi(x)\| = \infty$ this diffeomorphism is global by the Hadamard-Cacciopoli theorem [DMGZ94]. In particular, $\Phi(M_0)$ is a compact connected d-dimensional submanifold of $\mathbb{R}^n$ without boundary. In addition, by Taylor's theorem, Φ is Lipschitz with constant $\sup_x \|d_x\Phi\|_{\mathrm{op}} \leq (1 + \sup_x \|I_n - d_x\Phi\|_{\mathrm{op}}) \leq 11/10$, Φ^{-1} is Lipschitz with constant $\sup_x \left\|d_x\Phi^{-1}\right\|_{\mathrm{op}} \leq (1 - \sup_x \|I_n - d_x\Phi\|_{\mathrm{op}})^{-1} \leq 10/9$, and $d\Phi$ is Lipschitz with constant $\sup_x \left\|d_x^2\Phi\right\|_{\mathrm{op}} \leq 1/(4\mathrm{rch}_{\min}) \leq 1/(2\mathrm{rch}_{M_0})$. Hence, [Fed59, Theorem 4.19] yields

$$\mathrm{rch}_{\Phi(M)} \geq \frac{(2\mathrm{rch}_{M_0})(1 - \sup_x \|I_n - d_x\Phi\|_{\mathrm{op}})^2}{\sup_x \|d_x^2\Phi\|_{\mathrm{op}} (2\mathrm{rch}_M) + (1 + \sup_x \|I_n - d\Phi\|_{\mathrm{op}})} \geq (2\mathrm{rch}_{M_0})/2 \geq \mathrm{rch}_{\min}.$$

As a result, we have $\Phi(M_0) \in \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$. For the last claim, we use the properties of the Hausdorff measure $\mathcal{H}^d$ under Lipschitz maps [ACLZ17, Lemma 6] to get

$$\begin{aligned}
\mathcal{H}^d(\Phi(M)) &\leq \sup_x \|d_x\Phi\|_{\mathrm{op}}^d \, \mathcal{H}^d(M) \\
&\leq (1 + 1/(10d))^d \, \mathcal{H}^d(M) \\
&\leq 2\mathcal{H}^d(M),
\end{aligned}$$

and symmetrically,

$$\begin{aligned}
\mathcal{H}^d(M) &\leq \sup_x \left\|d_x\Phi^{-1}\right\|_{\mathrm{op}}^d \, \mathcal{H}^d(\Phi(M)) \\
&\leq \frac{1}{(1 - 1/(10d))^d} \mathcal{H}^d(\Phi(M)) \\
&= 2\mathcal{H}^d(\Phi(M)),
\end{aligned}$$

which concludes the proof. □

Among the smooth perturbations $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ nearly preserving $\mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$, the following localized bump-like functions will be of particular interest for deriving lower bounds.

Lemma H.1. *Let $\delta, \eta > 0$ be positive reals. Fix $p_1, \ldots, p_N \in \mathbb{R}^n$ be such that $\|p_i - p_j\| > 2\delta$ for all $i \neq j \in \{1, \ldots, N\}$. Given a family of unit vectors $\mathbf{w} = (w_i)_{1\leq i\leq N} \in (\mathbb{R}^n)^N$, we let $\Phi_{\mathbf{w}}$ be the function that maps any $x \in \mathbb{R}^n$ to*

$$\Phi_{\mathbf{w}}(x) = x + \eta \left(\sum_{i=1}^N \phi\left(\frac{x - p_i}{\delta}\right) w_i \right),$$

where $\phi : \mathbb{R}^n \to \mathbb{R}$ the real-valued bump function defined by

$$\begin{aligned}
\phi : \mathbb{R}^n &\longrightarrow \mathbb{R} \\
y &\longmapsto \exp\left(-\|y\|^2/(1 - \|y\|^2)\right) \mathbb{1}_{\mathrm{B}(0,1)}(y).
\end{aligned}$$

Then $\Phi_{\mathbf{w}}$ is $\mathcal{C}^\infty$ smooth, $\lim_{\|x\|\to\infty} \|\Phi_{\mathbf{w}}(x)\| = \infty$, and $\Phi_{\mathbf{w}}$ satisfies $\sup_{x\in\mathbb{R}^n} \|x - \Phi_{\mathbf{w}}(x)\| \leq \eta$,

$$\sup_{x\in\mathbb{R}^n} \|I_n - d_x\Phi_{\mathbf{w}}\|_{\mathrm{op}} \leq \frac{5\eta}{2\delta} \text{ and } \sup_{x\in\mathbb{R}^n} \left\|d_x^2\Phi_{\mathbf{w}}\right\|_{\mathrm{op}} \leq \frac{23\eta}{\delta^2}.$$

Proof of Lemma H.1. Straightforward calculations show that the real-valued map $\phi : \mathbb{R}^n \longrightarrow \mathbb{R}$ is $\mathcal{C}^\infty$ smooth over $\mathbb{R}^n$, equals to 0 outside $\mathrm{B}(0,1)$, and satisfies $0 \leq \phi \leq 1$, $\phi(0) = 1$,

$$\sup_{y \in \mathrm{B}(0,1)} \|d_y \phi\| \leq 5/2 \text{ and } \sup_{y \in \mathrm{B}(0,1)} \left\|d_y^2 \phi\right\|_{\mathrm{op}} \leq 23.$$

By composition and linear combination of $\mathcal{C}^\infty$ smooth functions, $\Phi_{\mathbf{w}}$ is therefore $\mathcal{C}^\infty$ smooth. Also, $\Phi_{\mathbf{w}}$ coincides with the identity map outside the compact set $\cup_{i=1}^N \mathrm{B}(p_i, \delta)$. Furthermore, for $i \neq j \in \{1, \ldots, N\}$, $\mathrm{B}(p_i, \delta) \cap \mathrm{B}(p_j, \delta) = \emptyset$, since $\|p_i - p_j\| > 2\delta$. Therefore, if $x \in \mathrm{B}(p_i, \delta)$, we have $\Phi_{\mathbf{w}}(x) = x + \eta\phi\left(\frac{x - p_i}{\delta}\right) w_i$. This directly gives $\sup_{x \in \mathbb{R}^n} \|x - \Phi_{\mathbf{w}}(x)\| \leq \eta$, and by chain rule,

$$\begin{aligned}
\sup_{x \in \mathbb{R}^n} \|I_n - d_x \Phi_{\mathbf{w}}\|_{\mathrm{op}} &= \max_{1 \leq i \leq N} \sup_{x \in \mathrm{B}(p_i, \delta)} \eta \left\| d_x \left(\phi \left(\frac{\cdot - p_i}{\delta} \right) w_i \right) \right\|_{\mathrm{op}} \\
&= \max_{1 \leq i \leq N} \sup_{y \in \mathrm{B}(0,1)} \left\| w_i (d_y \phi)^\top \right\|_{\mathrm{op}} \frac{\eta}{\delta} \\
&= \sup_{y \in \mathrm{B}(0,1)} \|d_y \phi\| \frac{\eta}{\delta} \\
&\leq \frac{5\eta}{2\delta},
\end{aligned}$$

and

$$\begin{aligned}
\sup_{x \in \mathbb{R}^n} \left\|d_x^2 \Phi_{\mathbf{w}}\right\|_{\mathrm{op}} &= \max_{1 \leq i \leq N} \sup_{y \in \mathrm{B}(0,1)} \left\|w_i d_y^2 \phi\right\|_{\mathrm{op}} \frac{\eta}{\delta^2} \\
&\leq \frac{23\eta}{\delta^2},
\end{aligned}$$

which concludes the proof. □

H.2 Building a Large-Volume Submanifold with Small Euclidean Diameter

The proofs of Theorems 5.5 and 5.6 will involve the construction of submanifolds $M \subseteq \mathbb{R}^n$ with prescribed and possibly large volume $\mathcal{H}^d(M)$. Informally, this will enable us to build hypotheses and packings with large cardinality by local variations of it (see Propositions H.3 and H.5) under nearly minimal assumptions on $f_{\min}$ (which can be seen as an inverse volume, for uniform distributions). For the reasons mentioned in Section 2.2.3, one easily checks that the volume of $M \in \mathrm{B}(0, R) \sqcap \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$ can neither be too small nor too large, when $\mathrm{rch}_{\min}$ and R are fixed (Proposition B.5). Conversely, this section is devoted to prove the *existence* of submanifolds $M \in \mathrm{B}(0, R) \sqcap \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$ that nearly achieve the minimum and maximum possible such volumes given by Proposition B.5.

H.2.1 The Statement

Namely, the goal of Appendix H.2 is to prove the following result.

Proposition H.2. *Assume that* $\mathrm{rch}_{\min} \leq R/36$. *Writing* $C'_d = 9(2^{2d+1}\sigma_{d-1})$, *let* $\mathcal{V} > 0$ *be such that*

$$1 \leq \frac{\mathcal{V}}{C'_d \mathrm{rch}_{\min}^d} \leq \max_{1 \leq k \leq n} \left(\frac{R}{48 \mathrm{rch}_{\min} \sqrt{k}} \right)^k.$$

Then there exists $M_0 \in \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$ *such that* $M_0 \subseteq \mathrm{B}(0, R)$ *and*

$$\mathcal{V}/24 \leq \mathcal{H}^d(M_0) \leq \mathcal{V}.$$

Informally, in codimension one (i.e. $D = d+1$), the manifold M_0 of Proposition H.2 can be though of as the boundary of the offset of a Hilbert curve in $\mathrm{B}(0,R)$ of prescribed length. This intuition, however, is only limited to codimension one, and requires extra technical developments for general $d < D$.

Proof of Proposition H.2. Consider the discrete grid G_0 in $\mathbb{R}^n$ centered at $0 \in \mathbb{R}^n$, with vertices $(24\mathrm{rch}_{\min}\mathbb{Z}^n) \cap \mathrm{B}(0,R/2)$, and composed of hypercubes of side-length $24\mathrm{rch}_{\min}$. By considering a k_0-dimensional sub-grid parallel to the axes, we see that the grid G_0 contains a square grid G with side cardinality $\kappa = \left\lceil \frac{R/2}{24\mathrm{rch}_{\min}\sqrt{k_0}} \right\rceil$, where k_0 belongs to $\mathrm{argmax}_{1\leq k\leq n} \left(\frac{R}{48\mathrm{rch}_{\min}\sqrt{k}}\right)^k$. Let us write $\ell = \lfloor \mathcal{V}/(C'_d \mathrm{rch}_{\min}^d) \rfloor$. By assumption on $\mathcal{V}$, $\mathrm{rch}_{\min}$ and R, we have

$$1 \leq \ell \leq \frac{\mathcal{V}}{C'_d \mathrm{rch}_{\min}^d} \leq \max_{1\leq k\leq n} \left(\frac{R}{48\mathrm{rch}_{\min}\sqrt{k}}\right)^k \leq \kappa^{k_0}.$$

Hence, Lemma H.4 asserts that there exists a connected open simple path $L_n(\ell)$ in $G \subseteq G_0$ with length $|L_n(\ell)| = \ell$. Furthermore, Lemma H.3 applied with reach parameter $2\mathrm{rch}_{\min}$ provides us with a closed d-dimensional submanifold M'_0 of class $C^{1,1}$ such that $M'_0 = M(L_n(\ell)) \subseteq G^{12\mathrm{rch}_{\min}} \subseteq \mathrm{B}(0,R/2)^{12\mathrm{rch}_{\min}} \subseteq \mathrm{B}(0,2R/3)$ since $\mathrm{rch}_{\min} \leq R/36$, reach $\mathrm{rch}_{M'_0} \geq 2\mathrm{rch}_{\min}$. Furthermore, writing $C_d = 9(2^d\sigma_{d-1})$ for the constant of Lemma H.3, we also have

$$\begin{aligned}
\mathcal{H}^d(M'_0) &\leq (C_d(2\mathrm{rch}_{\min})^d)|L_n(\ell)| \\
&\leq (C_d(2\mathrm{rch}_{\min})^d)\frac{\mathcal{V}}{C'_d\mathrm{rch}_{\min}^d} \\
&\leq \mathcal{V}/2,
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{H}^d(M'_0) &\geq (C_d(2\mathrm{rch}_{\min})^d/3)|L_n(\ell)| \\
&\geq (C_d(2\mathrm{rch}_{\min})^d/3)\frac{\mathcal{V}}{2C'_d\mathrm{rch}_{\min}^d} \\
&= \mathcal{V}/12,
\end{aligned}$$

where we used that $\lfloor t \rfloor \geq t/2$ for all $t \geq 1$ To conclude the proof, we use the density of $\mathcal{C}^2$ submanifolds in the space of $\mathcal{C}^{1,1}$ submanifolds to obtain a closed d-dimensional submanifold M_0 of class $\mathcal{C}^2$ such that $\mathrm{rch}_{M_0} \geq \mathrm{rch}_{M'_0}/2 \geq \mathrm{rch}_{\min}$, $\mathrm{d_H}(M_0, M'_0) \leq \mathrm{rch}_{\min}$ (and hence $M_0 \subset \mathrm{B}(0, 2R/3 + \mathrm{rch}_{\min}) \subset \mathrm{B}(0,R)$), and $1/2 \leq \mathcal{H}^d(M_0)/\mathcal{H}^d(M'_0) \leq 2$ (and hence $\mathcal{V}/24 \leq \mathcal{H}^d(M_0) \leq \mathcal{V}$). □

H.2.2 Widget Gluing: From Paths on the Discrete Grid to Manifolds

Lemma H.2. *Given* $\mathrm{rch}_{\min} > 0$ *and* $d \geq 1$*, there exist four* d*-dimensional* $\mathcal{C}^{1,1}$*-submanifolds with boundary:*

$$M_E, M_S, M_{TB} \subseteq [-6\mathrm{rch}_{\min}, 6\mathrm{rch}_{\min}]^{d+1} \text{ and } M_{NB} \subseteq [-6\mathrm{rch}_{\min}, 6\mathrm{rch}_{\min}]^{d+2},$$

called respectively end, straight, tangent bend *and* normal bend *widgets (see Figure 3), that:*

- *are smooth:* $\mathrm{rch}_{M_E}, \mathrm{rch}_{M_S}, \mathrm{rch}_{M_{TB}}, \mathrm{rch}_{M_{NB}} \geq \mathrm{rch}_{\min}$*;*
- *have the following topologies:*
  - M_E *is isotopic to a* d*-ball* $\mathrm{B}_d(0,1)$*,*
  - M_S*,* M_{TB} *and* M_{NB} *are isotopic to a* d*-cylinder* $\mathcal{S}^{d-1} \times [0,1]$*;*

- *are linkable: writing* $s = 6\mathrm{rch}_{\min}$*, we have*
  - *For the tip widget* M_E*:*
    * $M_E \cap \left([-s/2, s/2]^{d+1}\right)^c = M_E \cap \left([s/2, s] \times \mathbb{R}^d\right) = [s/2, s] \times \mathcal{S}^{d-1}(0, s/3)$.
  - *For the straight widget* M_S*:*
    * $M_S \cap \left([-s/2, s/2]^{d+1}\right)^c = M_S \cap \left(\left([-s, -s/2] \times \mathbb{R}^d\right) \cup \left([s/2, s] \times \mathbb{R}^d\right)\right)$,
    * $M_S \cap \left([-s, -s/2] \times \mathbb{R}^d\right) = [-s, -s/2] \times \mathcal{S}^{d-1}(0, s/3)$,
    * $M_S \cap \left([s/2, s] \times \mathbb{R}^d\right) = [s/2, s] \times \mathcal{S}^{d-1}(0, s/3)$.
  - *For the tangent bend widget* M_{TB}*:*
    * $M_{TB} \cap \left([-s/2, s/2]^{d+1}\right)^c = M_{TB} \cap \left(\left([-s, -s/2] \times \mathbb{R}^d\right) \cup \left(\mathbb{R}^d \times [-s, -s/2]\right)\right)$,
    * $M_{TB} \cap \left([-s, -s/2] \times \mathbb{R}^d\right) = [-s, -s/2] \times \mathcal{S}^{d-1}(0, s/3)$,
    * $M_{TB} \cap \left(\mathbb{R}^d \times [-s, -s/2]\right) = \mathcal{S}^{d-1}(0, s/3) \times [-s, -s/2]$.
  - *For the normal bend widget* M_{NB}*:*
    * $M_{NB} \cap \left([-s/2, s/2]^{d+2}\right)^c = M_{NB} \cap \left(\left([-s, -s/2] \times \mathbb{R}^d \times \{0\}\right) \cup \left(\{0\} \times \mathbb{R}^d \times [-s, -s/2]\right)\right)$,
    * $M_{NB} \cap \left([-s, -s/2] \times \mathbb{R}^d \times \{0\}\right) = [-s, -s/2] \times \mathcal{S}^{d-1}(0, s/3) \times \{0\}$,
    * $M_{NB} \cap \left(\{0\} \times \mathbb{R}^d \times [-s, -s/2]\right) = \{0\} \times \mathcal{S}^{d-1}(0, s/3) \times [-s, -s/2]$.

  Furthermore,

$$(C_d/3)\mathrm{rch}_{\min}^d \leq \mathcal{H}^d(M_E), \mathcal{H}^d(M_S), \mathcal{H}^d(M_{TB}), \mathcal{H}^d(M_{NB}) \leq C_d \mathrm{rch}_{\min}^d,$$

  where $C_d = 9(2^d \sigma_{d-1})$ *depends only on* d*.*

Proof of Lemma H.2. First notice that by homogeneity, we can carry out the construction in the unit hypercubes $[-1, 1]^{d+1}$ (respectively $[-1, 1]^{d+2}$) and conclude by applying an homothetic transformation. Indeed, for all closed set $K \subseteq \mathbb{R}^n$ and $\lambda \geq 0$, $\mathrm{rch}_{\lambda K} = \lambda \mathrm{rch}_K$ and $\mathcal{H}^d(\lambda K) = \lambda^d \mathcal{H}^d(K)$.

- End widget: the idea is to glue in a $\mathcal{C}^2$ way a half d-sphere with a d-cylinder. Namely, let us consider

$$M_E^{(0)} = \left(\mathcal{S}^d(0, 1/3) \cap \left([-1, 0] \times [-1, 1]^d\right)\right) \cup \left([0, 1] \times \mathcal{S}^{d-1}(0, 1/3)\right).$$

  Elementary calculations yield the intersections

$$M_E^{(0)} \cap \left([-1/2, 1/2]^{d+1}\right)^c = M_E^{(0)} \cap \left([1/2, 1] \times \mathbb{R}^d\right) = [1/2, 1] \times \mathcal{S}^{d-1}(0, 1/3).$$

  In addition, its medial axis is $\mathrm{Med}(M_E^{(0)}) = [0, 1] \times \{0\}^d$, so that

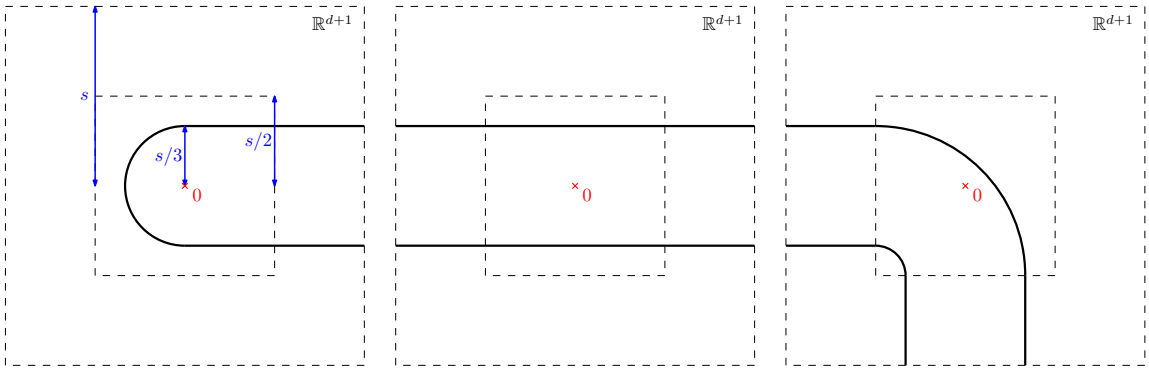
$$\mathrm{rch}_{M_E^{(0)}} = \inf_{z \in \mathrm{Med}(M_E^{(0)})} \mathrm{d}(z, M_E^{(0)}) = 1/3.$$

  Finally, $M_E^{(0)}$ is isotopic to the half d-sphere $\mathcal{S}^d(0, 1/3) \cap \left([-1, 0] \times [-1, 1]^d\right)$, or equivalently to a d-ball.

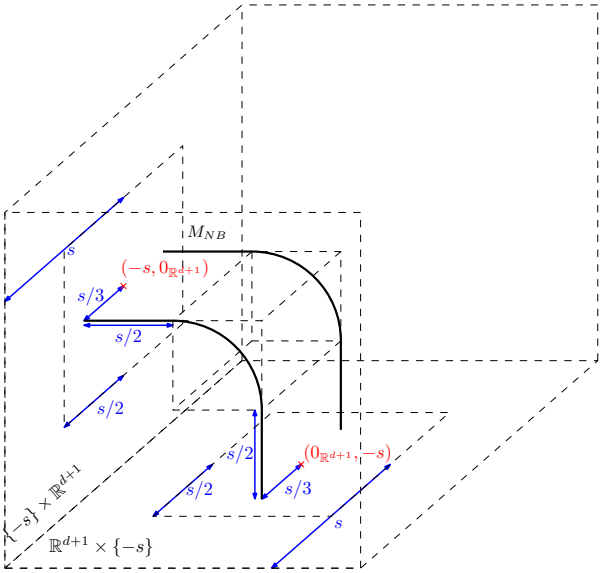
- Straight widget: a simple d-cylinder satisfies our requirements. Similarly as above, the set

$$M_S^{(0)} = [-1, 1] \times \mathcal{S}^{d-1}(0, 1/3)$$

  clearly is (isotopic to) a d-cylinder, has reach $\mathrm{rch}_{M_S^{(0)}} = 1/3$, and all the announced intersection properties with $s = 1$.



(a) End widget M_E. (b) Straight widget M_S. (c) Tangent Bend widget M_{TB}.



(d) Normal Bend widget M_{NB}.

Figure 3: The widgets built in Lemma H.2 and used in the proof of Lemma H.3.

- Tangent Bend widget: we will glue two orthogonal straight d-cylinders via a smoothly rotating $(d-1)$-sphere. More precisely, consider the d-cylinders $C_1 = \mathcal{S}^{d-1}(0, 1/3) \times [-1, -1/2]$ and $C_2 = [-1, -1/2] \times \mathcal{S}^{d-1}(0, 1/3)$. We will connect smoothly their tips, which are the $(d-1)$-spheres $S_1 = \mathcal{S}^{d-1}(0, 1/3) \times \{-1/2\} \subseteq C_1$ and $S_2 = \{-1/2\} \times \mathcal{S}^{d-1}(0, 1/3) \subseteq C_2$ of same radius. To this aim, take the trajectory of S_1 via the affine rotations of center $x_c = (-1/2, 0_{\mathbb{R}^{d-1}}, -1/2)$ and linear parts

$$R_\theta = \begin{pmatrix} \cos\theta & 0 & \cdots & 0 & -\sin\theta \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ \sin\theta & 0 & \cdots & 0 & \cos\theta \end{pmatrix} \in \mathbb{R}^{(d+1)\times(d+1)},$$

when θ varies in $[0, \pi/2]$. Hence, letting $f_\theta(x) = x_c + R_\theta(x - x_c)$, we have $f_0(S_1) = S_1$, $f_{\pi/2}(S_1) = S_2$. In addition, for all $\theta \in [0, \pi/2]$ and $x \in [-1/2, 1/2]^d \times \{-1/2\}$, we have $f_\theta(x) \in [-1/2, 1/2]^{d+1}$. Hence, letting

$$M_{TB}^{(0)} = C_1 \cup \left(\bigcup_{0 \le \theta \le \pi/2} f_\theta(S_1) \right) \cup C_2,$$

we directly get that $M_{TB}^{(0)}$ is isotopic to a d-cylinder, and that it satisfies all the announced intersection properties with $s = 1$. To conclude, by symmetry, the medial axis of this widget writes as

$$\mathrm{Med}(M_{TB}^{(0)}) = \{0\}^d \times [-1, -1/2] \cup \left(x_c + \bigcup_{t \geq 0} (-t, 0_{\mathbb{R}^{d-1}}, -t) \right) \cup [-1, -1/2] \times \{0\}^d,$$

so that straightforward calculations yield $\mathrm{rch}_{M_{TB}^{(0)}} = \min \left\{ 1/3, \mathrm{d}(x_c, M_{TB}^{(0)}) \right\} = 1/6$.

- Normal Bend widget: same as for the tangent bend widget, we glue the two orthogonal straight d-cylinders $C_1 = \{0\} \times \mathcal{S}^{d-1}(0, 1/3) \times [-1, -1/2]$ and $C_2 = [-1, -1/2] \times \mathcal{S}^{d-1}(0, 1/3) \times \{0\}$. via their respective tips, $S_1 = \{0\} \times \mathcal{S}^{d-1}(0, 1/3) \times \{-1/2\} \subseteq C_1$ and $S_2 = \{-1/2\} \times \mathcal{S}^{d-1}(0, 1/3)\{0\} \subseteq C_2$. To this aim, take trajectory of S_1 via the affine rotation of center $x_c = (-1/2, 0_{\mathbb{R}^d}, -1/2)$ and linear parts $R_\theta \in \mathbb{R}^{(d+2)\times(d+2)}$ for $\theta \in [0, \pi/2]$. As before, letting $f_\theta(x) = x_c + R_\theta(x - x_c)$, we have $f_0(S_1) = S_1$, $f_{\pi/2}(S_1) = S_2$. Also, for all $\theta \in [0, \pi/2]$ and $x \in \{0\} \times [-1/2, 1/2]^d \times \{-1/2\}$, we have $f_\theta(x) \in [-1/2, 1/2]^{d+1}$. Hence, letting

$$M_{NB}^{(0)} = C_1 \cup \left(\bigcup_{0 \leq \theta \leq \pi/2} f_\theta(S_1) \right) \cup C_2,$$

we get the announced results with $s = 1$, and in a similar way as above, $\mathrm{rch}_{M_{NB}^{(0)}} = \min\{1/3, 1/2\} = 1/3$.

Also one easily checks in all the four above cases that

$$\frac{\sigma_{d-1}/3^{d-2}}{3} \leq \mathcal{H}^d(M_E^{(0)}), \mathcal{H}^d(M_S^{(0)}), \mathcal{H}^d(M_{TB}^{(0)}), \mathcal{H}^d(M_{NB}^{(0)}) \leq \sigma_{d-1}/3^{d-2}.$$

Finally, letting

$$C^{-1} = \min \left\{ \mathrm{rch}_{M_E^{(0)}}, \mathrm{rch}_{M_S^{(0)}}, \mathrm{rch}_{M_{TB}^{(0)}}, \mathrm{rch}_{M_{NB}^{(0)}} \right\} = 1/6$$

and considering the dilations $M_E = (C\mathrm{rch}_{\min})M_E^{(0)}$, $M_S = (C\mathrm{rch}_{\min})M_S^{(0)}$, $M_{TB} = (C\mathrm{rch}_{\min})M_{TB}^{(0)}$ and $M_{NB} = (C\mathrm{rch}_{\min})M_{NB}^{(0)}$ yields the result by homogeneity, with $C_d = C^d\sigma_{d-1}/3^{d-2} = 9(2^d\sigma_{d-1})$. □

Lemma H.3. *Let G be a discrete grid in $\mathbb{R}^n$ composed of hypercubes of side-length $12\mathrm{rch}_{\min}$. Then any connected open simple path L in G (see Lemma H.4) defines a $\mathcal{C}^{1,1}$ d-dimensional closed submanifold, denoted by $M(L)$, such that:*

- $M(L) \subseteq G^{6\mathrm{rch}_{\min}}$;
- $M(L) \in \mathcal{M}_{\mathrm{rch}_{\min}}^{n,d}$;
- $C_d/3 \leq \dfrac{\mathcal{H}^d(M(L))}{|L|\mathrm{rch}_{\min}^d} \leq C_d$, *where C_d is the constant of Lemma H.2;*
- *If L and L' are two different such paths in G,*

$$\mathrm{d_H}(M(L), M(L')) > 2\mathrm{rch}_{\min}.$$

Remark H.1. The construction of Lemma H.3 shows that, given one discrete path L, one could actually define several different manifolds $M(L)$ with the same properties. We will not exploit this fact as the construction is enough for our purpose.

Proof of Lemma H.3. For short, we let $s = 6\mathrm{rch}_{\min}$. Let L be a fixed connected open simple path on G. If $|L| = 1$, take $M(L)$ to be a d-sphere of radius $2\mathrm{rch}_{\min}$ centered at the only vertex of L. Assuming now that $|L| \geq 2$, we will build $M(L)$ iteratively by adding appropriate widgets of Lemma H.2 along the consecutive vertices that L goes through. We pick one of the two degree 1 vertices (endpoints) of L arbitrarily, and denote the consecutive vertices of L as $x_1, \dots, x_{[L|-1}$.

(i) The path L has exactly one edge at x_0, called v_0^+, which is parallel to the axes of $\mathbb{R}^n$ since G is the square grid. In the cube $x_0 + [-6\mathrm{rch}_{\min}, 6\mathrm{rch}_{\min}]^n$, we define $M(L)$ to coincide with the End widget $M_E \times \{0\}^{n-(d+1)}$, rotated in the (e_1, v_0^+) plane so that $-e_1$ is sent on v_0^+. In this first cube, $M(L)$ hence presents a d-cylinder, obtained by a rotation of $[-s, -s/2] \times \mathcal{S}^{d-1}(0, s/3) \times \{0\}^{n-(d+1)}$ around x_0, and pointing towards v_0^+. Let us call this cylinder C_0^+.

(ii) Assume now that we have visited the consecutive vertices $x_0, \dots, x_{k-1}$ of L, for some $k \geq 1$, and that in the cube around x_{k-1}, $M(L)$ presents a cylinder C_{k-1}^+ in the direction v_{k-1}^+. If x_k is not the other endpoint of L, there are exactly two edges at x_k, represented by the axis-parallel vectors $v_k^- = (x_{k-1} - x_k) = -v_{k-1}^+$ and $v_k^+ = (x_{k+1} - x_k)$. There are three possible cases depending on the turn that L takes at x_k:

(a) If v_k^- and v_k^+ are aligned, take $M(L) \cap (x_k + [-s, s]^n)$ to coincide with the Straight widget $M_S \times \{0\}^{n-(d+1)}$, rotated in the $\{e_1, v_k^+\}$-plane so that e_1 is sent on v_k^+.

(b) If v_k^+ belongs to the $(d+1)$-plane spanned by C_{k-1}^+ but v_k^- and v_k^+ are not aligned, proceed similarly by rotating the Tangent Bend widget $M_{TB} \times \{0\}^{n-(d+1)}$ so that (e_1, e_{d+1}) is sent on $(-v_{k-1}^-, v_k^+)$.

(c) Otherwise, if v_k^+ does not belong to the $(d+1)$-plane spanned by C_{k-1}^+, then $\{v_k^+, C_{k-1}^+\}$ defines a $(d+2)$-plane. Hence, we proceed similarly by rotating the Normal Bend widget $M_{NB} \times \{0\}^{n-(d+2)}$ so that (e_1, e_{d+2}) is sent on $(-v_{k-1}^-, v_k^+)$. Note that this case can only occur if $n \geq d+2$.

(iii) If we reached the other endpoint of L ($k = |L| - 1$), add a rotated End widget oriented in the direction of C_{k-1}^+.

Now that the construction of $M(L)$ has been carried out, let us move to its claimed properties.

- By construction of the widgets and the fact that all of them are centered at points of the grid G, $M(L)$ is included in the offset of G of radius $6\mathrm{rch}_{\min}$.

- By induction on the length of the path, it is clear that the union of the straight and bend widgets (without the ends) is isotopic to a cylinder $\mathcal{S}^{d-1}(0,1) \times [0,1]$. As a result, adding the two end widgets at the endpoints of the path yields that $M(L)$ is isotopic to a d-dimensional sphere $\mathcal{S}^d(0,1)$. It is also clear that $M(L)$ connected, by connectedness of L. In particular, $M(L)$ is a compact connected d-dimensional submanifold of $\mathbb{R}^n$ without boundary.

  What remains to be proved is that $\mathrm{rch}_{M(L)} \geq \mathrm{rch}_{\min}$. To see this, notice that by construction, the widgets connect smoothly through sections of facing straight cylinders $C^{\pm} = \mathcal{S}^{d-1}(0, s/3) \times [0, \pm s/2] \times \{0\}^{n-(d+1)}$ (rotated), which are included in the boxes $[-s/2, s/2]^n$ centered a the midpoints of the grid. Apart from these connected ingoing and outgoing cylinders, the widgets are included in boxes $[-s/2, s/2]^n$, which are separated by a distance s. Hence, if two points $x, y \in M(L)$ are such that $\|y - x\| \leq s/2$, then they must belong to either the same widget or the same connecting cylinder $C^- \cup C^+$. As a result, from [Fed59,

Theorem 4.18] and the fact that $\mathrm{d}(y-x, T_x M(L)) \leq \|y-x\|$ for all $x \in M(L)$, we get

$$\begin{aligned}
\mathrm{rch}_{M(L)} &= \inf_{x \neq y \in M(L)} \frac{\|y-x\|^2}{2\mathrm{d}(y-x, T_x M(L))} \\
&= \min \left\{ \inf_{\substack{x,y \in M(L) \\ \|y-x\| \geq s/2}} \frac{\|y-x\|^2}{2\mathrm{d}(y-x, T_x M(L))}, \inf_{\substack{x \neq y \in M(L) \\ \|y-x\| \leq s/2}} \frac{\|y-x\|^2}{2\mathrm{d}(y-x, T_x M(L))} \right\} \\
&\geq \min \left\{ s/4, \min \left\{ \mathrm{rch}_{M_E}, \mathrm{rch}_{M_S}, \mathrm{rch}_{M_{TB}}, \mathrm{rch}_{M_{NB}} \right\} \right\} \\
&\geq \min \left\{ 6\mathrm{rch}_{\min}/4, \mathrm{rch}_{\min} \right\} \\
&= \mathrm{rch}_{\min},
\end{aligned}$$

which ends proving that $M(L) \in \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$.

- As $M(L)$ is the union of $|L|$ of the widgets defined in Lemma H.2, it follows

$$\begin{aligned}
\mathcal{H}^d(M(L)) &\leq |L| \max \left\{ \mathcal{H}^d(M_E), \mathcal{H}^d(M_S), \mathcal{H}^d(M_{TB}), \mathcal{H}^d(M_{NB}) \right\} \\
&\leq |L| C_d \mathrm{rch}^d_{\min},
\end{aligned}$$

  and similarly, as the intersection of the consecutive widgets (i.e. $(d-1)$-spheres) is $\mathcal{H}^d$-negligible, we have

$$\begin{aligned}
\mathcal{H}^d(M(L)) &\geq |L| \min \left\{ \mathcal{H}^d(M_E), \mathcal{H}^d(M_S), \mathcal{H}^d(M_{TB}), \mathcal{H}^d(M_{NB}) \right\} \\
&\geq |L| (C_d/3) \mathrm{rch}^d_{\min}.
\end{aligned}$$

- Let us now fix two different connected open simple paths L and L' in G. Since $L \neq L'$, L passes through a vertex, say $x_0 \in \mathbb{R}^n$, where L' doesn't. Regardless of the widget used at x_0 to build $M(L)$, this widget contains, up to rotation centered at x_0, the set $x_0 + \{-s/2\} \times \mathcal{S}^{d-1}(0, s/3) \times \{0\}^{n-(d+1)}$. As a result, $\mathrm{d}(x_0, M(L)) \leq \sqrt{(s/2)^2 + (s/3)^2}$. On the other hand, $M(L')$ does not intersect the cube $x_0 + [-s, s]^n$, so $\mathrm{d}(x_0, M(L')) \geq s$. Finally, we get

$$\begin{aligned}
\mathrm{d_H}(M(L), M(L')) &= \sup_{x \in \mathbb{R}^n} \left| \mathrm{d}(x, M(L')) - \mathrm{d}(x, M(L)) \right| \\
&\geq \left| \mathrm{d}(x_0, M(L')) - \mathrm{d}(x_0, M(L)) \right| \\
&\geq s - \sqrt{(s/2)^2 + (s/3)^2} \\
&= 6(1 - \sqrt{13}/6) \mathrm{rch}_{\min} \\
&> 2\mathrm{rch}_{\min},
\end{aligned}$$

  which concludes the proof.

□

H.2.3 Existence of Long Paths on the Grid

In order to complete the construction of Proposition H.2, we need the existence of paths of prescribed length over the n-dimensional discrete grid. Although standard, we include this construction for sake of completeness.

Lemma H.4. *Let $\kappa \geq 1$ be an integer and consider the square grid graph G_n on $\{1, \ldots, \kappa\}^n$. Then for all $\ell \in \{1, \ldots, \kappa^n\}$, there exists a connected open simple path $L_n(\ell)$ of length ℓ in G_n. That is, $L_n(\ell)$ is a subgraph of G_n such that:*

- *$L_n(\ell)$ is connected;*
- *$L_n(\ell)$ has vertex cardinality ℓ;*
- *if $\ell \geq 2$, $L_n(\ell)$ has maximum degree 2, and exactly two vertices with degree 1.*

Proof of Lemma H.4. For $\kappa = 1$, G_n consists of a single point, so that the result is trivial. We hence assume that $\kappa \geq 2$. Let us first build the paths $L_n = L_n(\kappa^n)$ by induction on n. For $n = 1$, simply take L_1 to be the full graph G_n. We orientate L_1 by enumerating its adjacent vertices in order: $L_1^{\rightarrow}[i] = i$ for all $1 \leq i \leq \kappa$. Given an orientation $L^{\rightarrow}$ of some path L in G_n, we also let $L^{\leftarrow}[i] = L^{\rightarrow}[|L| - i]$ denote its backwards orientation. Now, assume that we have built L_n for some $n \geq 1$, together with an orientation $L_n^{\rightarrow}$. To describe L_{n+1}, we list an orientation $L_{n+1}^{\rightarrow}$ of it: an edge of G_n hence belongs to L_n if an only if it joins two consecutive vertices in $L_{n+1}^{\rightarrow}$. Namely, for $1 \leq i \leq \kappa^n$, we let

$$\begin{aligned} L_{n+1}^{\rightarrow}[i] &= (L_n^{\rightarrow}[i], 1) \\ L_{n+1}^{\rightarrow}[\kappa^n + i] &= (L_n^{\leftarrow}[i], 2) \\ &\vdots \\ L_{n+1}^{\rightarrow}[(\kappa - 1)\kappa^n + i] &= (L_n^{\leftrightarrow}[i], \kappa)\,, \end{aligned}$$

where for the last line, $\leftrightarrow$ stands for $\rightarrow$ if κ is odd, and $\leftarrow$ otherwise. L_{n+1} clearly is connected and visits all the vertices $\{1, \ldots, \kappa\}^n$. Its edges all have degree two, except $(L_n^{\rightarrow}[1], 1)$ and $(L_n^{\leftrightarrow}[\kappa^n], \kappa)$ which have degree 1, which concludes the construction of $L_n = L_n(\kappa^n)$. To conclude the proof, take $L_n(\ell)$ $(1 \leq \ell \leq \kappa^n)$ to be the first ℓ consecutive vertices of $L_n^{\rightarrow}(\kappa^n)$. □

H.3 Informational Lower Bounds: Hypotheses for Le Cam's Lemma

This section is devoted to prove the two informational lower bounds Theorems 5.2 and 5.5. We will use the general informational lower bound from Theorem G.1 in the models $\{0\} \sqcup \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ and $\mathrm{B}(0, R) \sqcap \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ respectively, and parameter of interest $\theta(D) = \mathrm{Supp}(D)$ that lies in the metric space formed by the non-empty compact sets of $\mathbb{R}^n$ equipped with the metric $\rho = \mathrm{d_H}$.

H.3.1 Construction of the Hypotheses

First, we show how to build hypotheses, i.e probability distributions for Le Cam's Lemma (Theorem G.1). We present a generic construction in the manifold setting by perturbing a base submanifold M_0. Note that the larger the volume $\mathcal{H}^d(M_0)$, the stronger the result. See also Proposition H.5 for a result similar in spirit, and used to derive computational lower bounds instead of informational ones.

Proposition H.3. *For all $M_0 \in \mathcal{M}^{n,d}_{2\mathrm{rch}_{\min}}$, $x_0 \in M_0$ and $\tau \leq 1$, there exists a manifold $M_1 \in \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$ such that $x_0 \in M_1$, $\mathcal{H}^d(M_0)/2 \leq \mathcal{H}^d(M_1) \leq 2\mathcal{H}^d(M_0)$,*

$$\frac{\mathrm{rch}_{\min}}{2^{18}} \min\left\{\frac{1}{2^{22}d^2}, \left(\frac{\mathcal{H}^d(M_0)\tau}{\omega_d \mathrm{rch}_{\min}^d}\right)^{2/d}\right\} \leq \mathrm{d_H}(M_0, M_1) \leq \mathrm{rch}_{\min}/10,$$

and so that the uniform distributions D_0, D_1 over M_0, M_1 satisfy $\mathrm{TV}(D_0, D_1) \leq \tau/2$.

Proof of Proposition H.3. Let $p_0 \in M_0$ be an arbitrary point such that $\|p_0 - x_0\| \geq \mathrm{rch}_{\min}$. For instance, by taking the geodesic variation $p_0 = \gamma_{x_0,v_0}(2\mathrm{rch}_{\min})$, where $v_0 \in T_{x_0}M_0$ is a unit

tangent vector, a Taylor expansion of γ_{x_0,v_0} and Lemma 2.2 yields

$$\begin{aligned}\|p_0 - x_0\| &\geq \|2\mathrm{rch}_{\min} v_0\| - \|\gamma_{x_0,v_0}(\mathrm{rch}_{\min}) - (x_0 + 2\mathrm{rch}_{\min} v_0)\| \\ &\geq 2\mathrm{rch}_{\min} - (2\mathrm{rch}_{\min})^2/(2\mathrm{rch}_{M_0}) \\ &\geq \mathrm{rch}_{\min},\end{aligned}$$

since $\mathrm{rch}_{M_0} \geq 2\mathrm{rch}_{\min}$. Let us denote by $w_0 \in (T_{p_0}M_0)^\perp$ a unit normal vector of M_0 at p_0. For $\delta, \eta > 0$ to be chosen later, let Φ_{w_0} be the function that maps any $x \in \mathbb{R}^n$ to

$$\Phi_{w_0}(x) = x + \eta\phi\left(\frac{x - p_0}{\delta}\right) w_0,$$

where $\phi : \mathbb{R}^n \to \mathbb{R}$ is the real bump function $\phi(y) = \exp\left(-\|y\|^2/(1 - \|y\|^2)\right) \mathbb{1}_{\mathrm{B}(0,1)}(y)$ of Lemma H.1. We let $M_1 = \Phi_{w_0}(M_0)$ be the image of M_0 by Φ_{w_0}. Roughly speaking, M_0 and M_1 only differ by a bump of width δ and height η in the neighborhood of p_0. Note by now that Φ_{w_0} coincides with the identity map outside $\mathrm{B}(p_0, \delta)$ and in particular, $p_0 = \Phi_{w_0}(p_0) \in M_1$ as soon as $\delta \leq \mathrm{rch}_{\min}$.

Combining Proposition H.1 and Lemma H.1, we get that $M_1 \in \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$ and $\mathcal{H}^d(M_0)/2 \leq \mathcal{H}^d(M_1) \leq 2\mathcal{H}^d(M_0)$ as soon as

$$\frac{5\eta}{2\delta} \leq \frac{1}{10d} \text{ and } \frac{23\eta}{\delta^2} \leq \frac{1}{4\mathrm{rch}_{\min}}.$$

Under these assumptions, we have in particular that $\mathrm{d_H}(M_0, M_1) \leq \|\Phi_{w_0} - I_n\|_\infty \leq \eta \leq \mathrm{rch}_{\min}/10$. Also, by construction, $\Phi_{w_0}(p_0) = p_0 + \eta w_0$ belongs to M_1, so that

$$\mathrm{d_H}(M_0, M_1) \geq \mathrm{d}(p_0 + \eta w_0, M_0) = \eta,$$

since $w_0 \in (T_{p_0}M_0)^\perp$ [Fed59, Theorem 4.8 (12)]. Let us now consider the uniform probability distributions D_0 and D_1 over M_0 and M_1 respectively. These distributions have respective densities $f_i = \mathcal{H}^d(M_i)^{-1}\mathbb{1}_{M_i}$ ($i \in \{0, 1\}$) with respect to the d-dimensional Hausdorff measure $\mathcal{H}^d$ on $\mathbb{R}^n$. Furthermore, Φ_{w_0} is a global diffeomorphism that coincides with the identity map on $\mathrm{B}(p_0, \delta)^c$. As a result, since $\frac{5\eta}{2\delta} \leq \frac{1}{10d} \leq (2^{1/d} - 1)$, [AKC+19, Lemma D.2] yields that for $\delta \leq \mathrm{rch}_{\min}/2$,

$$\begin{aligned}\mathrm{TV}(D_0, D_1) &\leq 12 D_0\left(\mathrm{B}(0, \delta)\right) \\ &= 12\mathcal{H}^d(M_0 \cap \mathrm{B}(0, \delta))/\mathcal{H}^d(M_0) \\ &\leq 12(2^d\omega_d\delta^d)/\mathcal{H}^d(M_0),\end{aligned}$$

where we applied the upper bound of Lemma B.1 to get the last inequality, using that $\mathrm{rch}_{M_0} \geq 2\mathrm{rch}_{\min}$.

Finally, setting $\eta = \delta^2/(92\mathrm{rch}_{\min})$ yields a valid choice of parameters for all $\delta \leq \mathrm{rch}_{\min}/(2300d)$. Hence, we have shown that for all $\delta \leq \mathrm{rch}_{\min}/(2^{12}d) \leq \mathrm{rch}_{\min}/(2300d)$,

$$\mathrm{d_H}(M_0, M_1) \geq \frac{\delta^2}{92\mathrm{rch}_{\min}} \text{ and } \mathrm{TV}(D_0, D_1) \leq 12(2^d\omega_d\delta^d)/\mathcal{H}^d(M_0).$$

Equivalently, setting $\tau/2 = 12(2^d\omega_d\delta^d)/\mathcal{H}^d(M_0)$ and $\tau_{(0)} := 24\omega_d(\mathrm{rch}_{\min}/(2^{11}d))^d/\mathcal{H}^d(M_0)$, we have shown that for all $\tau \leq \tau_{(0)}$, there exists $M_1 \in \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$ such that

$$\mathrm{d_H}(M_0, M_1) \geq \frac{1}{92\mathrm{rch}_{\min}}\left(\frac{\mathcal{H}^d(M_0)\tau}{24(2^d\omega_d)}\right)^{2/d} \text{ and } \mathrm{TV}(D_0, D_1) \leq \tau/2.$$

We conclude the proof for $\tau \leq \tau_{(0)}$ by further bounding the term

$$\begin{aligned}
\mathrm{d_H}(M_0, M_1) &\geq \frac{1}{92\mathrm{rch}_{\min}} \left(\frac{\mathcal{H}^d(M_0)\tau}{24(2^d\omega_d)} \right)^{2/d} \\
&= \frac{\mathrm{rch}_{\min}}{368 \times 24^{2/d}} \left(\frac{\mathcal{H}^d(M_0)\tau}{\omega_d \mathrm{rch}_{\min}^d} \right)^{2/d} \\
&\geq \frac{\mathrm{rch}_{\min}}{2^{18}} \left(\frac{\mathcal{H}^d(M_0)\tau}{\omega_d \mathrm{rch}_{\min}^d} \right)^{2/d}.
\end{aligned}$$

Otherwise, if $\tau > \tau_{(0)}$, then the above construction applied with $\tau_{(0)}$ yields the existence of some $M_1 \in \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$ with the same properties, and

$$\mathrm{d_H}(M_0, M_1) \geq \frac{\mathrm{rch}_{\min}}{2^{18}} \left(\frac{\mathcal{H}^d(M_0)\tau_{(0)}}{\omega_d \mathrm{rch}_{\min}^d} \right)^{2/d} \quad \text{and} \quad \mathrm{TV}(D_0, D_1) \leq \tau_{(0)}/2 \leq \tau/2.$$

Summing up the two cases above, for all $\tau \leq 1$ we have exhibited some $M_1 \in \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$ with properties as above, $\mathrm{TV}(D_0, D_1) \leq \tau/2$ and

$$\begin{aligned}
\mathrm{d_H}(M_0, M_1) &\geq \frac{\mathrm{rch}_{\min}}{2^{18}} \left(\frac{\mathcal{H}^d(M_0) \min\left\{\tau, \tau_{(0)}\right\}}{\omega_d \mathrm{rch}_{\min}^d} \right)^{2/d} \\
&\geq \frac{\mathrm{rch}_{\min}}{2^{18}} \min\left\{ \frac{1}{2^{22}d^2}, \left(\frac{\mathcal{H}^d(M_0)\tau}{\omega_d \mathrm{rch}_{\min}^d} \right)^{2/d} \right\},
\end{aligned}$$

which concludes the proof.

□

Applying the technique of Proposition H.3 with manifolds M_0 having largest possible volume (typically of order $1/f_{\min}$) in the models $\{0\} \sqcup \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ and $\mathrm{B}(0, R) \sqcap \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ yields the following result. The proof follows the ideas of [AL19, Lemma 5]. To our knowledge, the first result of this type dates back to [GPPVW12a, Theorem 6].

Lemma H.5. • *Assume that* $f_{\min} \leq f_{\max}/4$ *and that*

$$2^{d+1}\sigma_d f_{\min} \mathrm{rch}_{\min}^d \leq 1.$$

Then for all $\tau \leq 1$*, there exist* $D_0, D_1 \in \{0\} \sqcup \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ *with respective supports* M_0 *and* M_1 *such that*

$$\mathrm{d_H}(M_0, M_1) \geq \frac{\mathrm{rch}_{\min}}{2^{20}} \min\left\{ \frac{1}{2^{20}d^2}, \left(\frac{\tau}{\omega_d f_{\min} \mathrm{rch}_{\min}^d} \right)^{2/d} \right\} \quad \textit{and} \quad \mathrm{TV}(D_0, D_1) \leq \tau/2.$$

• *Assume that* $\mathrm{rch}_{\min} \leq R/144$ *and* $f_{\min} \leq f_{\max}/96$*. Writing* $C'_d = 9(2^{2d+1}\sigma_{d-1})$*, assume that*

$$\min_{1 \leq k \leq n} \left(\frac{192\mathrm{rch}_{\min}\sqrt{k}}{R} \right)^k \leq 2^{d+1} C'_d f_{\min} \mathrm{rch}_{\min}^d \leq 1.$$

Then for all $\tau \leq 1$*, there exist* $D_0, D_1 \in \mathrm{B}(0, R) \sqcap \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ *with respective supports* M_0 *and* M_1 *such that*

$$\mathrm{d_H}(M_0, M_1) \geq \frac{\mathrm{rch}_{\min}}{2^{30}} \min\left\{ \frac{1}{2^{10}d^2}, \left(\frac{\tau}{\omega_d f_{\min} \mathrm{rch}_{\min}^d} \right)^{2/d} \right\} \quad \textit{and} \quad \mathrm{TV}(D_0, D_1) \leq \tau/2.$$

Proof of Lemma H.5. For both models, the idea is to first build a manifold $M_0 \in \mathcal{M}^{n,d}_{2\mathrm{rch}_{\min}}$ with prescribed volume close to $1/f_{\min}$, and then consider the variations of it given by Proposition H.3.

- Let M_0 be a d-dimensional sphere of radius $r_0 = \left(\frac{1}{2\sigma_d f_{\min}}\right)^{1/d}$ in $\mathbb{R}^{d+1} \times \{0\}^{n-(d+1)} \subseteq \mathbb{R}^n$ containing $x_0 = 0 \in \mathbb{R}^n$. By construction, $\mathrm{rch}_{M_0} = r_0 \geq 2\mathrm{rch}_{\min}$, so that $M_0 \in \mathcal{M}^{n,d}_{2\mathrm{rch}_{\min}}$, and one easily checks that $\mathcal{H}^d(M_0) = 1/(2f_{\min})$. For all $\tau \leq 1$, Proposition H.3 asserts that there exists a manifold $M_1 \in \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$ such that $x_0 \in M_1$, with volume
$$1/f_{\max} \leq 1/(4f_{\min}) \leq \mathcal{H}^d(M_0) \leq \mathcal{H}^d(M_1) \leq 2\mathcal{H}^d(M_0) \leq 1/f_{\min},$$
such that
$$\begin{aligned}\mathrm{d_H}(M_0, M_1) &\geq \frac{\mathrm{rch}_{\min}}{2^{18}} \min\left\{\frac{1}{2^{22}d^2}, \left(\frac{\tau}{2\omega_d f_{\min}\mathrm{rch}^d_{\min}}\right)^{2/d}\right\} \\ &\geq \frac{\mathrm{rch}_{\min}}{2^{20}} \min\left\{\frac{1}{2^{20}d^2}, \left(\frac{\tau}{\omega_d f_{\min}\mathrm{rch}^d_{\min}}\right)^{2/d}\right\},\end{aligned}$$
and with respective uniform distributions D_0 and D_1 over M_0 and M_1 that satisfy $\mathrm{TV}(D_0, D_1) \leq \tau/2$. Since the densities of D_0 and D_1 are constant and equal to $\mathcal{H}^d(M_0)^{-1}$ and $\mathcal{H}^d(M_1)^{-1}$ respectively, the bounds on the volumes of M_0 and M_1 show that D_0 and D_1 belong to $\{0\} \sqcup \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L = 0) \subseteq \{0\} \sqcup \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$, which concludes the proof.
- Let $M_0 \subseteq \mathbb{R}^n$ be a submanifold given by Proposition H.2 applied with parameters $\mathrm{rch}'_{\min} = 2\mathrm{rch}_{\min}$, $\mathcal{V} = 1/(2f_{\min})$ and $R' = R/2$. That is, $M_0 \in \mathcal{M}^{n,d}_{2\mathrm{rch}_{\min}}$ is such that $1/(48f_{\min}) \leq \mathcal{H}^d(M_0) \leq 1/(2f_{\min})$ and $M_0 \subseteq \mathrm{B}(0, R/2)$. For all $\tau \leq 1$, Proposition H.3 asserts that there exists a manifold $M_1 \in \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$ such that $\mathrm{d_H}(M_0, M_1) \leq \mathrm{rch}_{\min}/10$, with volume
$$1/f_{\max} \leq 1/(96f_{\min}) \leq \mathcal{H}^d(M_0)/2 \leq \mathcal{H}^d(M_1) \leq 2\mathcal{H}^d(M_0) \leq 1/f_{\min},$$
and
$$\begin{aligned}\mathrm{d_H}(M_0, M_1) &\geq \frac{\mathrm{rch}_{\min}}{2^{18}} \min\left\{\frac{1}{2^{22}d^2}, \left(\frac{\tau}{48\omega_d f_{\min}\mathrm{rch}^d_{\min}}\right)^{2/d}\right\} \\ &\geq \frac{\mathrm{rch}_{\min}}{2^{30}} \min\left\{\frac{1}{2^{10}d^2}, \left(\frac{\tau}{\omega_d f_{\min}\mathrm{rch}^d_{\min}}\right)^{2/d}\right\},\end{aligned}$$
and such that the respective uniform distributions D_0 and D_1 over M_0 and M_1 satisfy $\mathrm{TV}(D_0, D_1) \leq \tau/2$. Because $M_0 \subseteq \mathrm{B}(0, R/2)$ and $\mathrm{d_H}(M_0, M_1) \leq \mathrm{rch}_{\min}/10 \leq R/2$, we immediately get that $M_1 \subseteq \mathrm{B}(0, R/2 + R/2) = \mathrm{B}(0, R)$. As a result, this family clearly provides the existence of the announced ε-packing of $\left(\mathrm{B}(0, R) \sqcap \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}, \mathrm{d_H}\right)$. As above, the bounds on the volumes of M_0 and M_1 show that $D_0, D_1 \in \mathrm{B}(0, R) \sqcap \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L = 0) \subseteq \mathrm{B}(0, R) \sqcap \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$, which concludes the proof.

□

H.3.2 Proof of the Informational Lower Bounds for Manifold Estimation

With all the intermediate results above, the proofs of Theorem 5.2 and Theorem 5.5 follow straightforwardly.

Proof of Theorem 5.2 and Theorem 5.5. These are direct applications of Theorem G.1 for parameter of interest $\theta(D) = \mathrm{Supp}(D)$ and distance $\rho = \mathrm{d_H}$, with the hypotheses D_0, D_1 of the models $\{0\} \sqcup \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ and $\mathrm{B}(0, R) \sqcap \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L)$ given by Lemma H.5. □

H.4 Computational Lower Bounds: Packing Number of Manifold Classes

We now prove the computational lower bounds Theorems 5.3 and 5.6. For this, and in order to apply Theorem G.2, we build explicit packings of the manifold classes. To study the two models and the different regimes of parameters, we exhibit two types of such packings. The first ones that we describe (Proposition H.4) use translations of a fixed manifold M_0 in the ambient space, and are called ambient packings (see Appendix H.4.1). The second ones (Proposition H.5) use a local smooth bumping strategy based on a fixed manifold M_0, and are called intrinsic packings (see Appendix H.4.2). Finally, the proof of the computational lower bounds are presented in Appendix H.4.3.

H.4.1 Global Ambient Packings

To derive the first manifold packing lower bound, we will use translations in $\mathbb{R}^n$ and the following lemma.

Lemma H.6. *La K be a compact subset of $\mathbb{R}^n$. Given $v \in \mathbb{R}^n$, let $K_v = \{p + v, p \in K\}$ be the translation of K by the vector v. Then $\mathrm{d_H}(K, K_v) = \|v\|$.*

Proof of Lemma H.6. If $v = 0$, the result is straightforward, so let us assume that $v \neq 0$. Since K is compact, the map g defined for $p \in K$ by $g(p) = \langle v/\|v\|, p\rangle$ attains its maximum at some $p_0 \in K$. But by definition of K_v, $p_0 + v \in K_v$, so

$$\begin{aligned}
\mathrm{d_H}(K, K_v) &\geq \mathrm{d}(p_0 + v, K) \\
&= \min_{p \in K} \|(p_0 + v) - p\| \\
&\geq \min_{p \in K} \left\langle \frac{v}{\|v\|}, (p_0 + v) - p \right\rangle \\
&= \|v\| + \min_{p \in K} \left\langle \frac{v}{\|v\|}, p_0 - p \right\rangle \\
&= \|v\| \,.
\end{aligned}$$

On the other hand, for all $p \in K$ we have $p+v \in K_v$, yielding $\mathrm{d}(p, K_v) \leq \|v\|$, and symmetrically $\mathrm{d}(p + v, K) \leq \|v\|$. Therefore $\mathrm{d_H}(K, K_v) \leq \|v\|$, which concludes the proof. □

As a result, packings of sets in $\mathbb{R}^n$ naturally yields packings in the manifold space, by translating a fixed manifold $M_0 \subset \mathbb{R}^n$. With this remark in mind, we get the following ambient packing lower bound.

Proposition H.4. *Assume that $\mathrm{rch_{min}} \leq R/24$. Writing $C_d = 9(2^d \sigma_{d-1})$, let $\mathcal{V} > 0$ be such that*

$$1 \leq \frac{\mathcal{V}}{C_d \mathrm{rch}_{\min}^d} \leq \max_{1 \leq k \leq n} \left(\frac{R}{48 \mathrm{rch}_{\min} \sqrt{k}} \right)^k .$$

Then for all $\varepsilon \leq R/2$,

$$\log \mathrm{pk}_{\left(\mathrm{B}(0,R) \sqcap \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}, \mathrm{d_H}\right)}(\varepsilon) \geq n \log \left(\frac{R}{4\varepsilon} \right),$$

and such a packing can be chosen so that all its elements M have volume $\mathcal{V}/6 \leq \mathcal{H}^d(M) \leq \mathcal{V}$.

Proof of Proposition H.4. Let $z_1, \ldots, z_N \in \mathrm{B}(0, R/2)$ be a r-packing of $\mathrm{B}(0, R/2)$. From Proposition B.4, such a packing can be taken so that $N \geq (R/(4r))^n$. Applying Proposition H.2 with parameters $\mathrm{rch_{min}}$, $\mathcal{V}$ and $R' = R/2$, we get the existence of some $M_0 \in \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$ such

that $\mathcal{V}/6 \leq \mathcal{H}^d(M_0) \leq \mathcal{V}$ and $M_0 \subseteq \mathrm{B}(0, R/2)$. Note that for all $z \in \mathrm{B}(0, R/2)$, the translation $M_z = \{p + z, p \in M_0\}$ belongs to $\mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$, has the same volume as M_0, and satisfies $M_z \subseteq \mathrm{B}(0, R/2 + \|z\|) \subseteq \mathrm{B}(0, R)$. In addition, Lemma H.6 asserts that for all $z, z' \in \mathrm{B}(0, R/2)$, $\mathrm{d_H}(M_z, M_{z'}) = \|z - z'\|$. In particular, for all $i \neq j \in \{1, \ldots, N\}$, $\mathrm{d_H}(M_{z_i}, M_{z_j}) = \|z_i - z_j\| > 2r$. As a result, the family $\{M_{z_i}\}_{1 \leq i \leq N}$ provides us with an r-packing of $\left(\mathrm{B}(0, R) \sqcap \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}, \mathrm{d_H}\right)$ with cardinality $N \geq (R/(4r))^n$, and composed of submanifold with volume $\mathcal{V}/6 \leq \mathcal{H}^d(M) \leq \mathcal{V}$, which concludes the proof. □

H.4.2 Local Intrinsic Packings

In the same spirit as Proposition H.3 for informational lower bounds, the following result allows to build packings of manifold classes by small perturbations of a base submanifold M_0. Note, again, that the larger the volume $\mathcal{H}^d(M_0)$, the stronger the result.

Proposition H.5. *For all $M_0 \in \mathcal{M}^{n,d}_{2\mathrm{rch}_{\min}}$ and $r \leq \mathrm{rch}_{\min}/(2^{34}d^2)$, there exists a family of submanifolds $\{M_s\}_{1 \leq s \leq \mathcal{N}} \subseteq \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$ with cardinality $\mathcal{N}$ such that*

$$\log \mathcal{N} \geq n \frac{\mathcal{H}^d(M_0)}{\omega_d \mathrm{rch}^d_{\min}} \left(\frac{\mathrm{rch}_{\min}}{2^{19} r}\right)^{d/2},$$

and that satisfies:

- *M_0 and $\{M_s\}_{1 \leq s \leq \mathcal{N}}$ have a point in common: $M_0 \cap \left(\cap_{1 \leq s \leq \mathcal{N}} M_s\right) \neq \emptyset$.*
- *For all $s \in \{1, \ldots, \mathcal{N}\}$,*

$$\mathrm{d_H}(M_0, M_s) \leq 23r \text{ and } \mathcal{H}^d(M_0)/2 \leq \mathcal{H}^d(M_s) \leq 2\mathcal{H}^d(M_0).$$

- *For all $s \neq s' \in \{1, \ldots, \mathcal{N}\}$ $\mathrm{d_H}(M_s, M_{s'}) > 2r$.*

Proof of Proposition H.5. For $\delta \leq \mathrm{rch}_{\min}/8$ to be chosen later, let $\{p_i\}_{1 \leq i \leq N}$ be a maximal δ-packing of M_0. From Proposition B.3, this maximal packing has cardinality $N \geq \frac{\mathcal{H}^d(M_0)}{\omega_d (4\delta)^d}$.

Let $\eta > 0$ be a parameter to be chosen later. Given a family of unit vectors $\mathbf{w} = (w_i)_{1 \leq i \leq N} \in (\mathbb{R}^n)^N$ normal at the p_i's, i.e. $w_i \in (T_{p_i} M)^\perp$ and $\|w_i\| = 1$, we let $\Phi_{\mathbf{w}}$ be the function defined in Lemma H.1, that maps any $x \in \mathbb{R}^n$ to

$$\Phi_{\mathbf{w}}(x) = x + \eta \left(\sum_{i=1}^N \phi\left(\frac{x - p_i}{\delta}\right) w_i\right),$$

where $\phi : \mathbb{R}^n \to \mathbb{R}$ is the real bump function $\phi(y) = \exp\left(-\|y\|^2/(1 - \|y\|^2)\right) \mathbb{1}_{\mathrm{B}(0,1)}(y)$ of Lemma H.1. We let $M_{\mathbf{w}} = \Phi_{\mathbf{w}}(M_0)$ be the image of M_0 by $\Phi_{\mathbf{w}}$. The set $M_{\mathbf{w}} \subseteq \mathbb{R}^n$ hence coincides with M_0, except in the δ-neighborhoods of the p_i's, where it has a bump of size η towards direction w_i. Note by now that up to rotations of its coordinates, the vector $\mathbf{w} = (w_i)_{1 \leq i \leq N}$ belongs to $\mathcal{S}^{n-d}(0, 1)^N$. Combining Proposition H.1 and Lemma H.1, we see that $M_{\mathbf{w}} \in \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$ and $\mathcal{H}^d(M_0)/2 \leq \mathcal{H}^d(M_{\mathbf{w}}) \leq 2\mathcal{H}^d(M_0)$ as soon as

$$\frac{5\eta}{2\delta} \leq \frac{1}{10d} \text{ and } \frac{23\eta}{\delta^2} \leq \frac{1}{4\mathrm{rch}_{\min}}.$$

In the rest of the proof, we will work with these two inequalities holding true. In particular, because $\|\Phi_{\mathbf{w}} - I_n\|_\infty \leq \eta$, we immediately get that $\mathrm{d_H}(M_0, M_{\mathbf{w}}) \leq \eta$. We note also that all the $\Phi_{\mathbf{w}}$'s coincide with the identity map on (say) $M_0 \cap \partial \mathrm{B}(p_1, \delta)$, so that $M_0 \cap \left(\cap_{\mathbf{w}} M_{\mathbf{w}}\right)$ contains $M_0 \cap \partial \mathrm{B}(x_1, \delta)$ and is hence non-empty.

We now take two different families of unit normal vectors $\mathbf{w}$ and $\mathbf{w}'$ (i.e. $w_i, w_i' \in (T_{p_i}M_0)^\perp$ and $\|w_i\| = \|w_i'\| = 1$ for $1 \le i \le N$), and we will show that their associated submanifolds $M_{\mathbf{w}}$ and $M_{\mathbf{w}'}$ are far away in Hausdorff distance as soon as $\max_{1\le i\le N} \|w_i - w_i'\|$ is large enough. To this aim, we first see that by construction, $\Phi_{\mathbf{w}}(p_i) = p_i + \eta w_i \in \Phi_{\mathbf{w}}(M_0) = M_{\mathbf{w}}$ for all $i \in \{1, \dots, N\}$. In particular,

$$\mathrm{d_H}(M_{\mathbf{w}}, M_{\mathbf{w}'}) \ge \max_{1\le i\le N} \mathrm{d}(p_i + \eta w_i, M_{\mathbf{w}'}).$$

Let us fix a free parameter $\lambda_i \in [0,1]$ to be chosen later. As $\|\Phi_{\mathbf{w}'} - I_n\|_\infty \le \eta$, we can write for all $i \in \{1, \dots, N\}$ that

$$\begin{aligned}
\mathrm{d}(p_i + \eta w_i, M_{\mathbf{w}'}) &= \mathrm{d}\left(p_i + \eta w_i, \Phi_{\mathbf{w}'}(M_0)\right) \\
&= \mathrm{d}\left(p_i + \eta w_i, \Phi_{\mathbf{w}'}(M_0 \setminus \mathrm{B}(p_i, \lambda_i\delta))\right) \wedge \mathrm{d}\big(p_i + \eta w_i, \Phi_{\mathbf{w}'}(M_0 \cap \mathrm{B}(p_i, \lambda_i\delta))\big) \\
&\ge (\lambda_i\delta - \eta) \wedge \mathrm{d}\big(p_i + \eta w_i, \Phi_{\mathbf{w}'}(M_0 \cap \mathrm{B}(p_i, \lambda_i\delta))\big).
\end{aligned}$$

Further investigating the term $\mathrm{d}\left(p_i + \eta w_i, \Phi_{\mathbf{w}'}(M_0 \cap \mathrm{B}(p_i, \lambda_i\delta))\right)$, we see that for all $x \in M_0 \cap \mathrm{B}(p_i, \lambda_i\delta) \subseteq \mathrm{B}(p_i, \delta)$, $\Phi_{\mathbf{w}'}(x) = x + \eta\phi\left(\frac{x-p_i}{\delta}\right) w_i$. But from [Fed59, Theorem 4.18], $\mathrm{rch}_{M_0} \ge 2\mathrm{rch}_{\min}$ ensures that any $x \in M_0 \cap \mathrm{B}(p_i, \lambda_i\delta)$ can be written as $x = p_i + v + u$, where $v \in T_{p_i}M_0$ with $\|v\| \le \lambda_i\delta$, and $u \in \big(T_{p_i}M_0\big)^\perp$ with $\|u\| \le (\lambda_i\delta)^2/(4\mathrm{rch}_{\min})$. As a result, we have

$$\begin{aligned}
&\mathrm{d}\left(p_i + \eta w_i, \Phi_{\mathbf{w}'}(M_0 \cap \mathrm{B}(p_i, \lambda_i\delta))\right) \\
&\ge \min_{\substack{v\in T_{p_i}M_0, \|v\|\le\lambda_i\delta \\ u\in(T_{p_i}M_0)^\perp, \|u\|\le(\lambda_i\delta)^2/(4\mathrm{rch}_{\min})}} \left\| v + u + \eta\left(\phi\left(\frac{v+u}{\delta}\right) w_i' - w_i\right)\right\|.
\end{aligned}$$

But in the above minimum, v is orthogonal to u, w_i and w_i', so

$$\left\| v + u + \eta\left(\phi\left(\frac{v+u}{\delta}\right) w_i' - w_i\right)\right\| \ge \left\| u + \eta\left(\phi\left(\frac{v+u}{\delta}\right) w_i' - w_i\right)\right\|.$$

Additionally, $\phi\left(\frac{v+u}{\delta}\right)$ ranges in (a subset of) $[0,1]$ since $0 \le \phi \le 1$. In particular,

$$\begin{aligned}
\mathrm{d}\left(p_i + \eta w_i, \Phi_{\mathbf{w}'}(M_0 \cap \mathrm{B}(p_i, \lambda_i\delta))\right) &\ge \min_{\substack{u\in(T_{p_i}M_0)^\perp, \|u\|\le(\lambda_i\delta)^2/(4\mathrm{rch}_{\min}) \\ 0\le t\le 1}} \left\| u + \eta\left(t w_i' - w_i\right)\right\| \\
&\ge \min_{0\le t\le 1} \eta \left\| t w_i' - w_i\right\| - \frac{(\lambda_i\delta)^2}{4\mathrm{rch}_{\min}} \\
&= \eta \left\| \left(0 \vee \langle w_i, w_i'\rangle\right) w_i' - w_i\right\| - \frac{(\lambda_i\delta)^2}{4\mathrm{rch}_{\min}} \\
&\ge \eta \frac{\|w_i' - w_i\|}{2} - \frac{(\lambda_i\delta)^2}{4\mathrm{rch}_{\min}},
\end{aligned}$$

where the second line follows from triangle inequality, and the last two from elementary calculations. Putting everything together, we have shown that for all $\lambda_1, \dots, \lambda_N \in [0,1]$,

$$\mathrm{d_H}(M_{\mathbf{w}}, M_{\mathbf{w}'}) \ge \max_{1\le i\le N} \left\{ (\lambda_i\delta - \eta) \wedge \left(\eta \frac{\|w_i' - w_i\|}{2} - \frac{(\lambda_i\delta)^2}{4\mathrm{rch}_{\min}}\right)\right\}.$$

One easily checks that under the above assumptions on the parameters,

$$\lambda_i := \frac{\sqrt{\sqrt{2}\mathrm{rch}_{\min} \|w_i' - w_i\| \eta}}{\delta}$$

provides valid choices of $\lambda_i \in [0,1]$. Plugging these values in the previous bound yields

$$\mathrm{d_H}(M_{\mathbf{w}}, M_{\mathbf{w}'}) \geq \max_{1\leq i\leq N} \left\{ \left(\sqrt{\sqrt{2}\mathrm{rch}_{\min} \|w_i' - w_i\| \eta} - \eta \right) \wedge \left(\eta \frac{\|w_i' - w_i\|}{8} \right) \right\},$$

so that if we further assume that $\|w_i' - w_i\| \geq 4\sqrt{2}\eta/\mathrm{rch}_{\min}$, we obtain

$$\begin{aligned}\mathrm{d_H}(M_{\mathbf{w}}, M_{\mathbf{w}'}) &\geq \max_{1\leq i\leq N} \left\{ \eta \wedge \left(\eta \frac{\|w_i' - w_i\|}{8} \right) \right\} \\ &= \frac{\eta}{8} \max_{1\leq i\leq N} \|w_i' - w_i\|,\end{aligned}$$

where the last line follows from $\|w_i - w_i'\| \leq \|w_i\| + \|w_i'\| \leq 2$.

Setting $\eta = \delta^2/(92\mathrm{rch}_{\min})$, which is a value that satisfies all the requirements above as soon as $\delta \leq \mathrm{rch}_{\min}/(2300d)$, we have built a family of submanifolds $\{M_{\mathbf{w}}\}_{\mathbf{w}}$ of $\mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$ indexed by $\mathbf{w} \in \mathcal{S}^{n-d}(0,1)^N$, such that $\mathcal{H}^d(M_0)/2 \leq \mathcal{H}^d(M_{\mathbf{w}}) \leq 2\mathcal{H}^d(M_0)$, and which are guaranteed to satisfy

$$\mathrm{d_H}(M_{\mathbf{w}}, M_{\mathbf{w}'}) > \frac{\delta^2}{8(92\mathrm{rch}_{\min})} \times \frac{1}{4} \geq 2\left(\frac{\delta^2}{2082\mathrm{rch}_{\min}}\right),$$

provided that $\max_{1\leq i\leq N} \|w_i' - w_i\| > 1/4 = 2/8$. As a result, if we consider $(1/8)$-packings of the unit spheres $\mathcal{S}_{(T_{p_i}M_0)^\perp}(0,1) = \mathcal{S}^{n-d}(0,1)$ for $i \in \{1,\dots,N\}$, then for all $\delta \leq \mathrm{rch}_{\min}/(2300d)$, it naturally defines a $\left(\frac{\delta^2}{2082\mathrm{rch}_{\min}}\right)$-packing of $\mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$ with cardinality $\mathcal{N}$ a least

$$\mathcal{N} \geq \mathrm{pk}_{\mathcal{S}^{n-d}(0,1)}(1/8)^N \geq \mathrm{pk}_{\mathcal{S}^{n-d}(0,1)}(1/8)^{\frac{\mathcal{H}^d(M_0)}{\omega_d(4\delta)^d}},$$

and which consists of elements $M_{\mathbf{w}}$ such that $\mathcal{H}^d(M_0)/2 \leq \mathcal{H}^d(M_{\mathbf{w}}) \leq 2\mathcal{H}^d(M_0)$ and $\mathrm{d_H}(M_0, M_{\mathbf{w}}) \leq \eta = \delta^2/(92\mathrm{rch}_{\min})$. In particular, by setting $r = \frac{\delta^2}{2082\mathrm{rch}_{\min}}$, then for all $0 < r \leq \mathrm{rch}_{\min}/(2^{34}d^2)$, we have exhibited a r-packing of $\mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$ of cardinality $\mathcal{N}$ with

$$\log \mathcal{N} \geq \frac{\mathcal{H}^d(M_0)}{\omega_d(4\sqrt{2082\mathrm{rch}_{\min} r})^d} \log \mathrm{pk}_{\mathcal{S}^{n-d}(0,1)}(1/8),$$

composed of submanifolds having volume as above, and $\mathrm{d_H}(M_0, M_{\mathbf{w}}) \leq 2082r/92 \leq 23r$. From Proposition B.4, $\log \mathrm{pk}_{\mathcal{S}^{n-d}(0,1)}(1/8) \geq (n-d)\log 2$. Finally, by considering the cases $d \leq n/2$ and $d \geq n/2$, one easily checks that $(n-d) \geq n/(2d)$. In all, we obtain the announced bound

$$\begin{aligned}\log \mathcal{N} &\geq \frac{\mathcal{H}^d(M_0)}{\omega_d(4\sqrt{2082\mathrm{rch}_{\min} r})^d} \frac{n\log 2}{2d} \\ &\geq n \frac{\mathcal{H}^d(M_0)}{\omega_d \mathrm{rch}^d_{\min}} \left(\frac{\mathrm{rch}_{\min}}{2^{19} r}\right)^{d/2},\end{aligned}$$

which yields the announced result. □

Applying the technique of Proposition H.5 with manifolds M_0 having a large prescribed volume $\{0\} \sqcup \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$ and $\mathrm{B}(0,R) \sqcap \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$ respectively yields the following result.

Proposition H.6. *Let $\mathcal{V} > 0$ and $\varepsilon \leq \mathrm{rch}_{\min}/(2^{34}d^2)$.*

- *Assume that*

$$1 \leq \frac{\mathcal{V}}{2^{d+1}\sigma_d \mathrm{rch}^d_{\min}}.$$

Then,

$$\log \mathrm{pk}_{\left(\{0\}\sqcup\mathcal{M}^{n,d}_{\mathrm{rch}_{\min}},\mathrm{d_H}\right)}(\varepsilon) \geq n\frac{\mathcal{V}}{\omega_d \mathrm{rch}^d_{\min}}\left(\frac{\mathrm{rch}_{\min}}{2^{21}\varepsilon}\right)^{d/2}.$$

Furthermore, this packing can be chosen so that all its elements M satisfy

$$\mathcal{V}/4 \leq \mathcal{H}^d(M) \leq \mathcal{V}.$$

- *Assume that* $\mathrm{rch}_{\min} \leq R/144$. *Writing* $C'_d = 9(2^{2d+1}\sigma_{d-1})$, *assume that*

$$1 \leq \frac{\mathcal{V}}{2^{d+1}C'_d \mathrm{rch}^d_{\min}} \leq \max_{1\leq k\leq n}\left(\frac{R}{192\mathrm{rch}_{\min}\sqrt{k}}\right)^k.$$

Then,

$$\log \mathrm{pk}_{\left(\mathrm{B}(0,R)\sqcap\mathcal{M}^{n,d}_{\mathrm{rch}_{\min}},\mathrm{d_H}\right)}(\varepsilon) \geq n\frac{\mathcal{V}}{\omega_d \mathrm{rch}^d_{\min}}\left(\frac{\mathrm{rch}_{\min}}{2^{31}\varepsilon}\right)^{d/2}.$$

Furthermore, this packing can be chosen so that all its elements M satisfy

$$\mathcal{V}/96 \leq \mathcal{H}^d(M) \leq \mathcal{V}.$$

Proof of Proposition H.6. For both models, the idea is to first build a manifold $M_0 \in \mathcal{M}^{n,d}_{2\mathrm{rch}_{\min}}$ with prescribed volume close to $\mathcal{V}$, and then consider the variations of it given by Proposition H.5.

- Let M_0 be the centered d-dimensional sphere of radius $r_0 = \left(\frac{\mathcal{V}/2}{\sigma_d}\right)^{1/d}$ in $\mathbb{R}^{d+1}\times\{0\}^{n-(d+1)} \subseteq \mathbb{R}^n$. By construction, $\mathrm{rch}_{M_0} = r_0 \geq 2\mathrm{rch}_{\min}$, so that $M_0 \in \mathcal{M}^{n,d}_{2\mathrm{rch}_{\min}}$. Furthermore, one easily checks that $\mathcal{H}^d(M_0) = \mathcal{V}/2$. From Proposition H.5, there exists a family of submanifolds $\{M_s\}_{1\leq s\leq\mathcal{N}} \subseteq \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$ with cardinality $\mathcal{N}$ such that

$$\begin{aligned}\log\mathcal{N} &\geq n\frac{\mathcal{V}/2}{\omega_d \mathrm{rch}^d_{\min}}\left(\frac{\mathrm{rch}_{\min}}{2^{19}\varepsilon}\right)^{d/2}\\ &\geq n\frac{\mathcal{V}}{\omega_d \mathrm{rch}^d_{\min}}\left(\frac{\mathrm{rch}_{\min}}{2^{21}\varepsilon}\right)^{d/2},\end{aligned}$$

  that all share a point $x_0 \in \cap_{1\leq s\leq\mathcal{N}} M_s$, and such that $\mathrm{d_H}(M_s, M_{s'}) > 2\varepsilon$ for all $s \neq s' \in \{1,\dots,\mathcal{N}\}$, with volumes $\mathcal{V}/4 = \mathcal{H}^d(M_0)/2 \leq \mathcal{H}^d(M_s) \leq 2\mathcal{H}^d(M_0) = \mathcal{V}$. As a result, the family given by the translations $M'_s = M_s - x_0$ clearly provides the existence of the announced ε-packing of $\left(\{0\}\sqcup\mathcal{M}^{n,d}_{\mathrm{rch}_{\min}},\mathrm{d_H}\right)$.

- Let $M_0 \subseteq \mathbb{R}^n$ be a submanifold given by Proposition H.2 applied with parameters $\mathrm{rch}'_{\min} = 2\mathrm{rch}_{\min}$, $\mathcal{V}' = \mathcal{V}/2$ and $R' = R/2$. That is, $M_0 \in \mathcal{M}^{n,d}_{2\mathrm{rch}_{\min}}$ is such that $\mathcal{V}/48 \leq \mathcal{H}^d(M_0) \leq \mathcal{V}/2$ and $M_0 \subseteq \mathrm{B}(0, R/2)$. From Proposition H.5, there exists a family of submanifolds $\{M_s\}_{1\leq s\leq\mathcal{N}} \subseteq \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}}$ with cardinality $\mathcal{N}$ such that

$$\begin{aligned}\log\mathcal{N} &\geq n\frac{\mathcal{V}/48}{\omega_d \mathrm{rch}^d_{\min}}\left(\frac{\mathrm{rch}_{\min}}{2^{19}\varepsilon}\right)^{d/2}\\ &\geq n\frac{\mathcal{V}}{\omega_d \mathrm{rch}^d_{\min}}\left(\frac{\mathrm{rch}_{\min}}{2^{31}\varepsilon}\right)^{d/2},\end{aligned}$$

with $\mathrm{d_H}(M_0, M_s) \leq 23\varepsilon$ and $\mathrm{d_H}(M_s, M_{s'}) > 2\varepsilon$ for all $s \neq s' \in \{1, \ldots, \mathcal{N}\}$, and volumes $\mathcal{V}/96 \leq \mathcal{H}^d(M_0)/2 \leq \mathcal{H}^d(M_s) \leq 2\mathcal{H}^d(M_0) \leq \mathcal{V}$. Because $M_0 \subseteq \mathrm{B}(0, R/2)$ and $\mathrm{d_H}(M_0, M_s) \leq 23\varepsilon$ for all $s \in \{1, \ldots, \mathcal{N}\}$, we immediately get that $M_s \subseteq \mathrm{B}(0, R/2 + 23\varepsilon) \subseteq \mathrm{B}(0, R)$. As a result, this family clearly provides the existence of the announced ε-packing of $\big(\mathrm{B}(0, R) \sqcap \mathcal{M}^{n,d}_{\mathrm{rch_{min}}}, \mathrm{d_H}\big)$.

□

H.4.3 Proof of the Computational Lower Bounds for Manifold Estimation

We are now in position to prove the computational lower bounds presented in this work. First, we turn to the infeasibiliy result of manifold estimation using statistical queries in the unbounded model $\mathcal{D}^{n,d}_{\mathrm{rch_{min}}}(f_{\min}, f_{\max}, L)$ (Proposition 2.1).

Proof of Proposition 2.1. Since $\sigma_d f_{\min} \mathrm{rch}^d_{\min} \leq 1$, the uniform probability distribution D_0 over the centered unit d-sphere $M_0 \subseteq \mathbb{R}^{d+1} \times \{0\}^{n-(d+1)}$ of radius $\mathrm{rch_{min}}$ belongs to $\mathcal{D}^{n,d}_{\mathrm{rch_{min}}}(f_{\min}, f_{\max}, L)$. Given a unit vector $v \in \mathbb{R}^n$, the invariance of the model by translation yields that the uniform distributions D_k over $M_k = \{p + (3k\varepsilon)v, p \in M_0\}$, for $k \in \mathbb{Z}$, also belong to $\mathcal{D}^{n,d}_{\mathrm{rch_{min}}}(f_{\min}, f_{\max}, L)$. But for all $k \neq k' \in \mathbb{Z}$, $\mathrm{d_H}(M_k, M_{k'}) = 3|k - k'|\varepsilon > 2\varepsilon$. Hence, writing

$$\mathcal{M} = \left\{\mathrm{Supp}(D), D \in \mathcal{D}^{n,d}_{\mathrm{rch_{min}}}(f_{\min}, f_{\max}, L)\right\},$$

we see that the family $\{M_k\}_{k \in \mathbb{Z}}$ forms an infinite ε-packing of $(\mathcal{M}, \mathrm{d_H})$. From Theorem G.2, we get that the statistical query complexity of manifold estimation over the model $\mathcal{D}^{n,d}_{\mathrm{rch_{min}}}(f_{\min}, f_{\max}, L)$ with precision ε is infinite, which concludes the proof. □

We finally come to the proofs of the computational lower bounds over the fixed point model $\{0\} \sqcup \mathcal{D}^{n,d}_{\mathrm{rch_{min}}}(f_{\min}, f_{\max}, L)$ (Theorem 5.3) and the bounding ball model $\mathrm{B}(0, R) \sqcap \mathcal{D}^{n,d}_{\mathrm{rch_{min}}}(f_{\min}, f_{\max}, L)$ (Theorem 5.6).

Proof of Theorem 5.3 and Theorem 5.6. For both results, the idea is to exhibit large enough ε-packings of $\mathcal{M} = \{\mathrm{Supp}(D), D \in \mathcal{D}\}$, and apply Theorem G.2. In each case, the assumptions on the parameters $f_{\min}$, $f_{\max}$, $\mathrm{rch_{min}}$ and d ensure that the uniform distributions over the manifolds given by the packings of Proposition H.6 (and Proposition H.4 for Theorem 5.6) applied with $\mathcal{V} = 1/f_{\min}$ belong to the model, and hence that $\mathcal{M}$ contain these packings.

- To prove Theorem 5.3, let us write

$$\mathcal{M}_0 := \left\{\mathrm{Supp}(D), D \in \{0\} \sqcup \mathcal{D}^{n,d}_{\mathrm{rch_{min}}}(f_{\min}, f_{\max}, L)\right\} \subseteq \{0\} \sqcup \mathcal{M}^{n,d}_{\mathrm{rch_{min}}}.$$

  From Theorem G.2, any randomized SQ algorithm estimating $M = \mathrm{Supp}(D)$ over the model $\{0\} \sqcup \mathcal{D}^{n,d}_{\mathrm{rch_{min}}}(f_{\min}, f_{\max}, L)$ with precision ε and with probability of success at least $1 - \alpha$ must make at least

$$q \geq \frac{\log\big((1-\alpha)\mathrm{pk}_{(\mathcal{M}_0, \mathrm{d_H})}(\varepsilon)\big)}{\log(1 + \lfloor 1/\tau \rfloor)}$$

  queries to $\mathrm{STAT}(\tau)$. Furthermore, let $\{M_i\}_{1 \leq i \leq \mathcal{N}}$ be an ε-packing of $\{0\} \sqcup \mathcal{M}^{n,d}_{\mathrm{rch_{min}}}$ given by Proposition H.6, that we apply with volume $\mathcal{V} = 1/f_{\min}$. Recall that these manifolds are guaranteed to have volumes $1/(4f_{\min}) \leq \mathcal{H}^d(M_i) \leq 1/f_{\min}$. From the assumptions on the parameters of the model, we get that the uniform distributions $\big\{D_i := \mathbb{1}_{M_i}\mathcal{H}^d/\mathcal{H}^d(M_i)\big\}_{1 \leq i \leq \mathcal{N}}$ over the M_i's all belong to $\{0\} \sqcup \mathcal{D}^{n,d}_{\mathrm{rch_{min}}}(f_{\min}, f_{\max}, L)$. In particular, the family $\{M_i\}_{1 \leq i \leq \mathcal{N}}$ is also an ε-packing of $\mathcal{M}_0$, and therefore

$$\log\big(\mathrm{pk}_{(\mathcal{M}_0, \mathrm{d_H})}(\varepsilon)\big) \geq \log \mathcal{N} \geq n \frac{1}{\omega_d f_{\min} \mathrm{rch}^d_{\min}} \left(\frac{\mathrm{rch_{min}}}{2^{21}\varepsilon}\right)^{d/2},$$

  which yields the announced result.

- Similarly, to prove Theorem 5.6, write

$$\mathcal{M}_R := \left\{ \mathrm{Supp}(D), D \in \mathrm{B}(0,R) \sqcap \mathcal{D}^{n,d}_{\mathrm{rch}_{\min}}(f_{\min}, f_{\max}, L) \right\} \subseteq \mathrm{B}(0,R) \sqcap \mathcal{M}^{n,d}_{\mathrm{rch}_{\min}},$$

and apply Theorem G.2 to get

$$q \geq \frac{\log\big((1-\alpha)\mathrm{pk}_{(\mathcal{M}_R,\mathrm{d_H})}(\varepsilon)\big)}{\log(1+\lfloor 1/\tau \rfloor)}.$$

The assumptions on the parameters ensure that the packings exhibited in Proposition H.4 and Proposition H.6 applied with volume $\mathcal{V} = 1/f_{\min}$ are included in $\mathcal{M}_R$, so that

$$\log\big(\mathrm{pk}_{(\mathcal{M}_R,\mathrm{d_H})}(\varepsilon)\big) \geq n \max\left\{ \log\left(\frac{R}{4\varepsilon}\right), \frac{1}{\omega_d f_{\min} \mathrm{rch}^d_{\min}} \left(\frac{\mathrm{rch}_{\min}}{2^{31}\varepsilon}\right)^{d/2} \right\},$$

which concludes the proof.

□

References

[ACLZ17] Ery Arias-Castro, Gilad Lerman, and Teng Zhang. Spectral clustering based on local PCA. *J. Mach. Learn. Res.*, 18:Paper No. 9, 57, 2017.

[ACP13] Ery Arias-Castro and Bruno Pelletier. On the convergence of maximum variance unfolding. *J. Mach. Learn. Res.*, 14:1747–1770, 2013.

[AGY20] Srinivasan Arunachalam, Alex B. Grilo, and Henry Yuen. Quantum statistical query learning. *CoRR*, abs/2002.08240, 2020.

[AK21] Eddie Aamari and Alexander Knop. Statistical query complexity of manifold estimation. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2021, page 116–122, New York, NY, USA, 2021. Association for Computing Machinery.

[AKC+19] Eddie Aamari, Jisu Kim, Frédéric Chazal, Bertrand Michel, Alessandro Rinaldo, and Larry Wasserman. Estimating the reach of a manifold. *Electron. J. Stat.*, 13(1):1359–1399, 2019.

[AL18] Eddie Aamari and Clément Levrard. Stability and minimax optimality of tangential Delaunay complexes for manifold reconstruction. *Discrete Comput. Geom.*, 59(4):923–971, 2018.

[AL19] Eddie Aamari and Clément Levrard. Nonasymptotic rates for manifold, tangent space and curvature estimation. *Ann. Statist.*, 47(1):177–204, 2019.

[Alm86] F. Almgren. Optimal isoperimetric inequalities. *Indiana Univ. Math. J.*, 35(3):451–547, 1986.

[AS21] Yariv Aizenbud and Barak Sober. Non-Parametric Estimation of Manifolds from Noisy Data. *arXiv e-prints*, page arXiv:2105.04754, May 2021.

[BBI01] Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001.

[BD98] Shai Ben-David and Eli Dichterman. Learning with restricted focus of attention. *J. Comput. Syst. Sci.*, 56(3):277–298, 1998.

[BFJ+94] Avrim Blum, Merrick Furst, Jeffrey Jackson, Michael Kearns, Yishay Mansour, and Steven Rudich. Weakly learning DNF and characterizing statistical query learning using fourier analysis. In *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing*, STOC '94, page 253–262, New York, NY, USA, 1994. Association for Computing Machinery.

[BG14] Jean-Daniel Boissonnat and Arijit Ghosh. Manifold reconstruction using tangential Delaunay complexes. *Discrete Comput. Geom.*, 51(1):221–267, 2014.

[BHHS22] Clément Berenfeld, John Harvey, Marc Hoffmann, and Krishnan Shankar. Estimating the reach of a manifold via its convexity defect function. *Discrete Comput. Geom.*, 67(2):403–438, 2022.

[BKW19] Jean-Daniel Boissonnat, Siargey Kachanovich, and Mathijs Wintraecken. Sampling and Meshing Submanifolds in High Dimension. working paper or preprint, November 2019.

[BLW19] Jean-Daniel Boissonnat, André Lieutier, and Mathijs Wintraecken. The reach, metric distortion, geodesic convexity and the variation of tangent spaces. *J. Appl. Comput. Topol.*, 3(1-2):29–58, 2019.

[Byl94] Tom Bylander. Learning linear threshold functions in the presence of classification noise. In *Proceedings of the Seventh Annual ACM Conference on Computational Learning Theory, COLT 1994, New Brunswick, NJ, USA, July 12-15, 1994*, pages 340–347, 1994.

[CF97] Antonio Cuevas and Ricardo Fraiman. A plug-in approach to support estimation. *Ann. Statist.*, 25(6):2300–2312, 1997.

[dC92] Manfredo Perdigão do Carmo. *Riemannian geometry.* Mathematics: Theory & Applications. Birkhäuser Boston, Inc., Boston, MA, 1992. Translated from the second Portuguese edition by Francis Flaherty.

[Dey07] Tamal K. Dey. *Curve and surface reconstruction: algorithms with mathematical analysis*, volume 23 of *Cambridge Monographs on Applied and Computational Mathematics.* Cambridge University Press, Cambridge, 2007.

[DFT+15] Dana Dachman-Soled, Vitaly Feldman, Li-Yang Tan, Andrew Wan, and Karl Wimmer. Approximate resilience, monotonicity, and the complexity of agnostic learning. In Piotr Indyk, editor, *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2015, San Diego, CA, USA, January 4-6, 2015*, pages 498–511. SIAM, 2015.

[Div21a] Vincent Divol. Minimax adaptive estimation in manifold inference. *Electron. J. Stat.*, 15(2):5888–5932, 2021.

[Div21b] Vincent Divol. Reconstructing measures on manifolds: an optimal transport approach. *arXiv e-prints*, page arXiv:2102.07595, February 2021.

[DKS17] Ilias Diakonikolas, Daniel M. Kane, and Alistair Stewart. Statistical query lower bounds for robust estimation of high-dimensional gaussians and gaussian mixtures. In Chris Umans, editor, *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 73–84. IEEE Computer Society, 2017.

[DMGZ94] Giuseppe De Marco, Gianluca Gorni, and Gaetano Zampieri. Global inversion of functions: an introduction. *NoDEA Nonlinear Differential Equations Appl.*, 1(3):229–248, 1994.

[DMNS06] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography*, volume 3876 of *Lecture Notes in Comput. Sci.*, pages 265–284. Springer, Berlin, 2006.

[DV08] John Dunagan and Santosh S. Vempala. A simple polynomial-time rescaling algorithm for solving linear programs. *Math. Program.*, 114(1):101–114, 2008.

[DW21] David B Dunson and Nan Wu. Inferring Manifolds From Noisy Data Using Gaussian Processes. *arXiv e-prints*, page arXiv:2110.07478, October 2021.

[EGS03] Alexandre V. Evfimievski, Johannes Gehrke, and Ramakrishnan Srikant. Limiting privacy breaches in privacy preserving data mining. In *Proceedings of the Twenty-Second ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 9-12, 2003, San Diego, CA, USA*, pages 211–222. ACM, 2003.

[FCRP08] Maryam Fazel, Emmanuel Candès, Benjamin Recht, and Pablo Parrilo. Compressed sensing and robust recovery of low rank matrices. In *in Proc. 40th Asilomar Conf. Signals, Systems and Computers*, 2008.

[Fed59] Herbert Federer. Curvature measures. *Trans. Amer. Math. Soc.*, 93:418–491, 1959.

[Fed69] Herbert Federer. *Geometric measure theory.* Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York Inc., New York, 1969.

[Fel17] Vitaly Feldman. A general characterization of the statistical query complexity. In *Proceedings of Machine Learning Research*, volume 65, pages 785–830, Amsterdam, Netherlands, 2017.

[FGR+17] Vitaly Feldman, Elena Grigorescu, Lev Reyzin, Santosh S. Vempala, and Ying Xiao. Statistical algorithms and a lower bound for detecting planted cliques. *J. ACM*, 64(2):8:1–8:37, 2017.

[FGV21] Vitaly Feldman, Cristóbal Guzmán, and Santosh Vempala. Statistical query algorithms for mean vector estimation and stochastic convex optimization. *Math. Oper. Res.*, 46(3):912–945, 2021.

[FILN19] Charles Fefferman, Sergei Ivanov, Matti Lassas, and Hariharan Narayanan. Fitting a manifold of large reach to noisy data. *arXiv e-prints*, page arXiv:1910.05084, October 2019.

[FPV18] Vitaly Feldman, Will Perkins, and Santosh S. Vempala. On the complexity of random satisfiability problems with planted solutions. *SIAM J. Comput.*, 47(4):1294–1338, 2018.

[GPPVW12a] Christopher R. Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman. Manifold estimation and singular deconvolution under Hausdorff loss. *Ann. Statist.*, 40(2):941–963, 2012.

[GPPVW12b] Christopher R. Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman. Minimax manifold estimation. *J. Mach. Learn. Res.*, 13:1263–1291, 2012.

[Hat02] Allen Hatcher. *Algebraic topology.* Cambridge University Press, Cambridge, 2002.

[HTF09] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning.* Springer Series in Statistics. Springer, New York, second edition, 2009. Data mining, inference, and prediction.

[Jar19] DR. Ron Jarmin. Census bureau adopts cutting edge privacy protections for 2020 census, 2019. available at https://www.census.gov/newsroom/blogs/random-samplings/2019/02/census_bureau_adopts.html.

[JNS18] Noah M. Johnson, Joseph P. Near, and Dawn Song. Towards practical differential privacy for SQL queries. *Proc. VLDB Endow.*, 11(5):526–539, 2018.

[Kea98] Michael J. Kearns. Efficient noise-tolerant learning from statistical queries. *J. ACM*, 45(6):983–1006, 1998.

[KLN+11] Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam D. Smith. What can we learn privately? *SIAM J. Comput.*, 40(3):793–826, 2011.

[KZ15] Arlene K. H. Kim and Harrison H. Zhou. Tight minimax rates for manifold estimation under Hausdorff loss. *Electron. J. Stat.*, 9(1):1562–1582, 2015.

[LC87] William E. Lorensen and Harvey E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '87, page 163–169, New York, NY, USA, 1987. Association for Computing Machinery.

[Liu11] Yi-Kai Liu. Universal low-rank matrix recovery from pauli measurements. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc., 2011.

[LV07] John A. Lee and Michel Verleysen. *Nonlinear dimensionality reduction*. Information Science and Statistics. Springer, New York, 2007.

[NSW08] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.*, 39(1-3):419–441, 2008.

[PS22] Nikita Puchkin and Vladimir Spokoiny. Structure-adaptive manifold estimation. *Journal of Machine Learning Research*, 23(40):1–62, 2022.

[RS00] Sam T. Roweis and Lawrence K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[RSP+20] Ryan Rogers, Subbu Subramaniam, Sean Peng, David Durfee, Seunghyun Lee, Santosh Kumar Kancha, Shraddha Sahay, and Parvez Ahammad. Linkedin's audience engagements API: A privacy preserving data analytics system at scale. *CoRR*, abs/2002.05839, 2020.

[She08] Alexander A. Sherstov. Halfspace matrices. *Comput. Complex.*, 17(2):149–178, 2008.

[SVW16] Jacob Steinhardt, Gregory Valiant, and Stefan Wager. Memory, communication, and statistical queries. In Vitaly Feldman, Alexander Rakhlin, and Ohad Shamir, editors, *Proceedings of the 29th Conference on Learning Theory, COLT 2016, New York, USA, June 23-26, 2016*, volume 49 of *JMLR Workshop and Conference Proceedings*, pages 1490–1516. JMLR.org, 2016.

[Ten97] Joshua B. Tenenbaum. Mapping a manifold of perceptual observations. In Michael I. Jordan, Michael J. Kearns, and Sara A. Solla, editors, *Advances in Neural Information Processing Systems 10, [NIPS Conference, Denver, Colorado, USA, 1997]*, pages 682–688. The MIT Press, 1997.

[Val84] Leslie G. Valiant. A theory of the learnable. *Commun. ACM*, 27(11):1134–1142, 1984.

[Ver90] Karsten A. Verbeurgt. Learning DNF under the uniform distribution in quasi-polynomial time. In Mark A. Fulk and John Case, editors, *Proceedings of the Third Annual Workshop on Computational Learning Theory, COLT 1990, University of Rochester, Rochester, NY, USA, August 6-8, 1990*, pages 314–326. Morgan Kaufmann, 1990.

[Wai14] Martin J. Wainwright. Constrained forms of statistical minimax: computation, communication, and privacy. In *Proceedings of the International Congress of Mathematicians—Seoul 2014. Vol. IV*, pages 273–290. Kyung Moon Sa, Seoul, 2014.

[War65] Stanley L. Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965.

[Was18] Larry Wasserman. Topological data analysis. *Annu. Rev. Stat. Appl.*, 5:501–535, 2018.

[Yu97] Bin Yu. Assouad, fano, and le cam. In *Festschrift for Lucien Le Cam*, pages 423–435. Springer, 1997.

[YWS15] Y. Yu, T. Wang, and R. J. Samworth. A useful variant of the Davis-Kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015.