



**THÈSE DE DOCTORAT**  
**DE L'UNIVERSITÉ PSL**

Préparée à Université Paris-Dauphine

**STATISTICAL INFERENCE ON UNKNOWN MANIFOLDS**

**INFÉRENCE STATISTIQUE SUR DES VARIÉTÉS INCONNUES**

Soutenue par

**Clément BERENFELD**

Le 20 septembre 2022

École doctorale n°543

**École doctorale SDOSE**

Spécialité

**Mathématiques**

**Composition du jury :**

Catherine AARON	
Maitresse de conférence, U. Clermont Auvergne	<i>Examinatrice</i>
Frédéric CHAZAL	
Directeur de recherche, INRIA Saclay	<i>Rapporteur</i>
Elisabeth GASSIAT	
Professeure, Université Paris-Saclay	<i>Présidente</i>
Marc HOFFMANN	
Professeur, Université Paris-Dauphine	<i>Directeur de thèse</i>
Richard NICKL	
Professor, University of Cambridge	<i>Rapporteur</i>
Vincent RIVOIRARD	
Professeur, Université Paris-Dauphine	<i>Examineur</i>
Judith ROUSSEAU	
Professeure, University of Oxford	<i>Examinatrice</i>









*À mes petits frères.*

---

# Remerciements

À Marc, mille mercis pour ton encadrement toujours bienveillant à mon égard. Merci pour l'énergie que tu as su insuffler à cette thèse, pour ta patience envers mon caractère difficile, pour tous tes conseils avisés, pour tout le temps que tu m'as consacré, bref, pour tous ce que j'ai appris à tes côtés. Tu as été un directeur irréprochable sur tous les plans. L'apprenti chercheur que je suis te doit tout !

I'm infinitely grateful to Frédéric Chazal and Richard Nickl for having had the kindness of reporting this manuscript, and allowing its defense to take place. I am sincerely touched by the attention they paid to my work. Un grand merci également à Catherine Aaron, Elisabeth Gassiat, Vincent Rivoirard et Judith Rousseau, qui ont accepté de faire partie du jury de ma thèse. C'est un grand honneur qu'ils me font.

Huge thanks to all the people I had the pleasure to work with during this PhD: Eddie Aamari, Ery Arias-Castro, John Harvey, Clément Levrard, Paul Rosa, Judith Rousseau, and Krishnan Shankar, this manuscript is also yours. Our collaborations have been some incredibly nice and formative experiences, and, most importantly, working with you truly led me to undertake an academic career. Mention spéciale à Eddie et Ery, que j'ai la chance de suivre depuis que je les ai rencontrés à San Diego, et avec qui j'aurai toujours un immense plaisir de travailler et d'échanger, de maths ou d'autres choses. Merci à Luc Illien pour m'avoir donné goût à la sismologie, à Vincent Divol, pour les nombreux projets et groupes de lecture que nous avons lancés ensemble, à Jacques Féjoz, pour avoir tenté de m'introduire au fascinant monde des systèmes dynamiques.

Many thanks to those who welcomed me in their labs during the short academic visits I made in the course of my PhD. For the third time, thank you very much to Judith for these lovely stays in Oxford; thank you to Markus Reiß for the warm welcome in Berlin — and for incidentally making me meet with Alexandra — et bien sûr, merci beaucoup à Alexandra Carpentier pour cet accueil des plus chaleureux à Potsdam, qui m'a fait m'y sentir déjà comme chez moi.

---

Thank you to all the organizers of the seminars, workshops, and conferences I had the pleasure to attend. These were the spice of my PhD.

À mes collègues de Dauphine, merci pour tous ces moments passés autour d'un café ou d'un bon petit plat du CROUS. Mention spéciale à Jeanne et à Ryad, sans qui la motivation de me rendre à l'université m'aurait souvent manquée. Merci également à toute l'équipe du CEREMADE, permanents, administratifs et ingénieurs, qui ont fait de ce laboratoire un environnement de travail exceptionnellement agréable.

Merci à celles et ceux qui ont eu la générosité de me prêter un coin de bureau pour travailler : à Yann et Félicien, pour toutes les heures que j'ai passées clandestinement chez eux ; à Paul et Léa, pour l'accès illimité à la plus belle mezzanine de la capitale qu'ils ont bien voulu m'octroyer.

Merci à toute ma famille, pour l'intérêt sincère qu'ils ont toujours porté à mes travaux et leurs encouragements constants, pour tout l'amour que j'ai reçu d'eux et que j'espère leur avoir rendu. À mes petits frères, que j'ai eu la chance d'avoir sur Paris cette dernière année. À ma mère, mon père et Sylvie, qui prenaient soin de moi quand je faisais semblant de travailler. À mes mamies, mon parrain et ma marraine, à qui je donne peu de nouvelles mais à qui je pense beaucoup.

Merci enfin à tous mes amis, de Paris, de Corrèze ou d'ailleurs, sans la présence et le soutien desquels cette thèse n'aurait pas vu le jour.

Pointe de la Torche, septembre 2022.

# Résumé (en français)

En statistique, l'*hypothèse des variétés* suppose que les données observées se répartissent autour de structures de faible dimension, appelées variétés. Ce postulat permet d'expliquer pourquoi les algorithmes d'apprentissage fonctionnent bien même sur des données en grande dimension, et est naturellement satisfait pour de nombreux jeux de données issus de la vie réelle.

Nous présentons dans cette thèse quelques contributions aux problèmes d'estimation de deux quantités sous cette hypothèse : la densité de la distribution sous-jacente, et le reach de son support. Pour l'estimation du reach, nous élaborons des stratégies basées sur des invariants géométriques, avec d'une part la fonction de défaut de convexité, et d'autre part, des mesures de distortion métrique, desquels nous obtenons des vitesses de convergence optimales au sens minimax. Concernant l'estimation de la densité, nous proposons deux approches : l'une s'appuyant sur l'étude fréquentiste d'un estimateur à noyaux, et une approche bayésienne non-paramétrique se reposant sur des mélanges de gaussiennes. Nous montrons que ces deux méthodes sont optimales et adaptatives en la régularité de la densité.

Enfin, nous examinons le comportement de certaines mesures de centralité dans des graphes aléatoires géométriques, l'étude duquel, bien que sans lien avec l'hypothèse des variétés, a des implications méthodologiques et théoriques qui peuvent être intéressantes dans tout cadre statistique.

## Mots clés

Statistique non-paramétrique, statistique en grande dimension, apprentissage sur variété, inférence géométrique, estimation de densité, théorie minimax, estimation adaptative, estimation du reach, inférence bayésienne, mesure de centralité, profondeur statistique

---

# Abstract

In high-dimensional statistics, the *manifold hypothesis* presumes that the data lie near low-dimensional structures, called manifolds. This assumption helps explain why machine learning algorithms work so well on high-dimensional data, and is satisfied for many real-life data sets.

We present in this thesis some contributions regarding the estimation of two quantities in this framework: the density of the underlying distribution, and the reach of its support. For the problem of reach estimation, we suggest different strategies based on important geometric invariants — namely the convexity defect functions, and measures of metric distortions — from which we derive minimax-optimal rates of convergence. Regarding the problem of density estimation, we propose two approaches: one relying on the frequentist study of a kernel density estimator, and a Bayesian non-parametric approach based on location-scale mixtures of Gaussians. Both methods are shown to be optimal in most settings, and adaptive to the smoothness of the density.

Lastly, we examine the behavior of some centrality measures in random geometric graph, the study of which, although unrelated to the manifold hypothesis, bears methodological and theoretical implications that can be of interest in any statistical framework.

## Keywords

Nonparametric statistics, high-dimensional statistics, manifold learning, manifold hypothesis, geometric inference, density estimation, minimax theory, adaptive estimation, reach estimation, Bayesian inference, centrality measure, data depth

---



# Contents

<b>Remerciements</b>	<b>vii</b>
<b>Résumé (en français)</b>	<b>ix</b>
<b>Abstract</b>	<b>xi</b>
<b>1 General introduction</b>	<b>1</b>
1.1 The manifold hypothesis . . . . .	2
1.2 Structural assumptions . . . . .	5
1.3 Our results . . . . .	9
1.4 Organization of the manuscript . . . . .	17
<b>2 Reach estimation via the convexity defect function</b>	<b>19</b>
2.1 Introduction . . . . .	20
2.2 Geometry of the reach . . . . .	24
2.3 Geometrical framework . . . . .	26
2.4 Convexity defect functions . . . . .	27
2.5 Approximating the reach . . . . .	38
2.6 Minimax rates for reach estimators: Upper bounds . . . . .	41
2.7 Minimax rates for reach estimators: Lower bounds . . . . .	48
<b>3 Reach estimation via metric learning</b>	<b>55</b>
3.1 Introduction . . . . .	56
3.2 Geometric and Statistical Model . . . . .	59
3.3 Reach and Related Quantities . . . . .	61
3.4 Spherical Distortion Radius . . . . .	67
3.5 Optimal Metric Learning . . . . .	73
3.6 Optimal Reach Estimation . . . . .	77
3.7 Conclusion and Further Prospects . . . . .	80

---

<b>Appendices</b>	<b>81</b>
3.A Proofs of Section 3.3 . . . . .	81
3.B Proofs of Section 3.4 . . . . .	93
3.C Proofs of Section 3.5 . . . . .	96
3.D Proofs of Section 3.6 . . . . .	104
<b>4 Pointwise density estimation on submanifolds</b>	<b>107</b>
4.1 Introduction . . . . .	108
4.2 Manifold-supported probability distributions . . . . .	112
4.3 Density estimation at a fixed point . . . . .	118
4.4 Numerical illustration . . . . .	126
<b>Appendices</b>	<b>133</b>
4.A Appendix of Chapter 4 . . . . .	133
<b>5 Bayesian manifold density estimation</b>	<b>149</b>
5.1 Introduction . . . . .	150
5.2 Model : distributions concentrated near manifolds . . . . .	154
5.3 Main results . . . . .	162
5.4 Numerical experiments . . . . .	166
5.5 Proofs of the main results . . . . .	169
5.6 Discussion . . . . .	178
<b>Appendices</b>	<b>181</b>
5.A Some facts on submanifolds with bounded reach . . . . .	181
5.B Appendix to Section 5.2 . . . . .	189
5.C Appendix to Section 5.5.2: proof of Theorem 5.3.7 . . . . .	199
5.D Appendix to Section 5.5.1: proof of Theorem 5.3.2 . . . . .	205
5.E Appendix to Section 5.4 . . . . .	216
<b>6 From graph centrality to data depth</b>	<b>219</b>
6.1 Introduction . . . . .	220
6.2 Degree . . . . .	222
6.3 H-Index . . . . .	225
6.4 Coreness . . . . .	229
6.5 Numerical Simulations . . . . .	240
6.6 Concluding Remarks and Open Questions . . . . .	244
<b>Bibliography</b>	<b>245</b>

# Chapter 1

## General introduction

The present thesis tackles a few problems in nonparametric statistics formulated under the manifold hypothesis, meaning that the data is assumed to lay on or near sub-manifolds of the Euclidean space. More precisely, we aim at presenting contributions regarding the estimation of two quantities: the density of the underlying probability distribution, and the reach of its support. We will also discuss the behavior of centrality measures in random geometric graphs, the study of which, although unrelated to our manifold framework, bears general, methodological purposes that can transversally apply to most statistical frameworks.

Section 1.1 seeks to motivate the research that has been carried out in the past few years during the preparation of this thesis, and is designed to be understandable without specific mathematical knowledge. Section 1.2 introduces more precisely the statistical setting in which this work takes place. The main contributions are reviewed in Section 1.3. They are stated in a non-rigorous fashion and we refer to the mentioned chapters for precise definitions and results. The overall organisation of the manuscript is summarized in Section 1.4.

### Contents

---

<b>1.1</b>	<b>The manifold hypothesis</b>	<b>2</b>
<b>1.2</b>	<b>Structural assumptions</b>	<b>5</b>
<b>1.3</b>	<b>Our results</b>	<b>9</b>
1.3.1	Estimating the reach of a submanifold	9
1.3.2	Manifold density estimation	11
1.3.3	Beyond density estimation	15
<b>1.4</b>	<b>Organization of the manuscript</b>	<b>17</b>

---

## 1.1 The manifold hypothesis

Modern applications of statistical learning often revolves around high-dimensional data sets<sup>1</sup>, whether they be genomic data [MPRR04], images [BDSS21], texts [DGK02], high-resolution time series [PVI<sup>+</sup>16], etc. This raises challenges not only in term of computational resources, but also in term of designing efficient algorithms with solid, established guarantees. Unfortunately, the performance of such algorithms should, from a theoretical point of view, deteriorate as the number of dimensions grows large. This phenomenon is widely known as the *curse of dimensionality* [VF05], and stems from the degenerated geometric behavior of high-dimensional spaces: the volume of the balls plummets and collapses toward their boundaries, Gaussian distributions become heavy-tailed, etc — see for instance [Gir21, Chap. 1]. And yet, current algorithms, and sometimes even the simplest ones, manage to performs very well on high-dimensional data, highlighting an indisputable gap between the empirical performance bounds of such algorithms and their theoretical counterparts.

Let's take a very simple example to illustrate our point. In 1998, researchers LeCun et al. [LBBH98] created the MNIST data set from the preexisting (but unfit to statistical analysis) NIST data set. The new data set consisted of 60000 pictures of handwritten digits normalized to 28x28-pixels, grayscale images, as shown in Figure 1.1.

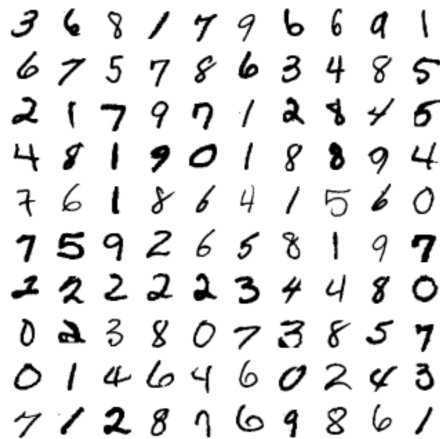


Figure 1.1 – Examples taken from the MNIST data set. Since the image are grayscale, the dimension  $D$  of the data is their number of pixels, which is  $D = 28 \times 28 = 784$ . Source: [LBBH98, Fig 4].

[LBBH98] then proceeded to test a number of different machine learning classification algorithms on this data set, aiming at automatically recognizing the digits

<sup>1</sup>The *dimension* of a piece of data is the number of parameters that describe it. For instance, the dimension of an image is its number of pixels (if it is a grayscale image) or three times its number of pixels (if it is a RGB image).

from their handwritten forms. Among the methods that served as a baseline for comparing the efficiencies of the algorithms, a  $k$ -nearest neighbors classification ( $k$ -NN) was performed. The  $k$ -NN algorithm is a very simple procedure that classifies a data point by a majority vote among its  $k$ -nearest neighbors in the data. This method is particularly interesting because it does not require any training to classify new data, and because very precise performance bounds are known for this method, see for instance [Sam12]. The last reference shows that, under mild assumptions, the optimal classifying accuracy of  $k$ -NN under  $n$  observations is  $n^{-4/(4+D)}$  (up to a constant), with  $D$  being the dimension of the data. Let us do a quick numerical application with the MNIST data set:

$$n^{-4/(4+D)} = 1.057... \quad \text{for } n = 60000 \text{ and } D = 784.$$

The macroscopic value of  $n^{-4/(4+D)}$  indicates that the  $k$ -NN classifier should perform poorly under such circumstances. This, however, comes in contradiction with the 95% accuracy observed by [LBBH98] for the same classifier. The theoretical number of observation  $n$  inferred from the bound  $n^{-4/(4+D)}$  that we would need to get as close as a 5% misclassification rate would be

$$n_{\text{theo}} \approx (100/5)^{(4+784)/4} = 20^{197} \approx 10^{256},$$

which, we can all agree, seems fairly larger than 60000 (and is an utterly unrealistic number for the size of a data set). In the other direction, one could wonder what would be the theoretical dimension of a data set of size 60000 that could achieve a misclassification rate of 5% with a  $k$ -NN classifier. That dimension is

$$D_{\text{theo}} \approx \frac{4 \log(60000)}{\log(100/5)} - 4 = 10.690... \approx 11, \tag{1.1}$$

which is very far below the actual dimension  $D = 784$ . This motivates the following assertion: the MNIST data, although living in a space of large dimension, is concentrated near a low-dimensional subset of the ambient space. The dimension of that subset will be called the *intrinsic* dimension of the data: it is the dimension that effectively plays a part in the performance of the algorithms, and can be seen as the number of implicit parameters that govern the data set. Regarding MNIST data set, [PZA<sup>+</sup>21] have estimated its intrinsic dimension using the method of [LB04]. They found the intrinsic dimension to lie between 7 and 13 depending on the tuning of some hyperparameter; our (improper) estimate (1.1) of the intrinsic dimension appears to fall exactly within that range.

Because we want to conduct an analysis in this framework, we need to add some structure to the underlying subset around which the data is gathered. The idea of assuming that this structure might be linear has been popular for a while, with the

rise of linear dimensionality reduction technics such as PCA [WEG87], or the study of linear models such as that single index models [HJS01], or multi-indices models [CHM11]. Going non-linear was thus only the next logical step. In the early 2000, the groundbreaking publications of [TSL00] and [RS00], which described a non-linear methods (Isomap and Local Linear Embedding) for reducing the dimensionality of the data, paved the way for the development of manifold learning, whose funding assumption is

**MH** *The data lie near a smooth, low-dimensional structure: a manifold.*

We will refer to Assumption (MH) as the *manifold hypothesis*. Ever since, many other non-linear dimensionality reduction methods have seen the light of the day — see for instance Laplacian Eigenmap [BN03], Maximum Variance Unfolding [WS06a], Uniform Manifold Approximation and Projection [MHM18] — see Figure 1.2 for a use of UMAP on the MNIST data set. All these technics came with an impressive number

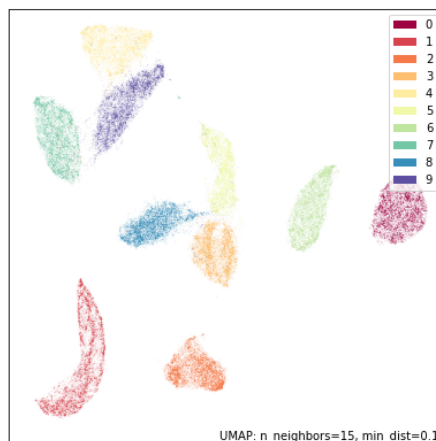


Figure 1.2 – UMAP representation of the MNIST data set. Each point in the graph represent one handwritten digit from the data set (see Figure 1.1), with the color standing for its value. The fact that this data set can be represented on a space of dimension two with a good inter-class separation certainly doesn't disprove the manifold hypothesis in this case. Source: [https://umap-learn.readthedocs.io/en/latest/densmap\\_demo.html](https://umap-learn.readthedocs.io/en/latest/densmap_demo.html).

of real-life applications, ranging, just to mention a few, from cosmology [BM21] (see Figure 1.3), seismology [KLM<sup>+</sup>20], to biology, with low-dimensional mapping of either cell trajectories [CSQ<sup>+</sup>19] or protein trajectories [VDL<sup>+</sup>21], or to medical science, with tumoral cells segmentation [XY15].

Let us finally mention that the manifold hypothesis is readily satisfied for a number of data set because of the geometric nature of the data itself. This concerns for instance low-dimensional data sets such as 3D point clouds of real-life objects like body scans

[Cot12], other space located-data such that locations of markers at the surface of cells [KPS14] — see Figure (1.4, Left) — or high-dimensional data in the case of sounds, texts [BN01], or images [DG05], see for instance Figure (1.4, Right) for a visual example.

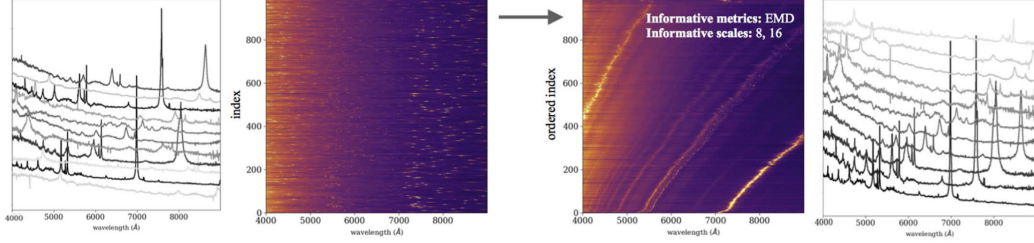


Figure 1.3 – Uncovering the one-dimensional structure of 1000 high-resolution spectra of quasars with the Sequencer. Source: [BM21, Fig 4].



Figure 1.4 – (Left) Point cloud obtained using single-molecule localization microscopy. Each dot represent the location of a protein at the surface of a cell. Source: [KPS14, Fig 1]. (Right) Data from the COIL-20 data set [NNM96]. Although high-dimensional, the data can be parametrized by the angle from which the picture is taken, and can therefore be faithfully represented on a submanifold of dimension 1. Source: [SJF14, Fig 3].

## 1.2 Structural assumptions

**Statistical model** Let us start with a few notations. For any closed subset  $K \subset \mathbb{R}^D$ , and any  $\delta > 0$ , we define the  $\delta$ -offset of  $K$  to be the set of all points of  $\mathbb{R}^D$  that are at distance less than  $\delta$  from  $K$  for the Euclidean distance, namely

$$K^\delta := \{x \in \mathbb{R}^D \mid d(x, K) \leq \delta\} = \bigcup_{z \in K} B(z, \delta),$$

where  $x \mapsto d(x, K)$  is the distance function to  $K$ , and where  $B(z, \delta)$  denotes the Euclidean open ball of center  $z$  and radius  $\delta$ . If  $K$  has Hausdorff dimension  $d \in \{1, \dots, D\}$ , we let  $\mu_K$  be the restriction to  $K$  of the Hausdorff  $d$ -dimensional measure on  $\mathbb{R}^D$  (see

[Fed69, Sec 2.10.2] for precise definitions regarding the Hausdorff dimension and measures). In particular, if  $K$  is an open subset of  $\mathbb{R}^D$  (such as  $K^\delta$ ), then  $\mu_K$  is simply the restriction to  $K$  of the Lebesgue measure on  $\mathbb{R}^D$ .

We consider a  $n$ -sample of iid random variables  $X_1, \dots, X_n$  drawn from a probability distribution  $P$  on  $\mathbb{R}^D$ . This distribution  $P$  will be always assume to satisfy one of the following assumptions, which are the formal counterparts of (MH):

**MH-0** *There exists a closed submanifold  $M \subset \mathbb{R}^D$  such that  $P \ll \mu_M$ ;*

**MH- $\delta$**  *There exists a closed submanifold  $M \subset \mathbb{R}^D$  such that  $P \ll \mu_{M^\delta}$ .*

The symbol  $\ll$  stands for absolute continuity of measures. The case (MH-0) corresponds to the ideal case where the data  $X_1, \dots, X_n$  lie exactly on the submanifold  $M$  and is very convenient as far as mathematical analysis is concerned, while the case (MH- $\delta$ ) encompasses the more realistic situation where the data lie  $\delta$ -close to  $M$  and is thus particularly applicable to real-life situations when the observations can be noisy. See Figure 1.5 for a visual representation of this situation. Note that in this framework, the submanifold  $M$  is unknown, so that its knowledge cannot be used in any estimation procedure.

This model naturally comes with numerous statistical challenges: since now the underlying distribution exhibit a rich geometric structure, it could seem particularly relevant to estimate its main features. In Section 1.1, we already briefly mentioned the problem of estimating the intrinsic dimension as tackled in [LB04], see also [FSA07, KRW16]. We could as well be interested in estimating the homology of the support [NSW08, BRS<sup>+</sup>12], its persistent homology [CGLM14], its boundary [AC20, AAL21], the geodesic length [TSL00, ACC20] or the support itself [GPPVW12a, AL19, Div21a].

**The minimax setting** Most of the time, we will adopt a minimax perspective. This means the following: given a quantity of interest  $\theta(P) \in \Theta$  for some parameter space  $\Theta$ , given a loss  $\ell : \Theta \times \Theta \rightarrow \mathbb{R}_+$  that quantifies the accuracy of estimators, and given a model  $\Sigma$  of probability distributions that fulfils either (MH-0) or (MH- $\delta$ ), we want to find an estimator  $\hat{\theta} := \hat{\theta}(X_1, \dots, X_n)$  such that its worst case risk on  $\Sigma$ ,

$$\sup_{P \in \Sigma} \mathbb{E}_{P^{\otimes n}}[\ell(\hat{\theta}, \theta(P))],$$

is as small as possible. By that, we mean that we require the worst case risk to decrease, as  $n$  grows large, at least as fast as the minimax risk, which is simply defined as the best worst case risk on  $\Sigma$ :

$$\inf_{\tilde{\theta}} \sup_{P \in \Sigma} \mathbb{E}_{P^{\otimes n}}[\ell(\tilde{\theta}, \theta(P))],$$



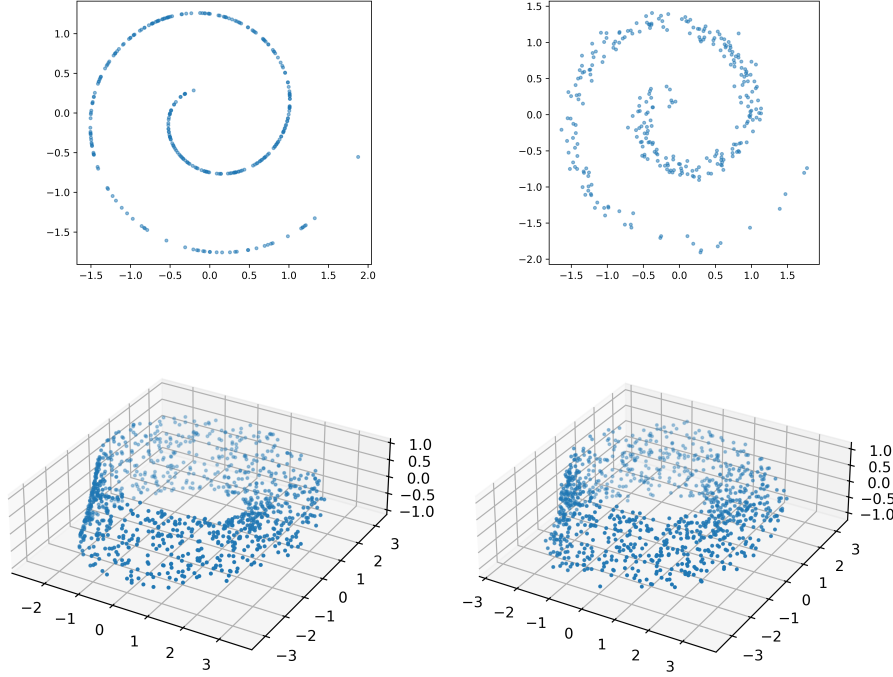


Figure 1.5 – Examples of synthetic data generated from distributions fulfilling either (MH-0) with intrinsic dimension 1 (Top Left) and 2 (Bottom Left) or (MH- $\delta$ ) with dimension 1 and  $\delta = 0.3$  (Top Right) and dimension 2 and  $\delta = 0.5$  (Bottom Right).

where the infimum is taken over all measurable estimators  $\tilde{\theta}$  of  $\theta(P)$ . The general strategy will be like this: first, we bound from above the worst case risk of our estimator  $\hat{\theta}$ . Then, we bound the minimax risk from below, using classical two-points or multiple-points argument [Yu97, Tsy08]. If the two bounds decrease with  $n$  at the same speed, our job is done.

**Reach constraints** Designing a relevant statistical model  $\Sigma$  is usually not an easy task. We want  $\Sigma$  to be general enough so that it covers a wide array of possibilities and yields estimators that are not too specifically subordinated to its features while being not too vast either so that doing statistical inference on  $\Sigma$  is not trivially impossible (meaning that the minimax risk is not bounded away from zero). To address that last issue we need in particular to discard from our models probability measures that are supported on manifolds that are too irregular, with the idea that no amount of observations will ever be able to account for that irregularity if the latter is not constrained. For this matter, we use the notion of reach as introduced by [Fed59], which is defined for any

closed subset  $A \subset \mathbb{R}^D$  as

$$\text{rch}(A) := \sup \{r \geq 0 \mid \forall x \in A^r, \exists! a \in A, d(x, A) = \|x - a\|\}.$$

The reach will encapsulate the regularity of a submanifold  $M$  on two aspects: first, it will measure how curved it is on a local level, and second, it will measure how close it is from self-intersecting. See Figure 1.6 for a visual interpretation of the situation.

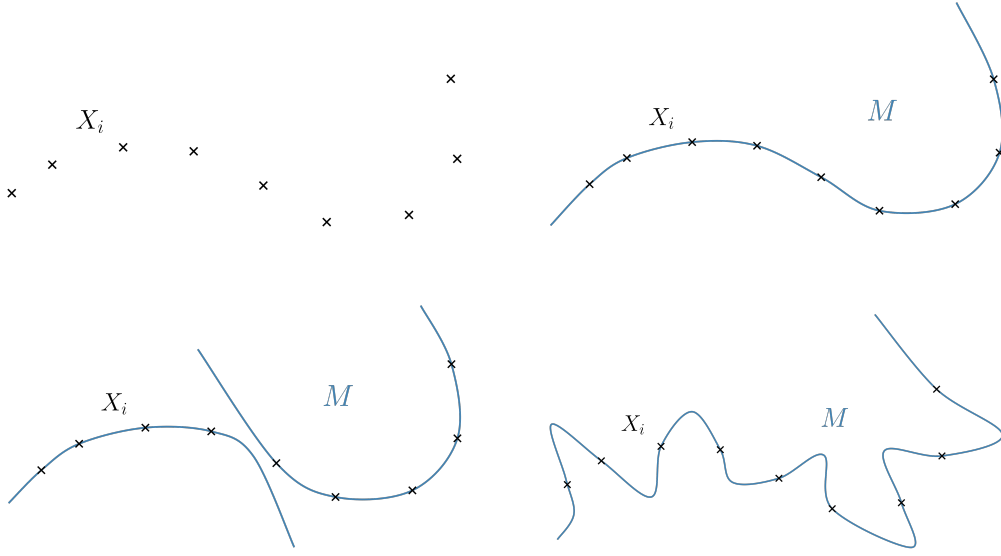


Figure 1.6 – A toy-example of point cloud  $X_1, \dots, X_n$  (Top Left) with three different kinds of interpolating manifolds. A reach constraint will tend to discard the last two (Bottom Left and Bottom Right) from the model.

Restraining the reach in our statistical models turns out to be unavoidable, with many quantities becoming statistically intractable without the presence of such a constraint. We show for instance in Chapter 4 (Theorem 4.2.10), that estimating the density in a pointwise fashion is impossible without constraining the reach. See for instance [AL19] for other impossibility results regarding the estimation of the support, tangent spaces and curvatures in unconstrained models.

In what follows, we will consider, under either (MH-0) or (MH- $\delta$ ), models of probability measures  $\Sigma_k^d$  such that any  $P \in \Sigma_k^d$  typically fulfils these three conditions:

- i) Its associated submanifold  $M$  is of dimension  $d$ ;
- ii) The reach of  $M$  is bounded from below by some constant;
- iii)  $M$  is of regularity  $\mathcal{C}^k$ .

We refer to Chapters 2 to 5 for precise definitions of such models. Their definitions vary

slightly from one chapter to another, but we will use one notation in this introduction for the sake of clarity.

### 1.3 Our results

In the remaining parts of this introduction, we will use the symbols  $\approx$ ,  $\lesssim$  and  $\gtrsim$  to denote equalities or inequalities up to constants and poly-log terms in  $n$ .

#### 1.3.1 Estimating the reach of a submanifold

Because the reach is such an important quantity in our geometric framework, it may seem relevant to be able to estimate it. To our knowledge, little has been done regarding that matter. In their pioneering work, [AKC<sup>+</sup>19] show that, under (MH-0),

$$n^{-\frac{1}{d}} \lesssim \inf_{\widehat{\text{rch}}} \sup_{P \in \Sigma_3^d} \mathbb{E}_{P^{\otimes n}} [|\widehat{\text{rch}} - \text{rch}(M)|] \lesssim n^{-\frac{2}{3d-1}}. \quad (1.2)$$

The estimator they proposed is based on an alternative definition of the reach, found in [Fed59, Thm 4.18]. Its rate of convergence is only optimal for  $d = 1$ , and leaves the door open for improvement in finding optimal estimators for arbitrary intrinsic dimensions and for support regularity  $k \geq 4$ .

**Contribution 1** Inspired by this challenge, we partly tackle this problem in

- ▷ [BHHS22] Clément Berenfeld, John Harvey, Marc Hoffmann, and Krishnan Shankar. Estimating the reach of a manifold via its convexity defect function. *Discrete & Computational Geometry*, 67(2):403–438, 2022,

by achieving the bounds (Theorems 2.2 and 2.35)

$$n^{-\frac{k-2}{d}} \lesssim \inf_{\widehat{\text{rch}}} \sup_{P \in \Sigma_k^d} \mathbb{E}_{P^{\otimes n}} [|\widehat{\text{rch}} - \text{rch}(M)|] \lesssim \begin{cases} n^{-\frac{1}{d}} & \text{for } k = 3, \\ n^{-\frac{k}{2d}} & \text{for } k \geq 4. \end{cases}$$

under (MH-0) or under (MH- $\delta$ ) provided that  $\delta \lesssim n^{-k/d}$ . The lower-bound is a generalization of the lower-bound (1.2) for  $k \geq 4$ , while the upper-bound is a substantial improvement of the upper-bound (1.2). The strategy is to take advantage of an elegant reformulation of the reach found in [AKC<sup>+</sup>19]:

$$\text{rch}(M) = \min \{ \text{curv}(M), \text{wfs}(M) \},$$

where  $\text{curv}(M)$  is the minimal curvature radius of  $M$ , and  $\text{wfs}(M)$ , is the *weak feature size* of  $M$ , a crucial geometric invariant introduced in [CL04]. The paired estimation of this two quantities go through the exploitation of fine properties of the *convexity defect*

function of  $M$ , as introduced in [ALS13]. The convexity defect function  $t \mapsto h_M(t)$  quantifies how far  $M$  is from being convex at scale  $t > 0$ , and captures many geometric features of  $M$ . For instance, we show (Proposition 2.12) that  $\text{curv}(M)$  appears in the development of  $h_M$  around zero:

$$h_M(t) = \frac{t^2}{2 \text{curv}(M)} + O(t^{4 \wedge k}).$$

We also show (Corollary 2.14), that  $t = \text{wfs}(M)$  is a discontinuity point for  $h_M$  whenever  $\text{wfs}(M) < \text{curv}(M)$ , see Figure 1.7 for an illustration of that phenomenon. These two properties makes the reach of  $M$  computable from its convexity defect function alone and enable a plug-in strategy for the estimation of the reach.

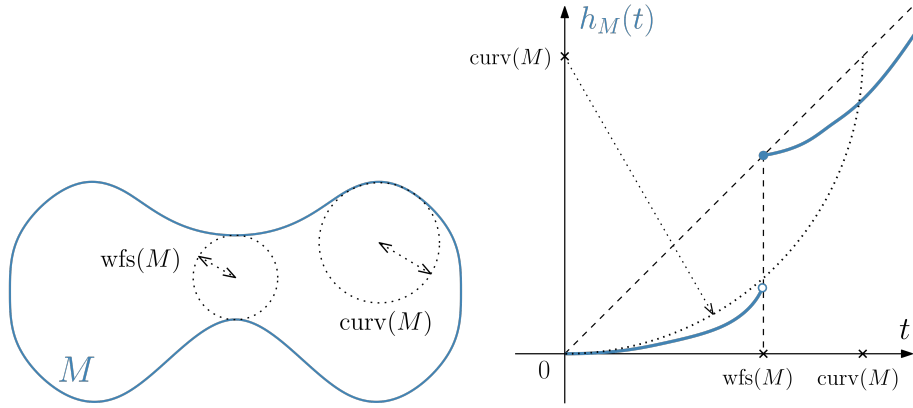


Figure 1.7 – (Left) A curve  $M$  and (Right) its convexity defect function  $h_M(t)$ , which, since  $\text{wfs}(M) < \text{curv}(M)$  in this diagram, presents a discontinuity at  $t = \text{wfs}(M)$ .

**Contribution 2** This last method however suffers two limitations. First, it fails at being optimal as soon as  $k \geq 5$ : this is because the estimator of  $\text{curv}(M)$  derived from the convexity defect function is suboptimal. Second, it relies on an estimator of the weak feature size, and this quantity is shown in Theorem 3.6 to be untractable statistically, making this approach less powerful. We circumvent these two disadvantages in

- ▷ [ABL22] Eddie Aamari, Clément Berenfeld, and Clément Levrard. Optimal reach estimation via metric learning. *In revision*, arXiv preprint arXiv:2207.06074, 2022,

by replacing the curvature estimator by an estimator stemming from the second fundamental form estimator of [AL19], and by proposing a new surrogate scale in place of the weak feature size. This new scale, coined the *spherical distortion radius* (Definition 3.13) and inspired by the work of [BLW19] and [CFM21], still carries important geometrical content regarding the geometry of the support. In short, this quantity

compare how the intrinsic distance over the support relates to the spherical distances, see Figure 1.8 for an illustration. The resulting estimator satisfies (Theorem 3.34)

$$n^{-\frac{k-2}{d}} \lesssim \inf_{\widehat{\text{rch}}} \sup_{P \in \Sigma_k^d} \mathbb{E}_{P^{\otimes n}} [|\widehat{\text{rch}} - \text{rch}(M)|] \lesssim n^{-\frac{k-2}{d}}$$

under (MH-0) or under (MH- $\delta$ ) provided that  $\delta \lesssim n^{-k/d}$ . This puts an end to the quest of an optimal estimation procedure for the reach.

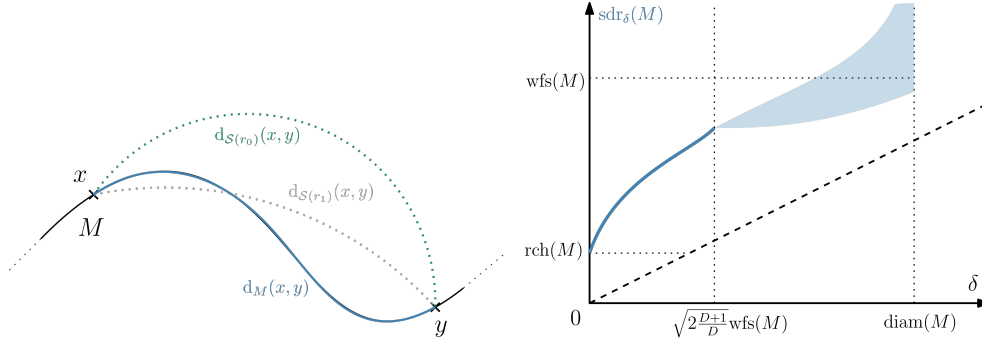


Figure 1.8 – (Left) Diagram of a submanifold  $M$  (a curve). In blue is the shortest path between two points  $x$  and  $y$ . In green (resp. grey) is the circle portion of radius  $r_0$  (resp.  $r_1$ ) going through  $x$  and  $y$ . The spherical distortion radius  $\text{sdr}_\Delta(M)$  find the biggest  $r$  such that  $d_M(x, y) \leq d_{S(r)}(x, y)$ , for all  $x, y \in M$  that are  $\Delta$ -apart. (Right) Plot of  $\Delta \mapsto \text{sdr}_\Delta(M)$ . We show (Proposition 3.14 and Theorem 3.22) that the SDR interpolates smoothly between the reach and the weak feature size of  $M$ .

Along the way, we also provide with optimal estimation bounds for the maximal curvature (Theorem 3.11), for the intrinsic metric (Theorems 3.25 and 3.28) and for the spherical distortion radius (Theorem 3.31). The optimality at each step of the process make our result adaptive to whether the value of the reach stems from a high curvature ( $\text{curv}(M) < \text{wfs}(M)$ ) or a bottleneck effect ( $\text{curv}(M) > \text{wfs}(M)$ ). On a submodel where the bottleneck effect predominates, we show that the same estimator automatically achieves the rate

$$n^{-\frac{k}{d}} \lesssim \inf_{\widehat{\text{rch}}} \sup_{P \in \Sigma_k^d} \mathbb{E}_{P^{\otimes n}} [|\widehat{\text{rch}} - \text{rch}(M)|] \lesssim n^{-\frac{k}{d}}.$$

### 1.3.2 Manifold density estimation

Density estimation is perhaps one of the most fundamental problems in nonparametric statistics. The stake is to find data-driven estimators of the underlying density from which the data is sampled, and to do it in a way that is optimal with respect to the regularity of the density. The practical purposes of density estimation are extremely diverse, ranging from descriptive and predictive analysis on spatial data

[Bit90, DNHTY14, And09] passing by a number of statistical learning tasks such as classification [CZ05], clustering [HG07], anomaly detection [RLSMG08], etc. See Figure 1.9 for a qualitative analysis of the density of a real-life data set in a manifold framework.

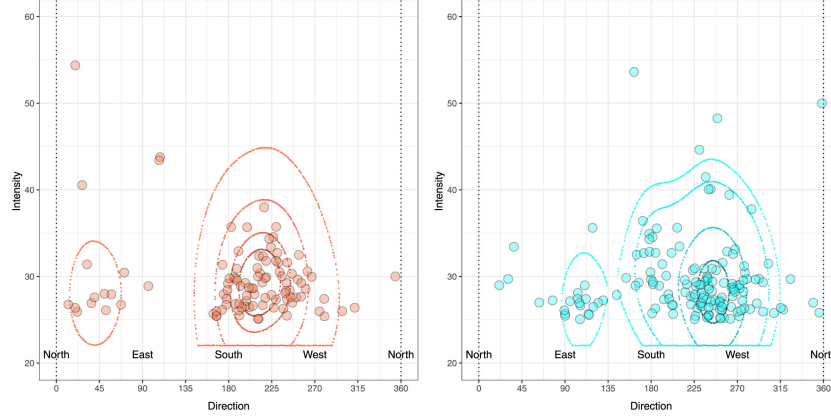


Figure 1.9 – Scatter plots of extreme wind measurements at the airport of Carrasco, Montevideo Uruguay during the hot season (Left) and the cold season (Right). The data lives on the cylinder  $S_1 \times \mathbb{R}$  because the first coordinate (the orientation of the wind) is periodic. In dotted lines are represented the level sets of the estimated densities, highlighting some crucial difference in the distributions of the winds between hot and cold seasons. Source: [CFM22, Fig 7].

As a long-established problem, density estimation has been studied in large and breadth in the multivariate, Euclidean case where the underlying probability distribution is absolutely continuous with respect the Lebesgue measure of the ambient space  $\mathbb{R}^D$ . Minimax estimators have been constructed for various regularity classes in this setting, see for instance textbooks [Tsy08, GN16], and adaptive procedures have been derived [LMS97, Low97, LMR17, BBM99, GL08, GN10], meaning that we can find estimators that do not depend on the underlying regularity of the density, but perform just as good as if they did. If the density is of regularity  $\beta_0 > 0$ , the minimax of estimation under reasonable losses has been shown to be  $n^{-\beta_0/(2\beta_0+D)}$ . This rate is reminiscent of the rate that we studied in Section 1.1, and suffers indeed from the curse of dimensionality, by becoming extremely slow as  $D$  grows large. That's enough to motivate an approach under the manifold hypothesis.

However, and surprisingly enough, little has been done in a manifold framework. Some estimation procedures exist in an abstract manifold setting [Hen90, Pel05] with sometimes involved adaptation results [KNP12], but these constructions always strongly rely on knowing the manifold perfectly, through the use of Fourier bases, needlet bases, volume density functions, etc.

Only recently have we seen the rise of results regarding density estimation in a framework where the supporting submanifold is unknown, see for instance [OG09, BS17] and the many new and exciting developments in the noiseless [WW20, Div21b, TY22] or noisy setting [HP21, CMG22], just to cite a few.

**Contribution 3** We define in

▷ [BH21] Clément Berenfeld and Marc Hoffmann. Density estimation on an unknown submanifold. *Electronic Journal of Statistics*, 15(1):2179–2223, 2021,

a new statistical model  $\Sigma_{k,\beta_0}^d$  on top of  $\Sigma_k^d$  by asking that the underlying density be  $\beta_0$ -Hölder in a sense defined in Chapter 4, Definition 4.2.9. For a pointwise loss, we find that, under (MH-0),

$$n^{-\frac{\beta_0}{2\beta_0+d}} \lesssim \inf_{\hat{f}(x_0)} \sup_{P \in \Sigma_{k,\beta_0}^d} \mathbb{E}_{P^{\otimes n}}[|\hat{f}(x_0) - f(x_0)|] \lesssim n^{-\frac{\beta_0}{2\beta_0+d}}, \quad (1.3)$$

provided that  $\beta_0 \leq k - 1$ , see Theorems 4.3.1 and 4.3.2. Notice that the ambient dimension  $D$  has been replaced by the intrinsic dimension  $d$  in the minimax rate, recovering the classical rate of estimating density in  $\mathbb{R}^d$ , and thus curbing the curse of dimensionality. This rate is achieved by a kernel estimator in the spirit of [Pel05, OG09] and only depends on the intrinsic dimension  $d$ , see Figure 1.10 for numerical study of the empirical performance of our estimator.

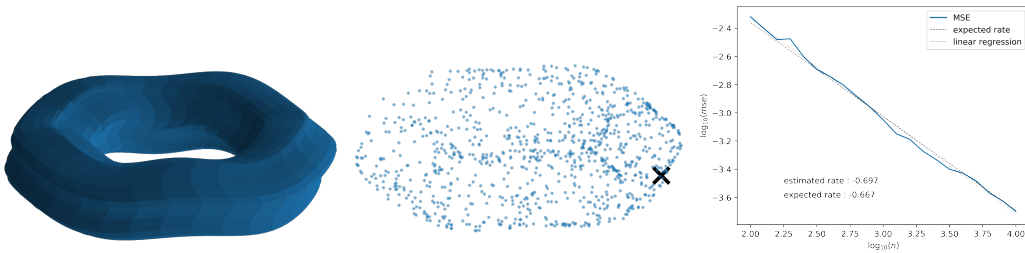


Figure 1.10 – (Left) A 2-dimensional submanifold in  $\mathbb{R}^3$ . It is endowed with a  $\beta_0$ -Hölder density with  $\beta_0 = 2$ , from which (Middle) a point cloud is sampled. (Right) The loss of the estimator is computed repeatedly for various value of  $n$ , and the averaged losses are represented in a double  $\log_{10}$ -scale, highlighting the rate found in (1.3).

A particular care has been given to the situation where the data is intrinsically one-dimensional. In this case, we build an estimator (Theorem 4.3.3) relying on an estimation of geodesic distances which achieves the rate (1.3) without the need of the constraint  $\beta_0 \leq k - 1$ .

Finally, using Lepski's method [LMS97] together with an estimator of the intrinsic dimension  $\widehat{d}$  yields a fully data-driven bandwidth  $\widehat{h}$  and an estimator which achieves the rates of (1.3) adaptively in  $d$  and  $\beta$ , see Theorem 4.3.4.

**Contribution 4** Another method considered in this manuscript is a Bayesian take on density estimation. The general idea behind Bayesian inference is to endow the space of probability measures on  $\mathbb{R}^D$  with some carefully chosen prior  $\Pi$ , and to update the prior, through the Bayes rule, as new data are observed. The updated prior is called the posterior distribution and is denoted  $\Pi(\cdot \mid X_1, \dots, X_n)$ . We will be interested in the concentration rates of the posterior distribution: it is the speed at which  $\Pi(\cdot \mid X_1, \dots, X_n)$  concentrates around  $P_0$  when the observations are sampled iid from  $P_0$ .

As in the frequentist world, Bayesian density estimation has been extensively studied in a Euclidean setting. From the trailblazing [Fer83] and [EW95], different methods have been proposed, such as using mixtures of Gaussians [GVDV07, KRVDV10, STG13], of Betas [Rou10], Gaussian processes [vdVvZ08], wavelet expansions [RR12], etc. And like before, the manifold case has received only little attention. We mention the existence of a Bayesian theory for abstract manifold developed in [CKP14] with concentration rates derived for density estimation, and the recent work of [MLD20] who tackle the problem by using mixture of Fisher-Gaussian kernels, but with no proven concentration rates.

We adress the problem of Bayesian density estimation in

- ▷ [BRR22] Clément Berenfeld, Paul Rosa, and Judith Rousseau. Estimating a density near an unknown manifold: a Bayesian nonparametric approach. *In revision*, ArXiv preprint arXiv:2205.15717, 2022.

by, first, introducing a new kind of priors that belong to the family of hybrid location-scale mixtures of Gaussians. These priors are flexible enough so that they can catch even very singular distributions (Theorem 5.3.7), while inheriting from the good contraction rates of the hybrid location-scale mixtures [NR17]. See Figure 1.11 for an illustration of the efficiency of these priors.

To quantify the rate of convergence on our family of priors, we build a model of densities under (MH- $\delta$ ) that are informally  $\beta_0$ -Hölder along the submanifold and  $\beta_1$ -Hölder normal to the manifold, see Definitions 5.2.1 and 5.2.3. Such a model encompasses in particular the real-life situation when a  $\beta_0$ -Hölder signal on the submanifold is blurred



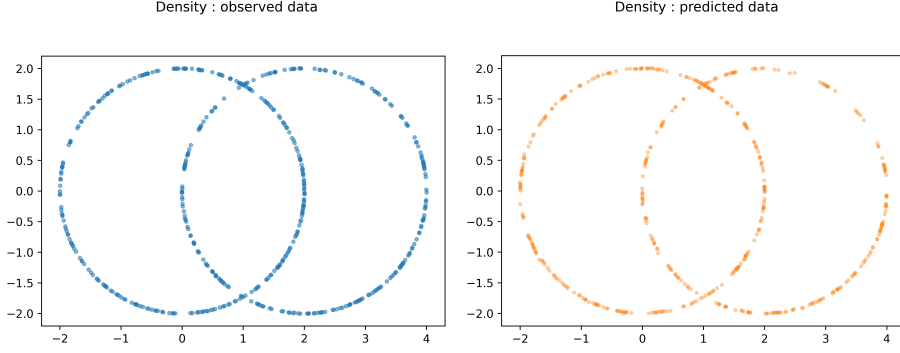


Figure 1.11 – (Left) 500 points drawn from an even mixture of uniform distributions on two circles with a noise of size  $\delta = 0.05$  and (Right) 500 points sampled from the (approximated) posterior distribution of our model. An asset of our method is that it can handle finite mixtures of manifold-supported probability distributions, therefore managing to capture densities supported on intricate geometric structures.

by an additive noise of regularity  $\beta_\perp$ . We obtain the concentration rate (Theorem 5.3.2)

$$\begin{aligned} \Pi(\|f - f_0\|_1 \geq \varepsilon_n \mid X_1, \dots, X_n) &\rightarrow 0 \quad \text{in } P_0^{\otimes n} \text{ - prob.} \\ \text{with } \varepsilon_n &\approx \frac{1}{\sqrt{n\delta^{\frac{D}{\alpha_0 - \alpha_\perp}}}} \vee n^{-\frac{\beta}{2\beta + D}}, \end{aligned}$$

where  $\beta$  is the effective regularity, defined through  $\frac{D}{\beta} = \frac{d}{\beta_0} + \frac{D-d}{\beta_\perp}$ , and where  $\alpha_0 = \beta/\beta_0$  and  $\alpha_\perp = \beta/\beta_\perp$ . For  $\delta$  not too small, this is the minimax rate for density estimation in a anisotropic setting. In particular, when the noise is assumed to be much more regular than the underlying  $\beta_0$ -Hölder density, the contraction rates become

$$\varepsilon_n \xrightarrow[\beta_\perp \rightarrow \infty]{\approx} \frac{1}{\sqrt{n\delta^d}} \vee n^{-\frac{\beta_0}{2\beta_0 + d}},$$

which, provided that  $\delta$  is not too small, is the minimax rate we found in (1.3) in a frequentist, noiseless setting. One of the strengths of this result is that our method does not depend on the knowledge of the regularities  $\beta_0$  and  $\beta_\perp$ , neither on the underlying manifold  $M$ , its dimension  $d$ , or the width of density  $\delta$ : the estimation is adaptive in these parameters.

### 1.3.3 Beyond density estimation

Density is a special case of a *depth*, a standard notion in multivariate analysis used as a mean to order multi-dimensional spaces — which typically lack unequivocal orderings as opposed to the real line — in a way that takes into account the underlying probability distribution. Typical depths includes the celebrated half-space depth [Tuk75], simplicial depths [Oja83], lens depths [KVL17, CFGM20], see [LSS06] for other notion of

depths. The density, also coined the *likelihood depth* [FM99], is a wonderful descriptor but its intrinsic locality makes it, as a depth, fail at taking into account some global features of the probability distribution.

**Contribution 5** Noticing that the density can be realized asymptotically as the rescaled degree of the vertices of a neighborhood graph built on top of the data, one could wonder what other kind of depths could emanate from taking large sample limits of other features of such graphs. This is the topic of

▷ [AACB21] Eddie Aamari, Ery Arias-Castro, and Clément Berenfeld. From graph centrality to data depth. *In revision*, arXiv preprint arXiv:2105.03122, 2021,

which aims at studying the asymptotic behavior of centrality measures [BE06] (to which the degree belongs) in random geometric graphs. The focus is made on two well-known notions of centrality, the H-index [Hir05] and the coreness [Sei83]. We prove that they converge (Theorems 6.3.5 and 6.4.12) towards their asymptotic counterparts, called the *continuous* H-index and coreness. The latter forms a spectrum of new depths which interpolate between the very local likelihood depth and the much more global continuous coreness. See Figure 1.12 for an illustration of this result.

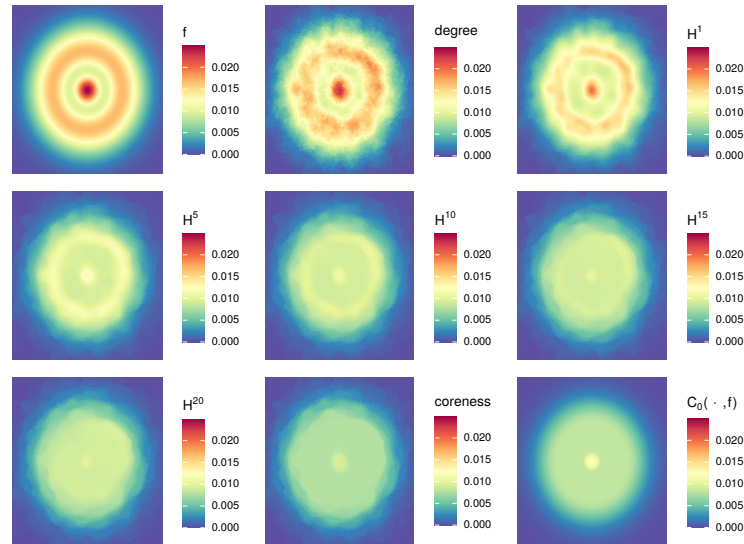


Figure 1.12 – (Top Left) A density function  $f$  on  $\mathbb{R}^2$  and (Bottom Right) its associated continuum coreness. In between are displayed the centralities (iterated H-indices and coreness) computed in the neighborhood graph built on top of  $n = 20000$  points sampled from  $f$ , showing a clear interpolation between  $f$  and its continuum coreness.

## 1.4 Organization of the manuscript

The rest of the manuscript is organized in chapters, whose content is taken from the aforementioned articles, written in the course of this thesis. The chapters, which are summarized below, are functioning individually and are fully independent notation-wise, and can therein be read in no particular order — although we might suggest that Chapter 2 be read before Chapter 3, and Chapter 4 before Chapter 5. There might be slight redundancies from one chapter to the others, for which we apologize in advance.

**Chapter 2** is based on [BHHS22]. It focuses on the problem of estimating the reach of a submanifold through the analysis of its convexity defect function. Non-asymptotic rates of estimation are derived on generic  $\mathcal{C}^k$  models, and are shown to be optimal in the very particular case of  $k = 3$  and  $k = 4$ , but not for  $k \geq 5$ .

**Chapter 3** is based on [ABL22] and bridges the gap left by the previous chapter by exhibiting a minimax optimal estimator of the reach under any smoothness assumption. It revolves around a new notion of geometric scale involving metric distortion. Rates of convergence are derived for this new scale and, in the process, we obtain new bounds for geodesic length estimation.

**Chapter 4** is based on [BH21], and tackle the problem of estimating a density that is singularly supported on a submanifold, in a pointwise fashion. We recover the usual minimax rate of estimating Hölder densities and we provide an adaptive kernel-based estimator with a data-driven bandwidth selection procedure.

**Chapter 5** is based on [BRR22] and lays some theoretical grounds for Bayesian analysis of density supported near submanifolds. We define a new notion of regularity for such densities and study the posterior contraction rates on these regularity classes for a new kind of hybrid location-scale Dirichlet process mixtures of Gaussians.

**Chapter 6** is based on [AACB21]. It studies the large-sample limits of well-known centrality measures in neighborhood graphs, with a focus on the H-index and the coreness.



## Chapter 2

# Reach estimation via the convexity defect function

The reach of a submanifold is a crucial regularity parameter for manifold learning and geometric inference from point clouds. This chapter relates the reach of a submanifold to its convexity defect function. Using the stability properties of convexity defect functions, we propose an estimator for the reach, based on a plug-in of an estimator of the support. A uniform expected loss bound over a  $\mathcal{C}^k$  model is found. Lower bounds for the minimax rate for estimating the reach over these models are also provided. The estimator almost achieves these rates in the  $\mathcal{C}^3$  and  $\mathcal{C}^k$  cases, with a gap given by a logarithmic factor. This chapter has been published in [BHHS22].

### Contents

---

<b>2.1</b>	<b>Introduction . . . . .</b>	<b>20</b>
2.1.1	Motivation . . . . .	20
2.1.2	Main results . . . . .	21
2.1.3	Organization of the paper . . . . .	22
<b>2.2</b>	<b>Geometry of the reach . . . . .</b>	<b>24</b>
2.2.1	The weak feature size . . . . .	24
2.2.2	The local reach . . . . .	25
<b>2.3</b>	<b>Geometrical framework . . . . .</b>	<b>26</b>
<b>2.4</b>	<b>Convexity defect functions . . . . .</b>	<b>27</b>
2.4.1	Upper bounds on the convexity defect function . . . . .	29
2.4.2	The convexity defect function near zero . . . . .	33
<b>2.5</b>	<b>Approximating the reach . . . . .</b>	<b>38</b>
<b>2.6</b>	<b>Minimax rates for reach estimators: Upper bounds . . . . .</b>	<b>41</b>
2.6.1	Estimating the local reach . . . . .	43
2.6.2	Estimating the global reach . . . . .	44
<b>2.7</b>	<b>Minimax rates for reach estimators: Lower bounds . . . . .</b>	<b>48</b>

---

## 2.1 Introduction

### 2.1.1 Motivation

The reach of a submanifold  $M \subseteq \mathbf{R}^D$  is a geometric invariant which measures how tightly the submanifold folds in on itself. Dating back to Federer [Fed59], it encodes both local curvature conditions as well as global ‘bottlenecks’ arising from two regions of the manifold that are far apart in the manifold’s intrinsic metric but are close in the ambient Euclidean metric. The reach is a key regularity parameter in the estimation of other geometric information. Methods and algorithms from topological data analysis often use the reach as a ‘tuning parameter’. The correctness of their results depends on setting this parameter correctly.

Statistical inference from point clouds has become an active area. In a probabilistic framework, a *reach condition*, meaning that the reach of the submanifold under study is bounded below, is usually necessary in order to obtain minimax inference results in manifold learning. These include: homology inference [NSW08, BRS<sup>+</sup>12], curvature [AL19], reach estimation itself [AKC<sup>+</sup>19] as well as manifold estimation [GPPVW12a, KRW16, AL19]. In this context, there is a risk of algorithms being applied as ‘black boxes’ without attention to their underlying assumptions. Efficient reach estimation would be a vital addition to this field, providing a so-called sanity test of other results.

In this direction, Aamari, Kim *et al.* paved the way: in [AKC<sup>+</sup>19], under some specific assumptions, an estimator of the reach has been proposed and studied when the observation is an  $n$ -sample of a smooth probability distribution supported on an unknown  $d$ -dimensional submanifold  $M$  of a Euclidean space  $\mathbf{R}^D$  together with the tangent spaces at each sampled point. For certain types of  $C^3$ -regularity models, the estimator, based on a representation of the reach in terms of points of  $M$  and its tangent spaces (Theorem 4.18 in [Fed59]) achieves the rate  $n^{-2/(3d-1)}$ . A lower bound for the minimax rate of convergence is given by  $n^{-1/d}$ . In the special case when the reach of  $M$  is attained at a bottleneck, the algorithm in [AKC<sup>+</sup>19] achieves this rate. However, in general, one does not know whether this condition is satisfied *a priori*.

In this paper, we continue the study of reach estimation by taking a completely different route: we use the relationship between the reach of a submanifold of  $\mathbf{R}^D$  and its *convexity defect function*. This function was introduced by Attali, Lieutier and Salinas in [ALS13] and measures how far a (bounded) subset  $X \subseteq \mathbf{R}^D$  is from being convex at a given scale. It is a powerful geometric tool that has other applications such as manifold reconstruction, see the recent work by Divol [Div21a]. By establishing certain new quantitative properties of the convexity defect function of a submanifold  $M \subseteq \mathbf{R}^D$  that relate to both its curvature and bottleneck properties, we show that the

convexity defect function can be used to compute the reach of a submanifold. From this we obtain a method which transforms an estimator of  $M$ , along with information on its error, into a new estimator of the reach.

The recent results of Aamari and Levrard in [AL19] provide an estimator of  $M$  which is optimal, to within logarithmic terms. Transforming this into an estimator of the reach, we obtain new convergence results over general  $C^k$ -regularity models ( $k \geq 3$ ). These rates improve upon the previous work of [AKC<sup>+</sup>19]. By establishing lower bounds for the minimax rates of convergence, we prove that our results are optimal up to logarithmic terms in the cases  $k = 3$  and  $k = 4$ .

### 2.1.2 Main results

We present here one of several possible definitions of the reach. Given a submanifold  $M \subseteq \mathbf{R}^D$ , consider its  $\delta$ -offset given by the open set  $M^\delta \subseteq \mathbf{R}^D$ , where

$$M^\delta = \bigcup_{p \in M} B_\delta(p).$$

Here  $B_\delta(p)$  denotes the open Euclidean ball centered at  $p$  with radius  $\delta$ . For small enough  $\delta$  (a uniform choice for such  $\delta$  exists in general only when  $M$  is compact), one has the property that for all  $y \in M^\delta \setminus M$ , there is a unique straight line from  $y$  to a point in  $M$  realizing the distance from  $y$  to  $M$ . In other words, the metric projection  $\pi: M^\delta \rightarrow M$  is well defined.

**Definition 2.1** (Federer [Fed59]). *The reach of a submanifold  $M$  is*

$$\sup \left\{ \delta \geq 0 : \text{The nearest point projection } \pi: M^\delta \rightarrow M \text{ is well defined} \right\}.$$

*We denote the reach by  $R(M)$  or simply  $R$  when the context is clear.*

Other equivalent characterizations of the reach exist. For example, in Section 2.4.1 below, we use the characterization from Theorem 4.18 in [Fed59]. More recently Theorem 1 in [BLW19] defined the reach in terms of the metric distortion.

Our main results are obtained for a statistical model which imposes certain standard regularity conditions on the manifolds being considered, requires that they are compact and connected, and also imposes conditions on the distributions being considered which have support on those manifolds. The set of distributions satisfying these constraints on  $C^k$  manifolds is denoted in the results below by  $\mathcal{P}^k$  and these constraints are elaborated upon in Sections 2.3 and 2.6.

**Theorem 2.2.** *For  $d$ -dimensional submanifolds of regularity  $C^k$  with  $k \geq 3$ , and for sufficiently large  $n$ , there exists an estimator  $\hat{R}$  explicitly constructed in Section 2.6 below*

that satisfies

$$\sup_{P \in \mathcal{P}^k} \mathbf{E}_{P^{\otimes n}}[|\widehat{R} - R|] \leq C \begin{cases} \left(\frac{\log(n)}{n-1}\right)^{1/d} & k = 3 \\ \left(\frac{\log(n)}{n-1}\right)^{k/(2d)} & k \geq 4, \end{cases}$$

where  $\widehat{R}$  denotes an estimator of the reach  $R = R(\mathbf{M})$  constructed from an  $n$ -sample  $(X_1, \dots, X_n)$  of independent random variables with common distribution  $P \in \mathcal{P}^k$ . The quantity  $C > 0$  depends on  $d, k$  and the regularity parameters that define the class  $\mathcal{P}^k$  and the notation  $\mathbf{E}_{P^{\otimes n}}[\cdot]$  refers to the expectation operator under the distribution  $P^{\otimes n}$  of the  $n$ -sample  $(X_1, \dots, X_n)$ .

We also provide a lower bound for the minimax convergence rate. In case  $k = 3, 4$ , our estimators are almost optimal, with a gap given by a  $\log(n)$  factor.

**Theorem 2.3.** *For certain values of the regularity parameters (depending only on  $d$  and  $k$ ), then*

$$\inf_{\widehat{R}} \sup_{P \in \mathcal{P}^k} \mathbf{E}_{P^{\otimes n}}[|\widehat{R} - R|] \geq cn^{-(k-2)/d},$$

where the infimum is taken over all the estimators  $\widehat{R} = \widehat{R}(X_1, \dots, X_n)$  and  $c > 0$  depends on  $d, k$  and the regularity parameters.

These results are of an entirely theoretical nature. The question of practical implementation remains, although it is not of primary interest for the paper. Starting from a point cloud  $X_1, \dots, X_n$ , one may implement the following protocol:

- Estimate  $\mathbf{M}$  from the data  $X_1, \dots, X_n$  by the best available manifold reconstruction method  $\widehat{\mathbf{M}}$ , or, indeed, by any other method.
- Compute  $h_{\widehat{\mathbf{M}}}$  (Definition 2.10) and derive  $\widehat{R}$  thanks to Definition 2.33.

The only inputs are  $\widehat{\mathbf{M}}$  and the regularity parameters that define the class  $\mathcal{P}^k$ . It is a common practice in statistics to assume some prior knowledge of the class in order to constrain the problem. However, the quantities  $C_{d,k,R_{\min}}$  and  $C$  in Theorem 2.34 are unknown, which creates difficulties in deriving the accuracy of the estimator  $\widehat{R}$  and, for example, calculating a confidence interval. This is common to every minimax result and could in practice be treated by estimating the variance of the estimator via any conventional computational method such as the bootstrap [ET94]. A more detailed discussion lies outside the scope of the present paper.

### 2.1.3 Organization of the paper



The paper is divided into two halves: a first half that is mainly geometric in flavor and a second half which employs mainly statistical techniques. To that end Sections 2.2, 2.3 and 2.4 describe the geometric setting of this paper in some detail, Section 2.5 discusses the approximation of the reach in a deterministic setting, while Sections 2.6 and 2.7 are devoted to showing that the new algorithm proposed to estimate the reach achieves the rates stated in Theorem 2.2 and to the proof of the lower bound for the minimax rate stated in Theorem 2.3.

Section 2.2: We elaborate on the geometry of the reach. We recall a dichotomy due to Aamari, Kim *et al.* [AKC<sup>+</sup>19] in Theorem 2.4 and we study in particular the distinction between *global reach* or *weak feature size* in Definition 2.5 and the *local reach* in Definition 2.6, according to the terminology of [AKC<sup>+</sup>19]. This is not apparent in the classical Definition 2.1 of Federer.

Section 2.3: A geometrical framework is given for studying reach estimation. We describe precisely a class  $\mathfrak{C}_{R_{\min}, \mathbf{L}}^k$  of submanifolds, following Aamari and Levrard [AL19]. Manifolds  $M$  in this class admit a local parametrization at all points  $p \in M$  by the tangent space  $T_p M$ , which is the inverse of the projection to the tangent space and satisfies certain  $\mathcal{C}^k$  bounds.

Section 2.4: This section is devoted to the study of the *convexity defect function*  $h_M$  of  $M$  as introduced in [ALS13] and its properties. We show how the local reach can be calculated from the values of  $h_M$  near the origin in Proposition 2.12 and how the weak feature size (the global reach) appears as a discontinuity point of  $h_M$  whenever it is smaller than the local reach. This is done by proving an upper bound on  $h_M$  in Proposition 2.13. Proposition 2.12 and 2.13 are central to the results of the paper.

Section 2.5: When we attempt to estimate the reach in later sections, we will not know  $M$  exactly. Instead, we will know it up to some statistical error coming from an estimator. Propositions 2.24 and 2.26 give approximations of the local reach and the weak feature size, respectively, calculated from some proxy  $\widetilde{M}$ . The errors of the approximations are given in terms of the Hausdorff distance  $H(M, \widetilde{M})$ .

Section 2.6: Building on the definitions in Section 2.3, a statistical framework is described within which we study reach estimation in a minimax setting. This defines a class  $\mathcal{P}^k$  of admissible distributions  $P$  over their support  $M$ , the submanifold of interest, which belongs to the class  $\mathfrak{C}_{R_{\min}, \mathbf{L}}^k$ . To apply the results of the previous section, we may use the Aamari–Levrard estimator [AL19]  $\widehat{M}$  of  $M$  from a sample  $(X_1, \dots, X_n)$  as the proxy  $\widetilde{M}$  for  $M$ . This estimator is almost optimal over the class  $\mathcal{P}^k$ . This yields estimators of the local reach and finally of the reach  $R(M)$  in Section 2.6. We then prove the upper bounds announced in Theorem 2.2 above in Theorems 2.31–2.34.

Section 2.7: Using the classical Le Cam testing argument we obtain minimax lower bounds as announced in Theorem 2.3.

## 2.2 Geometry of the reach

The reach of a submanifold  $M$ , which we will denote by  $R(M)$ , or simply  $R$ , is an unusual invariant. Definition 2.1 conceals what is almost a dichotomy – the reach of a submanifold can be realised in two very different ways. This is made precise by the following result.

**Theorem 2.4.** [AKC<sup>+</sup>19, Theorem 3.4] *Let  $M \subseteq \mathbb{R}^D$  be a compact submanifold with reach  $R(M) > 0$ . At least one of the following two assertions holds.*

- (Global case)  *$M$  has a bottleneck, that is, there exist  $q_1, q_2 \in M$  such that  $(q_1 + q_2)/2 \in \text{Med}(M)$  and  $\|q_1 - q_2\| = 2R(M)$ .*
- (Local case) *There exists  $q_0 \in M$  and an arc-length parametrized geodesic  $\gamma$  such that  $\gamma(0) = q_0$  and  $\|\gamma''(0)\| = 1/R(M)$ .*

Here,  $\text{Med}(M)$  is the medial axis of  $M$ , that subset of  $\mathbb{R}^D$  on which the nearest point projection on  $M$  is ill-defined, namely

$$\text{Med}(M) = \{z \in \mathbb{R}^D \mid \exists p, q \in M, p \neq q, d(z, p) = d(z, q) = d(z, M)\}.$$

We say that the result above is only ‘almost’ a dichotomy because it is possible for both conditions to hold simultaneously. The curve  $\gamma$  could be one half of a circle with radius  $R(M)$  joining  $q_1$  and  $q_2$ , for example, in which case the term ‘bottleneck’ might be considered a misnomer, or the points  $q_1$  and  $q_2$  might not lie on  $\gamma$  at all, so that the two assertions hold completely independently.

This situation invites us to consider two separate invariants. One, the *weak feature size*,  $R_{\text{wfs}}$ , is a widely studied invariant encoding large scale information such as bottlenecks. The second, which we will call the *local reach*,  $R_\ell$ , following [AKC<sup>+</sup>19], will encode curvature information. Theorem 2.4 states that the minimum of these two invariants is the reach,

$$R = \min \{R_\ell, R_{\text{wfs}}\}.$$

Note that, in Riemannian geometry, the local reach is referred to as the *focal radius* of  $M$ , while the reach itself is often referred to as the *normal injectivity radius* of  $M$ .

### 2.2.1 The weak feature size

The weak feature size is defined in terms of critical points of the distance function from  $M$  (in the sense of Grove and Shiohama; see for instance [Gro94], p. 360).

Consider the function,  $d_M : \mathbb{R}^D \rightarrow \mathbb{R}$  defined by  $d_M(y) = \inf_{p \in M} \|y - p\|$ . Note that  $M = d_M^{-1}(0)$ . Following [ALS13], let  $\Gamma_M(y) = \{x \in M : d_M(y, M) = \|x - y\|\}$ , i.e., those  $x$  in  $M$  realizing the distance between  $y$  and  $M$ . Then we define a generalized gradient as

$$\nabla_M(y) := \frac{y - \text{Center}(\Gamma_M(y))}{d_M(y, M)},$$

where  $\text{Center}(\sigma)$  is defined as the center of the smallest (Euclidean) ball enclosing the bounded subset  $\sigma \subseteq \mathbf{R}^D$ . This generalized gradient  $\nabla_{\mathbf{M}}$  for  $d_{\mathbf{M}}$  coincides with the usual gradient where  $d_{\mathbf{M}}$  is differentiable. We say that a point  $y \in \mathbf{R}^D \setminus \mathbf{M}$  is a *critical point* of  $d_{\mathbf{M}}$  if  $\nabla_{\mathbf{M}}(y) = \mathbf{0}$ .

For example, if  $y$  is the midpoint of a chord the endpoints of which meet the submanifold perpendicularly, then from  $y$  there are two shortest paths to  $\mathbf{M}$  which travel in opposite directions. It follows that  $y$  is a critical point.

**Definition 2.5.** *Given a submanifold  $\mathbf{M}$  of  $\mathbf{R}^D$  let  $\mathcal{C}$  denote the set of critical points of the distance function  $d_{\mathbf{M}}$ . The weak feature size, denoted  $R_{\text{wfs}}(\mathbf{M})$  or simply  $R_{\text{wfs}}$ , is then defined as  $R_{\text{wfs}} := \inf \{d_{\mathbf{M}}(y) : y \in \mathcal{C}\}$ .*

By Theorem 2.4, if the reach is realised globally then the first critical point will be the midpoint of a shortest chord which meets  $\mathbf{M}$  perpendicularly at both ends, and so the weak feature size is equal to the reach.

### 2.2.2 The local reach

In the local case, Theorem 2.4 tells us that the reach is determined by the maximum value of  $\|\gamma''\|$  over all arc-length parametrised geodesics  $\gamma$ . This can be formulated more concisely by considering instead the *second fundamental form*,  $\Pi$ , which measures how the submanifold  $\mathbf{M}$  curves in the ambient Euclidean space  $\mathbf{R}^D$ . We refer the reader to a standard text in Riemannian geometry such as [DCFF92] for a precise definition of the second fundamental form. Informally, the second fundamental form is defined as follows. For a pair of vector fields tangent to  $\mathbf{M}$ , the (Euclidean) derivative of one with respect to the other is not usually tangent to  $\mathbf{M}$ . In fact, the tangential component is the Levi–Civita connection of the induced (Riemannian) metric on  $\mathbf{M}$ . The normal, or perpendicular, component yields a symmetric, bilinear form, namely, the second fundamental form, denoted by  $\Pi_p$ . In particular, if the norm of  $\Pi_p$  is small then  $\mathbf{M}$  is nearly flat near  $p$  and if the norm is large then it is an area of high curvature.

**Definition 2.6.** *Given a submanifold  $\mathbf{M}$  of  $\mathbf{R}^D$  let  $\Pi_p$  denote the second fundamental form at  $p \in \mathbf{M}$ . The local reach of  $\mathbf{M}$ , denoted  $R_{\ell}(\mathbf{M})$  or simply  $R_{\ell}$  is the quantity*

$$R_{\ell} = \inf_{p \in \mathbf{M}} \left\{ \frac{1}{\|\Pi_p\|_{\text{op}}} \right\}.$$

We use the term ‘local reach’ here to reflect the fact that this quantity is generated entirely by the local geometry. In differential geometry literature the local reach is referred to as the *focal radius* of the submanifold.

## 2.3 Geometrical framework

We define a class of manifolds which are suitable for the task of reach estimation. This class is the same as that considered by Aamari and Levrard [AL19] for other problems in minimax geometric inference. The class is that of  $\mathcal{C}^k$  submanifolds, but with some additional regularity requirements. These guarantee the existence of a Taylor expansion of the embedding of the submanifold with bounded co-efficients, as well as a uniform lower bound on the reach.

**Definition 2.7.** (see [AL19]) For two fixed natural numbers  $d < D$  and for some  $k \geq 3$ ,  $R_{\min} > 0$ , and  $\mathbf{L} = (L_{\perp}, L_3, \dots, L_k)$ , we let  $\mathfrak{C}_{R_{\min}, \mathbf{L}}^k$  denote the set of  $d$ -dimensional, compact, connected submanifolds  $\mathbf{M}$  of  $\mathbf{R}^D$  such that:

$$(i) \ R(\mathbf{M}) \geq R_{\min};$$

(ii) For all  $p \in \mathbf{M}$ , there exists a local one-to-one parametrization  $\psi_p$  of the form:

$$\begin{aligned} \psi_p: B_{T_p \mathbf{M}}(0, r) &\subseteq T_p \mathbf{M} \rightarrow \mathbf{M}, \\ v &\mapsto p + v + \mathbf{N}_p(v) \end{aligned}$$

for some  $r \geq \frac{1}{4L_{\perp}}$ , with  $\mathbf{N}_p \in \mathcal{C}^k(B_{T_p \mathbf{M}}(0, r), \mathbf{R}^D)$  such that

$$\mathbf{N}_p(0) = 0, \quad d_0 \mathbf{N}_p = 0, \quad \|d_v^2 \mathbf{N}_p\|_{\text{op}} \leq L_{\perp},$$

for all  $\|v\| \leq \frac{1}{4L_{\perp}}$ ;

(iii) The differentials  $d_v^i \mathbf{N}_p$  satisfy  $\|d_v^i \mathbf{N}_p\|_{\text{op}} \leq L_i$  for all  $3 \leq i \leq k$  and  $\|v\| \leq \frac{1}{4L_{\perp}}$ .

We define subclasses of  $\mathfrak{C}_{R_{\min}, \mathbf{L}}^k$  as follows, using the gap  $R_{\ell} - R_{\text{wfs}}$  between the weak feature size and the local reach. For fixed values of  $R_{\min}$  and  $\mathbf{L}$ , we define

$$\mathcal{M}_0^k = \{\mathbf{M} \in \mathfrak{C}_{R_{\min}, \mathbf{L}}^k \mid R_{\text{wfs}}(\mathbf{M}) \geq R_{\ell}(\mathbf{M})\}$$

and

$$\mathcal{M}_{\alpha}^k = \{\mathbf{M} \in \mathfrak{C}_{R_{\min}, \mathbf{L}}^k \mid R_{\text{wfs}}(\mathbf{M}) \leq R_{\ell}(\mathbf{M}) - \alpha\}, \quad \alpha > 0.$$

Note that

$$\mathfrak{C}_{R_{\min}, \mathbf{L}}^k = \cup_{\alpha \geq 0} \mathcal{M}_{\alpha}^k.$$

Manifolds in  $\mathfrak{C}_{R_{\min}, \mathbf{L}}^k$  admit a second parametrization, one that represents the manifold locally as the graph of a function over the tangent space so that the first non-zero term in the Taylor expansion is of degree two and is given by the second fundamental form. These parametrizations in general satisfy weaker bounds than  $\mathbf{L}$ . The degree  $k$  Taylor polynomial then gives an algebraic approximation of the manifold, which will be very useful in later calculations. The following lemma from [AL19] describes the Taylor expansion of a local parametrization at every point  $p \in \mathbf{M}$ .

**Lemma 2.8.** [AL19, Lemma 2] Let  $k \geq 3$ ,  $\mathbf{M} \in \mathfrak{C}_{R_{\min}, \mathbf{L}}^k$  and  $r = \frac{1}{4} \min\{R_{\min}, L_{\perp}^{-1}\}$ . Then for all  $p \in \mathbf{M}$  there is a local one-to-one parametrization around  $p$ ,  $\Phi_p : U \rightarrow \mathbf{M}$ , for some  $U \subset T_p \mathbf{M}$ , which contains  $B(p, r) \cap \mathbf{M}$  in its image, satisfies  $\text{pr}_{T_p \mathbf{M}} \circ \Phi_p(v) = v$  on its domain, and takes the form

$$\Phi_p(v) = p + v + \frac{1}{2}T_2(v^{\otimes 2}) + \frac{1}{6}T_3(v^{\otimes 3}) + \dots + \frac{1}{(k-1)!}T_{k-1}(v^{\otimes(k-1)}) + \mathcal{R}_k(v),$$

where  $\|\mathcal{R}_k(v)\| \leq C\|v\|^k$ . Furthermore  $T_2 = \Pi_p$  and  $\|T_i\|_{\text{op}} \leq L'_i$ , where  $L'_i$  and  $C$  depends on  $d, k, R_{\min}$  and  $\mathbf{L}$ , and the terms  $T_2, \dots, T_{k-1}, \mathcal{R}_k$  are all normal to  $T_p \mathbf{M}$ .

**Definition 2.9.** We call the degree  $j$  truncation of the parametrization  $\Phi_p$  given in Lemma 2.8 the approximation of degree  $j$  to  $\mathbf{M}$  around  $p$  and write it

$$\Phi_p^j(v) = p + v + \frac{1}{2}T_2(v^{\otimes 2}) + \frac{1}{6}T_3(v^{\otimes 3}) + \dots + \frac{1}{j!}T_j(v^{\otimes j}).$$

## 2.4 Convexity defect functions

The convexity defect function, originally introduced by Attali, Lieutier and Salinas [ALS13], measures how far a subset  $\mathbf{X} \subseteq \mathbf{R}^D$  is from being convex at scale  $t$ . The goal of this section is to establish a relationship between the convexity defect function and the reach. The definition is valid for any compact subset of  $\mathbf{R}^D$ . In this section we will principally consider the case of a closed submanifold  $\mathbf{M}$  as before, but in the sequel we will need to know that this function can be defined in greater generality.

We recall the definition. Given a compact subset  $\sigma \subseteq \mathbf{X}$ , it is contained in a smallest enclosing closed ball in  $\mathbf{R}^D$ . We define  $\text{Rad}(\sigma)$  to be the radius of this ball. We denote by  $\text{Hull}(\sigma)$  the convex hull of  $\sigma$  in  $\mathbf{R}^D$ . Then we define the *convex hull of  $\mathbf{X}$  at scale  $t$*  to be the following subset of  $\mathbf{R}^D$ :

$$\text{Hull}(\mathbf{X}, t) = \bigcup_{\substack{\sigma \subseteq \mathbf{X} \\ \text{Rad}(\sigma) \leq t}} \text{Hull}(\sigma).$$

For two compact subsets  $A$  and  $B$  of  $\mathbf{R}^D$ , we define the asymmetric distance  $H(A|B) = \sup_{a \in A} d(a, B)$  so that  $H(A, B) = \max\{H(A|B), H(B|A)\}$  is the symmetric Hausdorff distance.

**Definition 2.10.** Given a compact subset  $\mathbf{X} \subseteq \mathbf{R}^D$ , we define the convexity defect function  $h_{\mathbf{X}} : \mathbf{R}_{\geq 0} \rightarrow \mathbf{R}_{\geq 0}$  by  $h_{\mathbf{X}}(t) = H(\text{Hull}(\mathbf{X}, t), \mathbf{X}) = H(\text{Hull}(\mathbf{X}, t) | \mathbf{X})$ .

We recall here from [ALS13] some useful properties of  $h_{\mathbf{X}}$ .

1.  $h_{\mathbf{X}}(0) = 0$ .
2.  $h_{\mathbf{X}}$  is non-decreasing on the interval  $[0, \text{Rad}(\mathbf{X})]$  and constant thereafter.

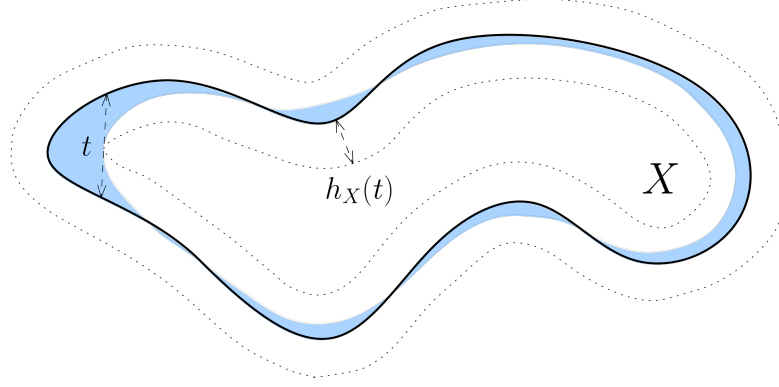


Figure 2.1 – The convex hull at scale  $t$ ,  $\text{Hull}(\mathbf{X}, t)$  (in blue), of a curve  $\mathbf{X}$  (in black). Enclosed between the dotted curves is the minimal tubular neighborhood around  $\mathbf{X}$  that contains  $\text{Hull}(\mathbf{X}, t)$  — its width is the convexity defect function  $h_{\mathbf{X}}(t)$ .

3. If  $\tilde{\mathbf{X}} \subseteq \mathbf{R}^D$  satisfies  $H(\mathbf{X}, \tilde{\mathbf{X}}) < \epsilon$ , where  $H$  is the Hausdorff distance, then  $h_{\tilde{\mathbf{X}}}(t - \epsilon) - 2\epsilon \leq h_{\mathbf{X}}(t) \leq h_{\tilde{\mathbf{X}}}(t + \epsilon) + 2\epsilon$  for any  $t \geq \epsilon$ .
4.  $h_{\mathbf{X}}(t) \leq t$  for all  $t \geq 0$ . Moreover,  $h_{\mathbf{X}}(t_0) = t_0$  if and only if  $t_0$  is a critical value of the distance function,  $d_{\mathbf{X}}$ .
5. If the reach,  $R = R(\mathbf{X}) > 0$ , then on  $[0, R)$  the function  $h_{\mathbf{X}}(t)$  is bounded above by a quarter-circle of radius  $R$  centered on  $(0, R)$ . In other words,  $h_{\mathbf{X}}(t) \leq R - \sqrt{R^2 - t^2}$  for  $t \in [0, R)$ .

From item 4 and the definition of the weak feature size in terms of critical points of the distance function, the following proposition is immediate.

**Proposition 2.11.** *If  $\mathbf{M}$  is a submanifold of  $\mathbf{R}^D$  then  $R_{\text{wfs}} = \inf \{t > 0 : h_{\mathbf{M}}(t) = t\}$ .*

We can also relate the local reach to the convexity defect function with the following proposition, which we will prove in Section 2.4.2.

**Proposition 2.12.** *Let  $k \geq 4$ . There exists a constant  $C$  (depending on  $R_{\min}, \mathbf{L}, d$  and  $k$ ) such that, for any sufficiently small non-negative real  $t$ ,  $t \leq t_{R_{\min}, \mathbf{L}, d, k}$ , and any  $\mathbf{M} \in \mathfrak{C}_{R_{\min}, \mathbf{L}}^k$ , we have*

$$\left| h_{\mathbf{M}}(t) - \frac{t^2}{2R_{\ell}} \right| \leq Ct^4.$$

*In case  $k = 3$ , there exists a constant  $C'$  (depending on  $R_{\min}, \mathbf{L}, d$ ) such that, for any sufficiently small non-negative real  $t$ ,  $t \leq t_{R_{\min}, \mathbf{L}, d}$ , and any  $\mathbf{M} \in \mathfrak{C}_{R_{\min}, \mathbf{L}}^k$ , we have*

$$\left| h_{\mathbf{M}}(t) - \frac{t^2}{2R_{\ell}} \right| \leq C't^3.$$

We will write, somewhat informally,

$$R_\ell = 1/h_M''(0).$$

The function  $h_M$  is not actually twice differentiable;  $h_M''(0)$  here is a ‘pointwise second derivative’. Since  $R = \min \{R_\ell, R_{\text{wfs}}\}$ , these two propositions show how the convexity defect function yields the reach.

Proposition 2.12 will be proven in Section 2.4.2, but first we need to refine the upper bound given in item 5 of the properties of  $h_X$  given after Definition 2.10 for the case where  $X$  is a submanifold.

### 2.4.1 Upper bounds on the convexity defect function

The two aspects of the reach relate to the convexity defect function in quite different ways, which naturally leads one to wonder which aspect of the reach is responsible for item 5 of the properties of  $h_X$  given after Definition 2.10. In this subsection we improve the upper bound by increasing the radius of the bounding circle from  $R$  to  $R_\ell$ , though the bound still only holds on the interval  $[0, R)$  (compare with Lemma 12 in [ALS13]). See Figure 2.2 for an illustration.

**Proposition 2.13.** *If  $M \in \mathfrak{C}_{R_{\min}, L}^k$  and  $R = R(M)$  is its reach, then on  $[0, R)$  the function  $h_M(t)$  is bounded above by a quarter-circle of radius  $R_\ell$  centered on  $(0, R_\ell)$ . In other words,  $h_M(t) \leq R_\ell - \sqrt{R_\ell^2 - t^2}$ .*

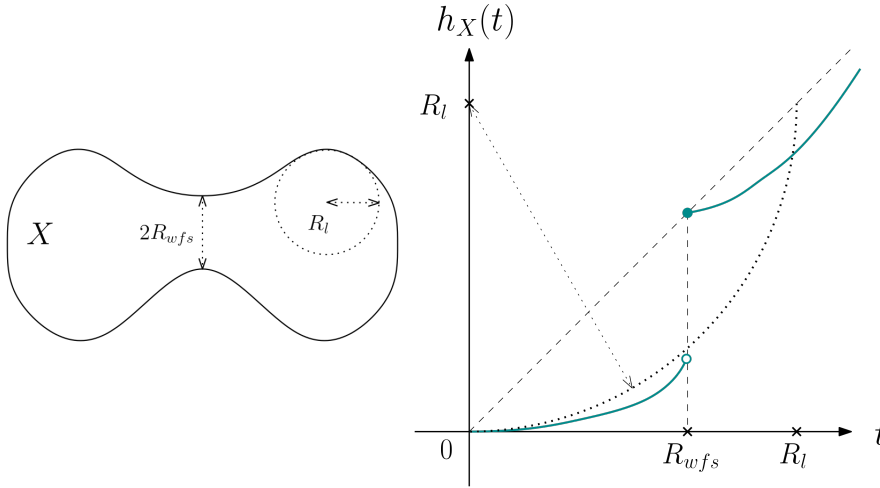


Figure 2.2 – A curve  $X$  (left) and its convexity defect function  $h_X(t)$  (right), which is below the quarter-circle of radius  $R_\ell$  for  $t < R(X) = R_{\text{wfs}}$ . Since  $R_{\text{wfs}} < R_\ell$ , we observe a discontinuity at  $t = R_{\text{wfs}}$ .

For submanifolds in the class  $\mathcal{M}_0^k$  (where  $R_{\text{wfs}} \geq R_\ell$ ), this result does not have any content. However, for manifolds in  $\mathcal{M}_\alpha^k$  i.e., manifolds for which  $R_{\text{wfs}} \leq R_\ell - \alpha$  for some  $\alpha > 0$ , the bound is sharper, with the following consequence.

**Corollary 2.14.** *If  $\mathbf{M} \in \mathcal{M}_\alpha^k$  for some  $\alpha > 0$ , then  $h_{\mathbf{M}}$  is discontinuous at  $R(\mathbf{M})$ .*

*Proof.* Since  $\alpha > 0$ , we have  $R(\mathbf{M}) = R_{\text{wfs}} < R_\ell$ . For  $t < R_{\text{wfs}}$  the bound  $h_{\mathbf{M}}(t) \leq R_\ell - \sqrt{R_\ell^2 - t^2}$  from Proposition 2.13 holds. On the other hand, for  $t = R_{\text{wfs}}$  we have  $h_{\mathbf{M}}(t) = t$ . Therefore the one-sided limit  $\lim_{t \nearrow R_{\text{wfs}}} h_{\mathbf{M}}(t) < h_{\mathbf{M}}(R_{\text{wfs}})$  and the function is discontinuous.  $\square$

The proof of Proposition 2.13 will require a few steps. We can focus our attention on the local reach by paying attention to sets of the form  $\mathbf{M}' = \mathbf{M} \cap B(z, r)$ , where  $z \in \mathbf{R}^D$ ,  $0 < r < R(\mathbf{M})$  and  $B(z, r)$  is a closed ball. Lemma 2.15 will show that subsets of this type have no bottlenecks. We would expect, then, that the reach of such a subset is generated by the local geometry. Lemma 2.18 quantifies this point: the reach of  $\mathbf{M}'$  is determined by the behaviour of the second fundamental form on  $\mathbf{M}'$ . The principal point of difficulty here relates to the boundary of the sets  $\mathbf{M}'$ . The proposition then follows from the fact that  $h_{\mathbf{M}}(t)$  can be bounded using the functions  $h_{\mathbf{M}'}(t)$  and so the bound is in fact determined by the second fundamental form, i.e. by  $R_\ell$  in particular.

**Lemma 2.15.** *Let  $A \subseteq \mathbf{R}^D$  be a compact set. Let  $0 < s < R(A)$ ,  $z \in \mathbf{R}^D$ , and  $A' = A \cap B(z, s)$ , where  $B$  is a closed ball. If  $A' \neq \emptyset$ , then  $A'$  cannot have any bottlenecks, i.e. there is no pair  $p, q \in A'$  with  $\|p - q\| = 2R(A')$  and  $(p + q)/2 \in \text{Med}(A')$ .*

*Proof.* Suppose for a contradiction that a bottleneck exists. Then it is a chord of length  $2R(A')$ . Since  $\text{diam } A' \leq 2s$  we obtain that  $2R(A') \leq 2s < 2R(A) \leq 2R(A')$ , the last inequality holding by [AL15, Lemma 5].  $\square$

We now consider the case where  $A = \mathbf{M}$ , a submanifold, and consider the intersections  $\mathbf{M}'$ . Our goal is to find the reach of the intersections,  $\mathbf{M}'$ , in order to bound  $h_{\mathbf{M}'}$  and hence  $h_{\mathbf{M}}$ . We will use the following characterisation of the reach due to Federer [Fed59]

$$\frac{1}{R(A)} = \sup_{p, q \in A} \frac{2d(q - p, C_p A)}{\|q - p\|^2},$$

where  $C_p A$  is the tangent cone at  $p$ , which Federer showed always exists for a set of positive reach. This quotient can be related to the second fundamental form as follows (cf. [AKC<sup>+</sup>19, Lemma 3.3]; and also work of Lytchak [Lyt04] for more general results).

**Lemma 2.16.** *Let  $k \geq 3$  and  $\mathbf{M} \in \mathfrak{C}_{R_{\min}, \mathbf{L}}^k$ . Let  $\mathbf{M}' = \mathbf{M} \cap B(z, r)$ , where  $z \in \mathbf{R}^D$ ,  $0 < r < R(\mathbf{M})$  and  $B$  is a closed ball. Then, provided  $\mathbf{M}'$  contains more than a single point, for any  $p \in \mathbf{M}'$  the norm of the second fundamental form is given by*

$$\|\Pi_p\|_{\text{op}} = \limsup_{\substack{q \rightarrow p \\ q \in \mathbf{M}'}} \frac{2d(q - p, C_p \mathbf{M}')}{\|q - p\|^2},$$

where  $C_p \mathbf{M}'$  is the tangent cone at  $p$  in  $\mathbf{M}'$ . In particular,  $1/R(\mathbf{M}') \geq \sup_{p \in \mathbf{M}'} \|\Pi_p\|_{\text{op}}$ .



*Proof.* We claim that  $\partial\mathbf{M}'$  is a  $C^k$  submanifold of  $\mathbf{M}$ . Consider the distance function to the central point  $z \in \mathbf{R}^D$ , say  $f(y) = d(z, y)$ . This function is smooth on  $\mathbf{R}^D \setminus z$  and its pull-back  $f|_{\mathbf{M}}$  is  $C^k$  on  $\mathbf{M} \setminus z$ . For any  $p \in \partial\mathbf{M}'$ ,  $f(p) = r$ . Note that  $r$  is a critical value of  $f|_{\mathbf{M}}$  precisely when the distance sphere  $\partial B(z, r)$  is tangent to  $\mathbf{M}$  at some  $p \in \mathbf{M}$ .

However, this cannot happen for  $r < R(\mathbf{M})$ . This is because  $r$  is less than the focal radius at  $p$  and so  $\mathbf{M}$  must lie in the exterior of  $B(z, r)$ . This in turn implies that  $\mathbf{M}' = \{p\}$  which contradicts the assumption that it is not a singleton. Therefore,  $r$  is a regular value of the  $C^k$  function  $f$  on  $\mathbf{M}$  and the pre-image  $\partial\mathbf{M}'$  is an embedded submanifold without boundary, as claimed.

As a consequence,  $\mathbf{M}'$  is an embedded submanifold of  $\mathbf{M}$  of full dimension with boundary. The tangent cone in  $\mathbf{R}^D$ ,  $C_p\mathbf{M}'$ , is given by  $T_p\mathbf{M}$  for  $p$  in the interior of  $\mathbf{M}'$  and by a half-space of  $T_p\mathbf{M}$  for  $p \in \partial\mathbf{M}'$ , namely

$$C_p\mathbf{M}' = T_p\mathbf{M} \cap \{u \mid \langle p - z, u \rangle \leq 0\},$$

where  $z$  is the center of the ball containing  $\mathbf{M}'$ . We now consider some other point  $q \in \mathbf{M}'$ ,  $q \neq p$ , and show that the projection of  $q$  to  $T_p\mathbf{M}$  lies in  $C_p\mathbf{M}'$ . Suppose  $p \in \partial\mathbf{M}' \subseteq \partial B$ . Consider the affine hyperplane  $H^{D-1}$  through  $p$  perpendicular to the line  $pz$ . Since  $q \in B$ ,  $q$  lies on the same side of  $H$  as  $z$  and therefore the projection of  $q$  to  $T_p\mathbf{M}$  lies in  $C_p\mathbf{M}'$ . If  $p \notin \partial\mathbf{M}'$  then  $T_p\mathbf{M} = C_p\mathbf{M}'$  and so this statement automatically holds.

Let us assume now that  $q$  is close to  $p$ , satisfying  $\|q - p\| \leq \frac{1}{4} \min\{R_{\min}, (L_{\perp})^{-1}\}$ , so that the projection of  $q$  to  $C_p\mathbf{M}'$  satisfies the conclusion of Lemma 2.8. In particular, if  $v$  is the projection of  $q$  onto  $T_p\mathbf{M}$ , we may write

$$q - p = v + \frac{1}{2}\Pi_p(v, v) + \mathcal{R}_3(v),$$

where the remainder  $\mathcal{R}_3(v)$  is of order  $O(\|v\|^3)$ . Therefore

$$d(q - p, C_p\mathbf{M}') = \left\| \frac{1}{2}\Pi_p(v, v) + \mathcal{R}_3(v) \right\|.$$

We can then calculate the Federer quotient,

$$\begin{aligned} \frac{2d(q - p, C_p\mathbf{M}')}{\|q - p\|^2} &= \frac{\left\| \frac{1}{2}\Pi_p(v, v) + \mathcal{R}_3(v) \right\|}{\left\| v \right\|^2 + \left\| \frac{1}{2}\Pi_p(v, v) + \mathcal{R}_3(v) \right\|^2} \\ &= \frac{1}{\frac{\|v\|^2}{\left\| \frac{1}{2}\Pi_p(v, v) + \mathcal{R}_3(v) \right\|} + \frac{1}{4} \left\| \frac{1}{2}\Pi_p(v, v) + \mathcal{R}_3(v) \right\|}}. \end{aligned}$$

As  $q \rightarrow p$  we see that  $v \rightarrow 0$ . In order to compute the  $\limsup$ , we may assume that a sequence of points  $q_i$  is chosen such that  $\left\| \frac{1}{2}\Pi_p(v_i, v_i) \right\|$  is maximized. Then, since all terms in the denominator go to zero except the ratio  $\frac{\|v_i\|^2}{\left\| \frac{1}{2}\Pi_p(v_i, v_i) \right\|}$ , we have

$$\limsup_{\substack{q \rightarrow p \\ q \in \mathbf{M}'}} \frac{2d(q - p, C_p\mathbf{M}')}{\|q - p\|^2} = \lim_{i \rightarrow \infty} \frac{\left\| \frac{1}{2}\Pi_p(v_i, v_i) \right\|}{\|v_i\|^2}.$$

We would like to claim that

$$\lim_{i \rightarrow \infty} \frac{\|\Pi_p(v_i, v_i)\|}{\|v_i\|^2} = \|\Pi_p\|_{\text{op}},$$

but recall that  $p$  may lie on the boundary of  $\mathbf{M}'$  and so we must check that a suitable sequence of points  $q_i \in \mathbf{M}'$  can be found. Since  $C_p\mathbf{M}'$  is a half-space and  $\Pi_p$  is a symmetric, bilinear form, there is some unit vector  $w \in C_p\mathbf{M}'$  so that  $\|\Pi_p(w, w)\| = \|\Pi_p\|_{\text{op}}$ . Then we can choose a sequence  $q_i \in \mathbf{M}'$  so that the projections of the  $q_i$  are  $t_i v_i$ , where the  $v_i$  are unit vectors in  $C_p\mathbf{M}'$  such that  $v_i \rightarrow w$  and the  $t_i$  are positive numbers with  $t_i \rightarrow 0$ . The existence of such a sequence is equivalent to the fact that  $w \in C_p\mathbf{M}'$ . The final statement then follows from

$$\|\Pi_p\|_{\text{op}} = \limsup_{\substack{q \rightarrow p \\ q \in \mathbf{M}'}} \frac{2d(q - p, C_p\mathbf{M}')}{\|q - p\|^2} \leq \sup_{p, q \in \mathbf{M}'} \frac{2d(q - p, C_p\mathbf{M}')}{\|q - p\|^2} = \frac{1}{R(\mathbf{M}')}. \quad \square$$

*Remark 2.17.* The regularity assumption of  $k \geq 3$  in the previous lemma may possibly be improved to  $k \geq 2$ . This stems from the assumption in Lemma 2.8 which in turn derives from the regularity assumption in [AL19, Lemma 2]. However, this is not needed in the sequel so we do not pursue this further.

**Lemma 2.18.** *Let  $k \geq 3$  and  $\mathbf{M} \in \mathfrak{C}_{R_{\min}, \mathbf{L}}^k$ . Let  $\mathbf{M}' = \mathbf{M} \cap B(z, r)$ , where  $z \in \mathbf{R}^D$ ,  $0 < r < R(\mathbf{M})$  and  $B$  is a closed ball. Then, provided  $\mathbf{M}'$  contains more than a single point, we have  $1/R(\mathbf{M}') = \sup_{p \in \mathbf{M}'} \|\Pi_p\|_{\text{op}}$ .*

*Proof.* We have already shown in Lemma 2.16 that  $1/R(\mathbf{M}') \geq \sup_{p \in \mathbf{M}'} \|\Pi_p\|_{\text{op}}$ . By Lemma 2.15,  $\mathbf{M}'$  does not contain any bottlenecks. It follows that the reach is attained in one of two ways and we examine each case.

*Case 1:* The reach of  $\mathbf{M}'$  is attained by a pair of points  $q, r \in \mathbf{M}'$  but  $\|q - r\| < 2R(\mathbf{M}')$ . In this case we apply [AKC<sup>+</sup>19, Lemma 3.2] to obtain, in  $\mathbf{M}'$ , an arc of a circle of radius  $R$  equal to the reach of  $\mathbf{M}'$ . Note that that lemma is stated for manifolds, but in fact the proof only requires a set of positive reach. Then, for any point  $p$  on the reach-attaining arc, we obtain that

$$\frac{1}{R(\mathbf{M}')} \leq \|\Pi_p\|_{\text{op}} \leq \sup_{s \in \mathbf{M}'} \|\Pi_s\|.$$

*Case 2:* The reach of  $\mathbf{M}'$  is attained at a single point, say  $p$ , in  $\mathbf{M}'$ . It follows, using Lemma 2.16 that

$$\frac{1}{R(\mathbf{M}')} = \limsup_{\substack{q \rightarrow p \\ q \in \mathbf{M}'}} \frac{2d(q - p, C_p\mathbf{M}')}{\|q - p\|^2} = \|\Pi_p\|_{\text{op}} \leq \sup_{s \in \mathbf{M}'} \|\Pi_s\|_{\text{op}}.$$

Combining the two cases, then, we also have that

$$\frac{1}{R(\mathbf{M}')} \leq \sup_{s \in \mathbf{M}'} \|\Pi_s\|_{\text{op}}$$

completing the proof.  $\square$

*Proof of Proposition 2.13.* Let  $M' = M \cap B(z, r)$ , where  $z \in \mathbf{R}^D$ ,  $0 < r < R(M)$  and  $B$  is a closed ball. Recall that on  $[0, R(M'))$  we have

$$h_{M'}(t) \leq R(M') - \sqrt{R(M')^2 - t^2}.$$

By Lemma 2.18, if  $M'$  is not a single point we have

$$\frac{1}{R_\ell} = \sup_{s \in M} \|\Pi_s\|_{\text{op}} \geq \sup_{s \in M'} \|\Pi_s\|_{\text{op}} = \frac{1}{R(M')},$$

and this entails the bound  $h_{M'}(t) \leq R_\ell - \sqrt{R_\ell^2 - t^2}$  on  $[0, R(M'))$ . If  $M'$  is a point then  $h_{M'}(t) = 0$  for all  $t$  and so the same bound holds.

Recalling that  $R(M') \geq R(M)$  for every  $M'$  with  $\text{Rad}(M') < R(M)$ , we have, for  $0 < t \leq r < R(M)$ ,

$$\sup_{M' = M \cap B(z, r)} h_{M'}(t) \leq R_\ell - \sqrt{R_\ell^2 - t^2}.$$

Now for every  $\sigma \subset M$  with  $\text{Rad}(\sigma) \leq t \leq r$ , there is some  $M' = M \cap B(z, r)$  with  $\sigma \subset M'$  and it follows that

$$h_M(t) \leq \sup_{M' = M \cap B(z, r)} h_{M'}(t).$$

Setting  $t = r$  and combining the two inequalities, we have, for  $0 < t < R(M)$ ,

$$h_M(t) \leq R_\ell - \sqrt{R_\ell^2 - t^2}. \quad \square$$

## 2.4.2 The convexity defect function near zero

We have seen in the previous section how, for  $M \subseteq \mathbf{R}^D$  a compact submanifold, the function  $h_M$  on  $[0, R)$  obeys an upper bound determined by  $R_\ell$ . We now study  $h_M$  in greater detail in a neighborhood of zero to obtain a Taylor polynomial, identifying  $R_\ell$  as the reciprocal of the ‘pointwise second derivative’,  $1/h_M''(0)$ . More formally, we prove Proposition 2.12, which states that, for any sufficiently small  $t$ ,

$$\left| h_M(t) - \frac{t^2}{2R_\ell} \right| \leq Ct^{k \wedge 4}.$$

Once more, we approach  $h_M$  by considering sets  $M'$ , which are the intersection of  $M$  with small closed balls.

We introduce a new function  $h_{M'}^{\text{loc}}(p, r_1, r_2; t)$ , which contains information on the convexity of  $M'$ . Lemma 2.19 shows how  $h_M$  can be determined from all the  $h_{M'}^{\text{loc}}(p, r_1, r_2; t)$ . Recall from Lemma 2.8 that such sets  $M'$  can be written as the graphs of functions over  $T_p M$  and that these functions have Taylor expansions.

Lemma 2.21 will set a lower bound on  $h^{\text{loc}}$  for the degree 3 approximation to  $M$  around  $p$ , which Lemma 2.23 translates to a lower bound on  $h_{M'}^{\text{loc}}(p, r_1, r_2; t)$  itself. Varying  $M'$  we obtain a lower bound on  $h_M(t)$  for small  $t$ , which we combine with the upper bound from Proposition 2.13 to prove the result.

**Lemma 2.19.** *Let  $B$  denote a closed ball, then, for any compact set  $\mathbf{X} \subset \mathbb{R}^D$  and any  $r_1, r_2, t > 0$  satisfying  $r_1 \geq 3t$  and  $r_2 \geq t + r_1$ , we have*

$$h_{\mathbf{X}}(t) = \sup_{p \in \mathbf{M}} h_{\mathbf{X}}^{\text{loc}}(p, r_1, r_2; t)$$

where

$$h_{\mathbf{X}}^{\text{loc}}(p, r_1, r_2; t) = h\left(\bigcup_{\substack{\sigma \subseteq \mathbf{X} \cap B(p, r_1) \\ \text{Rad } \sigma \leq t}} \text{Hull } \sigma \mid \mathbf{X} \cap B(p, r_2)\right).$$

*Proof.* We have immediately, for any  $p \in \mathbf{X}$  and any  $r_1, t > 0$

$$h_{\mathbf{X}}(t) = H\left(\bigcup_{\substack{\sigma \subseteq \mathbf{X} \\ \text{Rad } \sigma \leq t}} \text{Hull } \sigma \mid \mathbf{X}\right) \geq H\left(\bigcup_{\substack{\sigma \subseteq \mathbf{X} \cap B(p, r_1) \\ \text{Rad } \sigma \leq t}} \text{Hull } \sigma \mid \mathbf{X}\right)$$

and so all that is necessary is to check that

$$H\left(\bigcup_{\substack{\sigma \subseteq \mathbf{X} \cap B(p, r_1) \\ \text{Rad } \sigma \leq t}} \text{Hull } \sigma \mid \mathbf{X}\right) = H\left(\bigcup_{\substack{\sigma \subseteq \mathbf{X} \cap B(p, r_1) \\ \text{Rad } \sigma \leq t}} \text{Hull } \sigma \mid \mathbf{X} \cap B(p, r_2)\right) = h_{\mathbf{X}}^{\text{loc}}(p, r_1, r_2; t).$$

Let the asymmetric distance

$$h\left(\bigcup_{\substack{\sigma \subseteq \mathbf{X} \cap B(p, r_1) \\ \text{Rad } \sigma \leq t}} \text{Hull } \sigma \mid \mathbf{X}\right)$$

be realized by the data  $\sigma \subseteq \mathbf{X} \cap B(p, r_1)$ ,  $y \in \text{Hull } \sigma$ ,  $p' \in \mathbf{X}$ . We have  $d(p', y) \leq t$  and  $d(y, p) \leq r_1$  so that  $d(p', p) \leq r_1 + t \leq r_2$ . For the converse, just notice that if  $p' \in \mathbf{X}$  and  $\sigma \subset \mathbf{X}$ , then, for any  $y \in \text{Hull } \sigma$

$$\sigma \subset B(p', d(p', y) + 2\text{Rad}(\sigma))$$

yielding for the optimal data  $\sigma, y, p'$  the inclusion  $\sigma \subset B(p, 3t)$ .  $\square$

For a bilinear map  $S : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{D-d}$  and a trilinear map  $T : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{D-d}$ , we denote

$$M(S, T) = \{(v, S(v^{\otimes 2}) + T(v^{\otimes 3})) \mid v \in \mathbb{R}^d\} \subseteq \mathbb{R}^D$$

which is a  $C^\infty$  submanifold of  $\mathbb{R}^D$  of dimension  $d$ .

By setting  $S$  and  $T$  to be the coefficients of  $\Phi_p^3$ , the approximation of degree 3 to a manifold  $\mathbf{M}$  around  $p \in \mathbf{M}$  (see Definition 2.9), we can easily see that, near  $p$ ,  $M(S, T)$  is Hausdorff close to  $\mathbf{M}$ . This assumes that  $p = 0$  and that  $T_p \mathbf{M}$  is the subspace spanned by the first  $d$  co-ordinates. This assumption, which is used in the statement of the lemma below, is for convenience only. For each  $p \in \mathbf{M}$  there is an isometry of  $\mathbb{R}^D$  which causes it to be satisfied.

**Lemma 2.20.** Let  $\mathbf{M} \in \mathcal{C}_{R_{\min}, \mathbf{L}}^k$ . Suppose that  $p = 0 \in \mathbf{M}$  and  $T_p \mathbf{M} = \mathbf{R}^d \subseteq \mathbf{R}^D$ .

If  $k \geq 4$ , we have, for  $s \leq s_1$  with  $s_1$  depending only on  $R_{\min}, \mathbf{L}, k, d$ ,

$$H(\mathbf{M} \cap B(0, s), M(S, T) \cap B(0, s)) \leq C s^4,$$

where  $S$  and  $T$  are obtained from the degree 3 approximation  $\Phi_0^3$  given in Definition 2.9 by  $S = \frac{1}{2} d_0^2 \Phi_0^3 = \Pi_0$ ,  $T = \frac{1}{6} d_0^3 \Phi_0^3$  and the constant  $C = C_{R_{\min}, \mathbf{L}, k, d}$ .

When  $k = 3$  we can use the degree 2 approximation  $\Phi_0^2$  and pick  $T \equiv 0$ , to obtain

$$H(\mathbf{M} \cap B(0, s), M(S, 0) \cap B(0, s)) \leq C' s^3$$

*Proof.* Let us initially take  $s_1 = \min\{R_{\min}, L_{\perp}^{-1}\}/4$ . Then for any point  $q \in \mathbf{M} \cap B(0, s)$ , if  $v = \text{pr}_{T_0 \mathbf{M}}(q)$  then

$$q = \Phi_0(v) = v + S(v^{\otimes 2}) + T(v^{\otimes 3}) + \mathcal{R}(v),$$

where  $\Phi_0$  is the expansion given in Lemma 2.8 and  $\|\mathcal{R}(v)\| \leq \frac{L'_4}{24} \|v\|^4$ , unless  $k = 3$ . In case  $k = 3$ , if we wish to control the remainder we can only use the degree 2 polynomial approximation  $\Phi_0^2$ .

It is therefore clear that, for the point  $q = \Phi_0(v) \in \mathbf{M} \cap B(0, s)$ , there is a corresponding point  $\Phi_0^3(v) \in M(S, T)$  within the required distance and, conversely, for any point  $\Phi_0^3(v) \in M(S, T) \cap B(0, s)$ , there is a corresponding point  $\Phi_0(v) \in \mathbf{M}$ . The constant  $C$  may be chosen to be  $C = \frac{L'_4}{24}$ .

However, the corresponding point is not guaranteed to lie in the ball  $B(0, s)$ . In the next paragraph we establish that there is a vector  $v'$  very close to  $v$ , so that  $\Phi_0^3(v')$  or  $\Phi_0(v')$ , as appropriate, will be sufficiently close.

Let us continue to assume  $k \geq 4$ , since the case  $k = 3$  is essentially identical. We first consider the possibility that  $\|\Phi_0^3(v)\| \leq s$  but  $\|\Phi_0(v)\| > s$ . It is clear that, for sufficiently small  $s$ ,  $\|\Phi_0(v)\|^2 \leq s^2 + C_0 s^6$ , where  $C_0$  depends on  $R_{\min}, L_{\perp}, L_3$  and  $L_4$ . It follows that  $\|\Phi_0(v)\| \leq s + C_1 s^5$ . Consider now a vector  $v' = (1 - \lambda)v$ , with  $\lambda \approx 0$ , chosen so that  $\|\Phi_0(v')\| = s$ . For small enough  $s$  we have  $\lambda \leq C_2 s^4$ . It follows immediately that  $\Phi_0(v')$  lies within  $C_3 s^4$  of  $\Phi_0(v)$ , and hence within  $C s^4$  of  $\Phi_0^3(v)$ .

The case where  $\|\Phi_0(v)\| \leq s$  but  $\|\Phi_0^3(v)\| > s$  is dealt with similarly.  $\square$

The utility of  $M(S, T)$  is that, since it is algebraic, we can compute explicit bounds for  $h_{\mathbf{X}}^{\text{loc}}$ , where  $\mathbf{X} = M(S, T)$ .

**Lemma 2.21.** Let  $r_1 \leq r_2 \leq \frac{13^{1/4}}{2} \|T\|_{\text{op}}^{-1/2}$ , and let  $\mathbf{X} = M(S, T)$ . Then for any  $t \leq \min\left(\frac{1}{2} \|S\|_{\text{op}}^{-1}, \frac{2}{\sqrt{13}} r_1\right)$  we have

$$h_{\mathbf{X}}^{\text{loc}}(0, r_1, r_2; t) \geq \left(t - \frac{1}{2} t^5 \|T\|_{\text{op}}^2\right)^2 \|S\|_{\text{op}} \geq t^2 \|S\|_{\text{op}} - t^6 \|S\|_{\text{op}} \|T\|_{\text{op}}^2.$$

*Proof.* Let  $v$  be a unit norm vector in  $\mathbf{R}^d$  such that  $\|S(v^{\otimes 2})\| = \|S\|_{\text{op}}$ , and let  $z \leq \min(\frac{1}{2}\|S\|_{\text{op}}^{-1}, \frac{2}{\sqrt{13}}r_1)$ . Note that the upper bound on  $r_1$  gives a third upper bound for  $z$ , namely  $z \leq 13^{-1/4}\|T\|_{\text{op}}^{-1/2} \leq \|T\|_{\text{op}}^{-1/2}$ . We set

$$p_1 = (zv, S((zv)^{\otimes 2}) + T((zv)^{\otimes 3})) \text{ and } p_2 = (-zv, S((-zv)^{\otimes 2}) + T((-zv)^{\otimes 3}))$$

and denote the two-point set containing them by  $\sigma = \{p_1, p_2\}$ . In order to use  $\sigma$  to bound  $h_{\mathbf{X}}^{\text{loc}}$  we must (1) check  $\sigma \subseteq \mathbf{X} \cap B(0, r_1)$ , (2) find the radius of  $\sigma$  and (3) determine  $H(\text{Hull } \sigma | \mathbf{X} \cap B(0, r_2))$ .

Firstly, since  $\sigma \subseteq M(S, T)$ , it is enough to show that  $\|p_1\|^2, \|p_2\|^2 \leq r_1^2$ . Using all three bounds on  $z$ , we can check

$$\begin{aligned} \|p_1\|^2, \|p_2\|^2 &\leq z^2 + z^4\|S\|_{\text{op}}^2 + 2z^5\|S\|_{\text{op}}\|T\|_{\text{op}} + z^6\|T\|_{\text{op}}^2 \\ &\leq 2z^2 + 2z^3\|S\|_{\text{op}} + z^4\|S\|_{\text{op}}^2 \text{ by } z\|T\|_{\text{op}}^{1/2} < 1 \\ &\leq \frac{13}{4}z^2 \text{ by } z\|S\|_{\text{op}} \leq \frac{1}{2} \\ &\leq r_1^2 \text{ by } z \leq \frac{2}{\sqrt{13}}r_1. \end{aligned}$$

Secondly, we obtain the radius as

$$\begin{aligned} \text{Rad } \sigma &= \frac{1}{2}\sqrt{(2z)^2 + (2z^3\|T(v^{\otimes 3})\|)^2} \\ &= z\sqrt{1 + z^4\|T(v^{\otimes 3})\|^2} \\ &\leq z\left(1 + \frac{1}{2}z^4\|T\|_{\text{op}}^2\right) \text{ since } \sqrt{1+x} \leq 1 + \frac{1}{2}x \text{ for } x \geq 0 \\ &= z + \frac{1}{2}z^5\|T\|_{\text{op}}^2. \end{aligned}$$

Thirdly, we place a lower bound on  $H(\text{Hull } \sigma | \mathbf{X} \cap B(0, r_2))$ . Let  $q = \frac{1}{2}(p_1 + p_2) \in \text{Hull } \sigma$ . For any  $p = (w, S(w^{\otimes 2}) + T(w^{\otimes 3})) \in \mathbf{X}$  satisfying  $\|w\| \leq r_2$ , we have

$$\begin{aligned} d(q, p)^2 &= \|w\|^2 + \|S(w^{\otimes 2}) + T(w^{\otimes 3}) - z^2S(v^{\otimes 2})\|^2 \\ &\geq z^4\|S\|_{\text{op}}^2 + \|w\|^2(1 - 2z^2\|S\|_{\text{op}}^2 - 2z^2r_2\|S\|_{\text{op}}\|T\|_{\text{op}}). \end{aligned}$$

Since  $z\|S\|_{\text{op}} \leq 1/2$  we have  $2z^2\|S\|_{\text{op}}^2 \leq \frac{1}{2}$ . The same condition also allows us to see that  $2z^2r_2\|S\|_{\text{op}}\|T\|_{\text{op}} \leq zr_2\|T\|_{\text{op}} \leq \frac{1}{2}$ . It follows that

$$d(q, p)^2 \geq z^4\|S\|_{\text{op}}^2 = d(q, 0)^2$$

from which we obtain the bound  $h(\text{Hull } \sigma | \mathbf{X} \cap B(0, r_2)) \geq z^2\|S\|_{\text{op}}$ .

These three calculations yield  $h_{\mathbf{X}}^{\text{loc}}(0, r_1, r_2; z + \frac{1}{2}z^5\|T\|_{\text{op}}^2) \geq z^2\|S\|_{\text{op}}$ . Now we may reparametrize the argument by setting  $t = z + \frac{1}{2}z^5\|T\|_{\text{op}}^2$ . Obviously  $t \geq z$  so we can invert to obtain  $z = t - \frac{1}{2}t^5\|T\|_{\text{op}}^2 \geq t - \frac{1}{2}t^5\|T\|_{\text{op}}^2$  and so  $h_{\mathbf{X}}^{\text{loc}}(0, r_1, r_2; t) \geq (t - \frac{1}{2}t^5\|T\|_{\text{op}}^2)^2\|S\|_{\text{op}} \geq (t^2 - t^6\|T\|_{\text{op}}^2)\|S\|_{\text{op}}$ . If the bounds given in the statement hold for  $t$ , then they will also hold for the smaller value  $z$  and so the result is proved.  $\square$

We are now in a position to convert this bound for an algebraic approximation to  $\mathbf{M}$  into one for the small patch of  $\mathbf{M}$  itself.

We need a stability result first.

**Lemma 2.22.** *Let  $\mathbf{X}, \mathbf{Y}$  be two subset of  $\mathbb{R}^D$  and let  $r_1, r_2, t > 0$ . Then, if  $p \in \mathbf{X} \cap \mathbf{Y}$  and  $h(\mathbf{X} \cap B(p, r_2), \mathbf{Y} \cap B(p, r_2)) \leq \epsilon$ , we have*

$$h_{\mathbf{X}}^{\text{loc}}(0, r_1, r_2; t) \leq h_{\mathbf{Y}}^{\text{loc}}(p, r_1 + \epsilon, r_2; t + \epsilon) + 2\epsilon.$$

*Proof.* This is a straightforward adaptation of the proof of Lemma 5 in [ALS13]. Indeed let  $\sigma \subset \mathbf{X} \cap B(p, r_1)$  be such that  $\text{Rad } \sigma \leq t$ . Let  $\xi = \mathbf{Y} \cap B(p, r_2) \cap \sigma^\epsilon$ . Since  $h(\mathbf{X} \cap B(p, r_2), \mathbf{Y} \cap B(p, r_2)) \leq \epsilon$ ,  $\xi$  is not empty and satisfies  $h(\xi, \sigma) \leq \epsilon$ . Thus  $\xi \subset \mathbf{Y} \cap B(p, r_1 + \epsilon)$ , and furthermore, by Lemma 16 in [ALS13], we have  $\text{Rad } \xi \leq t + \epsilon$ . We conclude using that

$$\begin{aligned} \text{Hull } \sigma &\subset \text{Hull}(\xi^\epsilon) = (\text{Hull } \xi)^\epsilon \\ &\subset (\mathbf{Y} \cap B(p, r_2))^{h_{\mathbf{Y}}^{\text{loc}}(p, r_1 + \epsilon, r_2; t + \epsilon) + \epsilon} \\ &\subset (\mathbf{X} \cap B(p, r_2))^{h_{\mathbf{Y}}^{\text{loc}}(p, r_1 + \epsilon, r_2; t + \epsilon) + 2\epsilon}. \end{aligned} \quad \square$$

**Lemma 2.23.** *Let  $k \geq 4$ . There exists  $s_2 > 0$  depending only on  $R_{\min}, \mathbf{L}, k, d$  such that for any  $r_2 \leq s_2$  and for any  $r_1, t \geq 0$  such that*

$$C_0 r_2^4 \leq t \leq \frac{2}{\sqrt{13}} r_1$$

*for some constant  $C_0$  depending on  $R_{\min}, \mathbf{L}, k, d$ , we have, for all  $\mathbf{M} \in \mathfrak{C}_{R_{\min}, \mathbf{L}}^k$  and all  $p \in \mathbf{M}$ ,*

$$h_{\mathbf{M}}^{\text{loc}}(p, r_1, r_2; t) \geq \frac{1}{2} t^2 \|\Pi_p\|_{\text{op}} - C r_2^4$$

*where  $C$  is a constant depending on  $R_{\min}, \mathbf{L}, k, d$ .*

*In case  $k = 3$ , we have, for all  $\mathbf{M} \in \mathfrak{C}_{R_{\min}, \mathbf{L}}^k$  and all  $p \in \mathbf{M}$ ,*

$$h_{\mathbf{M}}^{\text{loc}}(p, r_1, r_2; t) \geq \frac{1}{2} t^2 \|\Pi_p\|_{\text{op}} - C' r_2^3$$

*where  $C'$  is a constant depending on  $R_{\min}, \mathbf{L}, d$ .*

*Proof.* By applying an isometry of  $\mathbf{R}^D$ , we may assume that  $p = 0$  and that  $T_p \mathbf{M} = \mathbf{R}^d \subseteq \mathbf{R}^D$ . The result will then follow from Lemmata 2.20 and 2.21 in addition to the Hausdorff stability property for  $h^{\text{loc}}$  (Lemma 2.22). Take  $r_2 > 0$  smaller than  $s_1$  (from Lemma 2.20) and then  $\frac{13^{1/4}}{2L_3^{1/2}}$  (from Lemma 2.21). In the case  $k \geq 4$ , where  $\Phi_p$  is the expansion described in Lemma 2.8,  $S = \frac{1}{2} d_0^2 \Phi_p = \Pi_p$ ,  $T = \frac{1}{6} d_0^3 \Phi_p$  and  $C_0$  is the constant from the statement of Lemma 2.20, we have

$$\begin{aligned}
 h_{\mathbf{M}}^{\text{loc}}(0, r_1, r_2; t) &\geq h_{M(S,T)}^{\text{loc}}(0, r_1 - C_0 s^4, r_2; t - C_0 r_2^4) - 2C_0 r_2^4 \\
 &\geq (t - C_0 r_2^4)^2 \|S\|_{\text{op}} - (t - C_0 r_2^4)^6 \|S\|_{\text{op}} \|T\|_{\text{op}}^2 - 2C_0 r_2^4 \\
 &\geq \frac{1}{2} \|\Pi_p\|_{\text{op}} t^2 - C r_2^4.
 \end{aligned}$$

where  $C$  depends only on  $R_{\min}, \mathbf{L}, d, k$ . The first inequality only holds if  $C_0 r_2^4 \leq t$ . In the case  $k = 3$  the result is obtained similarly.  $\square$

We conclude with the proof of Proposition 2.12.

*Proof of Proposition 2.12.* By taking

$$t \leq \frac{1}{4} s_2 \wedge (4^4 C_0)^{-1/3}$$

(from Lemma 2.23), and setting  $r_1 = 3t$  and  $r_2 = 4t$ , we have that

$$C_0 r_2^4 \leq t \leq \frac{2}{\sqrt{13}} r_1 \text{ and } t + r_1 \leq r_2$$

so that the hypotheses of both 2.19 and 2.23 hold. It is now immediate that if  $k \geq 4$

$$\begin{aligned}
 h_{\mathbf{M}}(t) &= \sup_{p \in \mathbf{M}} h_{\mathbf{M}}^{\text{loc}}(p, r_1, r_2; t) \\
 &\geq \sup_{p \in \mathbf{M}} \left( \frac{1}{2} \|\Pi_p\|_{\text{op}} t^2 - C r_2^4 \right) \\
 &= \frac{t^2}{2R_\ell} - 4^4 C t^4
 \end{aligned}$$

where  $C$  is a constant depending on  $R_{\min}, \mathbf{L}, d, k$ , while if  $k = 3$

$$h_{\mathbf{M}}(t) \geq \frac{1}{2R_\ell} t^2 - C' t^3,$$

where  $C'$  is a constant depending on  $R_{\min}, \mathbf{L}$ . On the other hand, Proposition 2.13 provides an upper bound which will hold for all  $t < R_{\min}$ :

$$\begin{aligned}
 h_{\mathbf{M}}(t) &\leq R_\ell - \sqrt{R_\ell^2 - t^2} \leq \frac{t^2}{2R_\ell} + \frac{t^4}{2R_\ell^3} \\
 &\leq \frac{t^2}{2R_\ell} + \frac{t^4}{2R_{\min}^3}.
 \end{aligned}$$

$\square$

## 2.5 Approximating the reach

Recall item 3 of the properties of  $h_{\mathbf{X}}$  given after Definition 2.10 which guarantees that the convexity defect function is stable with respect to perturbations of the manifold



which are small in the Hausdorff distance. This allows one to approximate the reach of a submanifold  $M \subseteq \mathbf{R}^D$  from a nearby subset  $\widetilde{M}$ .

Given a submanifold  $M$  and another subset  $\widetilde{M}$  (not necessarily a manifold) so that  $H(M, \widetilde{M}) < \epsilon$ , we can calculate the convexity defect function  $h_{\widetilde{M}}$ . This can then be used to approximate  $R_\ell = (h''_M(0))^{-1}$  and  $R_{\text{wfs}} = \inf \{t: h_M(t) = t, t > 0\}$ . We can approximate the local reach via

$$h''_M(0) \approx 2 \frac{h_{\widetilde{M}}(\Delta)}{\Delta^2}$$

for some choice of step size  $\Delta$ . Proposition 2.12 gives the following bound on the error.

**Proposition 2.24.** *Let  $M \in \mathfrak{C}_{R_{\min}, \mathbf{L}}^k$ . Let  $0 < \epsilon < \Delta < 1$  be such that  $\epsilon + \Delta$  is small enough to satisfy the hypotheses constraining the variable  $t$  in Proposition 2.12. Let  $\widetilde{M} \subseteq \mathbf{R}^D$  be such that  $H(M, \widetilde{M}) < \epsilon$ .*

*Then*

- *If  $k \geq 4$ ,  $\left| h''_M(0) - 2 \frac{h_{\widetilde{M}}(\Delta)}{\Delta^2} \right| \leq A\epsilon\Delta^{-2} + B\Delta^2$  and, in particular, if  $\Delta = \epsilon^{1/4}$ ,*

$$\left| h''_M(0) - 2 \frac{h_{\widetilde{M}}(\Delta)}{\Delta^2} \right| \leq (A + B)\epsilon^{1/2}$$

- *If  $k = 3$ ,  $\left| h''_M(0) - 2 \frac{h_{\widetilde{M}}(\Delta)}{\Delta^2} \right| \leq A\epsilon\Delta^{-2} + B\Delta$  and, in particular, if  $\Delta = \epsilon^{1/3}$ ,*

$$\left| h''_M(0) - 2 \frac{h_{\widetilde{M}}(\Delta)}{\Delta^2} \right| \leq (A + B)\epsilon^{1/3}$$

where the constants  $A$  and  $B$  depend only on  $R_{\min}, \mathbf{L}$ .

*Proof.* Set  $\kappa = h''_M(0)$  and  $\tilde{\kappa} = 2 \frac{h_{\widetilde{M}}(\Delta)}{\Delta^2}$ . Comparing  $M$  to  $\widetilde{M}$ , we obtain from stability that

$$2 \frac{h_M(\Delta - \epsilon) - 2\epsilon}{\Delta^2} \leq \tilde{\kappa} \leq 2 \frac{h_M(\Delta + \epsilon) + 2\epsilon}{\Delta^2}.$$

In the case  $k \geq 4$ , Proposition 2.12 states that  $|h_M(t) - \frac{\kappa}{2}t^2| \leq Ct^4$ , for some constant  $C$  depending only on  $R_{\min}, \mathbf{L}$ . It follows that

$$\frac{\kappa(\Delta - \epsilon)^2 - 2C(\Delta - \epsilon)^4 - 4\epsilon}{\Delta^2} \leq \tilde{\kappa} \leq \frac{\kappa(\Delta + \epsilon)^2 + 2C(\Delta + \epsilon)^4 + 4\epsilon}{\Delta^2}.$$

Expanding and using that  $\epsilon, \Delta < 1$ , we obtain

$$|\kappa - \tilde{\kappa}| \leq 2C\Delta^2 + (3\kappa + 30C + 4)\epsilon\Delta^{-2}.$$

Similarly, in the case  $k = 3$ , we obtain

$$|\kappa - \tilde{\kappa}| \leq 2C'\Delta + (3\kappa + 14C' + 4)\epsilon\Delta^{-2}$$

where  $C'$  is again a constant depending only on  $R_{\min}, \mathbf{L}$ . Since  $\kappa \leq 1/R_{\min}$ , the constants may be chosen to be  $A = \max\{3/R_{\min} + 30C + 4, 3/R_{\min} + 14C' + 4\}$  and  $B = \max\{2C, 2C'\}$ . They depend only on  $R_{\min}, \mathbf{L}$ .

Now set  $\Delta = \epsilon^p$  and seek the  $p$  yielding the fastest rate of convergence of the error bound to zero. Since the exponent in the first term increases with respect to  $p$  while that in the second decreases, the fastest rate is obtained by requiring the two exponents to be equal, so that  $p = 1/4$  for  $k \geq 4$  and  $p = 1/3$  for  $k = 3$ .  $\square$

At the weak feature size the convexity defect function satisfies  $h_{\mathbf{M}}(t) = t$ . The stability given by item 3 of the properties of  $h_{\mathbf{X}}$  given after Definition 2.10 guarantees that the graph of  $h_{\widehat{\mathbf{M}}}$  lies close to that of  $h_{\mathbf{M}}$ , but this alone cannot be used to approximate the first intersection of the graph of  $h_{\mathbf{M}}$  with the diagonal. The graph of  $h_{\mathbf{M}}$  could approach the diagonal very slowly before intersecting it, so that the error in approximating an intersection time based on the graph of  $h_{\widehat{\mathbf{M}}}$  is not necessarily small.

However, we are only interested in approximating the weak feature size if it yields the reach, i.e. when  $R_{\text{wfs}} < R_{\ell}$ . Corollary 2.14 guarantees the existence of a discontinuity in  $h_{\mathbf{M}}$  at  $R_{\text{wfs}}$ ; in this case the function  $h_{\mathbf{M}}$  must jump at  $R_{\text{wfs}}$  from being bounded above by a quarter circle of radius  $R_{\ell}$  to intersecting the diagonal. This feature makes it possible to bound the error in an approximation. We begin with a simple lemma.

**Lemma 2.25.** *Fix  $R > 0$ . Let the intersection points of the line  $y = x - 6\epsilon$  and the quarter-circle  $y = R - \sqrt{R^2 - x^2}$  be  $(x_0, y_0)$  and  $(x_1, y_1)$ . Then there is some  $\epsilon_0$ , which depends only on  $R$ , so that for  $0 < \epsilon < \epsilon_0$  the bounds  $x_0 \leq \frac{25}{4}\epsilon$  and  $x_1 \geq R - \frac{\epsilon}{4}$  hold.*

*Proof.* The equation  $x - 6\epsilon = R - \sqrt{R^2 - x^2}$  can be rearranged to give the quadratic  $2x^2 - (2R + 12\epsilon)x + (36\epsilon + 12R)\epsilon = 0$  with solutions

$$x = \frac{2R + 12\epsilon \pm \sqrt{(2R - 12\epsilon)^2 - 288\epsilon^2}}{4}.$$

For sufficiently small values of  $\epsilon$ , we have the bound

$$2R - 13\epsilon \leq 2R - 12\epsilon - \frac{288\epsilon^2}{4R - 24\epsilon} \leq \sqrt{(2R - 12\epsilon)^2 - 288\epsilon^2}$$

so that the solutions  $x_0$  and  $x_1$  are bounded by

$$\begin{aligned} x_0 &\leq \frac{2R + 12\epsilon - (2R - 13\epsilon)}{4} = \frac{25}{4}\epsilon \\ x_1 &\geq \frac{2R + 12\epsilon + (2R - 13\epsilon)}{4} = R - \frac{\epsilon}{4}. \end{aligned} \quad \square$$

It is clear from the proof that for any  $\delta > 0$  there is an  $\epsilon > 0$  so that the bounds can be taken to be  $(6 + \delta)\epsilon$  and  $R_{\ell} - \delta\epsilon$ . It is sufficient to proceed with  $\delta = 1/4$  and so we will do so.

**Proposition 2.26.** *Let  $\mathbf{M}$  be such that  $R(\mathbf{M}) > R_{\min}$  and let  $\epsilon < \frac{2}{9}R_{\min}$  be a positive number small enough that the conclusion of Lemma 2.25 holds for  $R = R_{\min}$ . Let  $\widetilde{\mathbf{M}} \subseteq \mathbf{R}^D$  be such that  $H(\mathbf{M}, \widetilde{\mathbf{M}}) < \epsilon$ .*

*Now suppose further that  $\mathbf{M}$  is such that  $R_\ell - R_{\text{wfs}} > \frac{9}{4}\epsilon$ . Then the value  $r = \inf \{t \geq \frac{22}{4}\epsilon : h_{\widetilde{\mathbf{M}}}(t) \geq t - 3\epsilon\}$  satisfies the bound  $|R_{\text{wfs}} - r| \leq \epsilon$ .*

*Proof.* We first claim that  $r \leq R_{\text{wfs}} + \epsilon$ . To see this, suppose that  $R_{\text{wfs}} + \epsilon < r$ . Then, by the definition of  $r$ , either  $R_{\text{wfs}} + \epsilon < \frac{22}{4}\epsilon$ , which by the assumption on  $\epsilon$  cannot happen, or  $h_{\widetilde{\mathbf{M}}}(R_{\text{wfs}} + \epsilon) < R_{\text{wfs}} - 2\epsilon$  in which case  $R_{\text{wfs}} = h_{\mathbf{M}}(R_{\text{wfs}}) \leq h_{\widetilde{\mathbf{M}}}(R_{\text{wfs}} + \epsilon) + 2\epsilon < R_{\text{wfs}}$ , which is a contradiction.

Now let us seek a lower bound for  $r$ , which relies on the fact that  $R = R_{\text{wfs}}$ . Note that  $h_{\mathbf{M}}(r + \epsilon) \geq h_{\widetilde{\mathbf{M}}}(r) - 2\epsilon \geq r - 5\epsilon$ . If the additional inequality

$$r - 5\epsilon \geq R_\ell - \sqrt{R_\ell^2 - (r + \epsilon)^2},$$

holds, so that  $h_{\mathbf{M}}(r + \epsilon) > R_\ell - \sqrt{R_\ell^2 - (r + \epsilon)^2}$ , then by Proposition 2.13 we would have  $r + \epsilon > R = R_{\text{wfs}}$ , providing the required lower bound  $r \geq R_{\text{wfs}} - \epsilon$  and completing the proof. By Lemma 2.25, this additional inequality holds whenever

$$\frac{25}{4}\epsilon \leq r + \epsilon \leq R_\ell - \frac{\epsilon}{4}.$$

The first bound is true by the definition of  $r$ . The second follows from the upper bound for  $r$  and the gap between  $R_{\text{wfs}}$  and  $R_\ell$ :  $r \leq R_{\text{wfs}} + \epsilon \leq R_\ell - \frac{5}{4}\epsilon$ .  $\square$

## 2.6 Minimax rates for reach estimators: Upper bounds

Every submanifold has a natural uniform probability distribution given by its volume measure. We consider probability distributions with density bounded above and below with respect to this volume measure. Recall the class of manifolds  $\mathfrak{C}_{R_{\min}, \mathbf{L}}^k$  studied by [AL19]:  $d$ -dimensional compact, connected, submanifolds of  $\mathbf{R}^D$  with a lower bound on the reach and admitting a local parametrization with bounded terms in the Taylor expansion (see Definition 2.7).

**Definition 2.27.** *For  $k \geq 3$ ,  $R_{\min} > 0$ ,  $\mathbf{L} = (L_1, L_2, \dots, L_k)$  and  $0 < f_{\min} \leq f_{\max} < \infty$ , we let  $\mathcal{P}_{R_{\min}, \mathbf{L}}^k(f_{\min}, f_{\max})$  denote the set of distributions  $P$  supported on some  $\mathbf{M} \in \mathfrak{C}_{R_{\min}, \mathbf{L}}^k$  which are absolutely continuous with respect to the volume measure  $\mu_{\mathbf{M}}$ , with density  $f$  taking values  $\mu_{\mathbf{M}}$ -a.s. in  $[f_{\min}, f_{\max}]$ .*

This will be abbreviated by  $\mathcal{P}^k$  where there is no ambiguity. We define the submodels  $\mathcal{P}_\alpha^k$  to be those distributions supported on elements of  $\mathcal{M}_\alpha^k$  (the classes defined in Section 2.3). These submodels are such that  $\mathcal{P}^k = \cup_{\alpha \geq 0} \mathcal{P}_\alpha^k$ .

The following lemma shows that the uniform lower bound,  $f_{\min}$ , on the density of elements of  $\mathcal{P}^k$  provides an upper bound  $R_{\max}$  for both  $R_\ell$  and  $R_{\text{wfs}}$ , which we will use in our estimators later in the section.

**Lemma 2.28.** *There exists  $R_{\max}$  depending on  $d, f_{\min}, R_{\min}$  so that, if  $P \in \mathcal{P}^k$  has support  $\mathbf{M}$ , then  $R_\ell, R_{\text{wfs}} \leq R_{\max}$ .*

*Proof.* Due to the relationship between curvature and volume, we have, by Point (3) on [Alm86, p. 2] that  $R_\ell \leq (\text{vol } M / \omega_d)^{1/d} \leq (f_{\min} \omega_d)^{-1/d}$ , where  $\omega_d$  is the volume of the  $d$ -dimensional sphere of radius 1.

Furthermore, Aamari and Levrard have shown [AL18, Lemma 2.2] that for some constant  $C$  depending only on dimension,  $\text{diam}(\mathbf{M}) \leq C(d) f_{\min}^{-1} R_{\min}^{1-d}$ . Since  $R_{\text{wfs}} \leq \frac{1}{2} \text{diam}(\mathbf{M})$  we have  $R_{\text{wfs}} \leq \frac{1}{2} C(d) f_{\min}^{-1} R_{\min}^{1-d}$ . Setting

$$R_{\max} := \max\{(f_{\min} \omega_d)^{-1/d}, \frac{1}{2} C(d) f_{\min}^{-1} R_{\min}^{1-d}\},$$

we have the result.  $\square$

In [AL19] the authors construct an estimator  $\widehat{\mathbf{M}}$  out of polynomial patches, from a sample  $(X_1, \dots, X_n)$  of random variables with common distribution  $P \in \mathcal{P}^k$ , supported on a submanifold  $\mathbf{M} \in \mathfrak{C}_{R_{\min}, \mathbf{L}}^k$ . That estimator has the following convergence property. (Note that the  $T_i^*$  referred to below are  $i$ -linear maps from  $T_p \mathbf{M}$  to  $\mathbf{R}^D$  which are the  $i$ th order terms in the Taylor expansion of the submanifold discussed in Section 2.3.)

**Theorem 2.29** (Theorem 6 in [AL19]). *Let  $k \geq 3$ . Set*

$$\theta = \left( C_{d,k} \frac{\log(n) f_{\max}^2}{(n-1) f_{\min}^3} \right)^{1/d}$$

*for  $C_{d,k}$  large enough. If  $n$  is large enough so that  $0 < \theta \leq \frac{1}{8} \min\{R_{\min}, L_\perp^{-1}\}$  and  $\theta^{-1} \geq C_{d,k, R_{\min}, \mathbf{L}} \geq \sup_{2 \leq i \leq k} |T_i^*|_{\text{op}}$ , then with probability at least  $1 - 2(\frac{1}{n})^{\frac{k}{d}}$ , we have*

$$H(\widehat{\mathbf{M}}, \mathbf{M}) \leq C^* \theta^k$$

*for some  $C^* > 0$ . In particular, for  $n$  large enough,*

$$\sup_{P \in \mathcal{P}^k} \mathbf{E}_{P^{\otimes n}} [H(\widehat{\mathbf{M}}, \mathbf{M})] \leq C \left( \frac{\log(n)}{n-1} \right)^{k/d},$$

*where  $C = C_{d,k, R_{\min}, \mathbf{L}, f_{\min}, f_{\max}}$ .*

Note that the estimator is dependent on the value of  $\theta \approx n^{-1/d}$  to within logarithmic terms, which serves as a bandwidth. The convergence rate of this estimator is very close to the currently established lower bound for estimating the reach  $R$ , which is  $n^{-k/d}$ ; see Theorem 2.35 in Section 2.7 below.

### 2.6.1 Estimating the local reach

**Definition 2.30.** We define an estimator for  $R_\ell(\mathbf{M})$ , the local reach of a submanifold  $\mathbf{M}$ , by

$$\widehat{R}_\ell = \min \left\{ \left( 2 \frac{h_{\widehat{\mathbf{M}}}(\Delta)}{\Delta^2} \right)^{-1}, R_{\max} \right\}$$

where  $\widehat{\mathbf{M}}$  is the Aamari–Levrard estimator of  $\mathbf{M}$  as discussed at the beginning of Section 2.6 above,  $\epsilon = C^* \theta^k$  as in Theorem 2.29,  $\Delta = \epsilon^{1/3}$  if  $k = 3$ , or  $\Delta = \epsilon^{1/4}$  if  $k \geq 4$ , and  $R_{\max}$  is as in Lemma 2.28.

**Theorem 2.31.** Let  $k \geq 3$ , let  $\theta$  be as in Theorem 2.29 and set  $\epsilon = C^* \theta^k$ . Then with probability at least  $1 - 2\left(\frac{1}{n}\right)^{\frac{k}{d}}$ , we have

$$|\widehat{R}_\ell - R_\ell| \leq C_{d,k,R_{\min},\mathbf{L},f_{\min}} \epsilon^{1/3},$$

and, where  $k \geq 4$ , the exponent is  $\epsilon^{1/2}$ . Moreover, for  $n$  large enough, we have

$$\sup_{P \in \mathcal{P}^k} \mathbf{E}_{P^{\otimes n}} [|\widehat{R}_\ell - R_\ell|] \leq C \left( \frac{\log(n)}{n-1} \right)^{\frac{k}{3d}},$$

or, for  $k \geq 4$ ,  $C \left( \frac{\log(n)}{n-1} \right)^{\frac{k}{2d}}$ , where  $C = C_{d,k,R_{\min},\mathbf{L},f_{\min},f_{\max}}$ .

A glance at the proof shows that we actually control  $|\widehat{R}_\ell^{-1} - R_\ell^{-1}|$  rather than  $|\widehat{R}_\ell - R_\ell|$ . This has no impact since  $R_\ell \leq R_{\max}$  is uniformly bounded and we do not seek fine control on  $C$ . Changing the parametrization  $R \mapsto 1/R$  in our statistical problem and estimating  $1/R$  instead of  $R$  would enable us to remove the projection onto  $[0, R_{\max}]$  that we use to define  $\widehat{R}_\ell$ .

*Proof.* By construction,  $\widehat{R}_\ell \leq R_{\max}$ , and it is also clear that

$$\left| \frac{1}{\widehat{R}_\ell} - \frac{1}{R_\ell} \right| \leq \left| 2 \frac{h_{\widehat{\mathbf{M}}}(\Delta)}{\Delta^2} - \frac{1}{R_\ell} \right|.$$

We derive

$$|\widehat{R}_\ell - R_\ell| = \widehat{R}_\ell R_\ell \left| \frac{1}{\widehat{R}_\ell} - \frac{1}{R_\ell} \right| \leq R_{\max}^2 \left| 2 \frac{h_{\widehat{\mathbf{M}}}(\Delta)}{\Delta^2} - \frac{1}{R_\ell} \right|.$$

The first statement of Theorem 2.31 is then a straightforward consequence of Proposition 2.24 together with Theorem 2.29. Next, we have

$$\begin{aligned} & \mathbf{E}_{P^{\otimes n}} [|\widehat{R}_\ell - R_\ell|] \\ & \leq C_{d,k,R_{\min},f_{\min},\mathbf{L}} \epsilon^{1/3} + 2R_{\max} P^{\otimes n} (|\widehat{R}_\ell - R_\ell| > C_{d,k,R_{\min},f_{\min},\mathbf{L}} \epsilon^{1/3}) \\ & \leq C_{d,k,R_{\min},f_{\min},\mathbf{L}} \epsilon^{1/3} + 4R_{\max} n^{-k/d} \end{aligned}$$

thanks to the first part of Theorem 2.31. This term is of order  $(\log n/n)^{k/3d}$ . For  $k \geq 4$ , we have the improvement to the exponent  $\epsilon^{1/2}$  and the order becomes  $(\log n/n)^{k/2d}$ , which establishes the second part of the theorem for all values of  $k \geq 3$  and completes the proof.  $\square$

For  $k = 3, 4$ , then, the constructed estimator is optimal up to a  $\log(n)$  factor as follows from Theorem 2.35 below.

### 2.6.2 Estimating the global reach

By the earlier discussion, it is not possible to give a convergence guarantee when estimating the weak feature size, i.e. the first positive critical value of  $d_M$ . However, in the case where  $R = R_{\text{wfs}}$ , that is, when  $R_{\text{wfs}} < R_\ell$ , this is possible. Accordingly, we now define an estimator for  $R_{\text{wfs}}$  and hence an estimator for the reach itself.

**Definition 2.32.** We define an estimator for  $R_{\text{wfs}}$ , the weak feature size of a submanifold  $M$ , by

$$\widehat{R}_{\text{wfs}} = \min \left\{ \inf \left\{ t \in \mathbf{R} : \frac{22}{4}\epsilon < t, \ h_{\widehat{M}}(t) \geq t - 3\epsilon \right\}, R_{\max} \right\},$$

where  $\widehat{M}$  is the Aamari–Levrard estimator of  $M$  as discussed at the beginning of Section 2.6 above,  $\epsilon = C^* \theta^k$  as in Theorem 2.29 and  $R_{\max}$  is as in Lemma 2.28.

Our estimator for the reach is then the lesser of the two individual estimators.

**Definition 2.33.** Let  $C^*, \theta$  be as in Theorem 2.29 and set  $\epsilon = C^* \theta^k$ . We define an estimator for  $R(M)$ , the reach of a submanifold  $M$ , by

$$\widehat{R} = \min \left\{ \widehat{R}_{\text{wfs}}, \widehat{R}_\ell \right\}.$$

Note that we could just as well use  $\widehat{R}_\ell$  in place of  $R_{\max}$  to cap the value of  $\widehat{R}_{\text{wfs}}$ , since we do not analyse the error in the case  $\widehat{R}_\ell < \widehat{R}_{\text{wfs}}$ . However, Definition 2.32 appears more natural as a stand-alone estimator of  $R_{\text{wfs}}$ .

**Theorem 2.34.** Let  $k \geq 3$ , let  $C^*, \theta$  be as in Theorem 2.29, and set  $\epsilon = C^* \theta^k$ , with  $\epsilon$  such that  $\frac{22}{4}\epsilon < \min(R_{\min}, 1)$ , which is always satisfied for large enough  $n \geq 1$ . Then with probability at least  $1 - 4n^{-k/d}$ , we have

$$|\widehat{R} - R| \leq C_{d,k,R_{\min},\mathbf{L}} \epsilon^{1/3},$$

and, where  $k \geq 4$ , the exponent is  $\epsilon^{1/2}$ . In particular, for  $n$  large enough,

$$\sup_{P \in \mathcal{P}^k} \mathbf{E}_{P^{\otimes n}} [|\widehat{R} - R|] \leq C \left( \frac{\log(n)}{n-1} \right)^{\frac{k}{3d}},$$

or, for  $k \geq 4$ ,  $C \left( \frac{\log(n)}{n-1} \right)^{\frac{k}{2d}}$ , where  $C = C_{d,k,\tau_{\min},\mathbf{L},f_{\min},f_{\max}}$ .

*Proof.* We will prove the result in three steps. In Step 1 we provide a bound in the case  $\widehat{R}_\ell < \widehat{R}_{\text{wfs}}$  which holds with high probability. Then in Step 2 we provide a bound in the complementary case  $\widehat{R}_\ell \geq \widehat{R}_{\text{wfs}}$ . Finally, in Step 3, we combine the two bounds, proving the first statement, and use it to obtain the bound on the expected loss. In the following, we use the letters  $C$  and  $C'$  to denote positive numbers that do not depend on  $n$  and that may vary at each occurrence.

*Step 1).* We have

$$\begin{aligned} |\widehat{R} - R| \mathbf{1}_{\{\widehat{R}_\ell < \widehat{R}_{\text{wfs}}\}} &= |\widehat{R}_\ell - \min(R_\ell, R_{\text{wfs}})| \mathbf{1}_{\{\widehat{R}_\ell < \widehat{R}_{\text{wfs}}\}} \\ &\leq |\widehat{R}_\ell - R_\ell| + |\widehat{R}_\ell - R_{\text{wfs}}| \mathbf{1}_{(R_{\text{wfs}} < R_\ell)} \mathbf{1}_{\{\widehat{R}_\ell < \widehat{R}_{\text{wfs}}\}} \\ &\leq 2|\widehat{R}_\ell - R_\ell| + |R_\ell - R_{\text{wfs}}| \mathbf{1}_{(R_{\text{wfs}} < R_\ell)} \mathbf{1}_{\{\widehat{R}_\ell < \widehat{R}_{\text{wfs}}\}} \end{aligned}$$

by triangle inequality. For  $C_1, C_2 > 0$ , introduce the events

$$\Omega_1 = \left\{ |\widehat{R}_\ell - R_\ell| \leq C_1 \epsilon^{1/3} \right\} \quad \text{and} \quad \Omega_2 = \left\{ H(\widehat{\mathbf{M}}, \mathbf{M}) \leq \epsilon \right\}.$$

On  $\{\widehat{R}_\ell < \widehat{R}_{\text{wfs}}\}$ , we have

$$\forall t \in \left[ \frac{22}{4} \epsilon, \widehat{R}_\ell \right] : h_{\widehat{\mathbf{M}}}(t) < t - 3\epsilon,$$

therefore, on  $\{\widehat{R}_\ell < \widehat{R}_{\text{wfs}}\} \cap \Omega_1$ , we infer that

$$\text{for all } t \in \left[ \frac{22}{4} \epsilon, R_\ell - C_1 \epsilon^{1/3} \right] : h_{\widehat{\mathbf{M}}}(t) < t - 3\epsilon.$$

By item 3 of the properties of the convexity defect function given after Definition 2.10, on  $\Omega_2$ , we have

$$h_{\widehat{\mathbf{M}}}(t) \geq h_{\mathbf{M}}(t - \epsilon) - 2\epsilon.$$

Putting the last two estimates together, we obtain on  $\{\widehat{R}_\ell < \widehat{R}_{\text{wfs}}\} \cap \Omega_1 \cap \Omega_2$  the bound

$$\forall t \in \left[ \frac{22}{4} \epsilon, R_\ell - C_1 \epsilon^{1/3} \right] : h_{\mathbf{M}}(t - \epsilon) < t - 3\epsilon + 2\epsilon$$

or equivalently

$$\forall t \in \left[ \left( \frac{22}{4} - 1 \right) \epsilon, R_\ell - C_1 \epsilon^{1/3} - \epsilon \right] : h_{\mathbf{M}}(t) < t.$$

Therefore  $h_{\mathbf{M}}(t) < t$  for  $t \leq R_\ell - C_1 \epsilon^{1/3} - \epsilon$  and this implies in turn

$$R_{\text{wfs}} \geq R_\ell - C_1 \epsilon^{1/3} - \epsilon.$$

We have thus proved

$$|R_\ell - R_{\text{wfs}}| \mathbf{1}_{(R_{\text{wfs}} < R_\ell)} \mathbf{1}_{\{\widehat{R}_\ell < \widehat{R}_{\text{wfs}}\}} \mathbf{1}_{\Omega_1 \cap \Omega_2} \leq (C_1 \epsilon^{1/3} + \epsilon) \leq C \epsilon^{1/3}.$$

Finally

$$|\widehat{R} - R| \mathbf{1}_{\{\widehat{R}_\ell < \widehat{R}_{\text{wfs}}\}} \mathbf{1}_{\Omega_1 \cap \Omega_2} \leq C \epsilon^{1/3}.$$

Step 2). We have

$$|\widehat{R} - R| \mathbf{1}_{\{\widehat{R}_\ell \geq \widehat{R}_{\text{wfs}}\}} \leq T_1 + T_2 + T_3,$$

with

$$\begin{aligned} T_1 &= |\widehat{R}_{\text{wfs}} - R_{\text{wfs}}| \mathbf{1}_{\left(R_{\text{wfs}} + \frac{9}{4}\epsilon < R_\ell\right)} \mathbf{1}_{\{\widehat{R}_\ell \geq \widehat{R}_{\text{wfs}}\}}, \\ T_2 &= |\widehat{R}_{\text{wfs}} - R_{\text{wfs}}| \mathbf{1}_{\left(R_{\text{wfs}} \leq R_\ell < R_{\text{wfs}} + \frac{9}{4}\epsilon\right)} \mathbf{1}_{\{\widehat{R}_\ell \geq \widehat{R}_{\text{wfs}}\}}, \\ T_3 &= |\widehat{R}_{\text{wfs}} - R_\ell| \mathbf{1}_{(R_\ell < R_{\text{wfs}})} \mathbf{1}_{\{\widehat{R}_\ell \geq \widehat{R}_{\text{wfs}}\}}. \end{aligned}$$

By Proposition 2.26, we have  $T_1 \leq \epsilon$  on  $\Omega_2$ . We turn to the term  $T_2$ . We have

$$h_{\widehat{\mathbf{M}}}(\widehat{R}_{\text{wfs}}) \geq \widehat{R}_{\text{wfs}} - 3\epsilon$$

on  $\{\widehat{R}_\ell \geq \widehat{R}_{\text{wfs}}\}$  by construction. Thanks to item 3 of the properties of the convexity defect function given after Definition 2.10, we also have

$$h_{\widehat{\mathbf{M}}}(\widehat{R}_{\text{wfs}}) \leq h_{\mathbf{M}}(\widehat{R}_{\text{wfs}} + \epsilon) + 2\epsilon \quad \text{on } \Omega_2$$

therefore

$$\widehat{R}_{\text{wfs}} - 5\epsilon \leq h_{\mathbf{M}}(\widehat{R}_{\text{wfs}} + \epsilon)$$

holds true on  $\{\widehat{R}_\ell \geq \widehat{R}_{\text{wfs}}\} \cap \Omega_2$ . Introduce now the event

$$\Omega_3 = \{\widehat{R}_{\text{wfs}} + \epsilon < R_{\text{wfs}}\}.$$

By Proposition 2.13, it follows that

$$\widehat{R}_{\text{wfs}} - 5\epsilon \leq R_\ell - \sqrt{R_\ell^2 - (\widehat{R}_{\text{wfs}} + \epsilon)^2}$$

on  $\{\widehat{R}_\ell \geq \widehat{R}_{\text{wfs}}\} \cap \Omega_2 \cap \Omega_3$ . Solving this inequality when  $R_\ell > \widehat{R}_{\text{wfs}} + \epsilon$  yields  $\widehat{R}_{\text{wfs}} \geq R_\ell - C\epsilon$  for some  $C > 0$  that depends on  $R_\ell$  only. Otherwise,  $R_\ell - \epsilon \leq \widehat{R}_{\text{wfs}}$  directly. Replacing  $C$  by  $\max\{1, C\}$ , we infer

$$R_\ell - C\epsilon \leq \widehat{R}_{\text{wfs}} \leq \widehat{R}_\ell \leq R_\ell + C_1\epsilon^{1/3}$$

on  $\{\widehat{R}_\ell \geq \widehat{R}_{\text{wfs}}\} \cap \Omega_1 \cap \Omega_2 \cap \Omega_3$  hence  $|\widehat{R}_{\text{wfs}} - R_\ell| \leq C\epsilon^{1/3}$  on that event. Combining this estimate with the condition  $|R_\ell - R_{\text{wfs}}| \leq \frac{9}{4}\epsilon$  in the definition of  $T_2$  implies

$$|\widehat{R}_{\text{wfs}} - R_{\text{wfs}}| \leq C\epsilon^{1/3} + \frac{9}{4}\epsilon.$$

We have thus proved

$$T_2 \mathbf{1}_{\cap_{i=1}^3 \Omega_i} \leq C\epsilon^{1/3} + \frac{9}{4}\epsilon \leq C'\epsilon^{1/3}.$$

On the complementary event  $\Omega_3^c = \{\widehat{R}_{\text{wfs}} + \epsilon \geq R_{\text{wfs}}\}$ , we have, on the one hand

$$R_{\text{wfs}} - \widehat{R}_{\text{wfs}} \leq \epsilon.$$



But on the other hand, on  $\{\widehat{R}_\ell \geq \widehat{R}_{\text{wfs}}\} \cap \Omega_1$ , we have

$$\begin{aligned} \widehat{R}_{\text{wfs}} - R_{\text{wfs}} &\leq \widehat{R}_\ell - R_{\text{wfs}} \\ &\leq R_\ell - R_{\text{wfs}} + C_1 \epsilon^{1/3} \\ &\leq \frac{9}{4} \epsilon + C_1 \epsilon^{1/3} \leq C \epsilon^{1/3} \end{aligned}$$

thanks to the condition  $|R_\ell - R_{\text{wfs}}| \leq \frac{9}{4} \epsilon$  in the definition of  $T_2$ . Combining these bounds, we obtain

$$T_2(1 - \mathbf{1}_{\Omega_3})\mathbf{1}_{\Omega_1} \leq C \epsilon^{1/3}.$$

Putting together this estimate and the bound  $T_2 \mathbf{1}_{\bigcap_{i=1}^3 \Omega_i} \leq C \epsilon^{1/3}$  we established previously, we derive

$$T_2 \mathbf{1}_{\Omega_1 \cap \Omega_2} \leq C \epsilon^{1/3}.$$

We finally turn to the term  $T_3$ . On  $\{\widehat{R}_{\text{wfs}} \geq R_\ell\}$  intersected with  $\{\widehat{R}_\ell \geq \widehat{R}_{\text{wfs}}\} \cap \Omega_1$ , we have

$$0 < R_\ell \leq \widehat{R}_{\text{wfs}} \leq \widehat{R}_\ell \leq R_\ell + C_1 \epsilon^{1/3}$$

which yields the estimate

$$|\widehat{R}_{\text{wfs}} - R_\ell| \leq C_1 \epsilon^{1/3} \text{ on } \{\widehat{R}_{\text{wfs}} \geq R_\ell\} \cap \{\widehat{R}_\ell \geq \widehat{R}_{\text{wfs}}\} \cap \Omega_1.$$

Alternatively, on the complementary event  $\{\widehat{R}_{\text{wfs}} < R_\ell\}$  intersected with  $\{\widehat{R}_\ell \geq \widehat{R}_{\text{wfs}}\} \cap \Omega_2$  we have  $\widehat{R}_{\text{wfs}} - 5\epsilon \leq R_\ell - \sqrt{R_\ell^2 - (\widehat{R}_{\text{wfs}} + \epsilon)^2}$  in the same way as for the term  $T_2$ , provided  $\widehat{R}_{\text{wfs}} + \epsilon < R_\ell$ . This implies  $\widehat{R}_{\text{wfs}} \geq R_\ell - C\epsilon$ . Otherwise  $\widehat{R}_{\text{wfs}} + \epsilon \geq R_\ell$  holds true. In any event, we obtain  $-C\epsilon \leq \widehat{R}_{\text{wfs}} - R_\ell$ . Since  $\widehat{R}_{\text{wfs}} - R_\ell \leq C_1 \epsilon^{1/3}$  on  $\Omega_1$ , we conclude

$$|\widehat{R}_{\text{wfs}} - R_\ell| \leq \epsilon + C_1 \epsilon^{1/3} \leq C \epsilon^{1/3} \text{ on } \{\widehat{R}_{\text{wfs}} < R_\ell\} \cap \{\widehat{R}_\ell \geq \widehat{R}_{\text{wfs}}\} \cap \Omega_1 \cap \Omega_2.$$

Combining these two bounds for  $|\widehat{R}_{\text{wfs}} - R_\ell|$ , we finally derive

$$T_3 \mathbf{1}_{\Omega_1 \cap \Omega_2} \leq C \epsilon^{1/3}.$$

Putting together our successive estimates for  $T_1, T_2$  and  $T_3$ , we have proved

$$|\widehat{R} - R| \mathbf{1}_{\{\widehat{R}_\ell \geq \widehat{R}_{\text{wfs}}\}} \mathbf{1}_{\Omega_1 \cap \Omega_2} \leq \epsilon + 2C \epsilon^{1/3} \leq C' \epsilon^{1/3}.$$

*Step 3).* Combining Step 1) and Step 2) yields

$$|\widehat{R} - R| \mathbf{1}_{\Omega_1 \cap \Omega_2} \leq C \epsilon^{1/3}.$$

By Theorem 2.31, we have  $P^{\otimes n}(\Omega_1) \geq 1 - 2n^{-k/d}$  as soon as  $C_1 \geq C_{d,k,R_{\min},f_{\min},L}$ . By Theorem 2.29, we have  $P^{\otimes n}(\Omega_2) \geq 1 - 2n^{-k/d}$ . The first estimate in Theorem 2.34 follows for  $k \geq 3$ . The improvement in the case  $k = 4$  is done in exactly the same way and we omit it.

Finally, integrating, we obtain

$$\begin{aligned}\mathbb{E}_{P^{\otimes n}}[|\widehat{R} - R|] &\leq C\epsilon^{1/3} + 2R_{\max}(P^{\otimes n}(\Omega_1^c) + P^{\otimes n}(\Omega_2^c)) \\ &\leq C\epsilon^{1/3} + 4R_{\max}n^{-k/d} \leq C'\epsilon^{1/3}\end{aligned}$$

and the second statement of Theorem 2.34 is proved for  $k \geq 3$ . The improvement in the case  $k = 4$  follows in similar fashion.  $\square$

## 2.7 Minimax rates for reach estimators: Lower bounds

We fix  $R_{\min}$ ,  $\mathbf{L}$ ,  $k$ ,  $f_{\min}$  and  $f_{\max}$  and recall the classes  $\mathcal{P}_{\alpha}^k$  which were defined in Section 2.6, parametrized by the gap  $\alpha \leq R_{\ell} - R_{\text{wfs}}$ . These sub-models are such that  $\mathcal{P}^k = \cup_{\alpha \geq 0} \mathcal{P}_{\alpha}^k$ .

**Theorem 2.35.** *If  $f_{\min}$  is small enough and  $f_{\max}$ ,  $\mathbf{L}$  are large enough (depending on  $R_{\min}$ , and on  $\alpha$  for the second statement), then we have the following lower bounds on the reach estimation problem*

$$\begin{aligned}\liminf_{n \rightarrow \infty} n^{(k-2)/d} \inf_{\widehat{R}} \sup_{P \in \mathcal{P}_0^k} \mathbb{E}_{P^{\otimes n}}[|\widehat{R} - R|] &\geq C_0 > 0 \quad \text{and} \\ \liminf_{n \rightarrow \infty} n^{k/d} \inf_{\widehat{R}} \sup_{P \in \mathcal{P}_{\alpha}^k} \mathbb{E}_{P^{\otimes n}}[|\widehat{R} - R|] &\geq C_{\alpha} > 0 \quad \forall \alpha > 0\end{aligned}$$

with  $C_0$  depending on  $R_{\min}$  and  $C_{\alpha}$  depending on  $R_{\min}$  and  $\alpha$ .

In particular, the minimax rate on the whole model  $\mathcal{P}^k$  is of order  $n^{-\frac{k-2}{d}}$ . To show the latter proposition, we will make use of Le Cam's Lemma, restated in our context.

**Lemma 2.36** (Le Cam Lemma, [Yu97]). *For any two  $P_1, P_2 \in \mathcal{P}$ , where  $\mathcal{P}$  is a model of manifold-supported probability measures, we have*

$$\inf_{\widehat{R}} \sup_{P \in \mathcal{P}} \mathbb{E}_{P^{\otimes n}}[|\widehat{R} - R|] \geq \frac{1}{2} |R_1 - R_2| (1 - \text{TV}(P_1, P_2))^n,$$

where  $\text{TV}$  denotes the total variation distance between measures and  $R_1$  (respectively  $R_2$ ) denotes the reach of the support of  $P_1$  (resp  $P_2$ ).

Therefore, one needs to compute the total variation distance between two given manifold-supported measures. When these measures are uniform over their support, we have the following convenient formula.

**Lemma 2.37.** *Let  $M_1, M_2$  be two compact  $d$ -dimensional submanifolds of  $\mathbb{R}^D$  and let  $P_1, P_2$  be the uniform distributions over  $M_1$  and  $M_2$ . Then we have*

$$\text{TV}(P_1, P_2) = \frac{\mathcal{H}^d(M_2 \setminus M_1)}{\text{vol } M_2} \quad \text{if} \quad \text{vol } M_2 \geq \text{vol } M_1,$$

where  $\mathcal{H}^d$  denotes the  $d$ -dimensional Hausdorff measure on  $\mathbb{R}^D$ .

*Proof.* First note that  $P_1$  and  $P_2$  are absolutely continuous with respect to  $\mathcal{H}^d$  with densities  $\frac{1}{\text{vol } M_1} \mathbb{1}_{M_1}$  and  $\frac{1}{\text{vol } M_2} \mathbb{1}_{M_2}$  respectively. Therefore, we have the following chain of equalities.

$$\begin{aligned}
 & \text{TV}(P_1, P_2) \\
 &= \frac{1}{2} \int \left| \frac{1}{\text{vol } M_1} \mathbb{1}_{M_1} - \frac{1}{\text{vol } M_2} \mathbb{1}_{M_2} \right| d\mathcal{H}^d \\
 &= \frac{\mathcal{H}^d(M_1 \setminus M_2)}{2 \text{vol } M_1} + \frac{\mathcal{H}^d(M_2 \setminus M_1)}{2 \text{vol } M_2} + \frac{1}{2} \mathcal{H}^d(M_1 \cap M_2) \left( \frac{1}{\text{vol } M_1} - \frac{1}{\text{vol } M_2} \right) \\
 &= \frac{1}{2} \left\{ 1 + \frac{\mathcal{H}^d(M_2 \setminus M_1) - \mathcal{H}^d(M_1 \cap M_2)}{\text{vol } M_2} \right\} \\
 &= \frac{\mathcal{H}^d(M_2 \setminus M_1)}{\text{vol } M_2}. \quad \square
 \end{aligned}$$

Before proving Theorem 2.35 we need to introduce the following technical result:

**Lemma 2.38.** *Let  $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}$  be a smooth function and let  $M = \{(v, \Phi(v)) \mid v \in \mathbb{R}^d\} \subseteq \mathbb{R}^{d+1}$  be its graph. The second fundamental form of  $M$  at the point  $x = (v, \Phi(v)) \in M$  is given by*

$$\Pi_x(u, w) = \frac{d^2\Phi(v)[\text{pr}(u), \text{pr}(w)]}{\sqrt{1 + \|\nabla\Phi(v)\|^2}}, \quad \text{for all } u, w \in T_x M$$

where  $\text{pr}$  is the linear projection to  $\mathbb{R}^d \subseteq \mathbb{R}^{d+1}$ .

*Proof.* We define  $\Psi : v \in \mathbb{R}^d \mapsto (v, \Phi(v)) \in \mathbb{R}^{d+1}$  so that  $M$  is the image of  $\mathbb{R}^d$  through the diffeomorphism  $\Psi$ . Let  $x \in M$  and let  $v \in \mathbb{R}^d$  be such that  $x = \Psi(v)$ . The tangent space  $T_x M$  is given by  $T_x M = \{d\Psi(v)[h] = (h, \langle h, \nabla\Phi(v) \rangle) \mid h \in \mathbb{R}^d\}$ , so that a normal vector field on  $M$  is given by

$$n(x) = \left( -\frac{\nabla\Phi(v)}{\sqrt{1 + \|\nabla\Phi(v)\|^2}}, \frac{1}{\sqrt{1 + \|\nabla\Phi(v)\|^2}} \right) \in \mathbb{R}^{d+1}.$$

For  $u \in T_x M$ , where  $h = \text{pr } u$ , we have

$$dn(x)[u] = \left( -\frac{H\Phi(v)h}{\sqrt{1 + \|\nabla\Phi(v)\|^2}}, 0 \right) - \frac{\langle H\Phi(v)h, \nabla\Phi(v) \rangle}{1 + \|\nabla\Phi(v)\|^2} n(x),$$

where  $H\Phi$  denotes the Hessian of  $\Phi$ . Now for  $w \in T_x M$  and  $\eta = \text{pr } w$ , we have

$$\begin{aligned}
 \Pi_x(u, w) &= -\langle dn(x)[u], w \rangle = \left\langle \left( -\frac{H\Phi(v)h}{\sqrt{1 + \|\nabla\Phi(v)\|^2}}, 0 \right), (\eta, \langle \eta, \nabla\Phi(v) \rangle) \right\rangle \\
 &= \left\langle \frac{H\Phi(v)h}{\sqrt{1 + \|\nabla\Phi(v)\|^2}}, \eta \right\rangle = \frac{d^2\Phi(v)[h, \eta]}{\sqrt{1 + \|\nabla\Phi(v)\|^2}}
 \end{aligned}$$

concluding the proof. □

We are now ready to prove Theorem 2.35.

*Proof of Theorem 2.35. Step 1: The case of  $\mathcal{P}_0^k$ .* Let  $M$  be the  $d$ -dimensional sphere in  $\mathbb{R}^{d+1}$  of radius  $r$  centered at  $-re_{d+1}$ , where  $e_{d+1} = (0, \dots, 0, 1)$ . We choose  $r$  to be such that  $r \geq 2R_{\min}$ . Since  $M$  is smooth, there exists  $\mathbf{L}^* \in \mathbb{R}^{k-2}$  (depending on  $r$ ) such that  $M \in \mathfrak{C}_{r, \mathbf{L}^*}^k$  and thus the uniform probability  $P$  on  $M$  is in  $\mathcal{P}_{r, \mathbf{L}^*}^k(a^*, a^*)$  (see Definition 2.27) with  $a^* = (r^d s_d)^{-1}$  and  $s_d$  being the volume of the unit  $d$ -dimensional sphere.

Let us now perturb  $M$  to  $M_\gamma$ , as illustrated in Figure 2.3. Define for any  $\gamma > 0$

$$\Phi_\gamma : \begin{cases} \mathbb{R}^{d+1} \rightarrow \mathbb{R}^{d+1} \\ z \mapsto z + \gamma^k \Psi(z/\gamma) e_{d+1}, \end{cases}$$

where  $\Psi(z) = \psi(\|z\|)$  and where  $\psi : \mathbb{R} \rightarrow \mathbb{R}$  is a smooth, even, non-trivial, positive map supported on  $[-1, 1]$ , decreasing on  $[0, 1]$ , and with  $\psi''(0) < 0$ . The above map is a global diffeomorphism as soon as  $\gamma^{k-1} \|\mathrm{d}\Psi\|_{\mathrm{op}, \infty} < 1$ . Moreover, we have  $\|\mathrm{d}\Phi_\gamma - I_D\|_{\mathrm{op}, \infty} = \gamma^{k-1} \|\mathrm{d}\Psi\|_{\mathrm{op}, \infty}$  and  $\|\mathrm{d}^j \Phi_\gamma\|_{\mathrm{op}, \infty} \leq \gamma^{k-j} \|\mathrm{d}^j \Psi\|$ , so that, provided  $\|\mathrm{d}^k \Psi\|$  is chosen small enough (depending on  $r$ ) and that  $\gamma$  is small enough (depending again on  $r$ ), then we can apply Proposition A.5 from the supplementary material in [AL19] to show that the submanifold  $M_\gamma = \Phi_\gamma(M)$  is in  $\mathfrak{C}_{r/2, 2\mathbf{L}^*}^k$ .

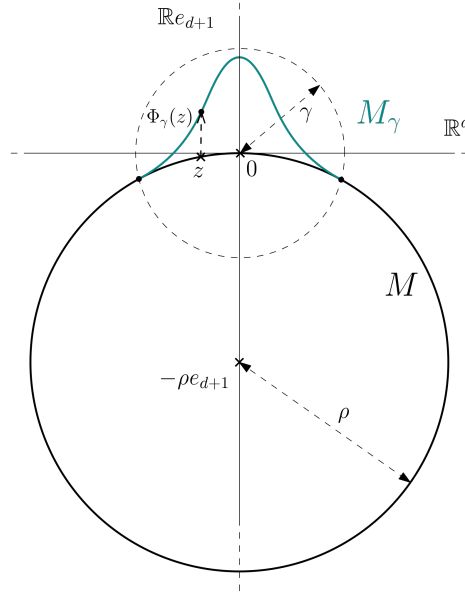


Figure 2.3 – The submanifolds  $M$  and  $M_\gamma$  used in the proof of the first part of the lower bound.

Then we have

$$\mathrm{vol} M_\gamma = \int_{M_\gamma} \mathrm{d} \mathrm{vol}_{M_\gamma}(x) = \int_M |\det \mathrm{d}\Phi_\gamma(z)| \mathrm{d} \mathrm{vol}_M(z).$$

Since  $\det d\Phi_\gamma(z) = 1 + \gamma^{k-1} \langle e_{d+1}, \nabla \Psi(z/\gamma) \rangle$  with  $\langle e_{d+1}, \nabla \Psi(z/\gamma) \rangle \geq 0$  (because  $\langle z, e_{d+1} \rangle \leq 0$ ), it follows that  $\text{vol } M \leq \text{vol } M_\gamma$  and that  $\text{vol } M_\gamma \leq 2 \text{vol } M$  for  $\gamma$  small enough (depending again on  $r$ ) so that the uniform distribution  $P_\gamma$  on  $M_\gamma$  is in  $\mathcal{P}_{r/2, 2L^*}^k(a^*/2, a^*)$ . If we assume that  $2L^* \leq L$ ,  $f_{\min} \leq a^*/2$  and  $a^* \leq f_{\max}$  (which we do from now on) then we immediately have  $P \in \mathcal{P}_0^k$  and  $P_\gamma \in \mathcal{P}_0^k$ , provided that  $R_{\text{wfs}}(M_\gamma) \geq R_\ell(M_\gamma)$ . We claim that the latter inequality holds.

Around 0, simple geometrical considerations show that  $M_\gamma$  can be viewed as the graph of the function

$$\xi_\gamma : \begin{cases} \mathbb{R}^d \rightarrow \mathbb{R} \\ v \mapsto \sqrt{r^2 - \|v\|^2} - r + \gamma^k \psi\left(\frac{r}{\gamma} \sqrt{2 - 2\sqrt{1 - \|v\|^2/r^2}}\right). \end{cases}$$

Writing  $\xi_\gamma(v) = \zeta_\gamma(\|v\|)$  with  $\zeta_\gamma : \mathbb{R} \rightarrow \mathbb{R}$ , a series of computations shows that

$$\zeta_\gamma''(0) = -\frac{1}{r} + r\gamma^{k-2}\psi''(0).$$

Setting  $c = -\psi''(0) > 0$  (which depends on  $r$ ) we have, using Lemma 2.38,

$$R_\ell(M_\gamma) \leq \frac{1}{|\zeta_\gamma''(0)|} = \frac{1}{\frac{1}{r} + cr\gamma^{k-2}} \leq r - \frac{1}{2}cr^2\gamma^{k-2}$$

as soon as  $cr^2\gamma^{k-2} \leq 1$ . Now let us turn to the control of  $R_{\text{wfs}}(M_\gamma)$ . We will show that the distance between any pair of bottleneck points is bounded below by  $2r$ . Let  $(x, y) \in M_\gamma$  be a pair of bottleneck points. First notice that  $x$  and  $y$  cannot lie simultaneously in  $B(0, \gamma)$  because  $M_\gamma \cap B(0, \gamma)$  can be seen as a graph. If  $x, y \in M_\gamma \setminus B(0, \gamma)$ , then  $d(x, y) = 2r$  necessarily. If, say,  $x \in B(0, \gamma)$  and  $y \in M_\gamma \setminus B(0, \gamma)$ , then the open segment  $(x, y)$  cross  $M$  at a single point  $z \in M$ . Therefore, we have that  $d(x, y) = d(x, z) + d(z, y)$ . But now since  $[x, y]$  is normal to  $M_\gamma$  at point  $y$ , we know that  $[z, y]$  is a diameter of  $M$  so that  $d(z, y) = 2r$  and thus  $d(x, y) \geq 2r$ . We have shown that  $R_{\text{wfs}}(M_\gamma) \geq r \geq R_\ell(M_\gamma)$  for  $\gamma$  small enough and thus  $M_\gamma \in \mathcal{M}_0^k$  and  $P_\gamma \in \mathcal{P}_0^k$ .

Now, by Lemma 2.37, we have that  $\text{TV}(P, P_\gamma) = \mathcal{H}^d(M_\gamma \setminus M) / \text{vol } M_\gamma \leq C\gamma^d$  for some constant  $C$  depending on  $r$ . Applying now Le Cam's Lemma (Lemma 4.A.4) and noting that  $R(M) - R(M_\gamma) \geq cr^2\gamma^{k-2}$ , we obtain

$$\inf_{\widehat{R}} \sup_{P \in \mathcal{P}_0^k} \mathbb{E}_{P^{\otimes n}}[|\widehat{R} - R|] \geq \frac{1}{2}cr^2\gamma^{k-2} \times (1 - C\gamma^d)^n.$$

Setting  $\gamma = 1/(Cn)^{1/d}$ , we know that for  $n$  large enough (depending on  $r$ ), we have

$$\inf_{\widehat{R}} \sup_{P \in \mathcal{P}_0^k} \mathbb{E}_{P^{\otimes n}}[|\widehat{R} - R|] \geq \frac{1}{8}cr^2(Cn)^{-(k-2)/d}.$$

Set  $r$  to be equal to  $2R_{\min}$  and the first statement of Theorem 2.35 follows.

*Step 2: The case of  $\mathcal{P}_\alpha^k$ .* We next turn to the second part of the theorem. We fix  $\alpha > 0$  and construct a manifold  $M \in \mathfrak{C}^k$  as follows. We consider the two parallel disks  $B(0, 2r) \subseteq \mathbb{R}^d \subseteq \mathbb{R}^{d+1}$  and  $B(2re_{d+1}, 2r) \subseteq 2re_{d+1} + \mathbb{R}^d \subseteq \mathbb{R}^{d+1}$ , with  $r \geq 2R_{\min}$ , and link them together so that  $M$  satisfies the following:

- $M$  is a smooth submanifold of  $\mathbb{R}^{d+1}$ ,
- $M$  has reach  $r$ , and  $(0, 2re_{d+1})$  is a reach attaining pair,
- $R_\ell(M) \geq r + \alpha$ .

See Figure 2.4 for a schematic notion of such  $M$ , visualized with  $d = 1$ .

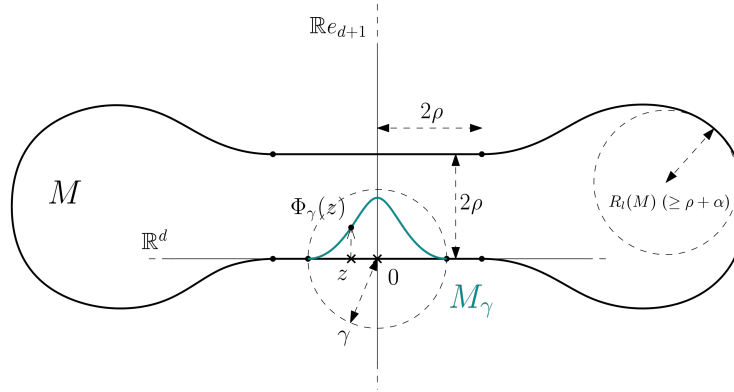


Figure 2.4 – The submanifolds  $M$  and  $M_\gamma$  used in the proof of the second part of the lower bound.

Furthermore, we know that there exists  $\mathbf{L}^*$  (depending on  $r$  and  $\alpha$ ) such that  $M \in \mathfrak{C}_{r, \mathbf{L}^*}^k$  and  $P \in \mathcal{P}_{r, \mathbf{L}^*}^k(a^*, a^*)$  where  $a^* = 1/\text{vol } M$  and where  $P$  is the uniform probability over  $M$ . We again consider the map

$$\Phi_\gamma : \begin{cases} \mathbb{R}^{d+1} \rightarrow \mathbb{R}^{d+1} \\ z \mapsto z + \gamma^k \Psi(z/\gamma) e_{d+1}. \end{cases}$$

Similarly to the first part of the theorem, for  $\gamma$  small enough (depending on  $\alpha$  and  $r$ ), we know that  $M_\gamma = \Phi_\gamma(M)$  is a smooth submanifold in  $\mathfrak{C}_{r/2, 2\mathbf{L}^*}^k$  and that the uniform distribution  $P_\gamma$  over  $M_\gamma$  lies in  $\mathcal{P}_{r/2, 2\mathbf{L}^*}^k(a^*/2, a^*)$ . Again, assuming that  $\mathbf{L} \geq 2\mathbf{L}^*$ ,  $f_{\min} \leq a^*/2$  and  $f_{\max} \geq 2a^*$ , we have that  $P \in \mathcal{P}_\alpha^k$  and, furthermore, that  $P_\gamma \in \mathcal{P}_\alpha^k$ , provided that  $R_\ell(M_\gamma) \geq R_{\text{wfs}}(M_\gamma) + \alpha$ . We claim that the latter inequality holds.

Since  $\Psi$  is maximal at 0, we know that  $(\gamma^k \psi(0) e_{d+1}, 2re_{d+1})$  is still a bottleneck pair, and thus  $R_{\text{wfs}}(M_\gamma) \leq r - c\gamma^k$  where we set  $c = -2\psi(0)$  (depending on  $\alpha$  and  $r$ ). For the curvature, notice that it is unchanged outside of  $B(0, \gamma)$  and that  $M_\gamma$  is just the graph of  $v \mapsto \gamma^k \Psi(v/\gamma)$  within this ball. Using Lemma 2.38, we thus have  $R_\ell(M_\gamma) \geq \min\{(r + \alpha), (C\gamma^{k-2})^{-1}\}$ , with  $C$  depending on  $\alpha$  and  $r$ , so that  $R_\ell(M_\gamma) \geq R_{\text{wfs}}(M_\gamma) + \alpha$  for  $\gamma$  small enough (depending on  $\alpha$  and  $r$ ), and therefore  $M_\gamma \in \mathcal{M}_\alpha^k$  and  $P_\gamma \in \mathcal{P}_\alpha^k$ .

Using Lemma 2.37, we have that  $\text{TV}(P, P_\gamma) = \mathcal{H}^d(M_\gamma \setminus M) / \text{vol } M_\gamma \leq \delta \gamma^d$  for some constant  $\delta$  depending on  $r$ . Applying now Le Cam's Lemma (Lemma 4.A.4) and noticing that  $R(M) - R(M_\gamma) \geq c\gamma^k$ , we get

$$\inf_{\widehat{R}} \sup_{P \in \mathcal{P}_0^k} \mathbb{E}_{P^{\otimes n}}[|\widehat{R} - R|] \geq \frac{1}{2} c \gamma^k \times (1 - \delta \gamma^d)^n.$$

Setting  $\gamma = 1/(\delta n)^{1/d}$ , we know that for  $n$  large enough (depending on  $r$  and  $\alpha$ ), we have

$$\inf_{\widehat{R}} \sup_{P \in \mathcal{P}_0^k} \mathbb{E}_{P^{\otimes n}}[|\widehat{R} - R|] \geq \frac{1}{8} c (\delta n)^{-k/d}.$$

Setting  $r = 2R_{\min}$  yields the result completing the proof of Theorem 2.35.  $\square$





# Chapter 3

## Reach estimation via metric learning

We study the estimation of the *reach*, an ubiquitous regularity parameter in manifold estimation and geometric data analysis. Given an i.i.d. sample over an unknown  $d$ -dimensional  $\mathcal{C}^k$ -smooth submanifold of  $\mathbb{R}^D$ , we provide optimal nonasymptotic bounds for the estimation of its reach. We build upon a formulation of the reach in terms of maximal curvature on one hand, and geodesic metric distortion on the other hand. The derived rates are adaptive, with rates depending on whether the reach of  $M$  arises from curvature or from a bottleneck structure. In the process, we derive optimal geodesic metric estimation bounds. This chapter is the result of the collaboration [ABL22], in revision.

### Contents

---

<b>3.1</b>	<b>Introduction</b>	<b>56</b>
3.1.1	Geometric Inference	56
3.1.2	Reach and Regularity	56
3.1.3	Metric Learning	57
3.1.4	Contribution and Outline	58
3.1.5	General Notation	59
<b>3.2</b>	<b>Geometric and Statistical Model</b>	<b>59</b>
<b>3.3</b>	<b>Reach and Related Quantities</b>	<b>61</b>
3.3.1	Characterizations and Relaxations of the Reach	61
3.3.2	Reach of Submanifolds	63
3.3.3	Plug-in Methods for Reach Estimation	64
<b>3.4</b>	<b>Spherical Distortion Radius</b>	<b>67</b>
3.4.1	Motivation and Definition	67
3.4.2	Stability Properties	70
<b>3.5</b>	<b>Optimal Metric Learning</b>	<b>73</b>
3.5.1	Unsupervised Distance Metric Learning	73

---

3.5.2	An optimal Approach of Metric Estimation . . . . .	75
<b>3.6</b>	<b>Optimal Reach Estimation</b> . . . . .	<b>77</b>
3.6.1	Optimal Spherical Distortion Radius Estimation . . . . .	77
3.6.2	Optimal Reach Estimation . . . . .	79
<b>3.7</b>	<b>Conclusion and Further Prospects</b> . . . . .	<b>80</b>

---

## 3.1 Introduction

### 3.1.1 Geometric Inference

Topological data analysis and geometric methods now constitute a standard toolbox in statistics and machine learning [Was18, CM21]. In this family of methods, data  $\mathbb{X}_n := \{X_1, \dots, X_n\}$  are usually seen as point clouds in high dimension, for which complex structural correlations give rise to an underlying structure that is neither full-dimensional, nor even linear. Dealing with non-linearity is very well understood through the prism of *non-parametric regression*. However, in absence of distinguished “covariate” and “response” variables (i.e. coordinates), regression does not make sense anymore. Hence, one needs to adopt a more global and coordinate-free approach: data are naturally viewed as lying on a submanifold  $M \subset \mathbb{R}^D$  of dimension  $d \ll D$ , where  $d$  corresponds to its *true* number of degrees of freedom.

This approach opens the way to the estimation of numerous geometric and topological quantities to describe data. Central to it is the manifold itself [GPPVW12b, GPPVW12a, KZ15, FILN19, Div21a, AS21, PS22], where error is most commonly measured in Hausdorff distance. Among many others, let us also mention the homology [BRS<sup>+</sup>12], persistent homology [CGLM14], differential quantities [AL19], intrinsic metric [ACC20] and regularity [AKC<sup>+</sup>19].

### 3.1.2 Reach and Regularity

Similarly to functional estimation, the theoretical study of nonparametric geometric problems naturally comes with regularity conditions. By far, the most ubiquitous regularity and scale parameter in this context is the *reach*. First introduced by H. Federer’s seminal paper [Fed59] on geometric measure theory, the reach  $\text{rch}(K) \in \mathbb{R}_+$  of a set  $K \subset \mathbb{R}^D$  measures how far  $K$  is from being convex [ALS13]. It hence provides a typical scale at which it shares most of the properties of a convex set. These properties include – among others – uniqueness of the projection map, contractibility of balls, and explicit formulas for the volume of thickenings (see [Fed59]). When  $K = M$  is a submanifold, the reach also assesses quantitatively how it deviates from its tangent spaces. Therefore, the reach also provides an upper bound on curvature (that is, a bound in  $\mathcal{C}^2$ ) and a minimal scale of possible quasi self-intersections [AL19].

For all these reasons, the reach practically appears in *all* geometric inference methods as a natural scale parameter, which either drives a bandwidth used in a localization method [GPPVW12b, AKC<sup>+</sup>19], a minimal regularity scale in a minimax study [KZ15], or a *signal* part in a signal-to-noise ratio [GPPVW12a, FILN19, AS21]. See [AKC<sup>+</sup>19] for more examples of its use. On the estimation side, the reach has already been studied under several angles.

- The formulation of  $\text{rch}(M)$  in terms of deviation to tangent spaces from [Fed59, Theorem 4.18] has been put to use through a plugin in [AL19]. The authors derived non-matching upper and lower bounds for the estimation of  $\text{rch}(M)$  over  $\mathcal{C}^3$  submanifolds. In addition to being suboptimal, the method of [AL19] requires the knowledge of tangent spaces, and is very sensitive to uncertainty on them (see [AL19, Section 6]).
- Extending the minimax study of [AL19], [BHHS22] took advantage of the so-called *convexity defect function* introduced by [ALS13] to propose another plugin strategy, with rates obtained over more general  $\mathcal{C}^k$ -smooth manifold classes. Despite still deriving non-matching upper and lower bounds, [BHHS22] managed to exhibit two different estimation rates, depending on whether the reach testifies of a high curvature zone (the so-called local case, with slow rates) or of a narrow bottleneck structure (global case, with faster rates). In this work, the derived rates are only suboptimal when the reach is achieved by curvature.
- More recently, [BLW19, Theorem 1] gave a new formulation of the reach in terms of geodesic distortion. Informally, they showed that  $\text{rch}(K)$  is the largest radius  $r \geq 0$  for which the geodesic distance  $d_K$  is smaller than the geodesic distance  $d_{S(r)}$  on a Euclidean ball of radius  $r$ . Based on this purely metric statement, [CFM21] proposed to plug-in a nearest-neighbor graph distance of the data in this formulation. This method provides a consistent estimator under very weak assumptions. Unfortunately, it fails to take advantage of high order regularity, when the reach is achieved by curvature (again).

With this analysis of possible estimation flaws in mind, this article proposes a two-step method. In short, we decouple the estimation of the local and global reaches [AKC<sup>+</sup>19], and estimate them separately via max-curvature estimation and geodesic distance estimation respectively.

### 3.1.3 Metric Learning

In the data analysis area, *metric learning* refers to the problem of finding a distance  $\widehat{d}$  over the space of observations  $\mathbb{X}_n \times \mathbb{X}_n$  that is relevant for a given task at stake [YJ06, SGH21]. For instance, in a supervised framework where one is provided with tuples of

allegedly *similar* or *dissimilar* observations, the goal is to find a distance that is small on the similar tuples and large on the dissimilar ones. There is a wide range of existing methods in the literature, ranging from parametric (LSI [XJRN02], MCML [GR05], LDML [GVS09] among others) to nonparametric (DMLMJ [NMDB17], kernel methods [KT03, CKTK10], to cite a few).

In an unsupervised setting, metric learning aims at finding a metric that takes into account the underlying geometry of the data. That is, it amounts to estimating of shortest path (or *geodesic*) distance. Often, this is done via a dimension reduction technique: any low-dimensional embedding of the data gives rise to a new distance over the data in the embedded space. Existing algorithms include PCA, t-SNE [HR02], MDS [CC08], Isomap [TSL00], or MVU [ACP13]. See [SGH21] for a thorough overview of the field.

Astonishingly, despite the variety of existing methods, we are not aware of any general minimax study of geodesic metric learning. Though, two major theoretical references seem to stand out:

- In [TSAY19], the authors use a neighborhood graph to estimate distances and derive convergence rates in the  $\mathcal{C}^2$  case, but only for nearby points.
- In [ACC20], estimation rates of geodesic distances are derived in the  $\mathcal{C}^2$  case using a reconstructing mesh. Lower bounds are also obtained, but in a fixed-design setting only.

We propose a simple plugin method, and show that estimating the geodesic metric is no harder than estimating the manifold itself in Hausdorff distance. This general strategy is also supported by a matching minimax lower bound.

### 3.1.4 Contribution and Outline

This article deals with the framework where data lies on an unknown  $d$ -dimensional  $\mathcal{C}^k$ -submanifold of  $\mathbb{R}^D$  (Section 3.2). The main contribution consists of nearly-tight minimax bounds for reach estimation (Section 3.6). Along the way, three major building blocks, interesting in their own rights, are developed thoroughly:

- Section 3.3: We propose a general plug-in strategy for estimating the reach of a manifold. It is based on curvature estimation on one hand, and on the estimation of an intermediate scale (framed between the reach and the weak feature size) on the other hand.
- Section 3.4: We define the so-called *spherical distortion radius* at scale  $\delta > 0$  and study its estimation. From the metric characterization of the reach from [BLW19], we notice that this purely metric quantity can be used to play the role of an

intermediate scale for reach estimation. We show that its stability properties make it well-suited to play the role of the intermediate scale of Section 3.3.

- Section 3.5: We propose a general plugin strategy for metric learning, and derive optimal geodesic metric estimation upper and lower bounds.

The proofs and the most technical points are deferred to the Appendix.

### 3.1.5 General Notation

In what follows,  $\mathbb{R}^D$  ( $D \geq 2$ ) is endowed with the Euclidean norm  $\|\cdot\|$ . The closed ball of radius  $r \geq 0$  centered at  $x \in \mathbb{R}^D$  is denoted by  $B(x, r)$ . If  $x \in T \subset \mathbb{R}^D$  is a linear subspace, we write  $B_T(x, r) := T \cap B(x, r)$  for the same ball in  $T$ . Throughout,  $c_\square, c'_\square, C_\square, C'_\square \geq 0$  denote generic constants that depend on  $\square$ , and that shall change from line to line to shorten notation. Similarly, universal constants shall generically be denoted by  $c, c', C, C' \geq 0$ .

## 3.2 Geometric and Statistical Model

Let us first present the models in which we will work throughout. As will be defined and discussed at length in Section 3.3.1, we let  $\text{rch}(K)$  denote the *reach* of a subset  $K \subset \mathbb{R}^D$  of the Euclidean space.

Building upon the standard regression setup, the following class is a good analog of Hölder classes of order  $k \geq 2$ , that is well adapted to submanifolds for stability reasons (see [AL19, Proposition 1]). Here, the analogy is to be understood as  $T_p M$  being the (local) covariate space, and  $\Psi_p$  being the regression function.

**Definition 3.1** ([AL19, Definition 1]). *Let  $k \geq 2$ ,  $\text{rch}_{\min} > 0$ , and  $\mathbf{L} = (L_2, L_3, \dots, L_k)$ . We let  $\mathcal{C}_{\text{rch}_{\min}, \mathbf{L}}^k$  denote the set of  $d$ -dimensional compact connected submanifolds  $M$  of  $\mathbb{R}^D$  with  $\text{rch}(M) \geq \text{rch}_{\min}$ , such that for all  $p \in M$ , there exists a local one-to-one parametrization  $\Psi_p$  of the form:*

$$\begin{aligned} \Psi_p: B_{T_p M}(0, r) &\longrightarrow M \\ v &\longmapsto p + v + \mathbf{N}_p(v) \end{aligned}$$

for some  $r \geq \frac{1}{4L_2}$ , with  $\mathbf{N}_p \in \mathcal{C}^k(B_{T_p M}(0, r), \mathbb{R}^D)$  such that for all  $\|v\| \leq \frac{1}{4L_2}$ ,

$$\mathbf{N}_p(0) = 0, \quad d_0 \mathbf{N}_p = 0, \quad \text{and} \quad \|d_v^j \mathbf{N}_p\|_{\text{op}} \leq L_j \text{ for all } j \in \{2, \dots, k\},$$

where  $d_v^j \mathbf{N}_p$  stands for the  $j$ th differential of  $\mathbf{N}_p$  at  $v$ , and  $\|\cdot\|_{\text{op}}$  for the Euclidean operator norm over tensors.

As explained in [AL19, Section 2.2], radii  $1/(4L_2)$  in local parametrizations have only been chosen for convenience. For  $k = 2$ , the existence of parametrizations  $\Psi_p$  is always guaranteed as soon as  $\text{rch}_{\min} > 0$  and  $L_2 \geq 2/\text{rch}_{\min}$  (see [AL19, Lemma 1]).

**Definition 3.2.** We let  $\mathcal{P}_{\text{rch}_{\min}, \mathbf{L}}^k(f_{\min}, f_{\max})$  denote the set of Borel probability distributions  $P$  on  $\mathbb{R}^D$  satisfying:

- Its support  $M := \text{supp}(P)$  belongs to  $\mathcal{C}_{\text{rch}_{\min}, \mathbf{L}}^k$ ;
- It has a density  $f$  with respect to the volume measure on  $M$ , such that

$$f_{\min} \leq f(x) \leq f_{\max} \quad \text{for all } x \in M.$$

On the estimation side, the uniform smoothness of the parametrizations in Definition 3.1 allows for estimation of the manifold via local polynomial fitting around sample points in the models  $\mathcal{P}_{\text{rch}_{\min}, \mathbf{L}}^k(f_{\min}, f_{\max})$ . Recall that the *Hausdorff distance* between two compact subsets  $K, K' \subset \mathbb{R}^D$  is defined by

$$d_H(K, K') := \max \left\{ \sup_{x \in K} d(x, K'), \sup_{x' \in K'} d(x', K) \right\}, \quad (3.1)$$

where for all  $u \in \mathbb{R}^D$ ,

$$d(u, K) := \min_{x \in K} \|x - u\| \quad (3.2)$$

stands for the *distance function* to  $K$ . The estimation rates over the models of Definition 3.2 have been studied in [AL19]. A key result that we will use is the following.

**Theorem 3.3** ([AL19, Theorem 6]). *There exists an estimator  $\widehat{M}$  such that for  $n$  large enough,*

$$\sup_{P \in \mathcal{P}_{\text{rch}_{\min}, \mathbf{L}}^k(f_{\min}, f_{\max})} \mathbb{E}_{P^n} [d_H(\widehat{M}, M)] \leq C_{d,k,\text{rch}_{\min}, \mathbf{L}, f_{\min}, f_{\max}} \left( \frac{\log n}{n} \right)^{k/d},$$

where in the supremum,  $M$  stands for  $\text{supp}(P)$ .

This rate is minimax optimal up to  $\log n$  factors [AL19, Theorem 7]. It can be achieved by a local polynomial patch estimator  $\widehat{M}$  (see (3.7) below) that we will use as a preliminary step towards reach estimation. Let us also mention here that these fitted local polynomials also allow for estimation of differential quantities of  $M$ , such as tangent spaces and curvature at sample points, with (minimax) convergence rates of order  $O(n^{-(k-1)/d})$  and  $O(n^{-(k-2)/d})$  respectively (see [AL19, Theorems 2 to 5]). This fact will be of key importance in Section 3.3.3, where estimating the maximal curvature of  $M$  will allow to estimate the so-called “local reach”.

### 3.3 Reach and Related Quantities

#### 3.3.1 Characterizations and Relaxations of the Reach

Let  $K$  be a compact subset of  $\mathbb{R}^D$ . Following the original definition of [Fed59], the *reach* of  $K$ , denoted by  $\text{rch}(K)$ , may be thought of as the largest radius of a neighborhood of  $K$  onto which the projection map  $\pi_K$  onto  $K$  is well-defined. More formally, define the *medial axis* of  $K$  by

$$\text{Med}(K) := \{u \in \mathbb{R}^D \mid \exists x_1 \neq x_2 \in K \quad \|u - x_1\| = \|u - x_2\| = d(u, K)\}.$$

The reach of  $K$  is then defined as the smallest distance between  $K$  and  $\text{Med}(K)$ .

**Definition 3.4.** For all closed  $K \subset \mathbb{R}^D$ , the reach of  $K$  is defined by

$$\text{rch}(K) := \min_{x \in K} d(x, \text{Med}(K)) = \inf_{u \in \text{Med}(K)} d(u, K).$$

Note that in full generality, the medial axis might not be a closed set, so that the infimum in Definition 3.4 may not be attained (for instance in the case where  $K$  is one-dimensional with a sharp edge). From a topological viewpoint, a key property of sets with positive reach is that the projection onto  $K$  induces continuous retractions from the *offset*  $K^r := \{u \in \mathbb{R}^D \mid d(u, K) \leq r\}$  onto  $K$ , whenever  $r < \text{rch}(K)$  [Fed59, Theorem 4.8]. This property is at the core of topologically consistent reconstruction procedures such as that of [BG14].

Sets with positive reach can also be thought of as generalizations of convex sets, characterized by the smoothness of their distance function. Indeed, based on the remark that  $x \mapsto d(x, K)$  is  $\mathcal{C}^1$  on  $\mathbb{R}^D \setminus K$  whenever  $K$  is convex, [CSW95] define *r-proximally-smooth* sets as the sets  $K$  such that  $d(\cdot, K)$  is  $\mathcal{C}^1$  over  $\{u \in \mathbb{R}^D \mid 0 < d(u, K) < r\}$ . Interestingly, for subsets of  $\mathbb{R}^D$ , *r-proximally smooth* sets are exactly sets with reach  $\text{rch}(K) \geq r$  [PRT00], so that the reach may be alternatively defined in terms of gradients of the distance function. To this aim, following [CL05], a generalized gradient function can be defined over  $\mathbb{R}^D \setminus K$ . For all  $x \in \mathbb{R}^D \setminus K$ , we write

$$\nabla d(x, K) := \frac{x - c_K(x)}{d(x, K)}, \quad (3.3)$$

where  $c_K(x)$  is the center of the smallest enclosing ball of the set  $\pi_K(\{x\})$  of nearest neighbors of  $x$  on  $K$ . Since  $c_K(x) = \pi_K(x)$  whenever  $x \notin \text{Med}(K)$ , the medial axis can actually be characterized as

$$\text{Med}(K) = \{x \in \mathbb{R}^D \setminus K \mid \|\nabla d(x, K)\| < 1\},$$

and the reach as

$$\text{rch}(K) = \sup \{r > 0 \mid 0 < d(x, K) < r \Rightarrow \|\nabla d(x, K)\| = 1\}.$$

This characterization of the reach allows for a straightforward relaxation. Namely, for a parameter  $\mu \in [0, 1]$ , the seminal paper [CCSL09] introduces the so-called  $\mu$ -medial axis as being

$$\text{Med}_\mu(K) := \{x \in \mathbb{R}^D \setminus K \mid \|\nabla d(x, K)\| \leq \mu\},$$

and the  $\mu$ -reach as

$$\text{rch}_\mu(K) := \inf_{u \in \text{Med}_\mu(K)} d(u, K). \quad (3.4)$$

It is clear that for all  $\mu < 1$ ,  $\text{rch}(K) \leq \text{rch}_\mu(K)$ , with  $\text{rch}(K)$  corresponding to the limit  $\text{rch}_{1-}(K)$ . Furthermore, this relaxation of the reach still yields enough regularity guarantees that the offsets  $K^r = \{u \in \mathbb{R}^D \mid d(u, K) \leq r\}$  are isotopic for all  $r \in (0, \text{rch}_\mu(K))$  [CCSL09, Lemma 2.1]. Hence, the condition that  $\text{rch}_\mu(K) > 0$  conveys enough regularity properties for many topological estimators to work [CCSM11].

Through this lens, the largest radius that ensures the topological stability of the offsets is the 0-reach, also called *weak-feature size*,

$$\text{wfs}(K) := \inf_{u \in \text{Med}_0(K)} d(u, K), \quad (3.5)$$

that is the distance from  $K$  to the set of critical points of  $d(\cdot, K)$ . As detailed in the following section, the weak-feature size plays a special role in the case where  $K$  is a manifold. Here come a few elementary properties of the weak feature size that we will use later on.

**Proposition 3.5.** *Let  $K \subset \mathbb{R}^D$  be compact.*

- (i) *If  $K$  is a closed submanifold of  $\mathbb{R}^D$ , then  $\text{wfs}(K) < +\infty$ ;*
- (ii) *If  $\text{wfs}(K) < +\infty$ , then for all  $\mu \in [0, 1]$ ,*

$$\text{rch}(K) \leq \text{rch}_\mu(K) \leq \text{wfs}(K) \leq \sqrt{\frac{D}{2(D+1)}} \text{diam}(K).$$

A proof is given in Section 3.A.1. Proposition 3.5 thus ensures that  $\text{wfs}(M)$  is uniformly bounded over the classes  $\mathcal{C}_{\text{rch}_{\min}, \mathbf{L}}^k$  introduced in Section 3.2. Since  $\text{wfs}(K)$  and  $\text{rch}_\mu(K)$  both measure a typical scale for topological stability, estimating them from sample could be of practical interest for topological inference. Unfortunately, the following negative result shows that this estimation problem is intractable, even over a well-behaved model of closed  $\mathcal{C}^k$ -submanifolds such as  $\mathcal{P}_{\text{rch}_{\min}, \mathbf{L}}^k(f_{\min}, f_{\max})$ .

**Theorem 3.6.** *Assume that  $f_{\min} \leq c_{d,k}/\text{rch}_{\min}^d$ ,  $f_{\max} \geq C_{d,k}/\text{rch}_{\min}^d$ , and  $L_j \geq C_{d,k}/\text{rch}_{\min}^{j-1}$  for all  $j \in \{2, \dots, k\}$ . Then there exists  $\tilde{c}_{d,k} > 0$  such that for all  $n \geq 1$  and  $\mu \in [0, 1]$ ,*

$$\inf_{\hat{r}_\mu} \sup_{P \in \mathcal{P}_{\text{rch}_{\min}, \mathbf{L}}^k(f_{\min}, f_{\max})} \mathbb{E}_{P^{\otimes n}} [|\hat{r}_\mu - \text{rch}_\mu(M)|] \geq \tilde{c}_{d,k} \text{rch}_{\min} > 0,$$

where  $\hat{r}_\mu$  ranges among all the possible estimators based on  $n$  samples.



An intuition behind Theorem 3.6 is that for all  $\mu < 1$ , the  $\mu$ -medial axis is an unstable structure. For certain manifolds  $M_0 \in \mathcal{C}_{\text{rch}_{\min}, \mathbf{L}}^k$ , one can find arbitrarily small perturbations of  $M_0$  whose  $\mu$ -medial axes remain at a fixed Hausdorff distance from  $\text{Med}_\mu(M_0)$ . See the proof of Theorem 3.6 in Section 3.A.2 for a precise statement of this intuition.

Despite the fact that  $\text{rch}(K) = \text{rch}_1(K)$ , this negative result indicates that we cannot leverage  $\mu$ -reach estimation to obtain quantitative bounds for reach estimation. We shall hence turn towards other reach-related quantities. In fact, the particular case where  $K = M$  is a manifold offers us several other characterizations of the reach, which suggest other estimation strategies.

### 3.3.2 Reach of Submanifolds

In what follows,  $M$  stands for a  $d$ -dimensional closed submanifold of  $\mathbb{R}^D$ . Note that [Fed59, Remarks 4.20 and 4.21] and [BLW19] assert that a closed submanifold with positive reach is at least of regularity  $\mathcal{C}^{1,1}$ , so that geodesics and tangent spaces are always defined in the usual differential sense. For the manifold case, the intuition of  $\text{rch}(M)$  as a generalized convexity parameter is further backed by [Fed59, Theorem 4.8]. Indeed, the inequality  $\langle x - \pi_C(x), \pi_C(x) - c \rangle \geq 0$  valid for all  $c \in C$  and  $x \in \mathbb{R}^D$  whenever  $C$  is convex, translates to  $\langle x - \pi_M(x), \pi_M(x) - y \rangle \geq -\|\pi_M(x) - y\|^2 \|x - \pi_M(x)\| / (2 \text{rch}(M))$  being valid for all  $y \in \mathbb{R}^D$  and  $x \in \mathbb{R}^D$  such that  $d(x, M) < \text{rch}(M)$ . This leads to the following characterization of the reach, in the manifold case.

**Theorem 3.7** ([Fed59, Theorem 4.18]). *For a submanifold  $M \subset \mathbb{R}^D$  without boundary,*

$$\text{rch}(M) = \inf_{p \neq q \in M} \frac{\|p - q\|^2}{2d(q - p, T_p M)},$$

where  $T_p M$  denotes the tangent space of  $M$  at  $p$ .

This result provides a natural plugin estimator, proposed by [AKC<sup>+</sup>19], which consists in replacing  $M$  and  $T_p M$  by suitable estimators of them. A key result from [AKC<sup>+</sup>19] is a description of how the infimum in Theorem 3.7 is achieved, possibly asymptotically.

**Theorem 3.8** ([AKC<sup>+</sup>19, Theorem 3.4]). *Let  $M \subset \mathbb{R}^D$  be a compact  $\mathcal{C}^2$  submanifold without boundary. Then,*

$$\text{rch}(M) = \text{wfs}(M) \wedge R_\ell(M),$$

where denoting by  $\mathbb{I}_p : T_p M \times T_p M \rightarrow T_p M^\perp$  the second fundamental form of  $M$  at  $p \in M$ ,

$$R_\ell(M) := \min_{p \in M} \|\mathbb{I}_p\|_{\text{op}}^{-1}$$

stands for the minimal curvature radius of  $M$ .

This result conveys the following intuition in the manifold case: the infimum in the right-hand side of Theorem 3.7 may be attained:

- **Local case:** Asymptotically, for pairs of points  $(p, q)$  converging to a maximal curvature point in some direction, so that  $\text{rch}(M) = R_\ell(M)$ .
- **Global case:** For a pair of points  $(p, q)$  belonging to parallel areas of  $M$ , forming a bottleneck zone, so that  $\text{rch}(M) = \text{wfs}(M)$ .

This local/global dichotomy of the reach may also be retrieved in the recent characterization given by [BLW19] in terms of metric distortion.

**Theorem 3.9** ([BLW19, Theorem 1]). *Let  $K \subset \mathbb{R}^D$  be a closed subset. Then*

$$\text{rch}(K) = \sup \left\{ r > 0 \mid \forall p, q \in K, \|p - q\| < 2r \Rightarrow d_K(p, q) \leq 2r \arcsin \left( \frac{\|p - q\|}{2r} \right) \right\},$$

where  $d_K : K \times K \rightarrow \bar{\mathbb{R}}_+$  stands for the shortest-path (or geodesic) distance on  $K$ .

Recall that, for all  $p, q \in K$ , the distance  $d_K(p, q)$  is the infimum of the length of all the continuous path in  $K$  between  $p$  and  $q$ . As will be detailed in Section 3.4, the above result allows to characterize the reach in terms of metric distortion with respect to metrics on spheres of radii  $r$ . In the same spirit as Theorem 3.8, when  $K = M$  is a submanifold, the configurations of  $(p, q, r)$  in the supremum of Theorem 3.9 are limited by the same two local and global layouts:

- **Local case:** When  $p$  and  $q$  tend to a maximal curvature point in some direction, the geodesic distance  $d_K$  behaves like that of a sphere of radius  $R_\ell(M)$  at this point in this direction.
- **Global case:** When  $p$  and  $q$  are in parallel areas, their geodesic distance must be larger than the spherical distance of radius  $\|p - q\|/2$ .

### 3.3.3 Plug-in Methods for Reach Estimation

The characterizations of the reach given in Section 3.3.2 all lead to their associate plug-in estimators:

- Studying a  $\mathcal{C}^3$  model similar to  $\mathcal{P}_{\text{rch}_{\min}, \mathbf{L}}^3(f_{\min}, f_{\max})$ , [AKC<sup>+</sup>19] took advantage of the characterization with tangent spaces (Theorem 3.7) to conceive a reach estimator that converges at rate  $O(n^{-2/(3d-1)})$  in the local case ( $\text{rch}(M) = R_\ell(M)$ ), and  $O(n^{-1/d})$  in the global case ( $\text{rch}(M) = \text{wfs}(M)$ ).
- Based on the metric distortion characterization of Theorem 3.9, [CFM21] propose a reach estimator that is consistent whenever  $M$  has positive reach.

In light of Theorem 3.8, differences of convergence rates between the local and global case are to be expected. To quantify this intuition, [BHHS22] introduces subclasses of the model  $\mathcal{P}_{\text{rch}_{\min}, \mathbf{L}}^k(f_{\min}, f_{\max})$ , parametrized by the gap between  $R_\ell(M)$  and  $\text{wfs}(M)$ . They obtain the following lower bounds.

**Theorem 3.10** ([BHHS22, Theorem 7.1] and [AKC<sup>+</sup>19, Proposition 2.9]). *Let  $\alpha \in \mathbb{R}$ ,  $k \geq 2$ , and write*

$$\mathcal{P}_{\text{rch}_{\min}, \mathbf{L}, \alpha}^k(f_{\min}, f_{\max}) := \left\{ P \in \mathcal{P}_{\text{rch}_{\min}, \mathbf{L}}^k(f_{\min}, f_{\max}) \mid R_\ell(M) \geq \text{wfs}(M) + \alpha \right\},$$

*where  $M$  denotes  $\text{supp}(P)$ . Then, for all  $\text{rch}_{\min} > 0$  there exists small enough  $f_{\min}$  and large enough  $f_{\max}$ ,  $\mathbf{L}$  such that*

$$\inf_{\widehat{\text{rch}}} \sup_{P \in \mathcal{P}_{\text{rch}_{\min}, \mathbf{L}, \alpha}^k(f_{\min}, f_{\max})} \mathbb{E} |\widehat{\text{rch}} - \text{rch}(M)| \geq c_{\text{rch}_{\min}, d, k} \left( \frac{1}{n} \right)^{(k-2)/d}, \quad \text{if } \alpha \leq 0,$$

$$\inf_{\widehat{\text{rch}}} \sup_{P \in \mathcal{P}_{\text{rch}_{\min}, \mathbf{L}, \alpha}^k(f_{\min}, f_{\max})} \mathbb{E} |\widehat{\text{rch}} - \text{rch}(M)| \geq c_{\text{rch}_{\min}, d, k, \alpha} \left( \frac{1}{n} \right)^{k/d}, \quad \text{if } \alpha > 0.$$

These bounds indicate that estimating the reach is at least as hard as estimating the curvature in the local case ( $\text{rch}(M) = R_\ell(M)$ ), and at least as hard as estimating the manifold in the global case ( $\text{rch}(M) = \text{wfs}(M)$ ). We will prove in Section 3.6 that these rates are in fact minimax optimal up to  $\log n$  factors. This means that reducing reach estimation to curvature *and* manifold estimation is a good way to go, as it leads to optimal rates. To do so, following the idea behind Theorem 3.8, estimating  $R_\ell(M)$  – or some notion of local reach – and  $\text{wfs}(M)$  – or some notion of global reach – separately seems a sensible approach.

**Local Reach Estimation** For (max-)curvature estimation, the strategy that we adopt follows from the polynomial patches estimator proposed in [AL19]. Given a localization bandwidth  $h > 0$ , and a parameter  $t > 0$ , for all  $i \in \{1, \dots, n\}$ , we let  $\hat{\pi}_i : \mathbb{R}^D \rightarrow \mathbb{R}^D$  be an orthogonal projector of rank  $d$  and  $\hat{\mathbb{T}}_i^{(j)} : (\mathbb{R}^D)^{\otimes j} \rightarrow \mathbb{R}^D$  be symmetric tensors solutions of the least squares problem

$$\min_{\substack{\pi \\ \max_{2 \leq j \leq k-1} \|\mathbb{T}^{(j)}\| \frac{1}{j-1} \leq t}} P_{n-1}^{(i)} \left[ \left\| x - \pi(x) - \sum_{j=2}^{k-1} \mathbb{T}^{(j)} (\pi(x)^{\otimes j}) \right\|^2 \mathbb{1}_{B(0, h)}(x) \right], \quad (3.6)$$

where  $P_{n-1}^{(i)} := \frac{1}{n-1} \sum_{p \neq i} \delta_{X_p - X_i}$  denotes the empirical measure centered at point  $X_i$ . Following [AL19, Section 3], if  $h$  is taken to be of order  $\Theta((\log n/n)^{1/d})$ , that  $t$  is chosen such that  $t^k h \leq 1$ , and that  $\hat{T}_i := \text{Im}(\hat{\pi}_i)$  denotes the image of  $\hat{\pi}_i$  – which is a  $d$ -dimensional vector space by construction –, then the local patches

$$\widehat{\Psi}_i : B_{\hat{T}_i}(0, 7h/8) \longrightarrow \mathbb{R}^D$$

$$v \longmapsto X_i + v + \sum_{j=2}^{k-1} \hat{\mathbb{T}}_i^{(j)}(v^{\otimes j}) \quad (3.7)$$

are local  $O(h^k)$  approximations of  $M$  whenever  $n$  is large enough. Furthermore, for  $v \in B_{\hat{T}_i}(0, h/4)$ , we can estimate the curvature tensor at  $\pi_M(\hat{\Psi}_i(v))$  via the second derivative of  $\hat{\Psi}_i$  at  $v$ , expressed in local coordinates around  $\hat{\Psi}_i(v)$  given by a basis of  $\text{Im}(d_v \hat{\Psi}_i)$ . To summarize, for all  $v \in B_{\hat{T}_i}(0, h/4)$ , (3.7) provides a  $d$ -dimensional space  $\hat{T}_{i,v} := \text{Im}(d_v \hat{\Psi}_i)$ , as well as a symmetric bilinear map

$$\hat{\mathbb{T}}_{i,v}^{(2)} : \hat{T}_{i,v} \times \hat{T}_{i,v} \rightarrow \hat{T}_{i,v}^\perp,$$

that is provably close to  $\Pi_{\pi_M(\hat{\Psi}_i(v))}$ . The precise definition of  $\hat{\mathbb{T}}_{i,v}^{(2)}$  is given in Section 3.A.3. A minimal curvature radius (i.e. maximal curvature) estimator may then be computed as the minimal curvature radius of all the polynomial patches around sample points, that is

$$\widehat{R}_\ell := \min_{1 \leq i \leq n} \min_{v \in B_{\hat{T}_i}(0, h/4)} \|\hat{\mathbb{T}}_{i,v}^{(2)}\|_{\text{op}}^{-1}. \quad (3.8)$$

Provided  $M$  is uniformly well approximated by  $\bigcup_{i=1}^n \hat{\Psi}_i(B_{\hat{T}_i}(0, h/4))$ , the convergence rate of  $\widehat{R}_\ell$  towards  $R_\ell(M)$  will follow from uniform curvature bounds, similar to the pointwise ones from [AL19, Theorem 4]. We are able to prove the following.

**Theorem 3.11.** *Let  $k \geq 3$  and  $P \in \mathcal{P}_{\text{rch}_{\min}, \mathbf{L}}^k(f_{\min}, f_{\max})$ . Write  $h = \left(C_{d,k} \frac{f_{\max}^2 \log n}{f_{\min}^3 n}\right)^{1/d}$ . Then for  $n$  large enough, with probability larger than  $1 - 2n^{-k/d}$ , we have*

$$|\widehat{R}_\ell - R_\ell(M)| \leq C_{d,k,\mathbf{L},\text{rch}_{\min}} R_\ell^2(M) \sqrt{\frac{f_{\max}}{f_{\min}}} h^{k-2}.$$

We refer to Section 3.A.3 for a proof of this result. In particular, the estimator  $\widehat{R}_\ell$  achieves the rate of the lower bound from Theorem 3.10 in the case where  $\text{rch}(M) = R_\ell(M)$  (i.e.  $\alpha \leq 0$ ), up to  $\log n$  factors.

**Global Reach Estimation** To complete the construction of an estimator of  $\text{rch}(M)$ , building an estimator of  $\text{wfs}(M)$  could be a possibility. However, Theorem 3.6 shows that building an estimator of the weak feature size with a uniform convergence rates over  $\mathcal{P}_{\text{rch}_{\min}, \mathbf{L}}^k(f_{\min}, f_{\max})$  is hopeless. Nonetheless, it is important to note that a uniform estimation rate of  $\text{wfs}(M)$  over  $\mathcal{P}^k$  is not necessary to obtain uniform convergence rate for  $\text{rch}(M)$ . Indeed, an estimator  $\widehat{\text{wfs}}$  of  $\text{wfs}(M)$  that exhibits an optimal uniform convergence rate whenever  $\text{wfs}(M) \leq R_\ell(M)$ , and that is provably larger than  $R_\ell(M)$  otherwise, is enough to build an optimal reach estimator when combined with  $\widehat{R}_\ell$ . This is the case, for instance, of the weak feature size estimator of [BHHS22] based on the so-called *convexity defect function*.

Based on this remark, we adopt a more general strategy, by seeking for an intermediate geometric scale  $\theta(M)$  (or *feature size*) such that for all  $M \in \mathcal{C}_{\text{rch}_{\min}, \mathbf{L}}^k$ ,

$$\text{rch}(M) \leq \theta(M) \leq \text{wfs}(M).$$

In such a case, Theorem 3.8 extends trivially, with  $\text{wfs}(M)$  replaced by  $\theta(M)$ .

**Proposition 3.12.** *Assume that  $\theta : \mathcal{C}_{\text{rch}_{\min}}^2 \rightarrow \mathbb{R}_+$  is such that  $\text{rch}(M) \leq \theta(M) \leq \text{wfs}(M)$  for all  $M \in \mathcal{C}_{\text{rch}_{\min}}^2$ . Then,*

$$\text{rch}(M) = \theta(M) \wedge R_\ell(M).$$

Given such an intermediate scale parameter of interest  $\theta(M)$ , and assuming that a consistent estimator  $\widehat{\theta}$  of  $\theta(M)$  is available, one can naturally consider the plugin  $\widehat{\text{rch}} := \widehat{R}_\ell \wedge \widehat{\theta}$ . For free, Proposition 3.12 yields that  $\theta(M) \mathbb{1}_{R_\ell(M) > \text{rch}(M)} = \text{rch}(M) \mathbb{1}_{R_\ell(M) > \text{rch}(M)}$ , so that

$$|\text{rch}(M) - \widehat{\text{rch}}| \leq |\widehat{R}_\ell - R_\ell(M)| \mathbb{1}_{R_\ell(M) \leq \text{rch}(M)} + |\widehat{\theta} - \theta(M)| \mathbb{1}_{R_\ell(M) > \text{rch}(M)}, \quad (3.9)$$

as soon as  $|R_\ell(M) - \widehat{R}_\ell| + |\theta(M) - \widehat{\theta}| \leq |R_\ell(M) - \theta(M)|$ . In addition, such a quantity would provide a local scale that is of interest for further topological inference, as exposed in Section 3.3.1.

According to Theorem 3.6, taking  $\theta(M)$  to be related to the medial axis characterization of the reach – such as the  $\mu$ -reach, or the  $\lambda$ -reach defined in [CL05] – is likely to lead to an unsolvable statistical problem, because of the inherent instability of the medial axis. Hence, we rather build upon the metric distortion characterization of the reach given by Theorem 3.9, and provide a better-behaved intermediate scale  $\theta(M)$ : the *spherical distortion radius*.

## 3.4 Spherical Distortion Radius

### 3.4.1 Motivation and Definition

Based on Theorem 3.9, we now build a geometrically stable feature size that measures the maximum radius (or scale) at which the geodesic distance can be compared to the corresponding spherical distance. To be more precise, for  $x, y \in \mathbb{R}^D$  and  $r > 0$ , we define the *spherical distance*  $d_{S(r)}(x, y)$  – or *great-circle distance* – as the distance between  $x$  and  $y$  when seen as both lying on a sphere of radius  $r$ . That is,

$$d_{S(r)}(x, y) := \begin{cases} 2r \arcsin\left(\frac{\|x-y\|}{2r}\right) & \text{if } \|x-y\| \leq 2r, \\ +\infty & \text{otherwise} \end{cases}$$

Note that the map  $r \mapsto d_{S(r)}(x, y)$  is decreasing on  $[\|x-y\|/2, \infty)$  and that

$$d_{S(r)}(x, y) = \frac{1}{2} \pi \|x-y\| \text{ for } r = \frac{\|x-y\|}{2} \text{ and } d_{S(r)}(x, y) \xrightarrow{r \rightarrow \infty} \|x-y\|.$$

Then, Theorem 3.9 can be rewritten as

$$\text{rch}(K) = \sup \left\{ r > 0 \mid \forall x, y \in K, \|x-y\| < 2r \Rightarrow d_K(x, y) \leq d_{S(r)}(x, y) \right\}.$$

It should be noted that  $d_{S(r)}$  is not formally a distance on  $K$  (unless  $K$  is a subset of a sphere of radius  $r$ ), but this is of little importance in what follows.

Based on the same idea that motivates the introduction of the  $\mu$ -reach, we intend to discard curvature effects to obtain some notion of global reach. In the metric characterization of the reach from Theorem 3.9, this can be done by the supremum restricting to points that are not too close.

**Definition 3.13.** Let  $K$  be a compact subset of  $\mathbb{R}^D$ ,  $d$  a distance on  $K$  and  $\delta > 0$ . The spherical distortion radius of the metric space  $(K, d)$  at scale  $\delta$  is defined by

$$\text{sdr}_\delta(K, d) := \sup \{ r > 0 \mid \forall x, y \in K, \delta \leq \|x - y\| < 2r \Rightarrow d(x, y) \leq d_{S(r)}(x, y) \}.$$

In words, the spherical distortion radius at scale  $\delta > 0$  is the largest radius  $r$  for which the distance  $d$  is bounded above by the spherical distance at radius  $r$ , when restricted to points that are at least  $\delta$ -apart for the Euclidean distance.

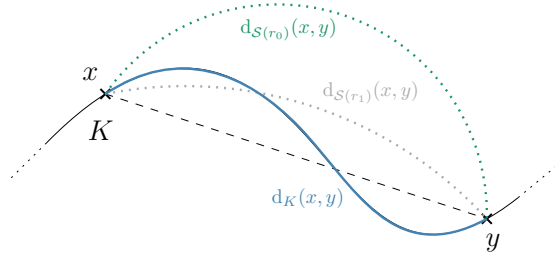


Figure 3.1 – A curve  $K$  in the plane. In blue is the shortest path between two points  $x$  and  $y$ , whose length is  $d_K(x, y)$ . In green (resp. grey) is the circle portion of radius  $r_0$  (resp.  $r_1$ ) going through  $x$  and  $y$ . The layout is chosen so that  $r_0 \leq r_1$  and  $d_{S(r_1)}(x, y) \leq d_K(x, y) \leq d_{S(r_0)}(x, y)$ .

By construction,  $\text{sdr}_\delta(K, d) \geq \delta/2$  for all  $\delta > 0$ . Furthermore, whenever  $\delta$  is strictly greater than  $\text{diam } K$ , then no pairs of points in  $x, y \in K$  satisfies  $\|x - y\| \geq \delta$  so that  $\text{sdr}_\delta(K) = +\infty$ . On the other hand, if  $\delta = 0$ , then the spherical distortion radius of  $(K, d_K)$ , coincides with the reach of  $K$  (Theorem 3.9). In fact, Proposition 3.14 below confirms that the spherical distortion radius interpolates between the reach and the weak feature size.

**Proposition 3.14.** For all closed  $K \subset \mathbb{R}^D$  and all metric  $d$  on  $K$ , the map  $\delta \mapsto \text{sdr}_\delta(K, d)$  is non-decreasing. Furthermore, for  $d = d_K$ ,

$$\text{rch}(K) \leq \text{sdr}_\delta(K, d_K) \leq \text{wfs}(K) \quad \text{for all } 0 \leq \delta \leq \sqrt{\frac{2(D+1)}{D}} \text{wfs}(K).$$

A proof of Proposition 3.14 is given in Appendix 3.B.1.

**Example 3.15.** As a toy example, let us study the spherical distortion radius of the wedge shape  $K_\alpha = \mathcal{L}_1 \cup \mathcal{L}_2$  where  $\mathcal{L}_1$  and  $\mathcal{L}_2$  are two half-line originated from a common point

$z \in \mathbb{R}^D$  (see Figure 3.2). We let  $\alpha \in (0, \pi)$  be the angle between these two lines. In this context, we have  $\text{rch}(K_\alpha) = 0$ , and it is easy to see that  $\text{wfs}(K_\alpha) = \infty$ . Furthermore, the usual interpolations between the reach and the weak feature size exhibit a very degenerate behavior in the presence of an angular configuration such as this one, with for instance

$$\text{rch}_\mu(K_\alpha) = \begin{cases} 0 & \text{if } \mu \geq \sin(\alpha/2), \\ \infty & \text{if } \mu < \sin(\alpha/2). \end{cases}$$

On the contrary, we show hereafter that the spherical distortion radius interpolates non-trivially between  $\text{rch}(K_\alpha)$  and  $\text{wfs}(K_\alpha)$  in this case, giving rise to a new family of relevant characteristic scales even for non-smooth subsets  $K_\alpha$ .

To see this, take  $x \in \mathcal{L}_1$  and  $y \in \mathcal{L}_2$ , and denote by  $a := \|x - z\|$  and  $b := \|y - z\|$ . The intrinsic distance  $d_{K_\alpha}(x, y)$  is given by  $a + b$  while  $\|x - y\|^2 = a^2 + b^2 - 2ab \cos(\alpha)$ . Now the solution of the minimization problem

$$\min \{a^2 + b^2 - 2ab \cos(\alpha) \mid a + b = d_{K_\alpha}(x, y)\}$$

is given by  $a = b = d_{K_\alpha}(x, y)/2$  and equals  $d_{K_\alpha}^2(x, y) \sin^2(\alpha/2)$ . The spherical distortion radius of  $K_\alpha$  at scale  $\delta$  is thus the largest  $r$  such that

$$\frac{\delta}{\sin(\alpha/2)} \leq 2r \arcsin\left(\frac{\delta}{2r}\right). \quad (3.10)$$

Since the right-hand side above ranges between  $\delta$  and  $\delta\pi/2$ , we distinguish two cases:

- If  $\sin(\alpha/2) < 2/\pi$ , then no  $r$  can fulfill (3.10). Hence,  $\text{sdr}_\delta(K_\alpha, d_{K_\alpha}) = \delta/2$ .
- Otherwise  $\sin(\alpha/2) \geq 2/\pi$ , in which case the largest  $r$  is given by the equality  $\varphi(2r/\delta) = 1/\sin(\alpha/2)$ , where  $\varphi(u) := u \arcsin(1/u)$  is a bijection between  $[1, \infty)$  and  $(1, \pi/2]$ .

All in all, it holds

$$\text{sdr}_\delta(K_\alpha, d_{K_\alpha}) = \begin{cases} \delta/2 & \text{if } \alpha < \alpha_* \\ (\delta/2)\varphi^{-1}(1/\sin(\alpha/2)) & \text{if } \alpha \geq \alpha_* \end{cases}$$

where  $\alpha_* = 2 \arcsin(2/\pi) < \pi/2$ . Note that compared to  $\text{rch}_\mu(K_\alpha)$ , there is no discontinuity in  $\text{sdr}_\delta(K_\alpha, d_{K_\alpha})$  as  $\alpha$  varies.

Example 3.15 above carries the intuition that the spherical distortion radius seems somehow stable with respect to Hausdorff perturbations, contrary to the  $\mu$ -reach. We quantify this intuition in the following section.

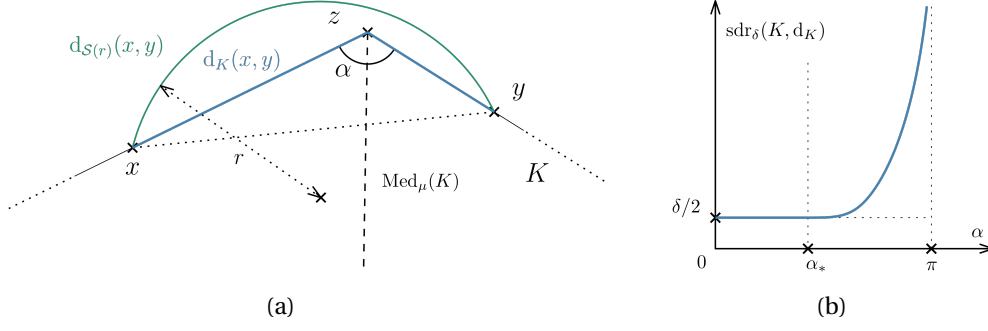


Figure 3.2 – (a) Diagram of  $K_\alpha = \mathcal{L}_1 \cup \mathcal{L}_2$  with an angle  $\alpha$  between the two half-lines. The shortest path between  $x$  and  $y$  is drawn in blue. In dashed the  $\mu$ -medial axis for  $\mu > \sin(\alpha/2)$ , showing in particular that  $\text{rch}_\mu(K_\alpha) = 0$  in this case. (b) Plot of the function  $\alpha \mapsto \text{sdr}_\delta(K_\alpha, d_{K_\alpha})$ , which operates a smooth interpolation between  $\delta/2$  and  $\infty$ .

### 3.4.2 Stability Properties

In this section, we will be comparing different metric spaces on subsets of  $\mathbb{R}^D$ . Let  $K$  and  $K'$  be two subsets of  $\mathbb{R}^D$ , endowed with distances  $d$  and  $d'$  respectively. We intend to prove that  $\text{sdr}_\delta(K, d)$  and  $\text{sdr}_\delta(K', d')$  are close whenever  $(K, d)$  and  $(K', d')$  are close, and that  $(K, d)$  has good properties. The notion of proximity between  $K$  and  $K'$  will be measured in Hausdorff distance (see (3.1)). It remains to define a notion of proximity between  $d$  and  $d'$ , which is called the *mutual distortion*.

**Definition 3.16.** Let  $(K, d)$  and  $(K', d')$  be two metric subspaces of  $\mathbb{R}^D$ . The metric distortion of  $d'$  relative to  $d$  at scale  $\delta > 0$  is

$$D_\delta(d'|d) := \sup_{\substack{x', y' \in K' \\ \|x' - y'\| \geq \delta}} \frac{d'(x', y')}{d(\text{pr}_K(\{x'\}), \text{pr}_K(\{y'\}))}.$$

where  $\text{pr}_K$  is the (possibly multivalued) closest-point projection onto  $K$  for the ambient Euclidean distance, and where

$$d(\text{pr}_K(\{x'\}), \text{pr}_K(\{y'\})) := \inf \{d(x, y) \mid x \in \text{pr}_K(\{x'\}), y \in \text{pr}_K(\{y'\})\}.$$

We adopt the convention  $D_\delta(d'|d) = 0$  if  $\delta > \text{diam}(K')$ . The mutual distortion of  $d$  and  $d'$  is then defined as

$$D_\delta(d, d') := \max \{D_\delta(d'|d), D_\delta(d|d')\}.$$

The mutual distortion defined above allows to compare distances on different spaces, while taking into account their respective embeddings in  $\mathbb{R}^D$ . A small distortion  $D_\delta(d, d')$  means that, if  $a, b \in K$  and  $x, y \in K'$  are two couples of points that are  $\delta$ -separated and such that  $x$  and  $a$ , and  $y$  and  $b$  are respectively close to each other,



then  $d(a, b)$  and  $d'(x, y)$  should be close as well. This definition of *mutual distortion* between metric subspaces of  $\mathbb{R}^D$  is related to the existing notion *metric distortion* of an embedding. See for instance [Bou85] or more recently [CVvL18] which deals with distortion measures in a statistical framework. It is nonetheless significantly different, in particular because the usual notion of distortion is invariant through re-scaling of either  $d$  or  $d'$ . In our framework, invariance with respect to scaling is an undesirable property, since we want to estimate the reach, which is itself a scale factor (or feature size).

*Remark 3.17.* When  $K = K'$ , the mutual distortion can be seen as the bi-Lipschitz coefficient of  $\text{Id} : (K, d) \rightarrow (K, d')$  at scale  $\delta$ , meaning that for all  $x, y \in K$

$$\|x - y\| \geq \delta \Rightarrow \frac{1}{L} d'(x, y) \leq d(x, y) \leq L d'(x, y),$$

where  $L = D_\delta(d, d')$ . In particular, a mutual distortion that is close to 1 means that  $(K, d)$  is quasi-isometric to  $(K, d')$ , at scale  $\delta$ .

If the two subspaces  $K$  and  $K'$  are too far apart, then it makes no sense to compare two distances  $d$  and  $d'$  defined on them, and one could expect the mutual distortion to explode. This is will typically the case when  $d_H(K, K') \geq \delta$ .

It is clear from the definition that using the notion of relative metric distortion defined above, the spherical distortion radius of  $K$  may be expressed as

$$\text{sdr}_\delta(K, d) = \sup \{r > 0 \mid D_\delta(d|_{d_{S(r)}}) \leq 1\}.$$

This point supports the idea that the relative metric distortion we defined is a suitable notion of proximity to assess stability of the spherical distortion radius, as exposed by the following proposition.

**Proposition 3.18.** *Let  $\delta_0 > 0$  and  $\varepsilon, \nu > 0$ . Assume that both  $d_H(K', K) \leq \varepsilon$  and  $D_{\delta_0}(d'|d) \leq 1 + \nu$ . Define*

$$\xi(r) := 384(1 + \pi) \frac{r^4}{\delta_0^4} \text{ for all } r \geq 0.$$

*Then, for all  $\delta \geq \delta_0$ , letting  $\Upsilon := (\delta\nu) \vee \varepsilon$  and  $r_1 := \text{sdr}_{\delta+2\varepsilon}(K', d')$ , if  $\xi(r_1)\Upsilon < r_1$ , then*

$$\text{sdr}_\delta(K, d) \leq \text{sdr}_{\delta+2\varepsilon}(K', d') + \xi(r_1)\Upsilon.$$

A proof of Proposition 3.18 is given in Appendix 3.B.2. Note that the condition  $d_H(K', K) \leq \varepsilon$  may be relaxed via  $d_H(K'|K) \leq \varepsilon$ , where  $d_H(K'|K) := \sup_{x \in K'} d(x, K)$ . Also, under the assumptions of Proposition 3.18, let us remark that if  $\text{sdr}_{\delta+2\varepsilon}(K', d')$  is finite, then so is  $\text{sdr}_\delta(K, d)$  with  $\text{sdr}_\delta(K, d) \leq 2 \text{sdr}_{\delta+2\varepsilon}(K', d')$ .

Proposition 3.18 can be symmetrized to get the following two-sided control.

**Corollary 3.19.** *Let  $0 < \delta_0 < \delta_1$  and  $\varepsilon, \nu > 0$ . Assume that both  $d_H(K', K) \leq \varepsilon$  and  $D_{\delta_0}(d', d) \leq 1 + \nu$ . Then, for any  $\delta \in (\delta_0 + 2\varepsilon, \delta_1 - 2\varepsilon)$ , it holds*

$$\text{sdr}_{\delta-2\varepsilon}(K, d) - \xi_0 \Upsilon \leq \text{sdr}_{\delta}(K', d') \leq \text{sdr}_{\delta+2\varepsilon}(K, d) + \xi_0 \Upsilon$$

*with  $\xi_0 := \xi(2 \text{sdr}_{\delta_1}(K, d))$  and  $\Upsilon := (\nu \delta) \vee \varepsilon$ , provided that  $\xi_0 \Upsilon \leq 2 \text{sdr}_{\delta_1}(K, d)$ .*

Corollary 3.19 is proven in Appendix 3.B.2. It ensures that the spherical distortion radius enjoys an interleaving property. That is the SDR of  $(K, d)$  at scale  $\delta$  may be framed by the SDR of an approximation  $(K', d')$  at scales  $\delta \pm \varepsilon$ . This interleaving property is a common thread with the  $\mu$ -reach (see, e.g., [CCSL09, Theorem 3.4]) and the  $\lambda$ -reach ([CL05, Theorem 3]), that is not enough to ensure consistent estimation. In fact, for the two aforementioned quantities, consistency may be proved with the additional assumption of  $\mu \mapsto \text{rch}_{\mu}(K)$  (resp.  $\lambda \mapsto \lambda\text{-reach}$ ) are continuous at the targeted  $\mu$  (resp.  $\lambda$ ).

As opposed to the  $\mu$ -reach the  $\lambda$ -reach, the SDR is also stable with respect to its the scale parameter  $\delta$ . Next, we prove that  $\delta \mapsto \text{sdr}_{\delta}(K, d)$  is continuous over a fixed range  $(0, \Delta^*)$  under mild structural assumptions on  $(K, d)$ . These assumptions will be easily checked in the model  $\mathcal{C}_{\text{rch}_{\min}, \mathbf{L}}^k$ , hence ensuring consistency of the subsequent reach estimator.

**Assumption 3.20.** *We say that  $K \subset \mathbb{R}^D$  is spreadable if there exist  $\Delta_0 > 0$ ,  $\varepsilon_0 > 0$ , and  $C_0 > 0$  such that for all  $x, y \in K$  such that  $\|x - y\| \leq \Delta_0$  and all  $\varepsilon \leq \varepsilon_0$ , there exists a point  $a \in K$  such that either*

- $\|a - y\| \leq \varepsilon$  and  $\|x - a\| \geq \|x - y\| + C_0 \varepsilon$ , or
- $\|a - x\| \leq \varepsilon$  and  $\|y - a\| \geq \|x - y\| + C_0 \varepsilon$ .

Assumption 3.20 requires that every point  $y$  of  $K$  may be locally pushed away from any (close enough) point  $x \in K$ . In particular, this means that  $K$  is nowhere discrete. In the manifold case, this pushing may be carried out using the exponential map (see Proposition 3.30).

**Assumption 3.21.** *We say that  $(K, d)$  is sub-Euclidean if there exist  $C_1 > 0$  and  $\Delta_1 > 0$  such that for all  $x, y \in K$  such that  $\|x - y\| \leq \Delta_1$ , we have  $d(x, y) \leq C_1 \|x - y\|$ .*

Assumption 3.21 requires that the distance locally compares with the ambient Euclidean distance. This essentially means that the identity map  $(K, d) \rightarrow (K, \|\cdot\|)$  is locally Lipschitz. Such an assumption is automatically fulfilled whenever  $K$  has positive reach and  $d = d_K$  (see [Fed59]), with explicit constants in the manifold case (see Proposition 3.30). Whenever these two conditions are met, the spherical distortion radius of  $(K, d)$  can be proved to be locally Lipschitz in  $\delta$ .

**Theorem 3.22.** *Assume that the metric space  $(K, d)$  fulfills Assumptions 3.20 and 3.21. Then  $\delta \mapsto \text{sdr}_\delta(K, d)$  is locally Lipschitz on  $(0, \Delta^*)$  where*

$$\Delta^* := \min \{ \Delta_0, \Delta_1, \sup \{ \delta \geq 0 \mid \text{sdr}_\delta(K, d) < \infty \} \}.$$

*More precisely, for all  $0 < \delta_0 < \delta_1 < \Delta^*$ , the map  $\delta \mapsto \text{sdr}_\delta(K, d)$  is  $L_0$ -Lipschitz on  $[\delta_0, \delta_1]$  with*

$$L_0 := \frac{192r_1^3}{C_0\delta_0^3} \left( C_1 + \pi \frac{r_1}{\delta_0} \right),$$

*where  $r_1 := \text{sdr}_{\delta_1}(K, d)$ .*

A proof of Theorem 3.22 can be found in Appendix 3.B.2. Not only does it ensure that the spherical distortion radius at scale  $\delta$  is continuous with respect to  $\delta$ , that is enough to guarantee consistency, but it also allows to control its variation via an explicit local Lipschitz constant. Combined with Corollary 3.19, this allows to convert a bound between  $(K, d)$  and  $(K', d')$  in terms of Hausdorff distance and metric distortion into a bound on the SDR's at scale  $\delta$ .

**Theorem 3.23.** *Let  $(K, d)$  fulfill Assumptions 3.20 and 3.21, and let  $(K', d')$  be such that  $d_H(K, K') \leq \varepsilon$  and  $D_{\delta_0}(d, d') \leq 1 + \nu$  for some  $\delta_0 < \Delta^*$ . Then, for all  $\delta_1 \in (\delta_0, \Delta^*)$  and  $\delta \in (\delta_0 + 2\varepsilon, \delta_1 - 2\varepsilon)$ , provided that  $\xi_0 \Upsilon \leq 2 \text{sdr}_{\delta_1}(K, d)$ , we have*

$$|\text{sdr}_\delta(K, d) - \text{sdr}_\delta(K', d')| \leq \zeta_0 \Upsilon,$$

*with  $\Upsilon = (\delta\nu) \vee \varepsilon$  and  $\zeta_0 = \xi_0 + 2L_0$ , where  $\xi_0$  is defined in Corollary 3.19,  $L_0$  is defined in Theorem 3.22.*

We refer to Appendix 3.B.2 for a proof of this result and to Figure 3.3 for a diagram of the scales at play. Note that the constant  $\zeta_0$  only depends on  $\delta_0$  and features of  $(K, d)$ , that the assumptions are required on  $(K, d)$  only, and that the constraint on  $\varepsilon$  depends only on  $(K, d)$  as well.

The estimation of  $K$  is a now well-understood in the manifold case (see [AL19]). To obtain guarantees on the estimation of  $\text{sdr}_\delta(K, d_K)$ , it hence remains to investigate the estimation of  $d_K$ . This is the aim of the following section.

## 3.5 Optimal Metric Learning

### 3.5.1 Unsupervised Distance Metric Learning

As explained in the introduction, various learning tasks lead to the problem of estimation the shortest-pat distance  $d_K$ , via an estimator  $\hat{d}$  on a sample of  $K \subset \mathbb{R}^D$ . Though,

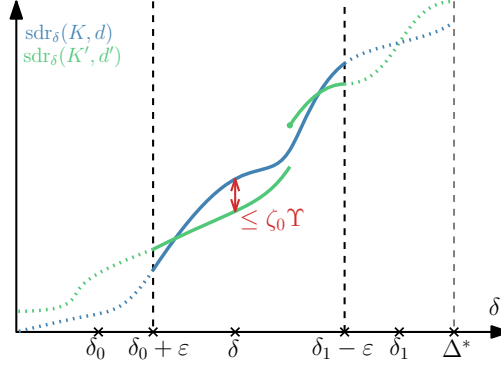


Figure 3.3 – Plot of  $\delta \mapsto \text{sdr}_\delta$  for  $(K, d)$  and  $(K', d')$  in the context of Theorem 3.23. On the interval  $(\delta_0 + \varepsilon, \delta_1 - \varepsilon)$ , the two functions do not differ of more than  $\zeta_0 \Upsilon$ . Even though  $(K', d')$  might not be well-behaved, the regularity of  $\delta \mapsto \text{sdr}_\delta(K, d)$  (Theorem 3.22) is sufficient to insure stability.

there is no canonical choice of loss for measuring the proximity of  $\widehat{d}$  to  $d_K$ . One could consider for instance the empirical sup-loss

$$\ell_n(\widehat{d}|d_K) := \sup_{x \neq y \in \mathbb{X}_n} \left| 1 - \frac{\widehat{d}(x, y)}{d_K(x, y)} \right|,$$

or the global sup-loss

$$\ell_\infty(\widehat{d}|d_K) := \sup_{x \neq y \in K} \left| 1 - \frac{\widehat{d}(x, y)}{d_K(x, y)} \right|.$$

It might seem counter-intuitive to ask an estimator  $\widehat{d}$  of  $d_K : K \times K \rightarrow \mathbb{R}_+$  to be defined on the whole set  $K \times K$ , while the this domain is unknown. It actually is easy to extend any metric estimator to the whole space  $\mathbb{R}^D \times \mathbb{R}^D$ . Indeed, given such a metric estimation procedure  $\widehat{d}_n : \mathbb{X}_n \times \mathbb{X}_n \rightarrow \mathbb{R}_+$  that outputs a distance  $\widehat{d}_n[\mathbb{X}_n](x, y)$  between any pair of points of  $\mathbb{X}_n$ , we can define  $\widehat{d}_n(x, y) := \widehat{d}_{n+2}[\mathbb{X}_n, x, y](x, y)$  for all  $(x, y) \in \mathbb{R}^D \times \mathbb{R}^D$ . Informally this means that one can treat any given tuple of points  $(x, y)$  as actual data points in the estimation process, and that we are only interested in the behavior of the later when  $x$  and  $y$  are in fact from  $K$ .

The losses  $\ell_n$  and  $\ell_\infty$  are naturally multiplicative, in particular because the usual notions of distortions are multiplicative by nature (see Section 3.4). Indeed, the sup-loss  $\ell_\infty(\widehat{d}|d_K)$  being smaller than  $\nu$  means that

$$\forall x, y \in K, \quad (1 - \nu)d_K(x, y) \leq \widehat{d}(x, y) \leq (1 + \nu)d_K(x, y),$$

which is the usual way to quantify if the intrinsic metric is well-estimated. See for instance [TSL00, ACC20]. When  $\nu$  is small, it yields that  $(K, \widehat{d})$  is quasi-isometric to  $(K, d_K)$ .

*Remark 3.24.* We emphasize the fact that the global sup-loss  $\ell_\infty$  and the mutual metric distortion  $D_\delta$  from Definition 3.16 are different in essence. Indeed, while the mutual

metric distortion  $D_\delta$  allows to compare different metrics on *different* subsets of  $\mathbb{R}^D$ , the sup-loss  $\ell_\infty$  compares two distances defined on the *same* subset.

However, the global sup-loss and the mutual distortion metric may be related as follows. Consider  $K$  endowed with either  $\widehat{d}$  or  $d_K$ . Denote by  $D_{0^+}(d_K, \widehat{d}) := \lim_{\delta \rightarrow 0} D_\delta(d_K, \widehat{d})$ . Then, straightforward computation entails

$$\ell_\infty(\widehat{d}|d_K) + 1 \leq D_{0^+}(d_K, \widehat{d}) \leq (1 - \ell_\infty(\widehat{d}|d_K))_+^{-1}.$$

Hence, the global sup-loss  $\ell_\infty(\widehat{d}|d_K)$  is somehow an additive counterpart to the mutual distortion  $D_{0^+}(\widehat{d}, d_K)$  in the case where  $K = K'$ . That is, when the support of the two metrics coincide in Definition 3.16, as already noticed in Remark 3.17.

When  $K = M$  is a  $\mathcal{C}^2$  submanifold of  $\mathbb{R}^D$  of dimension  $d$  with reach bounded below, methods using neighborhood graphs such as Isomap provably estimate  $d_M$  at rate  $O(n^{-2/3d})$  [ACLG19]. As we will show in Theorem 3.28, this rate is far from being optimal. To date, the best minimax lower bound in this setting is due to [ACC20], who obtain a rate of order  $\Omega(n^{-2/d})$  in the particular case of a deterministic design on  $\mathcal{C}^2$  submanifolds. Actually, we can extend the result of [ACC20] to our random design setting, and to general  $\mathcal{C}^k$  submanifolds with  $k \geq 2$ .

**Theorem 3.25.** *Assume that  $f_{\min} \leq c_{d,k}/\text{rch}_{\min}^d$  and  $f_{\max} \geq C_{d,k}/\text{rch}_{\min}^d$ , and  $L_j \geq C_{d,k}/\text{rch}_{\min}^{j-1}$  for all  $j \in \{2, \dots, k\}$ . Then for  $n$  large enough,*

$$\inf_{\widehat{d}} \sup_{P \in \mathcal{P}_{\text{rch}_{\min}^k, \mathbf{L}}(f_{\min}, f_{\max})} \mathbb{E}_{P^{\otimes n}}[\ell_\infty(\widehat{d}|d_M)] \geq \tilde{c}_{d,k,\text{rch}_{\min}} \left(\frac{1}{n}\right)^{k/d},$$

where the infimum is taken over all measurable estimator  $\widehat{d}$  of  $d_M$  based on  $n$  samples.

This theorem is proved in Appendix 3.C.1. As we shall prove shortly in Section 3.5.2, this lower-bound can be provided with a matching upper-bound up to  $\log n$  factors (Theorem 3.27), and is thus optimal.

### 3.5.2 An optimal Approach of Metric Estimation

The existing unsupervised methods for metric learning are known to either have no theoretical guarantees, or to have a sub-optimal rate for estimating the intrinsic metric. As stated before, Isomap reaches a rate of  $n^{-2/3d}$ , which is very far from the theoretical lower-bound  $n^{-k/d}$  shown in Theorem 3.25. Other methods, such as taking the shortest path distance over a Delaunay triangulation [ACC20], are shown to attain a precision of  $n^{-2/d}$  which is optimal for  $\mathcal{C}^2$ -model but not for  $k \geq 3$ . We propose here a fairly general approach that can output a family of minimax-optimal metric estimators. It relies on the following bound.

**Proposition 3.26.** *Let  $K \subset \mathbb{R}^D$  be a set of positive reach  $\text{rch}(K) > 0$ , and  $K' \subset \mathbb{R}^D$  be any set such that  $d_H(K', K) < \varepsilon \leq \text{rch}(K)/2$ . Then,*

$$\ell_\infty(d_{(K')^\varepsilon} | d_K) \leq \frac{2\varepsilon}{\text{rch}(K)},$$

where we recall that  $(K')^\varepsilon = \{u \in \mathbb{R}^D \mid d(u, K') \leq \varepsilon\}$ , so that  $K \subset (K')^\varepsilon$ .

Proposition 3.26 is proved in Appendix 3.C.2. It asserts that estimating geodesic distances of sets of positive reach is never harder than estimating the sets themselves in Hausdorff distance. Beyond the framework of closed manifold developed here, note that for the convex case  $\text{rch}(K) = \infty$ ,  $d_K$  coincides with the Euclidean metric, so that estimating  $d_K$  becomes trivial.

A significant consequence of Proposition 3.26 is that we can derive a consistent estimator of the intrinsic distance from any consistent estimator of the support, and with the same rate of convergence. In what follows, we write

$$d_{\max} := \frac{5^d}{\omega_d f_{\min} \text{rch}_{\min}^{d-1}}, \quad (3.11)$$

where  $\omega_d$  is the volume of the  $d$ -dimensional unit ball. In Lemma 3.C.2, the length  $d_{\max}$  is proved to be an upper bound on the geodesic diameter of the supports of any distribution in the model  $\mathcal{P}_{\text{rch}_{\min}, \mathbf{L}}^k(f_{\min}, f_{\max})$ .

**Theorem 3.27.** *Let  $k \geq 2$  and let  $\widehat{M}$  be an estimator satisfying*

$$\sup_{P \in \mathcal{P}_{\text{rch}_{\min}, \mathbf{L}}^k(f_{\min}, f_{\max})} P^{\otimes n}(d_H(\widehat{M}, M) \geq \varepsilon_n) \leq \eta_n,$$

for some positive sequences  $\varepsilon_n$  and  $\eta_n$  converging to 0. Then the metric estimator

$$\widehat{d}(x, y) := d_{\max} \wedge d_{(\widehat{M}_{x,y})^{\varepsilon_n}}(x, y) \quad \text{with} \quad \widehat{M}_{x,y} := \widehat{M} \cup \{x, y\},$$

which is defined for all  $x, y \in \mathbb{R}^D$ , satisfies

$$\sup_{P \in \mathcal{P}_{\text{rch}_{\min}, \mathbf{L}}^k(f_{\min}, f_{\max})} \mathbb{E}_{P^{\otimes n}}[\ell_\infty(\widehat{d} | d_M)] \leq \frac{2}{\text{rch}_{\min}} \varepsilon_n + \left(1 + \frac{d_{\max}}{\varepsilon_n}\right) \eta_n.$$

Theorem 3.27 is proved in Appendix 3.C.2. A particular advantage of this result is that it does not require the estimator  $\widehat{M}$  to have any geometric structure, nor to be regular in any sense. This contrasts sharply with [ACC20], which extensively uses the structural properties of the intermediate estimator  $\widehat{M}$ . Theorem 3.27 is much more versatile, since here,  $\widehat{M}$  could just as easily be anything as a point cloud, a metric graph, a triangulation, or a union of polynomial patches. For instance, taking

$\widehat{M} = \{X_1, \dots, X_n\}$  to be the observed data, we can take  $\varepsilon_n = C(\log n/n)^{1/d}$  for  $C$  large enough yields  $\eta_n \leq \varepsilon_n^2$  so that

$$\sup_{P \in \mathcal{P}_{\text{rch}_{\min}, \mathbf{L}}^2(f_{\min}, f_{\max})} \mathbb{E}_{P^{\otimes n}} [\ell_{\infty}(\widehat{d}|d_M)] \leq C_{\text{rch}_{\min}, d, f_{\min}} \left( \frac{\log n}{n} \right)^{1/d},$$

which is faster than the known rate of order  $O(n^{-2/3d})$  for Isomap (see for instance [ACC20, Eq (1.2)]). Now, taking  $\widehat{M}$  to be a minimax optimal estimator of  $M$  for the Hausdorff loss — as that of [AL19], for instance — and  $\varepsilon_n = C(\log n/n)^{k/d}$  for some large constant  $C > 0$  yields  $\eta_n \leq \varepsilon_n^2$  (see Lemma 3.A.4), and a metric estimator  $\widehat{d}$  that achieves the following rate.

**Theorem 3.28.** *Let  $\widehat{d}$  be the estimator described in Theorem 3.27 built on top of  $\widehat{M}$  described in Lemma 3.A.4. Then for  $n$  large enough,*

$$\sup_{P \in \mathcal{P}_{\text{rch}_{\min}, \mathbf{L}}^k(f_{\min}, f_{\max})} \mathbb{E}_{P^{\otimes n}} [\ell_{\infty}(\widehat{d}|d_M)] \leq C_{\text{rch}_{\min}, d, f_{\max}, f_{\min}, \mathbf{L}, k} \left( \frac{\log n}{n} \right)^{k/d}.$$

In virtue of Theorem 3.25, this rate is minimax optimal up to  $\log n$  factors.

## 3.6 Optimal Reach Estimation

### 3.6.1 Optimal Spherical Distortion Radius Estimation

Interesting as it is in its own right, we now investigate the estimation rates of the spherical distortion radius at scale  $\delta > 0$ . To obtain a minimax lower bound, we simply note that  $\text{sdr}_{\delta}(M, d_M)$  coincides with  $\text{rch}(M)$  whenever  $\text{rch}(M) = \text{wfs}(M)$  (Proposition 3.14). Hence, any lower bound for the estimation of  $\text{rch}(M)$  on a model over which  $\text{rch}(M) = \text{wfs}(M)$  yields a lower bound for the estimation of  $\text{sdr}_{\delta}(M, d_M)$ . In application of Theorem 3.10 with  $\alpha \geq 0$ , this immediately gives the following lower bound.

**Theorem 3.29.** *Assume that  $f_{\min} \leq c_{d,k}/\text{rch}_{\min}^d$  and  $f_{\max} \geq C_{d,k}/\text{rch}_{\min}^d$ , and  $L_j \geq C_{d,k}/\text{rch}_{\min}^{j-1}$  for all  $j \in \{2, \dots, k\}$ . Then for  $n$  large enough, for all  $\delta \in (0, \text{rch}_{\min})$ ,*

$$\inf_{\widehat{\text{sdr}}_{\delta}} \sup_{P \in \mathcal{P}_{\text{rch}_{\min}, \mathbf{L}}^k(f_{\min}, f_{\max})} \mathbb{E}_{P^{\otimes n}} [|\widehat{\text{sdr}}_{\delta} - \text{sdr}_{\delta}(M, d_M)|] \geq \tilde{c}_{\text{rch}_{\min}, d, k} n^{-k/d}.$$

where the infimum is taken over all measurable estimators  $\widehat{\text{sdr}}_{\delta}$  of  $\text{sdr}_{\delta}(M, d_M)$  based on  $n$  samples.

It turns out that this bound is optimal. To exhibit an estimator that achieves this rate, we take advantage of the Hausdorff and metric stability of the spherical distortion radius shown in Theorem 3.23. In order to apply it, we first need to check that Assumptions 3.20 and 3.21 are fulfilled for every manifolds in our models  $\mathcal{C}_{\text{rch}_{\min}, \mathbf{L}}^k$ .

**Proposition 3.30.** *Let  $M \subset \mathbb{R}^D$  be a submanifold with bounded reach  $\text{rch}(M) > 0$ . Then  $M$  satisfies Assumptions 3.20 and 3.21 with parameters*

$$\varepsilon_0 = \text{rch}(M)/4, \quad \Delta_0 = \text{rch}(M), \quad C_0 = 3/16, \quad \Delta_1 = \text{rch}(M)/2 \quad \text{and} \quad C_1 = 2.$$

Proposition 3.30 is proven in Appendix 3.D. In the vein of Theorem 3.27, and using the stability of the spherical distortion radius with respect to the pair  $(K, d)$ , we can now build an estimator of  $\text{sdr}_\delta(M, d_M)$  in a plug-in fashion over  $\mathcal{C}^k$  submanifolds. Recall that when  $M$  is in  $\mathcal{C}_{\text{rch}_{\min}, \mathbf{L}}^k$ , and  $\delta \in (0, \sqrt{2(D+1)/D} \text{wfs}(M))$ , then according to Propositions 3.14 and 3.5, and to Lemma 3.C.2,

$$0 < \text{rch}_{\min} \leq \text{rch}(M) \leq \text{sdr}_\delta(M, d_M) \leq \text{wfs}(M) \leq \sqrt{\frac{D}{2(D+1)}} \text{diam}(M) \leq s_{\max} < \infty,$$

where  $s_{\max} := \sqrt{D/(2(D+1))} d_{\max}$ , with  $d_{\max}$  being the constant introduced in (3.11).

**Theorem 3.31.** *Given  $k \geq 2$ , let  $\widehat{M}$  be an estimator satisfying*

$$\sup_{P \in \mathcal{P}_{\text{rch}_{\min}, \mathbf{L}}^k(f_{\min}, f_{\max})} P^{\otimes n}(\text{d}_H(M, \widehat{M}) \geq \varepsilon_n) \leq \eta_n$$

for some positive sequences  $\varepsilon_n, \eta_n$  converging to 0. Then, for any  $\delta \in (0, \text{rch}_{\min})$ , the estimator  $\widehat{\text{sdr}}_\delta := \text{sdr}_\delta(\widehat{M}, \widehat{d}) \wedge s_{\max}$ , where  $\widehat{d}$  is defined in Theorem 3.27, satisfies

$$\sup_{P \in \mathcal{P}_{\text{rch}_{\min}, \mathbf{L}}^k(f_{\min}, f_{\max})} \mathbb{E}_{P^{\otimes n}} |\widehat{\text{sdr}}_\delta - \text{sdr}_\delta(M, d_M)| \leq C \left( \frac{s_{\max}^4}{\delta^4} \varepsilon_n + s_{\max} \eta_n \right).$$

We refer to Appendix 3.D for a proof of this result.

**Remark 3.32.** In place of  $\widehat{d} = d_{\widehat{M}^{\varepsilon_n}}$ , one could actually plug any estimator  $\widehat{d}$  of the metric into Theorem 3.31. In light of the stability result of Theorem 3.23, as long as  $\widehat{d}$  satisfies

$$\sup_{P \in \mathcal{P}^k} P^{\otimes n} \left( D_\delta(\widehat{d}, d_M) \geq 1 + \frac{\varepsilon_n}{\delta} \right) \leq \eta_n,$$

the conclusion of Theorem 3.31 would still hold. This comes in handy, especially if one wants to input a computationally efficient distance estimator, such as shortest-path distance on a neighborhood graph [TSL00] or on Delaunay triangulations [ACC20].

Again, taking  $\widehat{M}$  to be a minimax optimal estimator for the Hausdorff loss [AL19] outputs an estimator  $\widehat{\text{sdr}}_\delta$  of the spherical distortion radius satisfying

**Theorem 3.33.** *For all  $\delta \in (0, \text{rch}_{\min})$ , with the construction of  $\widehat{\text{sdr}}_\delta$  above, we have that for  $n$  large enough,*

$$\sup_{P \in \mathcal{P}_{\text{rch}_{\min}, \mathbf{L}}^k(f_{\min}, f_{\max})} \mathbb{E}_{P^{\otimes n}} |\widehat{\text{sdr}}_\delta - \text{sdr}_\delta(M, d_M)| \leq C_{\text{rch}_{\min}, d, f_{\max}, f_{\min}, \mathbf{L}, k} \frac{1}{\delta^4} \left( \frac{\log n}{n} \right)^{k/d},$$

and this rate is optimal in regard of Theorem 3.29.



Note the presence of the factor  $1/\delta^4$  in the bound, which makes the rate diverge as  $\delta \rightarrow 0$ . This blowup is to be expected for the following reason. As  $\delta$  goes to 0, the spherical distortion radius goes to the reach  $\text{rch}(M)$  (Proposition 3.14). Since the estimation of  $\text{rch}(M)$  cannot be faster than  $n^{-(k-2)/d}$  (Theorem 3.10), the estimation rate of  $\text{sdr}_\delta(M, d_M)$  must deteriorate in some way as  $\delta \rightarrow 0$ .

### 3.6.2 Optimal Reach Estimation

In light of Proposition 3.12 and (3.9), it only remains to combine the maximal curvature estimator and the spherical distortion radius estimator to obtain an estimator of the reach. Naely, we let  $\widehat{M}$  be the minimax-Hausdorff estimator of Lemma 3.A.4. According to the very same Lemma 3.A.4, there exists  $c_{\text{rch}_{\min}, d, f_{\max}, f_{\min}, \mathbf{L}, k} > 0$  such that denoting by

$$\varepsilon_n := c_{\text{rch}_{\min}, d, f_{\max}, f_{\min}, \mathbf{L}, k} \left( \frac{\log n}{n} \right)^{k/d}, \quad (3.12)$$

there holds

$$\sup_{P \in \mathcal{P}_{\text{rch}_{\min}, \mathbf{L}}^k(f_{\min}, f_{\max})} P^{\otimes n}(\text{d}_H(\widehat{M}, M) \geq \varepsilon_n) \leq \varepsilon_n^2. \quad (3.13)$$

We also let  $\widehat{d}$  be the estimator of the intrinsic distance of Theorem 3.27 from  $\widehat{M}$  and  $\varepsilon_n$ . We let  $\widehat{\text{sdr}}_\delta := \text{sdr}_\delta(\widehat{M}, \widehat{d}) \wedge s_{\max}$  for some  $\delta \in (0, \text{rch}_{\min})$  as in Theorem 3.31. Finally, we write

$$\widehat{\text{rch}} := \widehat{R}_\ell \wedge \widehat{\text{sdr}}_\delta.$$

The following Theorem 3.34 is a straightforward consequence of Theorems 3.11 and 3.31, inserted in the plugin strategy of Proposition 3.12 and (3.9).

**Theorem 3.34.** *The estimator  $\widehat{\text{rch}}$  described above with  $\delta = \text{rch}_{\min}/2$  satisfies*

$$\sup_{P \in \mathcal{P}_{\text{rch}_{\min}, \mathbf{L}}^k(f_{\min}, f_{\max})} \mathbb{E}_{P^{\otimes n}} |\widehat{\text{rch}} - \text{rch}(M)| \leq C_{\text{rch}_{\min}, d, f_{\max}, f_{\min}, \mathbf{L}, k} \left( \frac{\log n}{n} \right)^{(k-2)/d},$$

and, for all  $\alpha > 0$ ,

$$\sup_{P \in \mathcal{P}_{\text{rch}_{\min}, \mathbf{L}, \alpha}^k(f_{\min}, f_{\max})} \mathbb{E}_{P^{\otimes n}} |\widehat{\text{rch}} - \text{rch}(M)| \leq C_{\text{rch}_{\min}, d, f_{\max}, f_{\min}, \mathbf{L}, k, \alpha} \left( \frac{\log n}{n} \right)^{k/d}.$$

As a conclusion, Theorems 3.10 and 3.34 assert that  $\widehat{\text{rch}}$  is minimax optimal, and that its rate of convergence adapts to whether  $\text{rch}(M)$  is attained by curvature (yielding the slower rate  $O(n^{-(k-2)/d})$ ) or by a bottleneck (yielding the faster rate  $O(n^{-k/d})$ ).

The computation of  $\widehat{\text{rch}}$  depends explicitly on the parameters of the models at two levels. First, in tuning the value of  $\varepsilon_n$  as in (3.12). Second, in choosing  $\delta \in (0, \text{rch}_{\min})$ . These two dependencies may be circumvented by picking

$$\widetilde{\varepsilon}_n = \log n \left( \frac{\log n}{n} \right)^{k/d},$$

and  $\delta_n = 1/\log n$ . Then, for  $n$  large enough, both (3.13) and  $\delta_n \in (0, \text{rch}_{\min})$  will be fulfilled. The price to pay for this way-around to calibration of constants limits to multiplicative  $\log n$  factors in the upper-bound of Theorem 3.34.

### 3.7 Conclusion and Further Prospects

We developed a general strategy for estimating the reach of a manifold  $M$ . It relies on two independent plugins, accountable for the estimation of the minimal curvature radius  $R_\ell(M)$  and any another set-defined feature size  $\theta(M)$  that lie between the reach and the weak feature size. We then introduced and studied the spherical distortion radius, the estimation of which reduces to geodesic distance estimation, itself reducing to set estimation in Hausdorff distance. All the derived results are minimax optimal, as testified by associated matching lower bounds up to  $\log n$  factors.

Geometrically, one should note that this overall method relies heavily on the local/global dichotomy of the reach for closed submanifolds [AKC<sup>+</sup>19]. Hence, it still remains unclear how to extend it to manifolds with boundary, even though their curvature and spherical distortion radius are likely to be estimated in a similar way [AAL21].

On the statistical side, a major extension of the results would consist in allowing for additive noise. Recent works obtained Hausdorff estimation rates for the support [FILN19, AS21, PS22] in such a noisy setting, so that the estimation of the spherical distortion radius inherits the same rates straightforwardly. In the same spirit as the iterated local polynomial fitting of [AS21], we expect that the same method could likewise lead to maximal curvature estimation.

Finally, since the main goal of this work was of minimax nature, we did not focus on the algorithmic properties of our estimators. As they stand,  $\widehat{R}_\ell$  and  $\widehat{\text{sdr}}$  both require to compute a supremum over the union of continuous patches  $\widehat{M}$ , which is computationally prohibitive. Actually, one can easily show that taking the same supremum over a discretization of  $\widehat{M}$  at scale  $O(n^{-\beta/d})$  – i.e.  $O(n^\beta)$  points in total – yields estimation rates of order  $O(n^{-(\beta \wedge (k-2))/d})$  for  $R_\ell(M)$ , and  $O(n^{-(\beta \wedge k)/d})$  for  $\text{sdr}_\delta(M, d_M)$ . This suggests a possible estimation-computation tradeoff which one could take advantage of. Yet, this is not a fully satisfactory solution, as  $\text{sdr}_\delta$  still requires to compute costly geodesic distances on a high-dimensional set. More globally, the quest for computationally efficient – yet optimal – geometric estimators in high dimensions is still in its infancy.

# Appendix

## Contents

---

<b>3.A Proofs of Section 3.3</b>	<b>81</b>
3.A.1 Comparing Reaches, Weak Feature Size and Diameter	81
3.A.2 Minimax Lower Bound for $\mu$ -Reach Estimation	82
3.A.3 Maximal Curvature Estimation	86
3.A.4 Proof of Lemma 3.A.3	91
<b>3.B Proofs of Section 3.4</b>	<b>93</b>
3.B.1 Comparing Reach, Weak Feature Size and Spherical Distortion Radius	93
3.B.2 Stability Properties of the Spherical Distortion Radius	93
<b>3.C Proofs of Section 3.5</b>	<b>96</b>
3.C.1 Minimax Lower Bound for Metric Learning	96
3.C.2 Plug-in Estimation for Metric Learning	103
<b>3.D Proofs of Section 3.6</b>	<b>104</b>

---

## 3.A Proofs of Section 3.3

### 3.A.1 Comparing Reaches, Weak Feature Size and Diameter

This Section is devoted to the Proof of Proposition 3.5, which goes as follows.

*Proof of Proposition 3.5.* For (i), recall that no closed compact submanifold can be contractible [Hat02, Theorem 3.26]. Furthermore, [Fed59, Theorem 4.8] and [CCSL09, Lemma 2.1] combined together yield that  $K^r$  is isotopic to  $K$  for all  $r < \text{wfs}(K)$ . On the other hand whenever  $r > \text{Rad}(K)$  where  $\text{Rad}(K)$  is the radius of the smallest ball enclosing  $K$ ,  $K^r$  is star-shaped with respect to any point of the non-empty intersection  $\cap_{x \in M} \mathbf{B}(x, r)$ . We conclude that  $\text{wfs}(K) \leq \text{Rad}(K)$ , Since  $\text{Rad}(K) < \infty$  because  $K$  is compact, we obtain  $\text{wfs}(K) < \infty$ .

For (ii), the first two inequalities come from the definition of  $\text{rch}_\mu(K)$  (see (3.4)). The rightmost comes Jung's Theorem [Fed69, Theorem 2.10.41], which asserts that  $\text{Rad}(K) \leq \sqrt{\frac{D}{2(D+1)}} \text{diam}(K)$ , and the fact that  $\text{wfs}(K) \leq \text{Rad}(K)$  whenever  $\text{wfs}(K)$  is finite (same argument as for (i)).  $\square$

### 3.A.2 Minimax Lower Bound for $\mu$ -Reach Estimation

This Section is devoted to the proof of Theorem 3.6. It builds upon the possible discontinuities of the map  $M \mapsto \text{Med}_\mu(M)$  in Hausdorff distance. The exhibition of such a discontinuity can be done in dimension  $d = 1$  and  $D = 2$ , and can then be generalized to arbitrary  $1 \leq d < D$  by using symmetry and rotation arguments.

The building block of the construction is the following arc of curve. For all  $\alpha \in (0, \pi/4]$ , write  $R_\alpha := 1/\sin(\alpha)$ . Let also  $C_\alpha : [0, 1] \rightarrow \mathbb{R}_+$  be defined as  $C_\alpha(t) := R_\alpha - \sqrt{R_\alpha^2 - t^2}$ , which graph is an arc of circle of radius  $R_\alpha$  and aperture  $\alpha$  (see Figure 3.A.1). To be able to glue up smoothly  $\alpha$ -turns like  $C_\alpha$  with straight lines, we smooth it as follows.

**Lemma 3.A.1.** *There exists  $G_\alpha : [0, 1] \rightarrow \mathbb{R}_+$  infinitely differentiable such that:*

1.  $G_\alpha^{(\ell)}(0) = 0$  for all  $\ell \geq 0$ ;
2.  $G_\alpha(1) = C_\alpha(1)$ ,  $G'_\alpha(1) = C'_\alpha(1)$  and  $G_\alpha^{(\ell)}(1) = 0$  for all  $\ell \geq 2$ ;
3.  $\|G_\alpha^{(\ell)}\|_\infty \leq C_\ell/R_\alpha$  for all  $\ell \geq 1$ ;
4.  $G_\alpha(t) < C_\alpha(t)$  for all  $t \in (0, 1)$ ;
5.  $G_\alpha$  is convex.

See Figure 3.A.1 for a diagram of such a  $G_\alpha$ . Let us first comment on the requirements on  $G_\alpha$ . Items 1 and 2 say that  $G_\alpha$  is a  $\mathcal{C}^k$  interpolation between the two tangent lines of two points of  $C_\alpha$  who are  $\alpha$ -apart in term of polar coordinate. Item 3 says that the graph of  $G_\alpha$ , once rescaled by  $1/R_\alpha$ , will be bounded in  $\mathcal{C}^k$ -norm for all  $k$ . Items 4 and 5 ensure well-behavior of the medial axes of our future construct (see Figure 3.A.2).

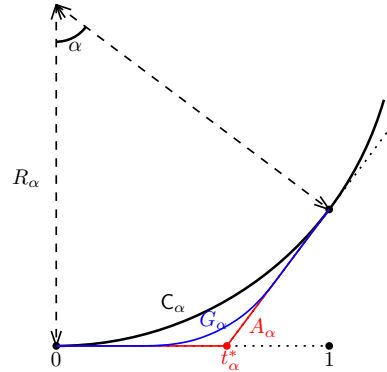


Figure 3.A.1 – Construction for Lemma 3.A.1: curves associated to  $C_\alpha$ ,  $A_\alpha$ , and  $G_\alpha$ .

*Proof of Lemma 3.A.1.* The following construction applies to general convex functions, although we restrict it to  $G_\alpha$  for simplicity. Consider the piecewise linear map  $A_\alpha$  given by the tangent lines of  $C_\alpha$  at  $t = 0$  and  $t = 1$ . That is, define  $A_\alpha(t)$  for all  $t \in \mathbb{R}$  by

$$\begin{aligned} A_\alpha(t) &:= \max \{C_\alpha(0) + C'_\alpha(0)t, C_\alpha(1) + (t-1)C'_\alpha(1)\} \\ &= \max \{0, C_\alpha(1) + (t-1)C'_\alpha(1)\}. \end{aligned}$$

As  $C_\alpha$  is strictly convex,  $A_\alpha < C_\alpha$  on  $\mathbb{R} \setminus \{0, 1\}$ . We also denote by  $t_\alpha^*$  the (unique) point of non-differentiability of  $A_\alpha$ , that is

$$t_\alpha^* := 1 - \frac{C_\alpha(1)}{C'_\alpha(1)} = R_\alpha \tan(\alpha/2).$$

Note by now that for all  $\alpha \in (0, \pi/4)$ ,  $1/2 \leq t_\alpha^* \leq 2 - \sqrt{2} \leq 6/10$ . Given  $h > 0$  to be chosen later, write  $K_h(t) := h^{-1}K(t/h)$ , where  $K(t) := c_0 \exp(-1/(1-t^2)) \mathbb{1}_{|t|<1}$  is a non-negative  $C^\infty$  kernel, and  $c_0$  is chosen so that  $\int_{\mathbb{R}} K = 1$ . Finally, consider the convolution

$$G_\alpha(t) := \int_{\mathbb{R}} K_h(x) A_\alpha(t-x) dx.$$

By smoothness of  $K_h$  and non-negativity of both  $K_h$  and  $A_\alpha$ ,  $G_\alpha = K_h * A_\alpha$  is infinitely differentiable and non-negative. Also, since  $A_\alpha$  is convex and  $K_h$  non-negative,  $G_\alpha$  is convex (Item 5). Furthermore, one easily checks that outside the interval  $[t_\alpha^* - h, t_\alpha^* + h]$ ,  $G_\alpha$  coincides with  $A_\alpha$ . Hence, if  $h \leq 1/4$ , we have  $[t_\alpha^* - h, t_\alpha^* + h] \subset [1/2 - 1/4, 6/10 + 1/4] = [1/4, 17/20]$ , so that Items 1 and 2 holds directly.

To check that  $G_\alpha < C_\alpha$  on  $(0, 1)$ , fix  $t \in (0, 1)$ . If  $t \notin [t_\alpha^* - h, t_\alpha^* + h]$ ,  $G_\alpha(t) = A_\alpha(t) < C_\alpha(t)$  by construction. If  $t \in [t_\alpha^* - h, t_\alpha^* + h]$ , we have  $G_\alpha(t) \leq G_\alpha(t_\alpha^* + h) = hC'_\alpha(1)$ . But on the other hand,  $C_\alpha(t) \geq C_\alpha(t_\alpha^* - h) > C_\alpha(1/4)$ . Hence, we do have  $G_\alpha(t) < C_\alpha(t)$  as soon as  $h \leq 1/100$ , since  $C_\alpha(1/4)/C'_\alpha(1) > 1/100$  for all  $\alpha \in (0, \pi/4)$ . This yields Item 4.

Finally, letting  $h = h_0 = 1/100$ , we obtain for all  $\ell \geq 1$  and  $t \in [0, 1]$ ,

$$|G_\alpha^{(\ell)}(t)| = |K_h^{(\ell)} * C_\alpha(t)| \leq \|K_h^{(\ell)}\|_\infty \|C_\alpha\|_\infty \leq C_\ell C_\alpha(1) \leq C_\ell / R_\alpha,$$

which yields Item 3 and concludes the proof.  $\square$

Given  $R > 0$ , we now let  $G_{\alpha,R}$  be the curve obtained by dilating homogeneously the graph of  $G_\alpha$  by a scale factor  $R/R_\alpha$ . We extend the construction of these smooth  $\alpha$ -turns for  $\alpha \in (\pi/4, \pi]$ : for this, we glue two  $G_{\alpha/2,R}$  or four  $G_{\alpha/4,R}$  to define  $G_{\alpha,R}$ .

**Proposition 3.A.2.** *Assume that for all  $j \in \{2, \dots, k\}$ ,  $L_j \geq C_{d,k} / \text{rch}_{\min}^{j-1}$  for  $C_{d,k} > 0$  large enough. Then for all  $\mu \in [0, 1)$  and  $\varepsilon > 0$  small enough, there exist  $M, M' \in C_{\text{rch}_{\min}, \mathbf{L}}^k$  such that:*

- $|\text{rch}_\mu(M) - \text{rch}_\mu(M')| \geq c_{d,k} \text{rch}_{\min}$  ;
- $c'_{d,k} \text{rch}_{\min}^d \leq \text{vol}_d(M) \wedge \text{vol}_d(M') \leq \text{vol}_d(M) \vee \text{vol}_d(M') \leq C''_{d,k} \text{rch}_{\min}^d$  ;

- $\text{vol}_d(M \triangle M') \leq C_{d,k}''' \text{rch}_{\min}^d \varepsilon$ .

*Proof of Proposition 3.A.2.* For small enough (and arbitrarily small)  $\varepsilon > 0$ , we let  $\alpha \in [0, \pi]$  be such that  $\sin((\alpha + \varepsilon)/2)^2 = 1 - \mu^2$ . Such an  $\alpha$  always exists since  $\mu^2 < 1$ . Given  $\Delta, R_0, R_1 > 0$  to be chosen later, we glue smooth turns from Lemma 3.A.1 with straight lines to create a  $\mathcal{C}^k$  closed curve in  $\mathbb{R}^2$ , as shown in Figure 3.A.2. Then, we obtain a  $\mathcal{C}^k$  closed  $d$ -dimensional submanifold  $M_\alpha$  of  $\mathbb{R}^{d+1}$ , with a symmetry of revolution with respect to the horizontal axis of Figure 3.A.2.

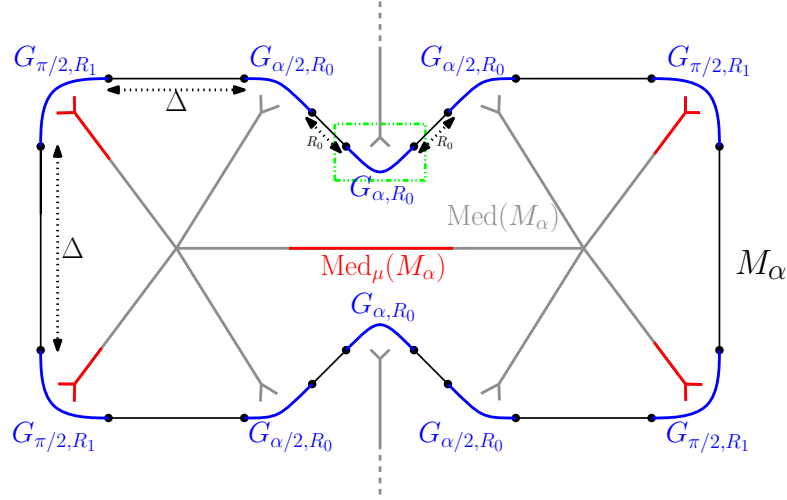


Figure 3.A.2 – Construction of  $M_\alpha$  in the proof of Proposition 3.A.2.

By construction, if  $\Delta \geq 8R_0$ , then  $M_\alpha$  has local parametrizations on top of its tangent spaces (see Definition 3.1) with  $L_j \leq C_{d,k}/(\Delta \wedge R_0 \wedge R_1)^{j-1}$  for all  $j \geq 2$ , and has volume  $\text{vol}_d(M_\alpha) \leq C_{d,k}(\Delta \vee R_0 \vee R_1)^d$  and  $\text{vol}_d(M_\alpha) \geq c_{d,k}(\Delta \wedge R_0 \wedge R_1)^d$ .

We now examine the structure of the medial axis and the reach of  $M_\alpha$ . If  $u \in \text{Med}(M_\alpha)$  is a point on the medial axis, rotational symmetry yields that two of its projections points must lie either:

- In a plane containing its horizontal axis of symmetry (i.e. Figure 3.A.2). As a result, its distance to  $M_\alpha$  cannot be smaller than the smallest reach of each of its parts  $G_{\pi/2, R_1}$ ,  $G_{\alpha/2, R_0}$  and  $G_{\alpha, R_0}$ , so that  $d(u, M_\alpha) \geq c_{d,k}R_0 \wedge R_1$ .
- In a  $d$ -plane orthogonal to the horizontal axis. By rotational invariance, this forces  $u$  to be on this axis of symmetry. As a result,  $d(u, M_\alpha) \geq \Delta/2 - 3R_0 \geq c_{d,k}\Delta$  since  $\Delta \geq 8R_0$ .

In all, we get  $\text{rch}(M_\alpha) \geq c_{d,k}(\Delta \wedge R_0 \wedge R_1)$ .

We now examine the  $\mu$ -reach of  $M_\alpha$ . By definition, if  $u \in \text{Med}_\mu(M_\alpha)$  has two nearest neighbors  $x, y \in M_\alpha$ , the angle between  $(u - x)$  and  $(u - y)$  must be at most  $2 \arcsin(\sqrt{1 - \mu^2})$ . As a result, a single branch of  $M_\alpha$  between the two arcs of  $G_{\alpha/2, R_0}$

cannot not generate any point of the  $\mu$ -medial axis, since  $\alpha$  has been chosen so that  $\alpha < 2 \arcsin(\sqrt{1 - \mu^2})$ . Hence, for  $\Delta, R_1$  large enough compared to  $R_0$ , we have  $\text{rch}_\mu(M_\alpha) \geq c'_{d,k}(\Delta \wedge R_1)$ .

Finally, we build  $M'_\alpha$  from  $M_\alpha$  by bumping the curve near  $G_{\alpha, R_0}$  as shown in Figure 3.A.3 (while still preserving the radial symmetry as before). The manifold  $M'_\alpha$  satisfies the same regularity conditions at  $M_\alpha$ . Furthermore,  $M_\alpha$  and  $M'_\alpha$  only differ on a set of volume  $\text{vol}_d(M_\alpha \triangle M'_\alpha) \leq C_{d,k}(\Delta \vee R_1)^{d-1}(R_0\varepsilon)$ .

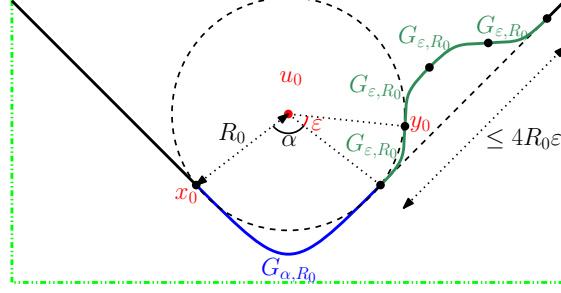


Figure 3.A.3 – Local bump of  $M'_\alpha$  for Proposition 3.A.2, in the boxed area of Figure 3.A.2.

With this extra bump, we create a point  $u_0 \in \text{Med}(M'_\alpha)$  that has two nearest neighbors  $x_0, y_0 \in M'_\alpha$  at distance  $R_0$ , with angle between  $(u_0 - x_0)$  and  $(u_0 - y_0)$  equal to  $\alpha' = \alpha + \varepsilon$ , which satisfies  $\sin(\alpha'/2)^2 = 1 - \mu^2$ . As a result,  $u_0 \in \text{Med}_\mu(M'_\alpha)$ , so that  $\text{rch}_\mu(M'_\alpha) \leq \|u_0 - y_0\| = R_0$ . In particular, we have

$$|\text{rch}_\mu(M_\alpha) - \text{rch}_\mu(M'_\alpha)| \geq c'_{d,k}(\Delta \wedge R_1) - R_0.$$

The proof is hence complete by setting  $M = M_\alpha$  and  $M' = M'_\alpha$ , with  $R_1 = \Delta = R_0/c'_{d,k}$  and  $R_0 = \text{rch}_{\min}/c_{d,k}$  for small enough  $c_{d,k}, c'_{d,k} > 0$ .  $\square$

*Proof of Theorem 3.6.* From Proposition 3.A.2, for  $\varepsilon > 0$  small enough, take  $M, M' \in \mathcal{C}_{\text{rch}_{\min}, \mathbf{L}}^k$  such that  $|\text{rch}_\mu(M) - \text{rch}_\mu(M')| \geq c_{d,k} \text{rch}_{\min}$ ,  $c'_{d,k} \text{rch}_{\min}^d \leq \text{vol}_d(M), \text{vol}_d(M') \leq C_{d,k} \text{rch}_{\min}^d$ , and  $\text{vol}_d(M \triangle M') \leq C'_{d,k} \text{rch}_{\min}^d \varepsilon$ . Let us denote by  $P$  and  $P'$  the uniform distributions over  $M$  and  $M'$  respectively. Elementary calculations directly yield that

$$\text{TV}(P, P') \leq \frac{\text{vol}_d(M \triangle M')}{\text{vol}_d(M) \vee \text{vol}_d(M')} \leq C''_{d,k} \varepsilon.$$

Furthermore, since  $c'_{d,k} \text{rch}_{\min}^d \leq \text{vol}_d(M) \wedge \text{vol}_d(M') \leq \text{vol}_d(M) \vee \text{vol}_d(M') \leq C_{d,k} \text{rch}_{\min}^d$ , we obtain that  $P, P' \in \mathcal{P}_{\text{rch}_{\min}, \mathbf{L}}^k(f_{\min}, f_{\max})$  as soon as  $f_{\min} \leq 1/(C_{d,k} \text{rch}_{\min}^d)$  and  $f_{\max} \geq 1/(c'_{d,k} \text{rch}_{\min}^d)$ . As a result, for all  $n \geq 1$ , Le Cam's Lemma [Yu97] yields

$$\begin{aligned} \inf_{\widehat{r}_\mu} \sup_{P \in \mathcal{P}_{\text{rch}_{\min}, \mathbf{L}}^k(f_{\min}, f_{\max})} \mathbb{E}_{P^{\otimes n}} [|\widehat{r}_\mu - \text{rch}_\mu(M)|] \\ \geq \frac{1}{2} |\text{rch}_\mu(M) - \text{rch}_\mu(M')| (1 - \text{TV}(P, P'))^n \\ \geq c_{d,k} \text{rch}_{\min} (1 - \varepsilon)^n. \end{aligned}$$

As this construction is valid for all  $\varepsilon > 0$  small enough, we obtain the result by letting  $\varepsilon$  tend to zero.  $\square$

### 3.A.3 Maximal Curvature Estimation

This section is devoted to the proof of Theorem 3.11. It is based on a careful investigation of the local polynomial fitting procedure described in [AL19]. First, recall that from [AL19, Lemma 2], if  $M \in \mathcal{C}_{\text{rch}_{\min}, \mathbf{L}}^k$ ,  $y \in M$  and  $y' \in B(y, \frac{L_2 \wedge \text{rch}_{\min}}{4}) \cap M$ , we may write

$$\begin{aligned} y' - y &= \pi_y^*(y' - y) + \mathbb{T}_y^{(2),*}(\pi_y^*(y' - y)^{\otimes 2}) + \dots + \mathbb{T}_y^{(k-1),*}(\pi_y^*(y' - y)^{\otimes k-1}) \\ &\quad + R_y^{(k)}(y' - y), \end{aligned} \quad (3.14)$$

where  $\pi_y^* := \pi_{T_y M}$ ,  $\mathbb{T}_y^{(j),*}$  are  $j$ -multilinear maps from  $T_y M$  to  $\mathbb{R}^D$ , and  $R_y^{(k)}$  satisfies

$$\|R_y^{(k)}(y' - y)\| \leq C t_*^{k-1} \|y' - y\|^k,$$

where  $t_* = \max_{2 \leq j \leq k, y \in M} \|\mathbb{T}_y^{(j),*}\|_{\text{op}}^{\frac{1}{j-1}} \leq C_{k,d,\text{rch}_{\min}, \mathbf{L}}$ . As assessed by [AL19, Lemma 2], the polynomial decomposition expressed in (3.14) allows to recover the curvature tensor via  $\mathbb{I}_y M = \mathbb{T}_y^{(2),*}$ . Following [AL19], we estimate this curvature tensor via the second term of the polynomial decomposition provided by local fit to data points (3.6). To this aim, a slight adaptation of [AL19, Lemma 3] is needed, that allows to translate quality of approximation in terms of Hausdorff distance to guarantees on the monomial terms.

**Lemma 3.A.3.** *Set  $h_0 = (\tau_{\min} \wedge L_2^{-1})/8$  and  $h \leq h_0$ . Let  $M \in \mathcal{C}_{\tau_{\min}, \mathbf{L}}^k$ ,  $x_0 = y_0 + z_0$ , with  $y_0 \in M$  and  $\|z_0\| \leq \sigma \leq h/4$ . Denote by  $\pi_{y_0}^*$  the orthogonal projection onto  $T_{y_0} M$ , and by  $\mathbb{T}_{y_0}^{(2),*}, \dots, \mathbb{T}_{y_0}^{(k-1),*}$  the multilinear maps given by (3.14).*

*Let  $x = y + z$  be such that  $y \in M$ ,  $\|z\| \leq \sigma \leq h/4$  and  $x \in B(x_0, h)$ . We also let  $\pi$  be an orthogonal projection, and  $\mathbb{T}^{(2)}, \dots, \mathbb{T}^{(k-1)}$  be multilinear maps that satisfy*

$$\begin{aligned} \left( \max_{2 \leq j \leq k-1} \|\mathbb{T}^{(j)}\|_{\text{op}}^{\frac{1}{j-1}} \right) \vee t^* &\leq t, \\ th &\leq \frac{1}{4}, \end{aligned}$$

*for some  $t \geq 0$ . Then it holds*

$$x - x_0 - \pi(x - x_0) - \sum_{j=2}^{k-1} \mathbb{T}^{(j)}(\pi(x - x_0)^{\otimes j}) = \sum_{j=1}^k \mathbb{T}_{y_0}^{(j),'}(\pi_{y_0}^*(y - y_0)^{\otimes j}) + R_{y_0}^{(k)}(x - x_0),$$

*where  $\mathbb{T}_{y_0}^{(j),'}$  are  $j$ -linear maps, and  $\|R_{y_0}^{(k)}(x - x_0)\| \leq C(\sigma + h^k(t_*^{k-1} + t^k h))$ , where  $C$  depends on  $d, k, \text{rch}_{\min}, L_2, \dots, L_k$ . Moreover, we have*

$$\begin{aligned} \mathbb{T}_{y_0}^{(1),'} &= (\pi_{y_0}^* - \pi), \\ \mathbb{T}_{y_0}^{(2),'} &= (\pi_{y_0}^* - \pi) \circ \mathbb{T}_{y_0}^{(2),*} + (\mathbb{T}_{y_0}^{(2),*} \circ \pi_{y_0}^* - \mathbb{T}^{(2)} \circ \pi), \end{aligned}$$

*and, if  $\pi = \pi_{y_0}^*$  and  $\mathbb{T}^{(j)} = \mathbb{T}_{y_0}^{(j),*}$  for all  $j \in \{2, \dots, k-1\}$ , then  $\mathbb{T}_{y_0}^{(j),'} = 0$  for all  $j \in \{1, \dots, k\}$ .*



The proof of Lemma 3.A.3 is deferred to Section 3.A.4. To ensure that our local curvature estimators allow to approximate the maximal curvature of  $M$ , we have to ensure that the sample covers  $M$  well enough. That is the aim of the following Lemma.

**Lemma 3.A.4** ([AL19, Appendix, Lemma B.7 & Section 5.1.4]). *Let  $P \in \mathcal{P}_{\text{rch}_{\min}, \mathbf{L}}^k(f_{\min}, f_{\max})$ . Write  $\mathbb{X}_n$  for an i.i.d.  $n$ -sample drawn from  $P$ . Let  $h = \left( C_{d,k} \frac{f_{\max}^2 \log n}{f_{\min}^3 n} \right)^{1/d}$ , for  $C_{d,k}$  large enough. Then, for  $n$  large enough so that  $h \leq \text{rch}_{\min}/4$ , with probability at least  $1 - 2\left(\frac{1}{n}\right)^{2k/d}$ , it holds*

$$\begin{aligned} d_H(M, \mathbb{X}_n) &\leq h/4, \\ d_H(M, \hat{M}) &\leq C_{d,k, \text{rch}_{\min}, \mathbf{L}} (t^*)^{k-1} \left( \frac{f_{\max}^{2+\frac{d}{2k}} \log n}{f_{\min}^{3+\frac{d}{2k}} n} \right)^{k/d}, \end{aligned}$$

where  $\hat{M}$  denotes the union of local polynomial patches

$$\hat{M} := \bigcup_{i=1}^n \hat{\Psi}_i \left( B_{\hat{T}_i}(0, 7h/8) \right)$$

defined by (3.6) and (3.7), and  $t_* = \max_{y \in M, 2 \leq j \leq k} \|\mathbb{T}_y^{(j),*}\|_{\text{op}}^{\frac{1}{j-1}} \leq C_{k,d, \text{rch}_{\min}, \mathbf{L}}$  as in Lemma 3.A.3.

Equipped with these two lemmas, we are in position to prove Theorem 3.11.

*Proof of Theorem 3.11.* Based on Lemma 3.A.4, for  $h = \left( C_{d,k} \frac{f_{\max}^2 \log n}{f_{\min}^3 n} \right)^{1/d}$ , given  $i \in \{1, \dots, n\}$ , we denote by  $\hat{\Psi}_i$  the polynomial estimator around  $X_i$  defined by

$$\hat{\Psi}_i(v) := X_i + v + \sum_{j=2}^{k-1} \hat{\mathbb{T}}_i^{(j)}(v^{\otimes j}),$$

for all  $v \in \hat{T}_i$ . Setting

$$\hat{M} := \bigcup_{i=1}^n \hat{\Psi}_i \left( B_{\hat{T}_i}(0, 7h/8) \right),$$

we have that with probability larger than  $1 - 2\left(\frac{1}{n}\right)^{\frac{2k}{d}}$ ,

$$d_H(\hat{M}, M) \leq C_{d,k, \text{rch}_{\min}, \mathbf{L}} (t_*)^{k-1} \left( \frac{f_{\max}^{2+\frac{d}{2k}} \log n}{f_{\min}^{3+\frac{d}{2k}} n} \right)^{\frac{k}{d}} := \varepsilon_1, \quad (3.15)$$

for  $n$  large enough, according to Lemma 3.A.4. In what follows we settle on the probability event of Lemma 3.A.4. In particular, denoting by

$$\hat{t} = \max_{1 \leq i \leq n} \max_{2 \leq j \leq k-1} \|\hat{\mathbb{T}}_i^{(j)}\|_{\text{op}}^{\frac{1}{j-1}},$$

note that [AL19, Section 5.1.2] ensures that  $\hat{t} \vee t_* \leq t \leq 1/(4h)$ , for some fixed  $t$ , provided  $n$  is large enough.

We let  $i \in \{1, \dots, n\}$ ,  $v \in B_{\hat{T}_i}(0, h/4)$ , and intend to approximate  $\mathbb{I}_{\pi_M(\hat{\Psi}_i(v))}$ . To do so, we consider the following polynomial expansion centered at  $v$ : for  $u \in B_{\hat{T}_i}(0, h/4)$ ,

$$\hat{\Psi}_i(v+u) - \hat{\Psi}_i(v) = u + \sum_{j=2}^{k-1} j \hat{\mathbb{T}}_i^{(j)}(v^{\otimes j-1} \otimes u) + \sum_{j=2}^{k-1} \sum_{r=j}^{k-1} \binom{r}{j} \hat{\mathbb{T}}_i^{(r)}(v^{\otimes r-j} \otimes u^{\otimes j}). \quad (3.16)$$

First we deduce from (3.16) an estimate for the tangent space at  $\pi_M(X_i + v)$ , as well as a coordinate system. Namely, we let

$$\begin{aligned} \hat{J}_{i,v}: \hat{T}_i &\longrightarrow \hat{J}_{i,v}(\hat{T}_i) \\ u &\longmapsto u + \sum_{j=2}^{k-1} j \hat{\mathbb{T}}_i^{(j)}(v^{\otimes j-1} \otimes u). \end{aligned}$$

Note that since  $th \leq 1/4$ , we have

$$\begin{aligned} \|\hat{J}_{i,v}(u) - u\| &\leq \sum_{j=2}^{k-1} j \left(\frac{th}{4}\right)^{j-1} \|u\| \\ &\leq \left(\sum_{j=1}^{\infty} j \left(\frac{th}{4}\right)^{j-1} - 1\right) \|u\| \\ &\leq \left(\left(\frac{1}{1 - \frac{th}{4}}\right)^2 - 1\right) \|u\| \leq \frac{\|u\|}{2}, \end{aligned}$$

so that  $\hat{J}_{i,v}$  is full-rank. In what follows we write  $\hat{T}_{i,v} := \text{Im}(\hat{J}_{i,v})$  and  $\hat{\pi}_{i,v} := \pi_{\hat{T}_{i,v}}$ . We now may express (3.16) in terms of the coordinate system given by  $\hat{T}_{i,v}$ :

$$\hat{\Psi}_i(v+u) - \hat{\Psi}_i(v) = \hat{J}_{i,v}(u) + \sum_{j=2}^{k-1} \tilde{\mathbb{T}}_{i,v}^{(j)}(\hat{J}_{i,v}(u)^{\otimes j}), \quad (3.17)$$

where the symmetric tensor of order  $j$  centered at  $v$ ,  $\tilde{\mathbb{T}}_{i,v}^{(j)}$ , is defined by

$$\tilde{\mathbb{T}}_{i,v}^{(j)}(w^{\otimes j}) := \sum_{r=j}^{k-1} \binom{r}{j} \hat{\mathbb{T}}_i^{(r)}(v^{\otimes r-j} \otimes \hat{J}_{i,v}^{-1}(w)^{\otimes j}),$$

for  $w \in \hat{T}_{i,v}$ . As well, since  $th \leq \frac{1}{4}$ , we may write

$$\begin{aligned} \|\tilde{\mathbb{T}}_{i,v}^{(j)}\|_{\text{op}} &\leq \sum_{r=j}^{k-1} \binom{r}{j} (3/2)^j t^{r-1} \left(\frac{h}{4}\right)^{r-j} \\ &\leq \left(\sum_{r=j}^{\infty} \binom{r}{j} \left(\frac{th}{4}\right)^{r-j}\right) (3/2)^j t^{j-1} \\ &\leq \left(\frac{1}{1 - \frac{th}{4}}\right)^j (3/2)^j t^{j-1} \leq (3/2)^{2j} t^{j-1}, \end{aligned}$$

so that  $\max_{2 \leq j \leq k-1} \|\tilde{\mathbb{T}}_{i,v}^{(j)}\|_{\text{op}}^{\frac{1}{j-1}} \leq \tilde{t} \leq \left(\frac{3}{2}\right)^4 t$ .

In particular, the bilinear form  $\tilde{\mathbb{T}}_{i,v}^{(2)} : \hat{T}_{i,v} \times \hat{T}_{i,v} \rightarrow \mathbb{R}^D$  may be expressed by

$$\tilde{\mathbb{T}}_{i,v}^{(2)}(w^{\otimes 2}) := \sum_{j=2}^{k-1} \binom{j}{2} \hat{\mathbb{T}}_i^{(j)}(v^{\otimes j-2} \otimes \hat{J}_{i,v}^{-1}(w)^{\otimes 2})$$

for all  $w \in \hat{T}_{i,v}$ . Our second fundamental form estimator at  $\pi_M(\hat{\Psi}_i(v))$  is then defined by

$$\hat{\mathbb{T}}_{i,v}^{(2)} := \tilde{\mathbb{T}}_{i,v}^{(2)} \circ \hat{\pi}_{i,v} - \hat{\pi}_{i,v} \circ \tilde{\mathbb{T}}_{i,v}^{(2)} \circ \hat{\pi}_{i,v},$$

where with a slight abuse of notation,  $\mathbb{T} \circ \pi(u) := \mathbb{T}(\pi(u)^{\otimes 2})$ . Note that composition with  $\hat{\pi}_{i,v}$  is performed to ensure that  $\hat{\mathbb{T}}_{i,v}^{(2)}$  ranges into  $\hat{T}_{i,v}^\perp$ .

Our final max-curvature estimator can now be defined as

$$\hat{R}_\ell^{-1} := \max_{1 \leq i \leq n} \max_{v \in \mathcal{B}_{\hat{T}_i}(h/4)} \left\| \hat{\mathbb{T}}_{i,v}^{(2)} \right\|_{\text{op}}.$$

First, we intend to show that, for a given  $v \in \mathcal{B}_{\hat{T}_i}(h/4)$ ,  $\hat{\mathbb{T}}_{i,v}^{(2)}$  is close to  $\mathbb{I}_{y_0}$ , for some  $y_0 \in M$ . To do so, we let  $u \in \mathcal{B}_{\hat{T}_i}(0, h/4)$ ,  $x := \hat{\Psi}_i(v+u)$ ,  $x_0 := \hat{\Psi}_i(v)$ , and  $\tilde{P}_{i,v}^{(r:k-1)} := \sum_{j=r}^{k-1} \tilde{\mathbb{T}}_{i,v}^{(j)}$ . Then, we have the decomposition

$$\begin{aligned} \hat{J}_{i,v}(u) &= \hat{\pi}_{i,v}(x - x_0) - \sum_{j=2}^{k-1} \hat{\pi}_{i,v} \circ \tilde{\mathbb{T}}_{i,v}^{(j)}(\hat{J}_{i,v}(u)^{\otimes j}) \\ &= \hat{\pi}_{i,v}(x - x_0) - \sum_{j=2}^{k-1} \hat{\pi}_{i,v} \circ \tilde{\mathbb{T}}_{i,v}^{(j)} \left[ \left( \hat{\pi}_{i,v}(x - x_0) - \hat{\pi}_{i,v} \circ \tilde{P}_{i,v}^{(2:k-1)}(\hat{J}_{i,v}(u)) \right)^{\otimes j} \right] \\ &= \hat{\pi}_{i,v}(x - x_0) + \sum_{j=2}^k \mathbb{T}_{i,v}^{(j),''}(\hat{\pi}_{i,v}(x - x_0)^{\otimes j}) + R_{i,v}^{(k)}(x - x_0), \end{aligned}$$

with  $\mathbb{T}_{i,v}^{(2),''} = -\hat{\pi}_{i,v} \circ \tilde{\mathbb{T}}_{i,v}^{(2)}$ , higher order tensors satisfying  $\left\| \mathbb{T}_{i,v}^{(j),''} \right\|_{\text{op}} \leq C_k \tilde{t}^{j-1} \leq C_k t^{j-1}$ , and remainder term  $\left\| R_{i,v}^{(k)} \right\| \leq C_k t^k h^{k+1}$ . Plugging the above inequalities into (3.17) yields

$$x - x_0 = \hat{\pi}_{i,v}(x - x_0) + \mathbb{T}_{i,v}^{(2)}(\hat{\pi}_{i,v}(x - x_0)^{\otimes 2}) + \sum_{j=3}^k \mathbb{T}_{i,v}^{(j)}(\hat{\pi}_{i,v}(x - x_0)^{\otimes j}) + R_{i,v}^{(k),'}(x - x_0), \quad (3.18)$$

with  $\mathbb{T}_{i,v}^{(2)} = \tilde{\mathbb{T}}_{i,v}^{(2)} - \hat{\pi}_{i,v} \circ \tilde{\mathbb{T}}_{i,v}^{(2)}$ ,  $\left\| \mathbb{T}_{i,v}^{(j)} \right\|_{\text{op}} \leq C_k t^{j-1}$ , and  $\left\| R_{i,v}^{(k),'}(x - x_0) \right\| \leq C_k t^k h^{k+1}$ .

Then, according to Lemma 3.A.4, there exists  $y_0 \in \mathcal{B}(X_i, \frac{8}{7 \times 4} h) \cap M$  such that  $\|y_0 - x_0\| \leq \varepsilon_1$ , where  $\varepsilon_1$  is defined by (3.15). We further have

$$\begin{aligned} \|v - \hat{\pi}_i(y_0 - X_i)\| &\leq \varepsilon_1 + \|\hat{\Psi}_i(v) - (X_i + v)\| \\ &\leq \varepsilon_1 + \left\| \sum_{j=2}^{k-1} \hat{\mathbb{T}}_i^{(j)}(v^{\otimes j}) \right\| \\ &\leq \varepsilon_1 + h/16 \\ &\leq h/8, \end{aligned}$$

since  $\hat{t}h \leq \frac{1}{4}$ , provided that  $\varepsilon_1 \leq h/16$  (satisfied for  $n$  large enough). Next, if  $z \in B(y_0, \frac{h}{8}) \cap M$ , we have

$$\begin{aligned} \|\hat{\pi}_i(z - X_i) - v\| &\leq \|\hat{\pi}_i(z - y_0)\| + \|v - \hat{\pi}_i(y_0 - X_i)\| \\ &\leq h/4, \end{aligned}$$

so that, writing  $x_z := \hat{\Psi}_i(\hat{\pi}_i(z - X_i))$ , it holds  $\|z - x_z\| \leq \varepsilon_1$  and (3.18) applies. Next, provided  $C_k t h < 1/4$  and  $C_k t \geq t_*$  (satisfied whenever  $n$  is large enough), Lemma 3.A.3 yields that

$$\begin{aligned} x_z - x_0 - \left( \hat{\pi}_{i,v}(x_z - x_0) + \mathbb{T}_{i,v}^{(2)}(\hat{\pi}_{i,v}(x_z - x_0)^{\otimes 2}) + \sum_{j=3}^k \mathbb{T}_{i,v}^{(j)}(\hat{\pi}_{i,v}(x_z - x_0)^{\otimes j}) \right) \\ = \sum_{j=1}^k \mathbb{T}_{i,v}^{(j),'}(\pi_y^*(z - y_0)^{\otimes j}) + R_{y_0}^{(k)}(x_z - x_0), \end{aligned}$$

so that

$$\left\| \sum_{j=1}^k \mathbb{T}_{i,v}^{(j),'}(\pi_y^*(z - y_0)^{\otimes j}) \right\| = \left\| R_{i,v}^{(k),'}(x_z - x_0) - R_{y_0}^{(k)}(x_z - x_0) \right\| \leq C_{k,d,\text{rch}_{\min},\mathbf{L}} \varepsilon_1,$$

according to (3.18) and Lemma 3.A.3, since  $t^k h \leq C_{k,d,\text{rch}_{\min},\mathbf{L}}$ . Using the development (3.16) and the inclusion  $B_{T_{y_0}M}(0, h/16) \subset \pi_{y_0}^*(B(y_0, h/8) \cap M - y_0)$  from [AL19, Lemma 2] then entails

$$\left\| \sum_{j=1}^k \mathbb{T}_{i,v}^{(j),'}(\pi_{y_0}^*(w)^{\otimes j}) \right\| \leq C_{k,d,\text{rch}_{\min},\mathbf{L}} \varepsilon_1,$$

for all  $w \in B_{T_{y_0}M}(0, h/16)$ . Proceeding as in [AL19, Proof of Theorem 2], we get

$$\begin{aligned} \left\| \mathbb{T}_{i,v}^{(1),'} \right\|_{\text{op}} &\leq C_{k,d,\text{rch}_{\min},\mathbf{L}} \varepsilon_1 h^{-1}, \\ \text{and } \left\| \mathbb{T}_{i,v}^{(2),'} \right\|_{\text{op}} &\leq C_{k,d,\text{rch}_{\min},\mathbf{L}} \varepsilon_1 h^{-2}. \end{aligned}$$

In turn, following [AL19, Proof of Theorem 4] entails

$$\left\| \hat{\mathbb{T}}_{i,v}^{(2)} \circ \hat{\pi}_{i,v} - \mathbb{T}_{y_0}^{(2),*} \circ \pi_{y_0}^* \right\|_{\text{op}} \leq C_{k,d,\text{rch}_{\min},\mathbf{L}} \varepsilon_1 h^{-2}.$$

Since  $\mathbb{I}_{y_0} = \mathbb{T}_{y_0}^{(2),*}$  ([AL19, Lemma 2]), we deduce that

$$\max_{1 \leq i \leq n} \max_{v \in B_{\hat{T}_j}(0, h/4)} \left\| \hat{\mathbb{T}}_{i,v}^{(2)} \circ \hat{\pi}_{i,v} \right\|_{\text{op}} \leq \max_{y \in M} \left\| \mathbb{I}_y \right\|_{\text{op}} + C_{k,d,\text{rch}_{\min},\mathbf{L}} \varepsilon_1 h^{-2}. \quad (3.19)$$

Conversely, since  $X_1, \dots, X_n$  is a  $(h/4)$ -covering of  $M$  onto the probability event described in Lemma 3.A.4, we deduce that for all  $y \in M$ , there exists  $i_0 \in \{1, \dots, n\}$  such that  $\|X_{i_0} - y\| \leq h/4$ . In particular, we have

$$v := \hat{\pi}_{i_0,v}(y - X_{i_0}) \in B_{\hat{T}_{i_0}}(0, h/4).$$

Proceeding as above similarly leads to

$$\left\| \hat{\mathbb{T}}_{i_0, v}^{(2)} \circ \hat{\pi}_{i_0, v} - \mathbb{I}_y \circ \pi_y^* \right\|_{\text{op}} \leq C_{k, d, \text{rch}_{\min}, \mathbf{L}} \varepsilon_1 h^{-2},$$

so that

$$\max_{y \in M} \left\| \mathbb{I}_y \right\|_{\text{op}} \leq \max_{1 \leq i \leq n} \max_{v \in \mathbf{B}_{\hat{T}_i}(0, h/4)} \left\| \hat{\mathbb{T}}_{i, v}^{(2)} \circ \hat{\pi}_{i, v} \right\|_{\text{op}} + C_{k, d, \text{rch}_{\min}, \mathbf{L}} \varepsilon_1 h^{-2}. \quad (3.20)$$

Combining (3.19) and (3.20) yields that for  $n$  large enough,

$$\left| \hat{R}_\ell - R_\ell(M) \right| \leq R_\ell(M)^2 C_{k, d, \text{rch}_{\min}, \mathbf{L}} \varepsilon_1 h^{-2},$$

which concludes the proof.  $\square$

### 3.A.4 Proof of Lemma 3.A.3

*Proof of Lemma 3.A.3.* We follow the proof of [AL19, Lemma 3]. Without loss of generality we take  $y_0 = 0$ , so that  $\|y\| \leq 3h/2$ . Let  $z' = z - z_0$ , so that  $\|z'\| \leq h/2$ . We write

$$\begin{aligned} x - x_0 - \pi(x - x_0) &= \sum_{j=2}^k \mathbb{T}^{(j)}(\pi(x - x_0)^{\otimes j}) \\ &= y + z' - \pi(y + z') - \sum_{j=2}^k \mathbb{T}^{(j)}((\pi(y) + \pi(z'))^{\otimes j}) \\ &= y + z' - \pi(y + z') - \sum_{j=2}^k \left[ \mathbb{T}^{(j)}(\pi(y)^{\otimes j}) + \sum_{r=0}^{j-1} \binom{j}{r} \mathbb{T}^{(j)}(\pi(y)^{\otimes r} \otimes \pi(z')^{\otimes j-r}) \right]. \end{aligned}$$

Since, for any  $j \geq 2$  and  $r \in \{0, \dots, j-1\}$ ,

$$\begin{aligned} \left\| \mathbb{T}^{(j)}(\pi(y)^{\otimes r} \otimes \pi(z')^{\otimes j-r}) \right\| &\leq t^{j-1} (3h/2)^r (2\sigma)^{j-r} \\ &\leq C_k \sigma t^{j-1} h^{j-1} \leq C_k \sigma, \end{aligned}$$

we may write

$$\begin{aligned} x - x_0 - \pi(x - x_0) &= \sum_{j=2}^k \mathbb{T}^{(j)}(\pi(x - x_0)^{\otimes j}) \\ &= y - \pi(y) - \sum_{j=2}^k \mathbb{T}^{(j)}(\pi(y)^{\otimes j}) + R^{(k), '}(x - x_0), \end{aligned} \quad (3.21)$$

where  $\|R^{(k), '}(x - x_0)\| \leq C_k \sigma$ . Next, (3.14) entails

$$\begin{aligned} y &= \pi_{y_0}^*(y) + \mathbb{T}_{y_0}^{(2), *}( \pi_{y_0}^*(y)^{\otimes 2} ) + \dots + \mathbb{T}_{y_0}^{(k-1), *}( \pi_{y_0}^*(y)^{\otimes k-1} ) \\ &\quad + R_{y_0}^{(k), ''}(y), \end{aligned}$$

with  $\|R_{y_0}^{(k),''}(y)\| \leq C_{k,d,\text{rch}_{\min},\mathbf{L}} t_*^{k-1} h^k$ . Denoting by

$$P_{y_0}^{*,(1:k-1)}(\pi_{y_0}^*(y)) := \pi_{y_0}^*(y) + \sum_{r=2}^{k-1} \mathbb{T}_{y_0}^{(r),*}(\pi_{y_0}^*(y)^{\otimes r}),$$

we deduce that

$$\begin{aligned} y - \pi(y) - \sum_{j=2}^k \mathbb{T}^{(j)}(\pi(y)^{\otimes j}) &= \\ &P_{y_0}^{*,(1:k-1)}(\pi_{y_0}^*(y)) + R_{y_0}^{(k),''}(y) - \pi\left(P_{y_0}^{*,(1:k-1)}(\pi_{y_0}^*(y)) + R_{y_0}^{(k),''}(y)\right) \\ &\quad - \sum_{j=2}^k \mathbb{T}^{(j)}\left(\pi\left[\left(P_{y_0}^{*,(1:k-1)}(\pi_{y_0}^*(y)) + R_{y_0}^{(k),''}(y)\right)\right]^{\otimes j}\right). \end{aligned}$$

Note that

$$\left\|\pi\left(R_{y_0}^{(k),''}(y)\right)\right\| \leq \left\|R_{y_0}^{(k),''}(y)\right\| \leq C_{k,d,\text{rch}_{\min},\mathbf{L}} t_*^{k-1} h^k.$$

Next, since  $\|y\| \leq 3h/2$ , it holds

$$\begin{aligned} \left\|P_{y_0}^{*,(1:k-1)}(\pi_{y_0}^*(y))\right\| &\leq \sum_{r=1}^{k-1} t_*^{r-1} \left(\frac{3h}{2}\right)^r \\ &\leq \frac{3h}{2} \frac{1}{1 - \frac{3t_*h}{2}} \leq 3h, \end{aligned}$$

so that, for all  $j \in \{2, \dots, k\}$ ,

$$\begin{aligned} &\left\|\mathbb{T}^{(j)}\left(\pi\left[\left(P_{y_0}^{*,(1:k-1)}(\pi_{y_0}^*(y)) + R_{y_0}^{(k),''}(y)\right)\right]^{\otimes j}\right) - \mathbb{T}^{(j)}\left(\pi\left[\left(P_{y_0}^{*,(1:k-1)}(\pi_{y_0}^*(y))\right)\right]^{\otimes j}\right)\right\| \\ &\leq t^{j-1} \sum_{r=1}^j \binom{j}{r} \left\|R_{y_0}^{(k),''}(y)\right\|^r (3h)^{j-r} \\ &\leq C_{k,d,\text{rch}_{\min},\mathbf{L}} t^{j-1} h^j \max_{1 \leq r \leq j} t_*^{(k-1)r} h^{(k-1)r} \\ &\leq C_{k,d,\text{rch}_{\min},\mathbf{L}} t^k h^{k+1}. \end{aligned}$$

Thus, we may write

$$\begin{aligned} y - \pi(y) - \sum_{j=2}^k \mathbb{T}^{(j)}(\pi(y)^{\otimes j}) &= P_{y_0}^{*,(1:k-1)}(\pi_{y_0}^*(y)) - \pi\left(P_{y_0}^{*,(1:k-1)}(\pi_{y_0}^*(y))\right) \\ &\quad - \sum_{j=2}^k \mathbb{T}^{(j)}\left(\pi\left[\left(P_{y_0}^{*,(1:k-1)}(\pi_{y_0}^*(y))\right)\right]^{\otimes j}\right) + R_{y_0}^{(k),'''}(y), \end{aligned}$$

where  $\left\|R_{y_0}^{(k),'''}(y)\right\| \leq C_{k,d,\text{rch}_{\min},\mathbf{L}} h^k (t_*^{k-1} + t^k h)$ . At last, for  $j \in \{2, \dots, k\}$ , and  $r_1, \dots, r_j \in \{1, \dots, k-1\}$  such that  $\sum_{s=1}^j r_s \geq k+1$ , we have

$$\begin{aligned} &\left\|\mathbb{T}^{(j)}\left(\bigotimes_{s=1}^j \pi\left(\mathbb{T}_{y_0}^{(r_s),*}(\pi_{y_0}^*(y)^{\otimes r_s})\right)\right)\right\| \leq t^{j-1} \prod_{s=1}^j t_*^{r_s-1} h^{r_s} \\ &\leq (th)^{\left(\sum_{s=1}^j r_s\right)-1} h \leq t^k h^{k+1}, \end{aligned}$$

where  $\mathbb{T}_{y_0}^{(1),*} = \pi_{y_0}^*$ , with a slight abuse of notation. Hence, it holds

$$\begin{aligned} y - \pi(y) - \sum_{j=2}^k \mathbb{T}^{(j)}(\pi(y)^{\otimes j}) &= (\pi_{y_0}^* - \pi \circ \pi_{y_0}^*)(y) + \mathbb{T}_{y_0}^{(2),*}(\pi_{y_0}^*(y)^{\otimes 2}) - \pi\left(\mathbb{T}_{y_0}^{(2),*}(\pi_{y_0}^*(y)^{\otimes 2})\right) \\ &\quad - \mathbb{T}^{(2)}\left((\pi \circ \pi_{y_0}^*(y))^{\otimes 2}\right) + \sum_{j=3}^k \mathbb{T}_{y_0}^{(j),'}(\pi_{y_0}^*(y)^{\otimes j}) + R_{y_0}^{(k),''''}(y), \end{aligned}$$

where  $\|R_{y_0}^{(k),''''}(y)\| \leq C_{k,d,\text{rch}_{\min},\mathbf{L}} h^k (t_*^{k-1} + t^k h)$ . Plugging the above equation into (3.21) gives the result.  $\square$

## 3.B Proofs of Section 3.4

### 3.B.1 Comparing Reach, Weak Feature Size and Spherical Distortion Radius

Let us prove Proposition 3.14.

*Proof of Proposition 3.14.* The monotonicity follows trivially from the definition, and since by [BLW19, Theorem 1],  $\text{sdr}_0(K, d_K) = \text{rch}(K, d_K)$ , there holds immediately that  $\text{sdr}_\delta(K, d_K) \geq \text{rch}(K)$  for any  $\delta \geq 0$ . Now take  $\delta \leq \sqrt{2(D+1)/D} \text{wfs}(K)$ , and take  $z$  a critical point of  $K$ , so that  $z \in \text{conv } \Gamma$  where  $\Gamma := \{x \in K \mid \|x - z\| = d(z, K)\}$ . Using Jung's theorem [Fed69, Theorem 2.10.41], there holds

$$\text{diam}(\Gamma) \geq \sqrt{\frac{2(D+1)}{D}} \text{Rad}(\Gamma) = \sqrt{\frac{2(D+1)}{D}} d(z, K) \geq \sqrt{\frac{2(D+1)}{D}} \text{wfs}(K) \geq \delta$$

so that there exists two points  $x, y \in \Gamma$  such that  $\|x - y\| \geq \delta$ . Furthermore, since the interior of  $B(z, \text{wfs}(K))$  contains no point of  $K$ , there holds

$$d_K(x, y) \geq d_{\mathcal{S}(\text{wfs}(K))}(x, y) > d_{\mathcal{S}(r)}(x, y), \text{ for all } r > \text{wfs}(K),$$

so that indeed  $\text{sdr}_\delta(K, d_K) \leq \text{wfs}(K)$ .  $\square$

### 3.B.2 Stability Properties of the Spherical Distortion Radius

We now move to the proofs of the stability properties of the SDR. As a first step, we will need the following lemma on geodesic distances over spheres.

**Lemma 3.B.1.** *Let  $r, \varepsilon > 0$  and take  $x, y, a, b \in K$  such that  $\|x - y\| < 2r$  and*

$$\|a - b\| \leq \left(1 + \frac{A\varepsilon}{r}\right) \|x - y\|$$

*for some  $A > 0$ . For all  $\lambda > 0$ , define*

$$\zeta_\lambda = \max \left\{ \frac{192r^3}{\|a - b\|^3} (\lambda + A\pi), 4A \right\}.$$

Then, for all  $\zeta \geq \zeta_\lambda$  such that  $\zeta\varepsilon \leq r$ , there holds

$$d_{S(r+\zeta\varepsilon)}(a, b) \leq d_{S(r)}(x, y) - \lambda\varepsilon.$$

*Proof of Lemma 3.B.1.* Notice that, denoting by  $\rho = \|x - y\|$ ,

$$d_{S(r)}(x, y) = 2r \arcsin\left(\frac{\rho}{2r}\right) = \rho \times \varphi(2r/\rho) \text{ with } \varphi(u) := u \arcsin(1/u).$$

The map  $\varphi$  is decreasing on  $[1, \infty)$  and, using the development of

$$\arcsin(u) = \sum_{n=0}^{\infty} (2n)! u^{2n+1} / (2^{2n} n!^2 (2n+1)),$$

we find that

$$\varphi'(u) = - \sum_{n=1}^{\infty} \frac{(2n)! \times 2n}{2^{2n} n!^2 (2n+1)} \frac{1}{u^{2n+1}} \leq -\frac{1}{3u^3}.$$

Notice furthermore that, by assumption

$$\begin{aligned} \frac{2(r+\zeta\varepsilon)}{\|a-b\|} &\geq \frac{2(r+\zeta\varepsilon)}{(1+A\varepsilon/r)\|x-y\|} = \frac{1+\zeta\varepsilon/r}{1+A\varepsilon/r} \frac{2r}{\|x-y\|} \\ &\geq \left(1 + \frac{\zeta\varepsilon}{2r}\right) \frac{2r}{\|x-y\|} \end{aligned}$$

where we used that  $A \leq \zeta/4$ ,  $\zeta\varepsilon \leq r$ , and that  $(1+u)/(1+u/4) \geq 1+u/2$  for  $|u| \leq 2$ . Now, as  $\varphi \leq \pi/2$  and that  $|\varphi'|$  is decreasing, we can write

$$\begin{aligned} d_{S(r+\zeta\varepsilon)}(a, b) &\leq \frac{A\varepsilon}{r} \|x-y\| \varphi(2(r+\zeta\varepsilon)/\|a-b\|) + \|x-y\| \varphi(2(r+\zeta\varepsilon)/\|a-b\|) \\ &\leq A\pi\varepsilon + d_{S(r)}(x, y) \\ &\quad - \|x-y\| \times |\varphi'| \left( \frac{2(r+\zeta\varepsilon)}{\|a-b\|} \right) \times \left( \frac{2(r+\zeta\varepsilon)}{\|a-b\|} - \frac{2r}{\|x-y\|} \right) \\ &\leq d_{S(r)}(x, y) + A\pi\varepsilon - \frac{\|a-b\|^3}{3(2(r+\zeta\varepsilon))^3} \zeta\varepsilon \\ &\leq d_{S(r)}(x, y) + \left( A\pi - \frac{\|a-b\|^3}{192r^3} \zeta \right) \varepsilon, \end{aligned}$$

and using  $\zeta \geq \zeta_\lambda$  ends the proof.  $\square$

We are now in position to prove Proposition 3.18 and Theorem 3.22.

*Proof of Proposition 3.18.* If  $r_1 = \infty$  there is nothing to show. Otherwise, notice that because  $r_1 \geq \delta_0/2$  by definition, there holds that

$$\xi(R) \geq \max \left\{ \frac{192R^3}{\delta_0^3} \left( 1 + \pi \frac{2R}{\delta_0} \right), \frac{8R}{\delta_0} \right\} \geq 1$$

for all  $R > r_1$ . Now, since  $\xi(r_1)\Upsilon < r_1$ , one can find  $R > r_1$  such that  $\xi(R)\Upsilon < R$ . By definition of  $r_1$ , there exist  $x, y \in K'$  such that  $\delta + 2\varepsilon \leq \|x - y\| < 2R$  and  $d_{S(R)}(x, y) <$



$d'(x, y)$ . Now, let  $a, b \in K$  be two closest points (in Euclidean distance) from  $x$  and  $y$  such that  $d(a, b) = d(\text{pr}_K(\{x\}), \text{pr}_K(\{y\}))$ . Then

$$\delta \leq \|a - b\| \leq \|x - y\| + 2\varepsilon < 2R + 2\Upsilon \leq 2(R + \xi(R)\Upsilon)$$

and

$$\|a - b\| \leq \|x - y\| + 2\varepsilon \leq \left(1 + \frac{2\Upsilon R}{\delta_0 R}\right) \|x - y\|.$$

We now can apply Lemma 3.B.1 with  $A = 2R/\delta_0$  and  $\lambda = 1$  to find that

$$\begin{aligned} d(a, b) &\geq \frac{1}{1 + \nu} d'(x, y) \\ &> \frac{1}{1 + \nu} d_{\mathcal{S}(R)}(x, y) \\ &\geq \frac{1}{1 + \nu} (d_{\mathcal{S}(R + \xi(R)\Upsilon)}(a, b) + \Upsilon) \\ &\geq \frac{1 + \Upsilon/\delta}{1 + \nu} d_{\mathcal{S}(R + \xi(R)\Upsilon)}(a, b), \end{aligned}$$

where the last inequality uses that  $d_{\mathcal{S}(R + \xi(R)\Upsilon)}(a, b) \geq \|a - b\| \geq \delta$ . At the end of the day, since  $\Upsilon \geq \delta\nu$ , we have  $d(a, b) > d_{\mathcal{S}(R + \xi(R)\Upsilon)}(a, b)$ , so that  $\text{sdr}_\delta(K, \delta) < R + \xi(R)\Upsilon$ . Taking  $R$  to  $r_1$  yields the result.  $\square$

*Proof of Theorem 3.22.* We take  $\varepsilon > 0$  such that

$$\varepsilon < C_0\varepsilon_0, \quad \varepsilon < (\delta_1 - \delta_0)/2, \quad \text{and} \quad \varepsilon < r_0/L_0,$$

and take  $\delta \in [\delta_0, \delta_1 - \varepsilon)$ . We write  $r_\delta := \text{sdr}_\delta(K, d)$  and  $r_{\delta+\varepsilon} := \text{sdr}_{\delta+\varepsilon}(K, d)$  for short. Recall that  $r_\delta \leq r_{\delta+\varepsilon}$ . Now take  $r \leq r_{\delta+\varepsilon} - L_0\varepsilon$ , and two points  $x, y \in K$  such that  $\delta \leq \|x - y\| < 2r$  (if there are none, then  $r \leq r_\delta$  automatically). If  $\|x - y\| \geq \delta + \varepsilon$ , then  $d(x, y) \leq d_{\mathcal{S}(r)}(x, y)$  because  $r \leq r_{\delta+\varepsilon}$ . If now  $\|x - y\| < \delta + \varepsilon$ , since  $\|x - y\| \leq \Delta_0$ , we can use Assumption 3.20 and find a point  $a \in K$  such that  $\|a - y\| \leq \varepsilon/C_0$  and  $\|x - a\| \geq \|x - y\| + \varepsilon \geq \delta + \varepsilon$ . Now, since  $r + L_0\varepsilon \leq r_{\delta+\varepsilon}$ , it holds  $d(x, a) \leq d_{\mathcal{S}(r + L_0\varepsilon)}(x, a)$ . Furthermore, notice that

$$\|x - a\| \leq \|x - y\| + \frac{1}{C_0}\varepsilon \leq \left(1 + \frac{r_1\varepsilon}{C_0\delta_0 r}\right) \|x - y\|.$$

Using Assumption 3.21 and Lemma 3.B.1 with  $A = r_1/(C_0\delta_0)$  and  $\lambda = C_1/C_0$ , we find

$$d(x, y) \leq d(x, a) + d(a, y) \leq d_{\mathcal{S}(r + L_0\varepsilon)}(x, a) + \frac{C_1}{C_0}\varepsilon \leq d_{\mathcal{S}(r)}(x, y),$$

so that in the end  $r \leq r_\delta$ . Taking  $r$  to  $r_{\delta+\varepsilon} - L_0\varepsilon$  yields that  $r_{\delta+\varepsilon} \leq r_\delta + L_0\varepsilon$ , ending the proof.  $\square$

Finally, Corollary 3.19 follows as a direct corollary of Proposition 3.18.

*Proof of Corollary 3.19.* Since  $\xi_0\varepsilon \leq 2\text{sdr}_{\delta_1}(K, d)$ , the radius  $\text{sdr}_{\delta_1}(K, d)$  is in particular finite so that, according to Proposition 3.18,  $\text{sdr}_{\delta}(K', d') \leq 2\text{sdr}_{\delta_1}(K, d)$  and, consequently,  $\xi_1\Upsilon \leq \text{sdr}_{\delta}(K', d')$  and  $\xi_2\Upsilon \leq \text{sdr}_{\delta+2\varepsilon}(K, d)$ , where  $\xi_1 = \xi(\text{sdr}_{\delta}(K', d'))$  and  $\xi_2 = \xi(\text{sdr}_{\delta+2\varepsilon}(K, d))$ . Applying Proposition 3.18 twice – which is possible, since  $\Upsilon \geq ((\delta - 2\varepsilon)\nu) \vee \varepsilon$  –, we thus find

$$\text{sdr}_{\delta-2\varepsilon}(K, d) - \xi_1\Upsilon \leq \text{sdr}_{\delta}(K', d') \leq \text{sdr}_{\delta+2\varepsilon}(K, d) + \xi_2\Upsilon,$$

and we conclude by noticing that both  $\xi_1$  and  $\xi_2$  are less than  $\xi_0$ .  $\square$

*Proof of Theorem 3.23.* Using Corollary 3.19 and Theorem 3.22, one find that

$$\begin{aligned} \text{sdr}_{\delta}(K', d') &\leq \text{sdr}_{\delta+2\varepsilon}(K, d) + \xi_0\Upsilon \\ &\leq \text{sdr}_{\delta}(K, d) + 2L_0\varepsilon + \xi_0\Upsilon \\ &\leq \text{sdr}_{\delta}(K, d) + \zeta_0\Upsilon, \end{aligned}$$

and likewise for the lower bound.  $\square$

### 3.C Proofs of Section 3.5

#### 3.C.1 Minimax Lower Bound for Metric Learning

We now turn towards the proof of Theorem 3.25. It relies on an adaptation of the classical Le Cam's argument [Yu97] to the asymmetric loss  $\ell_{\infty}$ .

**Lemma 3.C.1.** *Let  $x, y \in \mathbb{R}^D$  and let  $M_0$  and  $M_1$  be two submanifolds of  $\mathbb{R}^D$  such that  $x, y \in M_0 \cap M_1$  and the uniform distribution  $P_0$  (resp.  $P_1$ ) on  $M_0$  (resp.  $M_1$ ) is in  $\mathcal{P}_{\text{rch}_{\min}, \mathbf{L}}^k(f_{\min}, f_{\max})$ . Then if  $d_{M_0}(x, y) \leq d_{M_1}(x, y)$ ,*

$$\inf_{\widehat{d}} \sup_{P \in \mathcal{P}^k} \mathbb{E}_{P^{\otimes n}}[\ell_{\infty}(\widehat{d}|d_M)] \geq \frac{1}{2} \times \left| 1 - \frac{d_{M_0}(x, y)}{d_{M_1}(x, y)} \right| \times (1 - \text{TV}(P_0^{\otimes n}, P_1^{\otimes n})), \quad (3.22)$$

*Proof of Lemma 3.C.1.* For brevity, we write  $\mathcal{R}_n$  be the minimax risk appearing in the left-hand side of (3.22). First, we write

$$\begin{aligned} \mathcal{R}_n &\geq \inf_{\widehat{d}} \sup_{P \in \{P_0, P_1\}} \mathbb{E}_{P^{\otimes n}}[\ell_{\infty}(\widehat{d}|d_M)] \\ &\geq \inf_{\widehat{d}} \sup_{P \in \{P_0, P_1\}} \mathbb{E}_{P^{\otimes n}} \left[ \left| 1 - \frac{\widehat{d}(x, y)}{d_M(x, y)} \right| \right] \\ &\geq \frac{1}{2} \inf_{\widehat{d}} \left\{ \mathbb{E}_{P_0^{\otimes n}} \left[ \left| 1 - \frac{\widehat{d}(x, y)}{d_{M_0}(x, y)} \right| \right] + \mathbb{E}_{P_1^{\otimes n}} \left[ \left| 1 - \frac{\widehat{d}(x, y)}{d_{M_1}(x, y)} \right| \right] \right\} \\ &\geq \frac{1}{2} \inf_{\widehat{d}} \mathbb{E}_{P_0^{\otimes n}} \left[ \left( \left| 1 - \frac{\widehat{d}(x, y)}{d_{M_0}(x, y)} \right| + \left| 1 - \frac{\widehat{d}(x, y)}{d_{M_1}(x, y)} \right| \right) \times \left( 1 \wedge \frac{d_{P_1^{\otimes n}}}{d_{P_0^{\otimes n}}} \right) \right]. \end{aligned}$$

But now, using that  $d_{M_0}(x, y) \leq d_{M_1}(x, y)$ , a simple computation shows that the functional

$$\delta \mapsto \left| 1 - \frac{\delta}{d_{M_0}(x, y)} \right| + \left| 1 - \frac{\delta}{d_{M_1}(x, y)} \right|$$

is minimal for  $\delta = d_{M_0}(x, y)$  so that

$$\begin{aligned} \mathcal{R}_n &\geq \frac{1}{2} \mathbb{E}_{P_0^{\otimes n}} \left[ \left| 1 - \frac{d_{M_0}(x, y)}{d_{M_1}(x, y)} \right| \times \left( 1 \wedge \frac{dP_1^{\otimes n}}{dP_0^{\otimes n}} \right) \right] \\ &= \frac{1}{2} \times \left| 1 - \frac{d_{M_0}(x, y)}{d_{M_1}(x, y)} \right| \times (1 - \text{TV}(P_0^{\otimes n}, P_1^{\otimes n})), \end{aligned}$$

which ends the proof.  $\square$

*Proof of Theorem 3.25.* Without loss of generality, we set the analysis in  $\mathbb{R}^{d+1} \simeq \mathbb{R}^{d+1} \times \{0\}^{D-(d+1)} \subset \mathbb{R}^D$ .

**Submanifolds Construction** We let  $M_0 \subset \mathbb{R}^{d+1}$  be a submanifold of  $\mathcal{C}_{2\text{rch}_{\min}, \mathbf{L}/2}^k$  such that it contains the cylinder

$$\{(s, z) \in \mathbb{R}^2 \times \mathbb{R}^{d-1} \mid \|s\| = R \text{ and } \|z\| \leq 3R\}.$$

Such a manifold always exists as soon as  $R \geq 2\text{rch}_{\min}$  and  $L_j$  is large enough compared to  $1/R^{j-1}$ . For instance, one can design  $M_0$  as a hypersurface of revolution obtained based on patches the interpolating curves of Lemma 3.A.1.

In what follows, we denote any  $x \in \mathbb{R}^{d+1} = \mathbb{R}^d \times \mathbb{R}$  as  $x = (w, h) \in \mathbb{R}^d \times \mathbb{R}$ . With this notation, we define, for  $\varepsilon > 0$  and  $c > 0$  to be chosen later,

$$\Phi_\varepsilon(x) := x + c\varepsilon^k K(w/\varepsilon) e_{d+1} \quad \text{where } e_{d+1} = (0, \dots, 0, 1) \in \mathbb{R}^{d+1},$$

where  $K(w)$  equals  $\exp(-1/(1 - \|w\|^2)_+)$  for  $\|w\| < 1$  and 0 otherwise.

For  $\varepsilon \leq 1$  and  $c$  small enough,  $\Phi_\varepsilon$  is a diffeomorphism of  $\mathbb{R}^{d+1}$  with derivative bounded up to the order  $k$ . Using [AL19, Proposition A.4], we get that  $M_\varepsilon := \Phi_\varepsilon(M)$ , the image of  $M_0$  by  $\Phi_\varepsilon$ , belongs to  $\mathcal{C}_{\text{rch}_{\min}, \mathbf{L}}^k$  provided that  $c$  is small enough (depending on  $R$ ) and  $\varepsilon \leq cR$ .

Locally around the apex  $(0, R) \in \mathbb{R}^{d+1}$ ,  $M_0$  can be seen as the graph of  $\Psi_0(w) := \sqrt{R^2 - w_1^2}$ , defined on  $(-R, R) \times \mathbb{B}_{\mathbb{R}^{d-1}}(0, 3R)$ , while  $M_\varepsilon$  is the graph of

$$\Psi_\varepsilon(w) := \Psi_0(w) + c\varepsilon^k K(w/\varepsilon).$$

Finally, we let  $\bar{\Psi}_\varepsilon(w) := (w, \Psi_\varepsilon(w))$  and similarly define  $\bar{\Psi}_0$ . We refer to Figure 3.C.1 for a diagram of the situation.

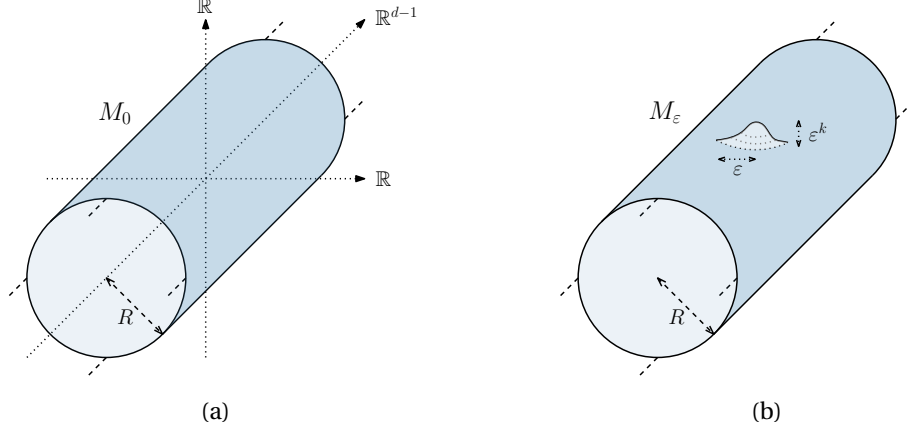


Figure 3.C.1 – (a) The cylindrical section of  $M_0$  used in the proof of Theorem 3.25, and (b) the perturbed submanifold  $M_\varepsilon$ .

**Shortest-Path Properties** In this section, we seek to derive a lower bound on the loss  $|1 - d_{M_0}(x, y)/d_{M_\varepsilon}(x, y)|$ , so as to apply Lemma 3.C.1. For this, we will consider well-chosen  $x, y \in M_0 \cap M_\varepsilon$  and derive a lower bound on  $d_{M_\varepsilon}(x, y) - d_{M_0}(x, y)$ .

We let  $\ell < R$ , and we pick  $x := \bar{\Psi}_0(-\ell e_1)$  and  $y := \bar{\Psi}_0(\ell e_1)$  where  $e_1 = (1, 0, \dots, 0) \in \mathbb{R}^d$ . By construction,  $x$  and  $y$  belong to  $M_0$ . Furthermore, provided that  $\ell \geq \varepsilon$ , there holds that  $x = \bar{\Psi}_\varepsilon(-\ell e_1)$  and  $y = \bar{\Psi}_\varepsilon(\ell e_1)$  so that  $x$  and  $y$  are also in  $M_\varepsilon$ . We let  $\gamma_\varepsilon : [-1, 1] \rightarrow M_\varepsilon$  be a shortest path in  $M_\varepsilon$  between  $x$  and  $y$ , parametrized at constant speed. We denote paths

$$w_\varepsilon := a_\varepsilon e_1 + b_\varepsilon := \text{pr}_{\mathbb{R}^d \times \{0\}}(\gamma_\varepsilon),$$

where  $b_\varepsilon \in \{0\} \times \mathbb{R}^{d-1}$ . We refer to Figure 3.C.2 for a diagram of the situation. Several observations are in order.

- Since  $w_\varepsilon(\pm 1) = \pm \ell e_1$ , we have  $a_\varepsilon(\pm 1) = \pm \ell$  and  $b_\varepsilon(\pm 1) = 0$ . Also, because  $\gamma_\varepsilon$  is a minimizing path,  $a_\varepsilon$  is nondecreasing, and  $\|b_\varepsilon\|_\infty \leq \varepsilon$  (see Figure 3.C.2).
- Because  $\gamma_\varepsilon$  has constant speed on  $[-1, 1]$ , there holds

$$\|\gamma'_\varepsilon(t)\| = \frac{1}{2} d_{M_\varepsilon}(x, y) \in [A_1 \ell, A_2 \ell], \quad \text{for all } t \in [-1, 1], \quad (3.23)$$

with  $A_1, A_2$  depending on  $R$  only, uniformly on small  $\varepsilon$ .

- $a_\varepsilon$  and  $b_\varepsilon$  are smooth and  $\gamma_\varepsilon = \bar{\Psi}_\varepsilon(w_\varepsilon)$ .
- Since  $M_\varepsilon$  is symmetric with respect to  $\{0\} \times \mathbb{R}^d$ , so should be the shortest path between  $x$  and  $y$ . This entails in particular that  $b_\varepsilon$  is even and that  $a_\varepsilon$  is odd;
- As  $\gamma_\varepsilon$  has constant speed and has a curvature bounded from above (as a shortest path in a bounded-curvature space), the ratio  $\|\gamma''_\varepsilon\|/\|\gamma'_\varepsilon\|^2$  is bounded in sup-norm

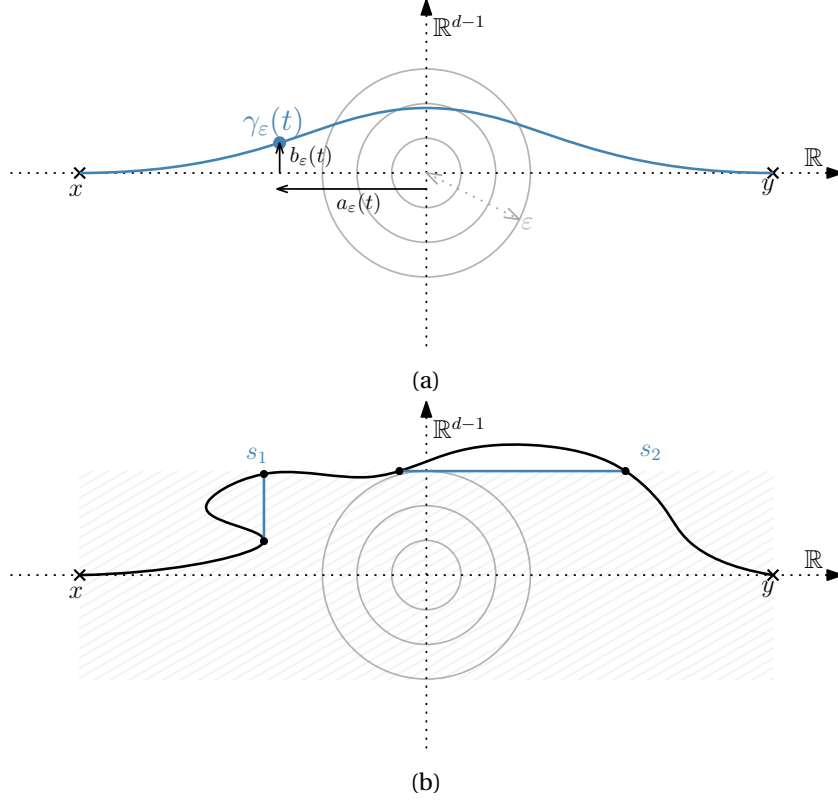


Figure 3.C.2 – (a) Top view of  $M_\varepsilon$  and of one of the shortest path between  $x$  and  $y$ , in blue. In light grey is represented the bump of size  $\varepsilon$ . (b) Same view of  $M_\varepsilon$  as (a), illustrating the fact that any shortest path must go from left to right (otherwise one can construct a shorter path, through  $s_1$  in the figure) and cannot go outside the shaded area (otherwise one can construct a shorter path, through  $s_2$  in the figure).

by a constant depending on  $R$  only. Therefore, there exists a constant  $B > 0$  depending on  $R$  only such that, uniformly on  $\varepsilon$  small enough,

$$\max \left\{ \|a'_\varepsilon\|_\infty/\ell, \|a''_\varepsilon\|_\infty/\ell^2, \|b'_\varepsilon\|_\infty/\ell, \|b''_\varepsilon\|_\infty/\ell^2 \right\} \leq B. \quad (3.24)$$

- By symmetry also,  $\gamma_\varepsilon$  crosses the hyperplane  $\{0\} \times \mathbb{R}^d$  orthogonally. As a consequence  $\langle \gamma'_\varepsilon(0), e_{d+1} \rangle = 0$ ,  $b'_\varepsilon(0) = 0$  and

$$a'_\varepsilon(0) = \|w'_\varepsilon(0)\| = \|\gamma'_\varepsilon(0)\| \in [A_1\ell, A_2\ell],$$

where  $A_1$  and  $A_2$  were introduced in (3.23).

- Finally, using (3.24), we deduce that there exists  $C > 0$  depending on  $R$  only such

that for all  $t \in [-1, 1]$ ,

$$\begin{cases} |a_\varepsilon(t) - a'_\varepsilon(0)t| & \leq C\ell^2 t^2, \\ |a'_\varepsilon(t) - a'_\varepsilon(0)| & \leq C\ell^2 t, \\ |a'_\varepsilon(t)a_\varepsilon(t) - a'_\varepsilon(0)^2 t| & \leq C\ell^3 t^2, \\ |b_\varepsilon(t) - b_\varepsilon(0)| & \leq C\ell t. \end{cases} \quad (3.25)$$

**Perturbative Expansion of the Geodesic Length** We let  $\gamma_0(t) := \bar{\Psi}_0(a_\varepsilon(t)e_1)$ . Although not constant-speed, monotonicity of  $a_\varepsilon$  implies that  $\gamma_0$  is the shortest path in  $M_0$  between  $x$  and  $y$ , and we get, using (3.24) and (3.25), that for some constant  $A_3$  depending on  $R$ ,

$$\frac{1}{2}A_1\ell \leq \|\gamma'_0(t)\| \leq A_3\ell \quad \text{if } \ell \leq \frac{A_1}{2C}, \quad (3.26)$$

which we will assume henceforth. Furthermore, the velocity of  $\gamma_\varepsilon$  writes

$$\begin{aligned} \gamma'_\varepsilon &= d\bar{\Psi}_\varepsilon(w_\varepsilon)[w'_\varepsilon] = d\bar{\Psi}_0(w_\varepsilon)[w'_\varepsilon] + c\varepsilon^{k-1}\langle \nabla K(w_\varepsilon/\varepsilon), w'_\varepsilon \rangle e_{d+1} \\ &= w'_\varepsilon + \langle \nabla \Psi_0(w_\varepsilon), w'_\varepsilon \rangle e_{d+1} + c\varepsilon^{k-1}\langle \nabla K(w_\varepsilon/\varepsilon), w'_\varepsilon \rangle e_{d+1} \\ &= a'_\varepsilon e_1 + b'_\varepsilon + \underbrace{(\langle \nabla \Psi_0(a_\varepsilon), a'_\varepsilon \rangle)}_{:=\nabla_0} + \underbrace{c\varepsilon^{k-1}\langle \nabla K(w_\varepsilon/\varepsilon), w'_\varepsilon \rangle}_{:=\nabla_1} e_{d+1}, \end{aligned}$$

where we used the fact that  $\Psi_0$  depends only on its first variable. We write the last term as  $(\nabla_0 + \nabla_1)e_{d+1}$ . Using that each three terms in the preceding development are orthogonal, we obtain

$$\|\gamma'_\varepsilon\|^2 = a'^2_\varepsilon + \|b'_\varepsilon\|^2 + (\nabla_0 + \nabla_1)^2 = \underbrace{a'^2_\varepsilon + \nabla_0^2}_{=:\|\gamma'_0\|^2} + \underbrace{\|b'_\varepsilon\|^2 + 2\nabla_0\nabla_1 + \nabla_1^2}_{:=Q_\varepsilon}, \quad (3.27)$$

and it only remains to study the last three terms, denoted by  $Q_\varepsilon$ . First, notice that using (3.24), one can find two constants  $D_0$  depending on  $R$  such that  $Q_\varepsilon \geq -D_0\varepsilon^2\ell^2$ . Together with (3.26), this yields that  $Q_\varepsilon/\|\gamma'_0\|^2 \geq -1$  for  $\varepsilon$  small enough (depending on  $R$ ). Likewise, we can show that  $Q_\varepsilon \leq D_1(\ell^2 + \ell^2\varepsilon^2 + \varepsilon^4)$ , for some constant  $D_1$  depending on  $R$ . This again yields

$$\frac{Q_\varepsilon}{\|\gamma'_0\|^2} \leq D_2 \quad \text{if } \varepsilon \leq D_3\ell, \quad (3.28)$$

for some constants  $D_2$  and  $D_3$  depending on  $R$  only. All in all, we have that  $Q_\varepsilon/\|\gamma'_0\| \in [-1, D_2]$ . Using that

$$\sqrt{1+z} \geq \begin{cases} 1+z & \text{if } z \in [-1, 0], \\ 1+D_4z & \text{if } z \in [0, D_2], \text{ with } D_4 = \frac{1}{D_2}(\sqrt{1+D_2}-1), \end{cases}$$

we can finally derive from (3.27) and (3.28) the following bound

$$\|\gamma'_\varepsilon\| = \|\gamma'_0\| \sqrt{1 + \frac{Q_\varepsilon}{\|\gamma'_0\|^2}} \geq \|\gamma'_0\| + \tau(Q_\varepsilon)Q_\varepsilon, \quad (3.29)$$

where

$$\tau(z) := \frac{2}{A_1 \ell} \mathbb{1}_{z < 0} + \frac{D_4}{A_3 \ell} \mathbb{1}_{z \geq 0},$$

and where we also used (3.26) to bound  $1/\|\gamma'_0\|$ . In particular, integrating (3.29) over  $[-1, 1]$  yields that

$$d_{M_\varepsilon}(x, y) \geq d_{M_0}(x, y) + \int_{-1}^1 \tau(Q_\varepsilon) Q_\varepsilon.$$

To obtain a more explicit bound, let us now study  $Q_\varepsilon$ . For this, first rewrite  $\nabla_0$  and  $\nabla_1$  more explicitly as

$$\nabla_0 = -\frac{a_\varepsilon a'_\varepsilon}{\sqrt{R^2 - a_\varepsilon^2}} \quad \text{and} \quad \nabla_1 = -2c\varepsilon^{k-2} \frac{K(w_\varepsilon/\varepsilon)}{(1 - \|w_\varepsilon/\varepsilon\|^2)^2} \langle w_\varepsilon, w'_\varepsilon \rangle.$$

Hence, noticing that  $\langle w_\varepsilon, w'_\varepsilon \rangle = a_\varepsilon a'_\varepsilon + \langle b_\varepsilon, b'_\varepsilon \rangle$ , one can write  $2\nabla_0 \nabla_1$  as  $P_0 + P_1$  with

$$\begin{cases} P_0 &= \varepsilon^{k-2} (a_\varepsilon a'_\varepsilon)^2 T_\varepsilon \\ P_1 &= \varepsilon^{k-2} T_\varepsilon a_\varepsilon a'_\varepsilon \langle b_\varepsilon, b'_\varepsilon \rangle \end{cases} \quad \text{with} \quad T_\varepsilon := \frac{4cK(w_\varepsilon/\varepsilon)}{\sqrt{R^2 - a_\varepsilon^2} (1 - \|w_\varepsilon/\varepsilon\|^2)^2}.$$

For  $\ell \leq A_1/4C$ , condition (3.25) together with  $a'_\varepsilon(0) \geq A_1 \ell/2$  imply that

$$\|w_\varepsilon(t)\| \geq |a_\varepsilon(t)| \geq \varepsilon \quad \text{for all } |t| \geq t_\varepsilon \quad \text{with } t_\varepsilon := \frac{4\varepsilon}{A_1 \ell},$$

so that in particular,  $T_\varepsilon(t) = 0$  for  $|t| \geq t_\varepsilon$ . Furthermore, notice that, provided that  $\ell$  is small before  $R$ ,  $T_\varepsilon$  is bounded by some constant  $E > 0$  depending on  $R$  only. Using again (3.25), we find that for  $\ell \leq A_1^2/8C$ , there holds

$$(a'_\varepsilon a_\varepsilon)^2(t) \geq \frac{1}{2} a'_\varepsilon(0)^4 t^2 - C^2 \ell^6 t^4 \geq \frac{1}{32} A_1^4 \ell^4 t^2, \quad (3.30)$$

$$\text{and} \quad |a'_\varepsilon a_\varepsilon|(t) \leq a'_\varepsilon(0)^2 |t| + C \ell^3 t^2 \leq 5A_1^2 \ell^2 |t| \quad \text{for all } t \in [-1, 1].$$

In particular, we find that

$$\begin{aligned} \int_{-1}^1 |P_1(t)| dt &\leq 5\varepsilon^{k-2} E A_1^2 \ell^2 \|b_\varepsilon\|_\infty \|b'_\varepsilon\|_\infty \int_{-t_\varepsilon}^{t_\varepsilon} |t| dt \\ &= 5\varepsilon^{k-2} E A_1^2 \ell^2 \|b_\varepsilon\|_\infty \|b'_\varepsilon\|_\infty t_\varepsilon^2 \\ &\leq 80BE \|b_\varepsilon\|_\infty \ell^2 \varepsilon^k, \end{aligned}$$

where we used (3.24) in the last inequality. On the other hand, letting  $t_0 \in (-1, 1)$  be a time at which  $\|b_\varepsilon(t_0)\| = \|b_\varepsilon\|_\infty$ , notice that

$$\begin{aligned} \int_{-1}^1 \|b'_\varepsilon\|^2 &= \int_{-1}^{t_0} \|b'_\varepsilon\|^2 + \int_{t_0}^1 \|b'_\varepsilon\|^2 \\ &\geq \frac{1}{1+t_0} \left\| \int_{-1}^{t_0} b'_\varepsilon \right\|^2 + \frac{1}{1-t_0} \left\| \int_{t_0}^1 b'_\varepsilon \right\|^2 \\ &= \left( \frac{1}{1+t_0} + \frac{1}{1-t_0} \right) \|b_\varepsilon(t_0)\|^2 \geq 2 \|b_\varepsilon\|_\infty^2. \end{aligned}$$

Integrating (3.27) and using that  $\nabla_1^2 \geq 0$  thus yields

$$\int_{-1}^1 \tau(Q_\varepsilon) Q_\varepsilon \geq 2\|b_\varepsilon\|_\infty (\tau_1 \|b_\varepsilon\|_\infty - 40\tau_2 BE\ell^2 \varepsilon^k) + \tau_1 \varepsilon^{k-2} \int_{-1}^1 (a_\varepsilon a'_\varepsilon)^2 T_\varepsilon. \quad (3.31)$$

where  $\tau_1$  is the smallest value of  $\tau$ , and  $\tau_2$  its greatest value. Now we distinguish on the value of  $\|b_\varepsilon(0)\|$ :

- If  $\|b_\varepsilon(0)\| \geq \varepsilon/2$ , then  $\|b_\varepsilon\|_\infty \geq \varepsilon/2$  and for  $\varepsilon$  small enough, we get, noticing that the last term in (3.31) is non-negative,

$$\int_{-1}^1 \tau(Q_\varepsilon) Q_\varepsilon \geq c_R \varepsilon (\varepsilon/2 - \varepsilon/4)/\ell \geq c_R \varepsilon^2/\ell.$$

- Otherwise, if  $\|b_\varepsilon(0)\| \leq \varepsilon/2$ , then, using (3.25), we find that

$$\begin{cases} \|b_\varepsilon(t)\| & \leq 3\varepsilon/4, \\ |a_\varepsilon(t)| & \leq \varepsilon/2, \end{cases} \quad \text{for all } |t| \leq t_\varepsilon^* \quad \text{with } t_\varepsilon^* := \min \left\{ \frac{\varepsilon}{4C\ell}, \frac{\varepsilon}{8A_2\ell}, \frac{2A_2}{C\ell} \right\}.$$

For  $\varepsilon$  small before  $R$ ,  $t_\varepsilon^*$  is of the form  $t_\varepsilon^* = G\varepsilon/\ell$  with  $G$  depending on  $R$  only. Furthermore, notice that for  $|t| \leq t_\varepsilon^*$ , there holds  $\|w_\varepsilon(t)\|^2 = |a_\varepsilon(t)|^2 + \|b_\varepsilon(t)\|^2 \leq 13\varepsilon^2/16$ . In particular,  $T_\varepsilon$  is lower-bounded on  $[-t_\varepsilon^*, t_\varepsilon^*]$  by a constant  $H$  depending on  $R$  only. Noticing that  $t_\varepsilon^* \leq t_\varepsilon$ , we can use the inequality in (3.30) to obtain

$$\int_{-1}^1 (a_\varepsilon a'_\varepsilon)^2 T_\varepsilon \geq \frac{1}{32} A_1^4 \ell^4 H \int_{-t_\varepsilon^*}^{t_\varepsilon^*} t^2 dt = \frac{1}{48} A_1^4 H G^3 \ell \varepsilon^3.$$

Finally, since  $z \mapsto z(z - \nu)$  is minimal on  $\mathbb{R}_+$  at  $z = \nu/2$  with minimal value  $-\nu^2/4$ , we find the bound

$$\int_{-1}^1 \tau(Q_\varepsilon) Q_\varepsilon \geq c_R \varepsilon^{k+1} - c'_R \ell^3 \varepsilon^{2k} \geq c_R \varepsilon^{k+1},$$

provided that  $\varepsilon$  is small enough before  $R$ .

In both cases, we find that  $\int_{-1}^1 \tau(Q_\varepsilon) Q_\varepsilon \geq c_R \varepsilon^{k+1}$ . Now integrating (3.29) gives

$$d_{M_\varepsilon}(x, y) \geq d_{M_0}(x, y) + c'_R \varepsilon^{k+1} > d_{M_0}(x, y).$$

Finally, (3.23) yields  $d_{M_\varepsilon}(x, y) \geq 2A_2\ell$  and letting  $\ell := (1 \vee D_3^{-1})\varepsilon$ , which we can from (3.28), finally gives

$$\left| 1 - \frac{d_{M_0}(x, y)}{d_{M_\varepsilon}(x, y)} \right| \geq c_R \varepsilon^k. \quad (3.32)$$

**Concluding with Le Cam's lemma** We apply Lemma 3.C.1 with  $M_0$  and  $M_1 := M_\varepsilon$  for  $\varepsilon$  properly chosen. Their volumes are bounded from above and below by something depending on  $R$  and  $d$  only, so that the uniform distribution on  $M_0$  and  $M_\varepsilon$  are in  $\mathcal{P}_{\text{rch}_{\min}, \text{L}}^k(f_{\min}, f_{\max})$  provided that  $f_{\min}$  and  $f_{\max}$  are respectively small enough and



large enough compared to  $1/R^d$ . Finally, we set  $R = 2 \text{rch}_{\min}$  and  $\varepsilon = (C_{\text{rch}_{\min}, d} n)^{-1/d}$ . For  $n$  large enough so that all previous controls are verified, Lemma 3.C.1 finally yields

$$\inf_{\widehat{d}} \sup_{P \in \mathcal{P}^k} \mathbb{E}_{P^{\otimes n}} [\ell_{\infty}(\widehat{d} | d_M)] \geq \frac{1}{2} c_{\text{rch}_{\min}} \varepsilon^k (1 - C_{\text{rch}_{\min}, d} n \varepsilon^d) \geq c_{\text{rch}_{\min}, d, k} n^{-k/d},$$

where the total variation was bounded using [BHHS22, Lemma 7].  $\square$

### 3.C.2 Plug-in Estimation for Metric Learning

We start by giving the proof of Proposition 3.26.

*Proof of Proposition 3.26.* Let  $x, y \in K$ . Notice that, since  $K \subset (K')^{\varepsilon}$ , there holds trivially that  $d_{(K')^{\varepsilon}}(x, y) \leq d_K(x, y)$ . For the converse inequality, let  $\gamma : [0, 1] \rightarrow \mathbb{R}^D$  be a continuous path in  $(K')^{\varepsilon}$  between  $x$  and  $y$ . Since  $\varepsilon < \text{rch}(K)/2$  the closest-point projection on  $K$  is well-defined on  $(K')^{\varepsilon} \subset K^{2\varepsilon}$  and we can consider  $\gamma_0 = \text{pr}_K \circ \gamma$ , which is a continuous path in  $K$ . For any subdivision  $0 = t_0 < t_1 < \dots < t_k = 1$ , there holds

$$\sum_{i=0}^{k-1} \|\gamma_0(t_{i+1}) - \gamma_0(t_i)\| \leq \frac{\text{rch}(K)}{\text{rch}(K) - 2\varepsilon} \sum_{i=0}^{k-1} \|\gamma(t_{i+1}) - \gamma(t_i)\|$$

where we used the fact that  $\text{pr}_K$  is  $\text{rch}(K)/(\text{rch}(K) - 2\varepsilon)$ -Lipschitz on  $K^{2\varepsilon}$  [Fed59, Thm 4.8 (8)]. Taking the supremum over all subdivision yields

$$d_K(x, y) \leq L(\gamma_0) \leq \frac{\text{rch}(K)}{\text{rch}(K) - 2\varepsilon} L(\gamma)$$

and then taking the infimum on all continuous path  $\gamma$  finally gives

$$d_{(K')^{\varepsilon}}(x, y) \geq \left(1 - \frac{2\varepsilon}{\text{rch}(K)}\right) d_K(x, y)$$

ending the proof.  $\square$

To prove Theorem 3.27, an intermediate result that bounds the intrinsic diameters of the supports in our statistical model is needed.

**Lemma 3.C.2.** *For any  $P \in \mathcal{P}_{\text{rch}_{\min}, \mathbf{L}}^k(f_{\min}, f_{\max})$ , if  $M = \text{supp}(P)$ , then*

$$\sup_{x, y \in M} d_M(x, y) \leq d_{\max}.$$

where  $d_{\max}$  is defined in Theorem 3.27.

*Proof of Lemma 3.C.2.* We let  $x_1, \dots, x_N$  be a  $\text{rch}_{\min}/4$ -packing of  $M$ . We let  $x, y \in M$ , and  $G$  be the neighborhood graph built on top of  $x, y, x_1, \dots, x_N$  with connectivity radius  $\text{rch}_{\min}/2$ . Using [NSW08, Theorem 6.3], denoting  $z_0 = x, z_1, \dots, z_k = y$  the shortest path between  $x$  and  $y$  in  $G$ , there holds

$$d_M(x, y) \leq \sum_{i=0}^{k-1} d_M(z_i, z_{i+1}) \leq \sum_{i=0}^{k-1} 2\|z_i - z_{i+1}\| \leq k \text{rch}_{\min}.$$

But now  $k \leq N - 1$  and

$$N \leq \frac{\text{vol}_d(M)}{\min_{x \in M} \text{vol}_d(M \cap \mathbb{B}(x, \text{rch}_{\min}/4))} \leq \frac{\text{vol}_d(M)}{(1 - 1/8^2)^{d/2} \omega_d \text{rch}_{\min}^d / 4^d},$$

where we used [NSW08, Lemma 5.3]. Noticing that  $\text{vol}_d(M) \leq 1/f_{\min}$ , we easily conclude.  $\square$

We are now in position to prove Theorem 3.27.

*Proof of Theorem 3.27.* We let  $\mathcal{A}_n := \{d_H(\widehat{M}, M) \leq \varepsilon_n\}$  denote the event where  $\widehat{M}$  is  $\varepsilon_n$ -precise in Hausdorff distance, and we take  $x, y \in M$ . On the event  $\mathcal{A}_n$ , for  $n$  large enough such that  $\varepsilon_n \leq \text{rch}_{\min}/2$ , Proposition 3.26 applies to  $K' = \widehat{M} \cup \{x, y\}$  and, together with Lemma 3.C.2, yields

$$\left| 1 - \frac{\widehat{d}(x, y)}{d_M(x, y)} \right| \leq \frac{2\varepsilon_n}{\text{rch}_{\min}}.$$

On  $\mathcal{A}_n^c$ , we distinguish whether  $\|x - y\| \leq \varepsilon_n$  or not. If so, then  $\widehat{d}(x, y) = \|x - y\| \leq d_M(x, y)$ . In the other case,  $d_M(x, y) \geq \|x - y\| \geq \varepsilon_n$  and  $d(x, y) \leq d_{\max}$  so that, in any case

$$\left| 1 - \frac{\widehat{d}(x, y)}{d_M(x, y)} \right| \leq 1 + \frac{\widehat{d}(x, y)}{d_M(x, y)} \leq 1 + \frac{d_{\max}}{\varepsilon_n},$$

for  $n$  large enough such that  $\varepsilon_n \leq d_{\max}$ . Patching these two bounds together yields

$$\mathbb{E}_{P^{\otimes n}}[\ell_{\infty}(\widehat{d}|d_M)] \leq \frac{2\varepsilon_n}{\text{rch}_{\min}} P^{\otimes n}(\mathcal{A}_n) + \left(1 + \frac{d_{\max}}{\varepsilon_n}\right) P^{\otimes n}(\mathcal{A}_n^c),$$

ending the proof.  $\square$

### 3.D Proofs of Section 3.6

We first prove that submanifolds of the model do fulfill Assumption 3.20 and Assumption 3.21.

*Proof of Proposition 3.30.* Assumption 3.21 is a simple consequence of [NSW08, Proposition 6.3] which yields fulfillment for  $\Delta_1 = \text{rch}(M)/2$  and  $C_1 = 2$ . For Assumption 3.20, take  $x, y \in M$  such that  $\|x - y\| \leq \text{rch}(M)$  and take  $\varepsilon < \text{rch}(M)/4$ . We consider  $a = \exp_y(v)$ , where

$$v = -\varepsilon \frac{\text{pr}_{T_y M}(x - y)}{\|\text{pr}_{T_y M}(x - y)\|}.$$

Thanks to [Fed59, Theorem 4.8 (7)], there holds

$$\begin{aligned} \|\text{pr}_{T_y M}(x - y)\|^2 &= \|x - y\|^2 - d^2(x - y, T_y M) \\ &\geq \|x - y\|^2 - \frac{\|x - y\|^4}{4\text{rch}^2(M)} \\ &\geq \frac{3}{4}\|x - y\|^2, \end{aligned}$$

and

$$\langle v, y - x \rangle = \varepsilon \frac{\langle x - y, \text{pr}_{T_y M}(x - y) \rangle}{\|\text{pr}_{T_y M}(x - y)\|} = \varepsilon \|\text{pr}_{T_y M}(x - y)\| \geq \frac{1}{2} \varepsilon \|x - y\|,$$

so that

$$\|x - y - v\|^2 \geq \|x - y\|^2 + \varepsilon \|x - y\| + \varepsilon^2 \geq \left( \|x - y\| + \frac{1}{2} \varepsilon \right)^2,$$

and thus  $\|x - y - v\| \geq \|x - y\| + \varepsilon/2$ . But now  $\|x - a\| \geq \|x - y - v\| - \|a - y - v\|$  and  $\|a - y - v\| \leq 5\varepsilon^2/4 \text{rch}(M)$  according to [AL19, Lemma 1]. All in all, we get that

$$\|x - a\| \geq \|x - y\| + \frac{1}{2} \varepsilon - \frac{5}{4 \text{rch}(K)} \varepsilon^2 \geq \|x - y\| + \frac{3}{16} \varepsilon,$$

ending the proof.  $\square$

To prove Theorem 3.31, a bound on the metric distortion between our distance estimator and  $d_M$  is needed, that easily follows from Proposition 3.26.

**Proposition 3.D.1.** *In the context of Proposition 3.26, we have that for all  $\delta > 4\varepsilon$ ,*

$$D_\delta(d_K, d_{(K')^\varepsilon}) \leq 1 + \frac{4\varepsilon}{(\delta - 4\varepsilon) \wedge \text{rch}(K)}.$$

*Proof of Proposition 3.D.1.* Proposition 3.26 already gives that  $D_\delta(d_K | d_{(K')^\varepsilon}) \leq 1 + 2\varepsilon/\text{rch}(K)$ . For the other control, notice that for any two  $x, y \in (K')^\varepsilon$  that are  $\delta$ -apart for the Euclidean distance, there holds denoting  $x_0 = \text{pr}_K(x)$  and  $y_0 = \text{pr}_K(y)$ ,

$$d_{(K')^\varepsilon}(x, y) \leq 4\varepsilon + d_K(x_0, y_0)$$

because the piecewise-defined path consisting of the segment  $[x, x_0]$  of the (or a near-minimizing) shortest-path between  $x_0$  and  $y_0$  in  $K$ , and of the segment  $[y_0, y]$ , is a continuous path in  $(K')^\varepsilon$  between  $x$  and  $y$  of length the RHS of the display above. Now notice that

$$d_K(x_0, y_0) \geq \|x_0 - y_0\| \geq \delta - 4\varepsilon,$$

which immediately yields  $D_\delta(d_{(K')^\varepsilon} | d_K) \leq 1 + \frac{4\varepsilon}{\delta - 4\varepsilon}$ .  $\square$

The rate of the plug-in SDR estimator follows straightforwardly.

*Proof of Theorem 3.31.* Let  $\mathcal{A}_n := \{d_H(M, \widehat{M}) \leq \varepsilon_n\}$ . On this event, we have  $D_\delta(\widehat{d}, d_M) \leq 1 + 8\varepsilon_n/\delta$  according to Proposition 3.D.1, so that applying Theorem 3.23 with  $\delta_0 = \delta/2$ ,  $\varepsilon = \varepsilon_n$  and  $\nu = 8\varepsilon_n/\delta$  yields  $|\widehat{\text{sdr}}_\delta - \text{sdr}_\delta(M, d_M)| \leq \zeta_0 \varepsilon_n$  with  $\zeta_0 \leq C s_{\max}^4 / \delta^4$ . We conclude that

$$\mathbb{E}_{P^{\otimes n}} |\widehat{\text{sdr}}_\delta - \text{sdr}_\delta(M, d_M)| \leq \zeta_0 \varepsilon_n P^{\otimes n}(\mathcal{A}_n) + 2s_{\max} P^{\otimes n}(\mathcal{A}_n^c),$$

which ends the proof.  $\square$



## Chapter 4

# Pointwise density estimation on submanifolds

In this chapter, we investigate density estimation from a  $n$ -sample in the Euclidean space, when the data is supported by an unknown submanifold of possibly unknown dimension. We study nonparametric kernel methods for pointwise loss, with data-driven bandwidths that incorporate some learning of the geometry via a local dimension estimator. When the density has Hölder smoothness, our estimator achieves an asymptotically minimax rate, provided that the underlying manifold is smooth enough. Following Lepski's principle, a bandwidth selection rule is shown to achieve smoothness adaptation. Finally, a numerical implementation is conducted on some case studies in order to confirm the practical feasibility of our estimators. This chapter has been published in [BH21].

### Contents

---

<b>4.1</b>	<b>Introduction . . . . .</b>	<b>108</b>
4.1.1	Motivation . . . . .	108
4.1.2	Main results . . . . .	110
4.1.3	Organisation of the paper . . . . .	111
<b>4.2</b>	<b>Manifold-supported probability distributions . . . . .</b>	<b>112</b>
4.2.1	Some material from geometry . . . . .	112
4.2.2	Hölder spaces on submanifolds . . . . .	113
4.2.3	The reach of a subset . . . . .	114
4.2.4	A statistical model for sampling on a unknown manifold . . . . .	115
4.2.5	Choice of a loss function and the reach assumption . . . . .	117
<b>4.3</b>	<b>Density estimation at a fixed point . . . . .</b>	<b>118</b>
4.3.1	Main results . . . . .	118
4.3.2	Kernel estimation . . . . .	119
4.3.3	The special case of one-dimensional submanifolds . . . . .	122
4.3.4	Smoothness adaptation . . . . .	123

4.3.5	Simultaneous adaptation to smoothness and dimension . . . .	124
<b>4.4</b>	<b>Numerical illustration . . . . .</b>	<b>126</b>
4.4.1	Simulation on one-dimensional submanifolds . . . . .	127
4.4.2	Simulation on two-dimensional submanifolds . . . . .	128
4.4.3	Adaptation . . . . .	130

## 4.1 Introduction

### 4.1.1 Motivation

Suppose we observe an  $n$ -sample  $(X_1, \dots, X_n)$  of size  $n$  distributed on an Euclidean space  $\mathbb{R}^D$  according to some density function  $f$ . We wish to recover  $f$  at some arbitrary point  $x_0 \in \mathbb{R}^D$  nonparametrically. If the smoothness of  $f$  at  $x_0$  measured in a strong sense is of order  $\beta$  – for instance by a Hölder condition or with a prescribed number of derivatives – then the optimal (minimax) rate for recovering  $f(x_0)$  is of order  $n^{-\beta/(2\beta+D)}$  and is achieved by kernel or projection methods, see *e.g.* the classical textbooks [Sil86, DG85] or [Tsy08, Sec. 1.2-1.3]. Extension to data-driven bandwidths [Bow84, Chi91] offers the possibility to adapt to unknown smoothness, see [GL08, GL11, GL14] for a modern mathematical formulation. More generally, recommended reference on adaptive estimation is the textbook by [GN16]. In many situations however, the dimension  $D$  of the ambient space is large, hitherto disqualifying such methods for practical applications. Opposite to the curse of dimensionality, a broad guiding principle in practice is that the observations  $(X_1, \dots, X_n)$  actually live on smaller dimensional structures and that the effective dimension of the problem is smaller if one can take advantage of the geometry of the data [FMN16a]. This classical paradigm probably goes back to a conjecture of [Sto82] that paved the way to the study of the celebrated single-index model in nonparametric regression, where a structural assumption is put in the form  $f(x) = g(\langle \vartheta, x \rangle)$ , where  $\langle \cdot, \cdot \rangle$  is the scalar product on  $\mathbb{R}^D$ , for some unknown univariate function  $g : \mathbb{R} \rightarrow \mathbb{R}$  and direction  $\vartheta \in \mathbb{R}^D$ . Under appropriate assumptions, the minimax rate of convergence for recovering  $f(x)$  with smoothness  $\beta$  drops to  $n^{-\beta/(2\beta+1)}$  and does not depend on the ambient dimension  $D$ , see *e.g.* [GL07, LS14] and the references therein. Also, in the search for significant variables, one postulates that  $f$  only depends on  $d < D$  coordinates, leading to the structural assumption  $f(x_1, \dots, x_D) = F(x_{i_1}, \dots, x_{i_d})$  for some unknown function  $F : \mathbb{R}^d \rightarrow \mathbb{R}$  and  $\{i_1, \dots, i_d\} \subset \{1, \dots, D\}$ . In an analogous setting, the minimax rate of convergence becomes  $n^{-\beta/(2\beta+d)}$  and this is also of a smaller order of magnitude than  $n^{-\beta/(2\beta+D)}$ , see [HL02] in the white noise model.

The next logical step is to assume that the data  $(X_1, \dots, X_n)$  live on a  $d$ -dimensional

submanifold  $M$  of the ambient space  $\mathbb{R}^D$ . When the manifold is known prior to the experiment, nonparametric density estimation dates back to [DG85] when  $M$  is the circle, and on a homogeneous Riemannian manifold by [Hen90], see also [Pel05]. Several results are known for specific geometric structures like the sphere or the torus involved in many applied situations: inverse problems for cosmological data [KK02, KKL09, KPNP11], in geology [HWC87] or flow calculation in fluid mechanics [ES00]. For genuine compact homogeneous Riemannian manifolds, a general setting for smoothness adaptive density estimation and inference has recently been considered by [KNP12], or even in more abstract metric spaces in [CGK<sup>+</sup>20]. See also [BKMP09, CKP14] and the references therein. A common strategy adapts conventional nonparametric tools like projection or kernel methods to the underlying geometry, via the spectral analysis of the Beltrami-Laplace operator on  $M$ . Under appropriate assumptions, this leads to exact or approximate eigenbases (spherical harmonics for the sphere, needlets and so on) or properly modified kernel methods, according to the Riemannian metric on  $M$ .

If the submanifold  $M$  itself is unknown, getting closer in spirit to a dimension reduction approach, the situation becomes drastically different:  $M$  hence its geometry is unknown, and considered as a nuisance parameter. In order to recover the density  $f$  at a given point  $x_0 \in \mathbb{R}^D$  of the ambient space, one has to understand the minimal geometry of  $M$  that must be learned from the data and how this geometry affects the optimal reconstruction of  $f$ . This is the topic of the paper.

We consider in the paper a seemingly unusual framework where the support of a distribution is unknown while the aim is to recover the density at a point  $x_0 \in \mathbb{R}^D$  which is known to be on the support. As mentioned above, this actually covers at least two situations:

- The data are high-dimensional and it is reasonable to believe that they actually lie on a smaller dimensional subset of the ambient space  $\mathbb{R}^D$ , which can be assumed to be a submanifold. In that case,  $x_0$  can be seen as an extraneous observation  $X$  from the density  $f$  (extracted for instance from the point cloud), and the analysis can be implicitly performed conditional on  $X = x_0$ ;
- The data naturally lie on a submanifold, like a spheroid for geological application, or a cell membrane in microbiology (see for instance [KPS14] who describe a technique that yields such a point cloud). In this case,  $x_0$  can be seen as an observation  $X$  like above, but there is also the situation where the statistician can know whether or not a given point  $x_0$  is within the support (for instance a point on a cell membrane, or a geographical location on the Earth surface) without knowing the geometric features of the latter and without needing to estimate

them.

### 4.1.2 Main results

We construct a class of compact smooth submanifolds of dimension  $d$  of the Euclidean space  $\mathbb{R}^D$ , without boundaries, that constitute generic models for the unknown support of the target density  $f$  that we wish to reconstruct. We further need a *reach condition*, a somehow unavoidable notion in manifold reconstruction that goes back to [Fed59]: it is a geometric invariant that quantifies both local curvature conditions and how tightly the submanifold folds on itself. It is related to the scale at which the sampling rate  $n$  can effectively recover the geometry of the submanifold, see Section 4.2.3 below. We consider regular manifolds  $M$  with reach bounded below that satisfy the following property:  $M$  admits a local parametrisation at every point  $x \in M$  by its tangent space  $T_x M$ , and this parametrisation is sufficiently regular. A natural candidate is given by the exponential map  $\exp_x : T_x M \rightarrow M \subset \mathbb{R}^D$ . More specifically, for some regularity parameter  $\alpha \geq 0$ , we require a certain uniform bound for the  $(\alpha + 1)$ -fold differential of the exponential map to hold, quantifying in some sense the regularity of the parametrisation in a minimax spirit, see Definition 4.2.4 below. Our approach is close to that of [AL19, Def. 1] that consider arbitrary parametrisations among those close to the inverse of the projection onto tangent spaces. Given a density function  $f : M \rightarrow [0, \infty)$  with respect to the volume measure on  $M$ , we have a natural extension of smoothness spaces on  $M$  by requiring that  $f \circ \exp_x : T_x M \rightarrow \mathbb{R}$  is a smooth map in any reasonable sense, see Section 4.2.2 below. This is line for instance with [Tri87] for the characterisation of function spaces on a Riemannian manifold.

Our main result is that in order to reconstruct  $f(x_0)$  efficiently at a point  $x_0 \in \mathbb{R}^D$  when  $f$  has smoothness  $\beta$  and lives on an unknown submanifold of smoothness  $\alpha$  and unknown dimension  $d < D$ , it is sufficient to consider estimators of the form

$$\widehat{f}_h(x_0) = \frac{1}{nh^{\widehat{d}(x_0)}} \sum_{i=1}^n K\left(\frac{x_0 - X_i}{h}\right), \quad x \in \mathbb{R}^D, \quad (4.1)$$

where  $K : \mathbb{R}^D \rightarrow \mathbb{R}$  is a certain kernel and  $\widehat{d}(x_0) = \widehat{d}(x_0, X_1, \dots, X_n)$  is an estimator of the local dimension of the support of  $f$  in the vicinity of  $x$  based on a scaling estimator as introduced in [FSA07]. We prove in Theorem 4.3.1 that following a classical bias-variance trade-off for the bandwidth  $h$ , the rate  $n^{-\alpha \wedge \beta / (2\alpha \wedge \beta + d)}$  is achievable for pointwise and global loss when the dimension of  $M$  is  $d$ , irrespectively of the ambient dimension  $D$ . In particular, it is noteworthy that in terms of manifold learning, only the dimension of  $M$  needs to be estimated. When  $\alpha \geq \beta$ , we also have a lower bound (Theorem 4.3.2) showing that our result is asymptotically minimax optimal. Moreover, by implementing Lepski's principle [Lep92], we are able to construct a data driven



bandwidth  $\widehat{h} = \widehat{h}(x_0, X_1, \dots, X_n)$  that achieves in Theorem 4.3.4 the rate  $n^{-\alpha \wedge \beta / (2\alpha \wedge \beta + d)}$  up to a logarithmic term — unavoidable in the case of pointwise loss due to the Lepski-Low phenomenon [Lep90, Low92]. When the dimension  $d$  is known, the estimator (4.1) has already been investigated in squared-error norm in [OG09] for a fixed manifold  $M$  and smoothness  $\beta = 2$ .

A remaining issue at this stage is to understand how the regularity of  $M$  can affect the minimax rates of convergence for smooth functions, *i.e.* when  $\alpha \leq \beta$ . We only have a partial answer to that question, when we restrict our attention to the one-dimensional case  $d = 1$ . When  $M$  is known, [Pel05] studied estimators of the form

$$\frac{1}{nh^d} \sum_{i=1}^n \frac{1}{\vartheta_{x_0}(X_i)} K\left(\frac{d_M(x_0, X_i)}{h}\right), \quad (4.2)$$

where  $K : \mathbb{R} \rightarrow \mathbb{R}$  is a radial kernel,  $d_M$  is the intrinsic Riemannian distance on  $M$  and the correction term  $\vartheta_{x_0}(X_i)$  is the volume density function on  $M$  [Bes78, p. 154] that accounts for the value of the density of the volume measure at  $X_i$  in normal coordinates around  $x_0$ , taking into account how the submanifold curves around  $X_i$ . By establishing in Lemma 4.3.9 that  $\vartheta_x$  is constant (and identically equal to one) when  $d = 1$ , we have another estimator by simply learning the geometry of  $M$  via its intrinsic distance  $d_M$  in (4.2). This can be done by efficiently estimating  $d_M$  in dimension  $d = 1$  thanks to the Isomap method as coined by [TSL00]. Therefore, in the special case when the dimension  $d$  of  $M$  is known and equal to 1, we are able to construct an estimator that achieves in Theorem 4.3.3 the rate  $n^{-\beta / (2\beta + 1)}$ , therefore establishing that in dimension  $d = 1$  at least, the regularity of the manifold  $M$  does not affect the minimax rate for estimating  $f$  even when  $M$  is unknown. However, the volume density function  $\vartheta_{x_0}$  is not constant as soon as  $d \geq 2$  and obtaining a global picture in higher dimensions remains an open and presumably challenging problem.

### 4.1.3 Organisation of the paper

In Section 4.2, we provide with all the necessary material and notation from classical geometry for the unfamiliar reader. Section 4.2.1 together with the construction of smoothness spaces – here Hölder spaces on a submanifold in Section 4.2.2. We elaborate in particular on the reach of a subset of the Euclidean space in Section 4.2.3 and construct a statistical model for sampling  $n$  data from a density  $f$  with regularity  $\beta$  living on an unknown submanifold  $M$  of unknown dimension  $d$  and smoothness  $\alpha$  in an ambient space of dimension  $D$  in Section 5.2. In this setting, we establish in Section 4.2.5 that a reach condition, *i.e.* assuming that the reach of  $M$  is bounded below, is necessary in order to reconstruct  $d$ . This is stated precisely in Theorem 4.2.10.

We give our main results in Section 4.3 and more specifically in Section 4.3.1. When the dimension  $d$  and the smoothness parameters  $\alpha$  of the unknown manifold  $M$  and the smoothness  $\beta$  of  $f$  are known, Theorem 4.3.1 states the existence of an estimator that achieves the rate  $n^{-\alpha \wedge \beta / (2\alpha \wedge \beta + d)}$  in expected pointwise loss, and Theorem 4.3.2 establishes that a minimax lower bound is  $n^{-\beta / (2\beta + d)}$ . Theorem 4.3.3 shows the existence of estimators in dimension  $d = 1$  that achieve the rate  $n^{-\beta / (2\beta + 1)}$ , which is therefore minimax in that case. Theorem 4.3.4 states the existence of smoothness and dimension adaptive estimators, when  $\alpha, \beta$  and  $d$  are unknown. Section 4.3.2 elaborates on special kernels upon which the estimators that achieve the aforementioned results are constructed, and their properties with respect to bias and variance analysis. The underlying geometry of  $M$  makes the usual orthogonality to non-constant polynomials of a certain degree (the order of the kernel) irrelevant, and a specific construction must be undertaken. Section 4.3.3 focuses on the case of one-dimensional submanifolds  $M$  when  $d = 1$ , where we explicitly construct a kernel estimator that achieves the minimax rate of convergence, revisiting the estimator (4.2) of [Pel05] and relying on the Isomap algorithm. In Section 4.3.4, we implement Lepski's algorithm on the bandwidth of our kernel estimators, following [LMS97]; this achieves smoothness adaptation w.r.t.  $\alpha \wedge \beta$ . Finally, in Section 4.3.5, we build an estimator of the dimension  $d$  of  $M$ , following ideas of [FSA07] and that enables us to obtain simultaneous adaptation w.r.t.  $\alpha \wedge \beta$  and  $d$  by plug-in.

Finally, numerical examples are developed in Section 4.4: we elaborate on examples of non-isometric embeddings of the circle and the torus in dimension 1 and 2 and explore in particular rates of convergence on Monte-Carlo simulations, illustrating how effective Lepski's method can be in that context. The proof are delayed until Appendix 4.A.

## 4.2 Manifold-supported probability distributions

### 4.2.1 Some material from geometry

We endow  $\mathbb{R}^D$  with its usual Euclidean product and norm, respectively denoted by  $\langle \cdot, \cdot \rangle$  and  $\| \cdot \|$ . We denote by  $B(x, r)$  the open ball of  $\mathbb{R}^D$  of center  $x$  and radius  $r$ . For any subspace  $H \subset \mathbb{R}^D$ , we set  $B_H(x, r) = H \cap B(x, r)$  for the open ball in  $H$  for the induced norm.

We recall some basic notions of geometry of submanifolds of the Euclidean space  $\mathbb{R}^D$  for the unfamiliar reader. We borrow material from the classical textbooks [GHL90] and [Lee06]. In all the paper, we consider  $C^\infty$  Riemannian manifolds  $(M, g)$  that we informally call *smooth*, which is an abstract manifold  $M$  endowed with a  $C^\infty$  atlas

[GHL90, Def 1.6 p.6] and a  $C^\infty$  metric  $g$  [GHL90, Def 2.1 p.52]. Such a manifold can always be embedded isometrically into some Euclidean space [GR70], meaning that the pull-back of the canonical Euclidean metric coincides with the metric of the manifold. When we consider a smooth submanifold  $M \subset \mathbb{R}^D$ , we mean that  $M$  is the image through such an embedding of a smooth abstract Riemannian manifold.

Since we quantitatively compare the smoothness of manifolds within a large class of models, we need to pick a canonical parametrisation. For this reason, we consider the *exponential map* [GHL90, Def 2.86 p.85]; for any smooth submanifold  $M \subset \mathbb{R}^D$  and any  $x \in M$ , it defines a smooth parametrisation

$$\exp_x : B_{T_x M}(0, \varepsilon) \rightarrow M$$

of  $M$  around  $x$ , provided that  $\varepsilon$  is chosen small enough [GHL90, Cor 2.89 p.86]. The supremum of all such  $\varepsilon$  is called the injectivity radius at  $x$  and is denoted  $\text{inj}_M(x)$ . When  $M$  is a closed subset of  $\mathbb{R}^D$ , the exponential maps are well defined on the whole tangent spaces. This is (one side of) the Hopf-Rinow theorem [Lee06, Thm 6.13 p.108].

Given a submanifold  $M$  of dimension  $d$ , the *volume measure* of  $M$ , denoted by  $\mu_M$ , is the restriction of the  $d$ -dimensional Hausdorff measure  $\mathcal{H}^d$  to  $M$ , see [Fed69, Sec 2.10.2 p.171] for a definition. It can be shown [EG92, Ex D p.102] that this definition coincides with the usual one of volume measure of a Riemannian manifold: if  $\psi : M \rightarrow \mathbb{R}$  is a continuous function with support in  $\exp_x(B_{T_x M}(0, \varepsilon))$  for  $\varepsilon$  smaller than  $\text{inj}_M(x)$ , we have

$$\mu_M(\psi) = \int_{B_{T_x M}(0, \varepsilon)} \psi \circ \exp_x(v) \sqrt{\det g^x(v)} dv,$$

with  $g_{ij}^x(v) = \langle d\exp_x(v)[e_i], d\exp_x(v)[e_j] \rangle$  and where  $(e_1, \dots, e_d)$  is an arbitrary orthonormal basis of  $T_x M$ . We refer to [GHL90, Sec 3.H.1 and Sec 3.H.2] for further details on the volume measure. The volume of  $M$ , denoted by  $\text{vol } M$ , is simply  $\mu_M(\mathbb{1})$ . It is finite when  $M$  is a compact submanifold of  $\mathbb{R}^D$ .

#### 4.2.2 Hölder spaces on submanifolds

Let  $M$  be a smooth submanifold of  $\mathbb{R}^D$  and let  $\rho > 0$ . We say that a vector-valued function  $\varphi : M \rightarrow \mathbb{R}^m$  with  $m \geq 1$  is  $\gamma$ -Hölder with  $\gamma > 0$  if for all  $x \in M$ , the map

$$\varphi \circ \exp_x : B_{T_x M}(0, \rho_x) \rightarrow \mathbb{R}^m \quad \text{where } \rho_x = \rho \wedge \text{inj}_M(x) \quad (4.3)$$

is  $\gamma$ -Hölder in the usual sense, namely

- (i)  $\varphi \circ \exp_x$  is  $k = \lceil \gamma - 1 \rceil$ -times differentiable;

(ii) and verifies

$$\forall v, w \in B_{T_x M}(0, \rho), \quad \|d^k(\varphi \circ \exp_x)(v) - d^k(\varphi \circ \exp_x)(w)\|_{\text{op}} \leq R \|v - w\|^\delta$$

with  $\delta = \gamma - k > 0$  and for some  $R > 0$ .

We denote by  $\mathcal{H}^\gamma(M, \mathbb{R}^m)$  the space of all such functions, and define for  $\varphi \in \mathcal{H}^\gamma(M, \mathbb{R}^m)$  the Hölder semi-norm

$$|\varphi|_\gamma = \sup_{x \in M} \sup_{v, w \in B_{T_x M}(0, \rho_x)} \frac{\|d^k(\varphi \circ \exp_x)(v) - d^k(\varphi \circ \exp_x)(w)\|_{\text{op}}}{\|v - w\|^\delta}.$$

The characterisation of the smoothness of a function  $\varphi : M \rightarrow \mathbb{R}^m$  through the exponential maps is a classical way to define functional spaces over Riemannian manifolds, see for instance [Tri87].

*Remark 4.2.1.* The functional set  $\mathcal{H}^\gamma(M, \cdot)$  also depends on the parameter  $\rho$ , merely introduced for a technical reason and dropped from the notation. When the manifold  $M$  has a reach (defined in the next Section 4.2.3) bounded from below by  $\tau > 0$ , a natural choice for  $\rho$  is  $\pi\tau$ , in light of the results of Proposition 4.2.6 below.

*Remark 4.2.2.* This definition of Hölder smoothness is also intrinsic, meaning that it does not depend on the way  $M$  is embedded into  $\mathbb{R}^D$ . If indeed  $\psi : N \rightarrow M$  is an isometry between  $N$  and  $M$ , then  $|\varphi|_\gamma = |\varphi \circ \psi|_\gamma$  for any  $\varphi : M \rightarrow \mathbb{R}^d$  and  $\gamma > 0$ .

### 4.2.3 The reach of a subset

One of the main concerns when dealing with observations sampled from a geometrically structured probability measure is to determine the suitable scale at which one should look at the data. Indeed, given finite-sized point cloud in  $\mathbb{R}^D$ , there are infinitely many submanifolds that interpolate the point cloud, see Figure 4.2.1 for an illustration. A popular notion of regularity for a subset of the Euclidean space is the *reach*, introduced by [Fed59].

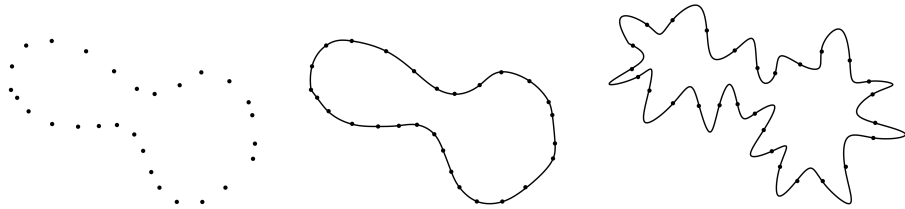


Figure 4.2.1 – An arbitrary points cloud (Left) for  $D = 2$ , and two smooth one-dimensional submanifolds passing through all its points (Middle, Right). A reach condition tends to discard the Right manifold as a likely candidate among all possible submanifolds the point cloud is sampled from.

**Definition 4.2.3.** Let  $K$  be a compact subset of  $\mathbb{R}^D$ . The reach  $\tau_K$  of  $K$  is the supremum of all  $r \geq 0$  such that the orthogonal projection  $\text{pr}_K$  on  $K$  is well-defined on the  $r$ -neighbourhood  $K^r$  of  $K$ , namely

$$\tau_K = \sup \left\{ r \geq 0 \mid \forall x \in \mathbb{R}^D, d(x, K) \leq r \Rightarrow \exists! y \in K, d(x, K) = \|x - y\| \right\}.$$

When  $M$  is a compact submanifold of  $\mathbb{R}^D$ , the reach  $\tau_M$  quantifies two geometric invariants: locally, it measures how curved the manifold is, and globally, it measures how close it is to intersect itself (the so-called bottleneck effect). See Figure 4.2.2 for an illustration of the phenomenon. A *reach condition*, meaning that the reach is bounded below, is necessary in order to obtain minimax inference results in manifold learning. These include: homology inference [NSW08, BRS<sup>+</sup>12], curvature [AL19] and reach estimation itself [AKC<sup>+</sup>19] as well as manifold estimation [GPPVW12a, AL19].

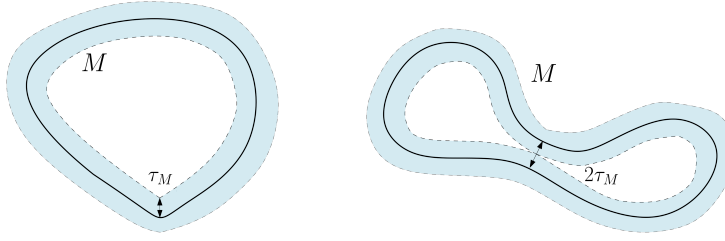


Figure 4.2.2 – For the first manifold  $M$  (Left), the value of the reach  $\tau_M$  comes from its curvature. For the second one (Right), the reach is equal to  $\tau_M$  because it is close to self intersecting (a bottleneck effect). The blue area represents the tubular neighbourhood over which the orthonormal projection on each manifold is well-defined.

#### 4.2.4 A statistical model for sampling on a unknown manifold

In the following, we fix a point  $x_0 \in \mathbb{R}^D$  in the ambient space. See Section 4.1.1 for a discussion on such a setting. Our statistical model is characterized by two quantities: the regularity of its support and the regularity of the density defined on this support. The support belongs to a class of submanifolds  $M$ , for which we need to fix some kind of *canonical* parametrisation. This is what [AL19] propose by asking the support  $M$  to admit a local parametrisation at all point  $x \in M$  by  $T_x M$ , and that this parametrisation is close to being the inverse of the projection over this tangent space. We follow this idea by imposing a constraint on the exponential map. In the following, we take  $\tau > 0$  and set  $\rho = \pi\tau$  in the definition of the Hölder spaces of Section 4.2.2.

**Definition 4.2.4.** Let  $1 \leq d < D$  be integers. We let  $\mathcal{C}_d(\tau)$  define the set of submanifolds  $M$  of  $\mathbb{R}^D$  that contain  $x_0$  and satisfying the following properties:

- (i) (Dimension)  $M$  is a smooth submanifold of dimension  $d$  without boundaries;

(ii) (Compactness)  $M$  is compact;

(iii) (Reach condition) We have  $\tau_M \geq \tau$ .

For  $\alpha \geq 1$  and  $L > 0$ , we define  $\mathcal{C}_{d,\alpha}(\tau, L)$  as the set of  $M \in \mathcal{C}_d(\tau)$  that fulfill the additional condition:

(iv) The inclusion map  $\iota_M : M \hookrightarrow \mathbb{R}^D$  is  $(\alpha + 1)$ -Hölder with  $|\iota_M|_{\alpha+1} \leq L$ .

**Remark 4.2.5.** The definitions above endow our model with *global* constraints, even though most of them can be stated in a *local* fashion, with properties of the support holding in a neighbourhood of our candidate point  $x \in \mathbb{R}^D$ . This meets two expectations:

- staying close to the existing manifold setting in statistics, like in [AL19];
- allowing for further developments, like estimation in global losses, such as  $L^p$ -norms, Wasserstein norms, or the sup-norm [WW20].

The reach condition  $\tau_M \geq \tau > 0$  in (iii) is essential in estimating consistently a density at a point in our setting, as shown in Theorem 4.2.10 in Section 4.2.5. Furthermore, a reach constraint enables one to benefit from several interesting geometric properties.

**Proposition 4.2.6.** *Let  $M$  be a compact smooth submanifold of  $\mathbb{R}^D$  with  $\tau_M \geq \tau$ . Then the injectivity radius  $\text{inj}_M$  is everywhere greater than  $\pi\tau$ .*

This result is a corollary of [AB06, Thm 1.3], as explained in [AL19, Lem A.1]. Pick  $M \in \mathcal{C}_d(\tau)$ . For any  $x \in M$ , the map  $v \mapsto \exp_x(v) - x$  is bounded above by  $\pi\tau$  on  $B_{T_x M}(0, \pi\tau)$ , since for any  $v \in B_{T_x M}(0, \pi\tau)$ , we have  $\|\exp_x(v) - x\| \leq d_M(\exp_x(v), x) = \|v\|$ , where  $d_M$  is the intrinsic distance on  $M$ . This uniform bound along with the Hölder condition (iv) allows one to obtain a uniform bound on the first derivatives of the exponential map.

**Lemma 4.2.7.** *For  $M \in \mathcal{C}_{d,\alpha}(\tau, L)$ , any  $x \in M$ , and any  $1 \leq j \leq \lceil \alpha \rceil$ , we have*

$$\sup_{v \in B_{T_x M}(0, \pi\tau/2)} \|d^j \exp_x(v)\|_{\text{op}} \leq L_j,$$

with  $L_j$  depending on  $d, \tau, L$  and  $\alpha$  only.

See Lemma 4.A.6 in the appendix for further details on the proof. In the light of this result, the model of Definition 4.2.4 is thus quite close to the one proposed by [AL19].

**Remark 4.2.8.** There is no obvious equivalence between  $M$  having a reach greater than  $\tau$ , and  $M$  being Hölder, in particular because the reach is a global quantity, while the Hölder smoothness property, as defined in Section 4.2.2, is a local feature. However, having a reach greater than  $\tau$  implies that  $\|d^2 \exp_x(0)\| \leq 2/\tau$ , in light of Proposition 4.A.1. For this reason, we always have that the submanifold  $M$  is at least 2-Hölder.

We are ready to define the class of density functions that we study, built upon submanifolds in the class  $\mathcal{C}_d(\tau)$ .

**Definition 4.2.9.** Let  $1 \leq d \leq D - 1$ ,  $\alpha \geq 1$ ,  $\beta > 0$ ,  $\tau > 0$ ,  $L > 0$ ,  $R > 0$  and  $0 \leq f_{\min} < f_{\max}$ . We define  $\Sigma_{\alpha,\beta}^d(\tau, L, f_{\min}, f_{\max}, R)$ , or  $\Sigma_{\alpha,\beta}^d$  for short, as the set of probability measures  $P$  on  $\mathbb{R}^D$  (endowed with its Borel  $\sigma$ -field) such that

- (i) There exists  $M_P \in \mathcal{C}_{d,\alpha}(\tau, L)$  such that  $\text{supp } P = M_P$ ;
- (ii) There exists a version of the Radon-Nikodym derivative  $\frac{dP}{d\mu_{M_P}}$ , denoted by  $f_P$ , that belongs to  $\mathcal{H}^\beta(M_P, \mathbb{R})$ ;
- (iii) This version satisfies  $f_{\min} \leq f_P \leq f_{\max}$  and  $|f_P|_\beta \leq R$ .

Some remarks: **1)** The support of any  $P \in \Sigma_{\alpha,\beta}^d(\tau, L, f_{\min}, f_{\max}, R)$  contains the candidate point  $x_0$  by construction, see Definition 4.2.4 where  $\mathcal{C}_{d,\alpha}(\tau, L)$  is defined. **2)** Condition (i) discards the possibility that  $f_P$  is zero on non-null subset of  $M$ ; in particular  $f_P$  is non zero around  $x_0$  (but can be zero at  $x_0$  nonetheless). This ensures that  $x_0$  does not lie too far from the data. An alternate definition is to impose a condition like  $P \ll \mu_M$ . This leads to the same results in the next sections, but with a slight ambiguity in the choice of  $M$ . **3)** The parameters in subscript or superscript  $(d, \alpha, \beta)$  control the rate of convergence of the estimation, while the parameters  $(\tau, L, f_{\min}, f_{\max}, R)$  control the pre-factor in the rates of convergence. For notational simplicity, we sometimes omit them when no confusion can be made.

#### 4.2.5 Choice of a loss function and the reach assumption

For  $P \in \Sigma_{\alpha,\beta}^d$  and a  $n$ -sample  $(X_1, \dots, X_n)$  drawn from  $P$ , our goal is to recover the value of  $f_P(x_0)$  thanks to an estimator  $\widehat{f}(x_0)$  built on top of the data  $(X_1, \dots, X_n)$ . We measure the accuracy of estimation by the maximal expected risk or order  $p$ , for  $p \geq 1$ , defined by

$$\sup_{P \in \Sigma_{\alpha,\beta}^d} \mathbb{E}_{P^{\otimes n}} [|\widehat{f}(x_0) - f_P(x_0)|^p]^{1/p}$$

We look for an estimator with the smallest possible maximal risk as the number of observations  $n$  goes to  $\infty$ . We first show that if we let  $\tau = 0$ , *i.e.* if we do not impose a reach condition, then it is impossible to estimate  $f_P(x_0)$  consistently as  $n \rightarrow \infty$  for any estimator, thus establishing that the reach assumption  $\tau > 0$  is unavoidable.

**Theorem 4.2.10.** In the setting of Definition 4.2.9, if we let  $\tau = 0$ , the following lower bound holds

$$\inf_{\widehat{f}(x)} \sup_{P \in \Sigma_{\alpha,\beta}^d} \mathbb{E}_{P^{\otimes n}} [|\widehat{f}(x_0) - f_P(x_0)|^p]^{1/p} \geq \frac{1}{2}(f_{\max} - f_{\min}) > 0,$$

where the infimum is taken over all estimators  $\widehat{f}(x_0)$  of  $f_P(x_0)$ .



The proof is given in Appendix 4.A.2. This result is in line with a reach condition  $\tau > 0$ , a customary necessary condition in a minimax reconstruction in geometric inference, when the manifold is unknown, see [NSW08, GPPVW12a, BRS<sup>+</sup>12, KRW16, AL19] and the references therein.

### 4.3 Density estimation at a fixed point

Recall that we fix a point  $x_0 \in \mathbb{R}^D$  where we wish to estimate  $f_P$ . Throughout the section, the symbols  $\lesssim$  and  $\gtrsim$  denote inequalities up to a constant that, unless specified otherwise, depends on the parameters  $d, \alpha, \beta, \tau, L, f_{\min}, f_{\max}, R$  and  $p$ . The expression *for  $n$  large enough* means for  $n$  bigger than a constant that depends on the same parameters.

#### 4.3.1 Main results

Let  $D \geq 2$ ,  $\tau > 0$ ,  $L > 0$ ,  $R > 0$ ,  $0 \leq f_{\min} < f_{\max}$  and  $p \geq 1$ . Recall that we write  $\Sigma_{\alpha, \beta}^d$  for short for  $\Sigma_{\alpha, \beta}^d(\tau, L, f_{\min}, f_{\max}, R)$  as defined in Definition 4.2.9. The main results of this section are the following.

**Theorem 4.3.1** (Upper bound). *For any  $1 \leq d \leq D - 1$ ,  $\alpha \geq 1$  and  $\beta > 0$ , there exists an estimator  $\widehat{f}(x_0)$  – explicitly constructed in Section 4.3.2 below – depending on  $\alpha, \beta$  and  $d$ , such that, for  $n$  large enough,*

$$\sup_{P \in \Sigma_{\alpha, \beta}^d} \mathbb{E}_{P^{\otimes n}} [|\widehat{f}(x_0) - f_P(x_0)|^p]^{1/p} \lesssim n^{-\alpha \wedge \beta / (2\alpha \wedge \beta + d)}.$$

The estimator of Theorem 4.3.1 is a kernel density estimator that depends on  $\alpha, \beta$  and  $d$  through the choice of the kernel and its order (in a certain sense specified below), together with its bandwidth. Its analysis is given in Section 4.3.2. The estimator is indeed optimal in a minimax sense, as soon as  $\alpha \geq \beta$ .

**Theorem 4.3.2** (Lower bound). *Let  $1 \leq d \leq D - 1$ ,  $\alpha \geq 1$  and  $\beta > 0$ . If  $L$  and  $f_{\max}$  are large enough and if  $f_{\min}$  is small enough (depending on  $\tau$ ), then*

$$\liminf_{n \rightarrow \infty} n^{\beta / (2\beta + d)} \inf_{\widehat{f}(x_0)} \sup_{P \in \Sigma_{\alpha, \beta}^d} \mathbb{E}_{P^{\otimes n}} [|\widehat{f}(x_0) - f_P(x_0)|^p]^{1/p} \geq C_* > 0$$

where  $C_*$  only depends on  $\tau$  and  $R$ .

See Appendix 4.A.3 for a proof. The rates from Theorem 4.3.1 and Theorem 4.3.2 agree, provided the underlying manifold  $M$  is regular enough, namely that  $\alpha \geq \beta$ . This probably covers most cases of interest in practice. However, when  $\alpha < \beta$  the question of optimality remains. We investigate in Section 4.3.3 below the simpler case  $d = 1$  and show that it is then possible to achieve the rate  $n^{-\beta / (2\beta + 1)}$ , at the extra cost of learning the geometry of  $M$  in a specific sense.



**Theorem 4.3.3** (One-dimensional case). *Let  $d = 1$  and  $\beta > 0$ . Assume that  $f_{\min} > 0$ . Then there exists an estimator  $\hat{f}^{1D}(x_0)$  – explicitly constructed in Section 4.3.3 below – depending on  $\beta$ , such that, for any  $\alpha \geq 1$  and for  $n$  large enough,*

$$\sup_{P \in \Sigma_{\alpha, \beta}^1} \mathbb{E}_{P^{\otimes n}} [|\hat{f}^{1D}(x_0) - f_P(x_0)|^p]^{1/p} \lesssim n^{-\beta/(2\beta+1)}.$$

The estimator described in Theorem 4.3.1 and Theorem 4.3.3 requires the specification of  $\alpha$ ,  $\beta$  and  $d$ , that are usually unknown in practice. We can circumvent this impediment by building an adaptative procedure with respect to these parameters. In Section 4.3.4 we adapt to the smoothness parameters  $\alpha$  and  $\beta$  by implementing Lepski’s method [Lep92]; in Section 4.3.5, we adapt to  $d$  by plugging-in a dimension estimator. We obtain the following result:

**Theorem 4.3.4** (Adaptation). *Let  $\ell \geq 0$ . Assume that  $f_{\min} > 0$ . Then, there exists an estimator  $\hat{f}^{\text{adapt}}(x_0)$  – explicitly constructed in Section 4.3.5 below – depending on  $\ell$  such that, for any  $\alpha, \beta$  in  $[0, \ell]$  and any  $1 \leq d \leq D - 1$ , we have, for  $n$  large enough,*

$$\sup_{P \in \Sigma_{\alpha, \beta}^d} \mathbb{E}_{P^{\otimes n}} [|\hat{f}^{\text{adapt}}(x_0) - f_P(x_0)|^p]^{1/p} \lesssim \left( \frac{\log n}{n} \right)^{\frac{\alpha \wedge \beta}{2\alpha \wedge \beta + d}}.$$

We were unable to obtain oracle inequalities in the spirit of the Goldenshluger-Lepski method, see [GL08, GL11, GL14], due to the non-Euclidean character of the support of  $f_P$ : our route goes along the more classical approach of [LMS97]. Obtaining oracle inequalities in this framework remain an open problem.

### 4.3.2 Kernel estimation

Classical nonparametric density estimation methods are based on kernel smoothing [Par62, Sil86]. In this section, we combine kernel density estimation with the minimal geometric features needed in order to recover efficiently their density. Since the intrinsic dimension  $d$  is not prone to change in this section, we further drop  $d$  in (most of) the notation. The proofs of this section can be found in Appendix 4.A.4.

Let  $K : \mathbb{R}^D \rightarrow \mathbb{R}$  be a smooth function vanishing outside the unit ball  $B(0, 1)$ . Given an  $n$ -sample  $(X_1, \dots, X_n)$  drawn from a distribution  $P$  on  $\mathbb{R}^D$ , we are interested in the behaviour of the kernel estimator

$$\hat{f}_h(x_0) = \frac{1}{nh^d} \sum_{j=1}^n K\left(\frac{X_j - x_0}{h}\right), \quad h > 0. \quad (4.4)$$

Note that the normalisation here is  $h^d$  and not  $h^D$  as one would set for a classical kernel estimator in  $\mathbb{R}^D$ . Our main result is that  $\hat{f}_h(x_0)$  behaves well when  $P$  is supported on a

$d$ -dimensional submanifold of  $\mathbb{R}^D$ .

We need some notation. For  $P \in \Sigma_{\alpha,\beta}^d$ , define

$$f_h(P, x_0) = \mathbb{E}_{P^{\otimes n}}[\widehat{f}_h(x_0)],$$

$$\mathcal{B}_h(P, x_0) = f_h(P, x_0) - f_P(x_0),$$

and

$$\widehat{\xi}_h(P, x_0) = \widehat{f}_h(x_0) - f_h(P, x_0),$$

that correspond respectively to the mean, bias and stochastic deviation of the estimator  $\widehat{f}_h(x_0)$ . We also introduce the quantity

$$\Omega(h) = \sqrt{\frac{2\omega}{nh^d}} + \frac{\|K\|_\infty}{nh^d} \quad \text{with} \quad \omega = 4^d \zeta_d \|K\|_\infty^2 f_{\max},$$

where  $\zeta_d$  is the volume of the unit ball in  $\mathbb{R}^d$ . The quantity  $\Omega(h)$  will prove to be a good majorant of the stochastic deviations of  $\widehat{f}_h(x_0)$ . The usual bias-stochastic decomposition of  $\widehat{f}_h(x)$  leads to

$$\mathbb{E}_{P^{\otimes n}}[|\widehat{f}_h(x_0) - f_P(x_0)|^p]^{1/p} \leq |\mathcal{B}_h(P, x_0)| + \left(\mathbb{E}_{P^{\otimes n}}[|\widehat{\xi}_h(P, x_0)|^p]\right)^{1/p}. \quad (4.5)$$

We study each term separately. The stochastic term can readily be bounded from above.

**Proposition 4.3.5.** *Let  $p \geq 1$ . There exists a constant  $c_p > 0$  depending on  $p$  only such that for any  $P \in \Sigma_{\alpha,\beta}^d$  and any  $h < \tau/2$ :*

$$\left(\mathbb{E}_{P^{\otimes n}}[|\widehat{\xi}_h(P, x_0)|^p]\right)^{1/p} \leq c_p \Omega(h).$$

Now we turn to the bias term. We need certain properties for the kernel  $K$ . More precisely, we assume that

**Assumption 4.3.6.** *The kernel  $K$  verifies*

- (i)  $K$  is smooth and supported on the unit ball  $B(0, 1)$ ;
- (ii) For any  $d$ -dimensional subspace  $H$  of  $\mathbb{R}^D$ , we have  $\int_H K(v) dv = 1$ .

One way to obtain Assumption 4.3.6 is to set  $\Lambda(x) = \exp(-1/(1 - \|x\|^2))$  for  $x \in B(0, 1)$  and  $\Lambda(x) = 0$  otherwise. Since  $\Lambda$  is rotationally invariant, its integral is the same over any  $d$ -subspace  $H$  of  $\mathbb{R}^D$ . Thus, with  $\lambda_d = \int_{H_0} \Lambda(v) dv$  where  $H_0 = \mathbb{R}^d \times \{0_{\mathbb{R}^{D-d}}\}$ , the function  $K(x) = \lambda_d^{-1} \Lambda(x)$  is a smooth kernel, supported on the unit ball of the ambient space  $\mathbb{R}^D$  that satisfies Assumption 4.3.6. In the following, we pick an arbitrary kernel  $K$  such that Assumption 4.3.6 is satisfied.

**Lemma 4.3.7.** For  $P \in \Sigma_{\alpha,\beta}^d$  and any  $h < \tau/2$ , setting  $k = \lceil \alpha \wedge \beta - 1 \rceil$ , we have

$$f_h(P, x_0) = f(x_0) + \sum_{j=1}^k h^j G_j(P, x_0) + R_h(P, x_0), \quad (4.6)$$

with  $|G_j(P, x_0)| \lesssim 1$  and  $|R_h(P, x_0)| \lesssim h^{\alpha \wedge \beta}$ .

The existence of such an expansion allows, by carefully choosing the kernel, to cancel the intermediate terms. Starting from a kernel  $K$  satisfying Assumption 4.3.6, we recursively define a sequence of smooth kernels  $(K^{(d,\ell)})_{\ell \geq 1}$ , simply denoted by  $K^{(\ell)}$  in this section, with support in  $B(0, 1)$  as follows (see Figure 4.3.1). For  $z \in \mathbb{R}^D$ , we put

$$\begin{cases} K^{(1)}(z) = K(z) \\ K^{(\ell+1)}(z) = 2^{1+d/\ell} K^{(\ell)}(2^{1/\ell} z) - K^{(\ell)}(z) \quad \forall \ell \geq 1. \end{cases} \quad (4.7)$$

A few remarks can be made: **1)** In a classical kernel density estimation framework, the integer  $\ell - 1$  plays the role of the order of the kernel. **2)** The assumption that  $K$  is compactly supported is seemingly quite strong. This is a way to make sure that the support of  $x \mapsto K((x - x_0)/h)$  is within the injectivity ball of the map  $\exp_{x_0}$  for any  $h < \pi\tau$ . **3)** The construction of  $K$  is simply an example of a Richardson's extrapolation as coined by [Ric11]. **4)** This construction somewhat differs from the classical constructions than can be found in textbooks such as [Tsy08]. There is one practical reason: we require that all the kernels satisfy Assumption 4.3.6; another reason that appears to be more intrinsically related to our model: since the Euclidean distance is only a second order approximation of the Riemannian distance on  $M$ , defining a kernel through orthogonality relations with respect to a family of polynomials is not sufficient in our framework.

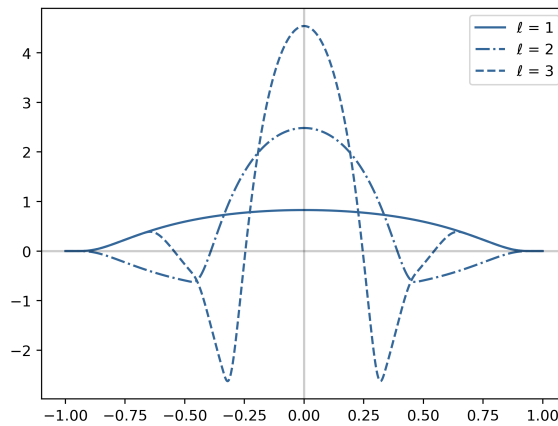


Figure 4.3.1 – Plots of the kernel  $K^{(d,\ell)}$  for  $d = 1$  and  $\ell = 1, 2, 3$ .

**Proposition 4.3.8.** Let  $\ell$  be an integer greater than  $\alpha$  and  $\beta$ , and let  $K^{(\ell)}$  be the kernel defined in (4.7) starting from a kernel  $K$  satisfying Assumption 4.3.6. Then, for any

$P \in \Sigma_{\alpha,\beta}^d$  and any bandwidth  $h < \tau/2$ , the estimator  $\widehat{f}_h(x_0)$  defined as in (4.4) using  $K^{(\ell)}$  is such that

$$|\mathcal{B}_h(P, x_0)| \lesssim h^{\alpha \wedge \beta}. \quad (4.8)$$

Patching together Proposition 4.3.5 and Proposition 4.3.8 yields

$$\sup_{P \in \Sigma_{\alpha,\beta}^d} \mathbb{E}_{P^{\otimes n}} [|\widehat{f}_h(x_0) - f_P(x_0)|^p]^{1/p} \lesssim \Omega(h) + h^{\alpha \wedge \beta},$$

for any  $h < \tau/2$ . Therefore the estimator  $\widehat{f}(x_0) = \widehat{f}_h(x_0)$  specified with  $h = n^{-1/(2\alpha \wedge \beta + d)}$  indeed satisfies the conclusion of Theorem 4.3.1, for  $n$  large enough.

### 4.3.3 The special case of one-dimensional submanifolds

The gap we observe between the two rates in Theorem 4.3.1 and Theorem 4.3.2 leads to the following question: does the regularity  $\alpha \geq 1$  of  $M$  have a genuine limiting effect in the estimation of  $f_P(x_0)$ , or does it rather reveal a weakness of the estimator described in Section 4.3.2? We do not have a definitive answer to this question except for  $d = 1$  i.e. when  $M$  is a closed curve in an Euclidean space. We can then show that the parameter  $\alpha$  does not interfere at all with the density estimation. The proofs of this section can be found in Appendix 4.A.5.

If  $d = 1$ , any submanifold  $M$  in  $\mathcal{C}_1(\tau)$  is a closed smooth injective curve that can be parametrized by a unit-speed path  $\gamma_M : [0, L_M] \rightarrow \mathbb{R}^D$  with  $\gamma_M(0) = \gamma_M(L_M)$  and with  $L_M = \text{vol } M$  being the length of the curve. In that case, the volume density function is trivial.

**Lemma 4.3.9.** *For  $M \in \mathcal{C}_1(\tau)$ , for any  $x \in M$  and any  $v \in T_x M$ , we have  $\det g^x(v) = 1$ .*

Thanks to Lemma 4.3.9, the estimator proposed by [Pel05] takes a simpler form, which we will try to take advantage of. Indeed, in the representation (4.2) of [Pel05], only  $d_M$  remains unknown. We now show how to efficiently estimate  $d_M$  thanks to the Isomap method as coined by [TSL00]. The analysis of this algorithm essentially comes from [BDSLT00] and is pursued in [ACLG19], but the bounds obtained there are manifold dependent. We thus propose a slight modification of their proofs in order to obtain uniform controls over  $\mathcal{C}_1(\tau)$ , and make use of the simplifications coming from the dimension 1. Indeed, for  $d = 1$ , we have the following simple and explicit formula for the intrinsic distance on  $M$ :

$$d_M(\gamma_M(s), \gamma_M(t)) = |t - s| \wedge (L_M - |t - s|) \quad \forall s, t \in [0, L_M].$$

The Isomap method can be described as follows: let  $\varepsilon > 0$ , and let  $\mathcal{G}_\varepsilon$  be the  $\varepsilon$ -neighbourhood graph built upon the data  $(X_1, \dots, X_n)$  and  $x_0$  — namely,  $\mathcal{G}_\varepsilon = (V, E)$

where  $V = (x_0, X_1, \dots, X_n)$ , and where  $E = \{(y, z) \in V \mid \|y - z\| \leq \varepsilon\}$ . For a path in  $\mathcal{G}_\varepsilon$  (meaning: a sequence of adjacent vertices)  $s = (p_0, \dots, p_m)$ , we define its length as  $L_s = \|p_1 - p_0\| + \dots + \|p_m - p_{m-1}\|$ . The distance between  $x$  and a vertex  $y$  in the graph  $\mathcal{G}_\varepsilon$  is then defined as

$$\widehat{d}_\varepsilon(x, y) = \min \{L_s \mid s \text{ path in } \mathcal{G}_\varepsilon \text{ connecting } x \text{ to } y\}, \quad (4.9)$$

and we set this distance to  $\infty$  if  $x$  and  $y$  are not connected. We are now ready to describe our estimators  $\hat{f}^{1D}(x_0)$ . For any  $h, \varepsilon > 0$ , we set

$$\hat{f}_h^{1D}(x_0) = \frac{1}{nh} \sum_{i=1}^n K^{1D}(\widehat{d}_\varepsilon(x_0, X_i)/h), \quad (4.10)$$

for some kernel  $K^{1D} : \mathbb{R} \rightarrow \mathbb{R}$ . Notice that the kernel  $K^{(1,\ell)}(\cdot) : \mathbb{R}^D \rightarrow \mathbb{R}$  defined in Section 4.3.2 starting from kernel  $K = \lambda_1^{-1} \Lambda$  can be put in the form  $K^{(1,\ell)}(x) = K^{(1,\ell)}(\|x\|)$  with  $K^{(1,\ell)}(\cdot)$  denoting thus (with a slight abuse of notation) both functions starting from either  $\mathbb{R}$  or  $\mathbb{R}^D$ . We choose this kernel in the next statement.

**Proposition 4.3.10.** *Assume that  $f_{\min} > 0$ . The estimator defined in (4.10) above and specified with  $K^{1D} = K^{(1,\ell)}(\cdot)$  satisfies the following property: for any  $\beta \in [0, \ell]$  and any  $\alpha \geq 1$ , we have*

$$\sup_{P \in \Sigma_{\alpha,\beta}^1} \mathbb{E}_{P^{\otimes n}} [|\hat{f}_h^{1D}(x_0) - f_P(x_0)|^p]^{1/p} \lesssim \frac{\varepsilon^2}{h^2} + \Omega(h) + h^\beta + \frac{1}{nh},$$

with

$$\varepsilon = \frac{32(p+1)}{f_{\min}} \times \frac{\log n}{n},$$

for  $h < \tau/4$  and  $n$  large enough.

The proof of Theorem 4.3.3 readily follows from Proposition 4.3.10 using the estimator  $\hat{f}^{1D} = \hat{f}_h^{1D}$  with  $h = n^{-1/(2\beta+1)}$ .

#### 4.3.4 Smoothness adaptation

We implement Lepski's algorithm, following closely [LMS97] in order to automatically select the bandwidth from the data  $(X_1, \dots, X_n)$ . We know from Section 4.3.2 that the optimal bandwidth on  $\Sigma_{\alpha,\beta}^d$  is of the form  $n^{-1/(2\alpha\wedge\beta)+d}$ . Hence, without prior knowledge of the value of  $\alpha$  and  $\beta$ , we can restrict our search for a bandwidth in a bounded interval of the form  $[h^-, 1]$  discretized as follow

$$\mathbb{H} = \{2^{-j}, \text{ for } 0 \leq j \leq \log_2(1/h^-)\}$$

We pick

$$h^- = \left( \frac{\|K\|_\infty}{2\omega} \right)^{1/d} \frac{1}{n^{1/d}};$$

this bandwidth is always smaller than the optimal bandwidth  $n^{-1/(2\alpha\wedge\beta+d)}$  on  $\Sigma_{\alpha,\beta}^d$  for  $n$  large, and is such that  $\Omega(h) \leq 2\sqrt{2\omega/(nh^d)}$  for all  $h \geq h^-$ . For  $h, \eta \in \mathbb{H}$ , we introduce the following quantities:

$$\begin{aligned}\lambda(h) &= 1 \vee \sqrt{\Theta d \log(1/h)}, \\ \psi(h, \eta) &= \Omega(h)\lambda(h) + \Omega(\eta)\lambda(\eta)\end{aligned}\tag{4.11}$$

where  $\Theta$  is a positive constant (to be specified). For  $h \in \mathbb{H}$  we define the subset of bandwidths  $\mathbb{H}(h) = \{\eta \in \mathbb{H}, \eta \leq h\}$ . The selection rule for  $h$  is the following:

$$\widehat{h}(x_0) = \max \left\{ h \in \mathbb{H} \mid \forall \eta \in \mathbb{H}(h), |\widehat{f}_h(x_0) - \widehat{f}_\eta(x_0)| \leq \psi(h, \eta) \right\},$$

and we finally consider the estimator

$$\widehat{f}(x_0) = \widehat{f}_{\widehat{h}(x_0)}(x_0),\tag{4.12}$$

where  $\widehat{f}_h(x_0)$  is defined at (4.4).

**Proposition 4.3.11.** *Assume  $\Theta > p$ . Let  $\ell \in \mathbb{N}$ , and let  $\widehat{f}(x_0)$  be the estimator defined in (4.12) using  $K^{(\ell)}$  originated from a kernel  $K$  satisfying Assumption 4.3.6. Then, for any  $\alpha, \beta \leq \ell$ , we have, for  $n$  large enough*

$$\sup_{P \in \Sigma_{\alpha,\beta}^d} \mathbb{E}_{P^{\otimes n}} [|\widehat{f}(x_0) - f_P(x_0)|^p]^{1/p} \lesssim \left( \frac{\log n}{n} \right)^{\alpha\wedge\beta/(2\alpha\wedge\beta+d)}.$$

The proof of Proposition 4.3.11 can be found in Appendix 4.A.6. Some remarks: **1)** Proposition 4.3.11 provides us with a classical smoothness adaptation result in the spirit of [LMS97]: the estimator  $\widehat{f}$  has the same performance as the estimator  $\widehat{f}_h$  selected with the optimal bandwidth  $n^{-1/(2\alpha\wedge\beta+d)}$ , up to a logarithmic factor on each model  $\Sigma_{\alpha,\beta}^d$  without the prior knowledge of  $\alpha \wedge \beta$  over the range  $[0, \ell]$ . **2)** The extra logarithmic term is the unavoidable payment for the Lepski-Low phenomenon [Lep90, Low92] when recovering a function in pointwise or in a uniform loss.

### 4.3.5 Simultaneous adaptation to smoothness and dimension

The estimators considered in Theorem 4.3.1 or Proposition 4.3.11 heavily rely on the intrinsic dimension  $d$  through the choice of a kernel satisfying Assumption 4.3.6, through the normalisation  $h^d$  and either through the choice of an optimal bandwidth  $h$ , or the selection procedure (4.11)-(4.12). We now show how to adapt to  $d$  considered as an unknown and nuisance parameter. The proofs of this section can be found in Appendix 4.A.7.

We redefine all the quantities introduced before as now depending on  $d$ . Namely, for  $h, \eta > 0$ , and a given family of kernel  $K(d; \cdot)$ , we set

$$\begin{aligned}\widehat{f}_h(d; x_0) &= \frac{1}{nh^d} \sum_{i=1}^n K(d; (X_i - x_0)/h), \\ \Omega(d; h) &= \sqrt{\frac{2\omega_d}{nh^d}} + \frac{\|K(d; \cdot)\|_\infty}{nh^d} \quad \text{with } \omega_d = 4^d \zeta_d \|K(d; \cdot)\|_\infty^2 f_{\max}, \\ \lambda(d; h) &= 1 \vee \sqrt{\Theta d \log(1/h)}, \\ \psi(d; h, \eta) &= v(d; h)\lambda(d; h) + v(d; \eta)\lambda(d; \eta), \\ h_d^- &= (\|K(d; \cdot)\|_\infty / 2\omega_d)^{1/d} n^{1/d}, \\ \mathbb{H}_d &= \{2^{-j}, \text{ for } 0 \leq j \leq \log_2(1/h_d^-)\},\end{aligned}$$

where  $\Theta$  is a constant. We also define

$$\widehat{h}(d; x_0) = \max \{h \in \mathbb{H}_d \mid \forall \eta \in \mathbb{H}_d(h), |\widehat{f}_h(d; x_0) - \widehat{f}_\eta(d; x_0)| \leq \psi(d; h, \eta)\} \quad (4.13)$$

with  $\mathbb{H}_d(h) = \{\eta \in \mathbb{H}_d, \eta \leq h\}$ . We are now left with the choice of kernel family  $K(d; \cdot)$ . For any  $1 \leq d \leq D-1$  and  $h > 0$ , we define

$$K^{(1)}(d; x) = \lambda_d^{-1} \Lambda(x)$$

where  $\Lambda$  and  $\lambda_d$  have been introduced in Section 4.3.2. We then pick an integer  $\ell \in \mathbb{N}$  and choose

$$K(d; \cdot) = K^{(d, \ell)}(d; \cdot) \quad (4.14)$$

where  $K^{(d, \ell)}(d; \cdot)$  is defined by recursion in (4.7) starting from the kernel  $K^{(1)}(d; \cdot)$ .

We assume that we have an estimator  $\widehat{d}$  of the dimension  $d$  of  $M$  with values in  $\{1, \dots, D\}$ . More precisely, we need the following property:

**Assumption 4.3.12.** *For any  $1 \leq d < D$  and all real numbers  $\alpha, \beta \in [0, \ell]$ , we have*

$$\sup_{P \in \Sigma_{\alpha, \beta}^d} P^{\otimes n}(\widehat{d} \neq d) \lesssim n^{-3p/2},$$

where  $p$  is the exponent of the loss function.

If we are given such a estimator of the dimension  $d$ , then we can built a estimator that adapts to this parameter.

**Proposition 4.3.13.** *Let  $\widehat{f}(x_0) = \widehat{f}_{\widehat{h}}(\widehat{d}, x_0)$  built with the kernel family (4.14), where  $\widehat{d}$  is a estimator satisfying Assumption 4.3.12 and where  $\widehat{h} = \widehat{h}(\widehat{d}, x_0)$  is defined at (4.13). Then, for any  $1 \leq d \leq D-1$ , and any  $\alpha, \beta \leq \ell$ , we have, for  $n$  large enough*

$$\sup_{P \in \Sigma_{\alpha, \beta}^d} \mathbb{E}_{P^{\otimes n}}[|\widehat{f}(x_0) - f_P(x_0)|^p]^{1/p} \lesssim \left(\frac{\log n}{n}\right)^{\alpha \wedge \beta / (2\alpha \wedge \beta + d)}.$$

It only remains to show that there exists an estimator  $\widehat{d}$  satisfying Assumption 4.3.12 to obtain Theorem 4.3.4.

There are various way to define such an estimator, see [FSA07] or even [KRW16] where an estimator with super-exponential minimax rate on a wide class of probability measures is constructed. For sake of completeness and simplicity, we will mildly adapt the work of [FSA07] to our setting. The resulting estimator will behave well as soon as we add the assumption that  $f_{\min} > 0$ .

**Definition 4.3.14.** *For a probability measure  $P$ , we write  $P_\eta = P(B(x_0, \eta))$  for any  $\eta > 0$ , and  $\widehat{P}_\eta = \widehat{P}_n(B(x_0, \eta))$  where  $\widehat{P}_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$  denotes the empirical measure of the sample  $(X_1, \dots, X_n)$ . Define*

$$\widehat{\delta}_\eta = \log_2 \widehat{P}_{2\eta} - \log_2 \widehat{P}_\eta,$$

and set  $\widehat{\delta}_\eta = D$  when  $\widehat{P}_\eta = 0$ . We define  $\widehat{d}_\eta$  to be the closest integer of  $\{1, \dots, D\}$  to  $\widehat{\delta}_\eta$ , namely  $\widehat{d}_\eta = \lfloor \widehat{\delta}_\eta + 1/2 \rfloor$ .

**Proposition 4.3.15.** *Assume that  $f_{\min} > 0$ . Then, for any  $1 \leq d \leq D - 1$ , and any  $\alpha \geq 1$ ,  $\beta > 0$ , the estimator  $\widehat{d} = \widehat{d}_\eta$  for  $\eta = n^{-1/(2D+2)}$  verifies for  $n$  large enough*

$$\sup_{P \in \Sigma_{\alpha, \beta}^d} P^{\otimes n}(\widehat{d} \neq d) \leq 4 \exp\left(-2n^{1-(d+1)/(D+1)}\right).$$

## 4.4 Numerical illustration

In this section we propose a few simulations to illustrate the results presented above. The goal is two-fold

- To highlight the rate obtained in Theorem 4.3.1 using estimator  $\widehat{f}_h(x_0)$ , in the case where  $\beta \leq \alpha$ , on arbitrary submanifold and for a carefully chosen bandwidth  $h$ ;
- To show the computational feasibility and performance of estimator  $\widehat{f}^{\text{adapt}}(x_0)$  described in Section 4.3.4.

For the sake of visualisation and simplicity, we focus on two typical examples of submanifold of  $\mathbb{R}^D$ , namely non-isometric embeddings of the flat circle  $\mathbb{T}^1 = \mathbb{R}/\mathbb{Z}$  and of the flat torus  $\mathbb{T}^2 = \mathbb{T}^1 \times \mathbb{T}^1$ . In particular, these embeddings will be chosen in such way that their images, as submanifolds of  $\mathbb{R}^D$ , are not homogeneous compact Riemannian manifolds, so that the work of [KNP12] for instance cannot be of use here.

For a given embedding  $\Phi : N \rightarrow M \subset \mathbb{R}^D$  where  $N$  is either  $\mathbb{T}^1$  or  $\mathbb{T}^2$ , we construct absolutely continuous probabilities on  $M$  by pushing forward probability densities



of  $N$  w.r.t. their volume measure. Indeed, if  $Q = g \cdot \mu_N$ , the push-forward measure  $P = \Phi_* Q$  has density  $f$  with respect to  $\mu_M$  given by

$$\forall x \in M, \quad f(x) = \frac{g(\Phi^{-1}(x))}{|\det d\Phi(\Phi^{-1}(x))|} \quad (4.15)$$

where the determinant is taken in an orthonormal basis of  $T_{\Phi^{-1}(x)}N$  and  $T_x M$ , so that, if  $\Phi$  is chosen smooth enough,  $f$  has the same regularity as  $g$ . If  $\Phi$  is an embedding of  $\mathbb{T}^1$ , we simply have  $|\det d\Phi(y)| = \|\Phi'(y)\|$  for all  $y \in \mathbb{T}^1$ . If now  $\Phi$  maps  $\mathbb{T}^2$  to  $M$ , we have

$$\begin{aligned} |\det d\Phi(y)| &= \sqrt{\det \langle d\Phi(y)[e_i], d\Phi(y)[e_j] \rangle_{1 \leq i, j \leq 2}} \\ &= \sqrt{\|d\Phi(y)[e_1]\|^2 \|d\Phi(y)[e_2]\|^2 - \langle d\Phi(y)[e_1], d\Phi(y)[e_2] \rangle^2} \end{aligned} \quad (4.16)$$

where  $(e_1, e_2)$  is an orthonormal basis of  $\mathbb{R}^2 \simeq T_y \mathbb{T}^2$ .

Strictly speaking, the probability measures  $P$  exhibited below are not elements of the models  $\Sigma_{\alpha, \beta}^d$ , but we know that they locally coincide with some  $\tilde{P} \in \Sigma_{\alpha, \beta}^d$  around our candidate point  $x$ , meaning that

$$P|_{B(x, r)} = \tilde{P}|_{B(x, r)} \quad \text{for some } r > 0.$$

This ensures that all the results displayed in Section 4.3 hold for  $P$  — see Remark 4.2.5 for a discussion on the local character of our setting.

#### 4.4.1 Simulation on one-dimensional submanifolds

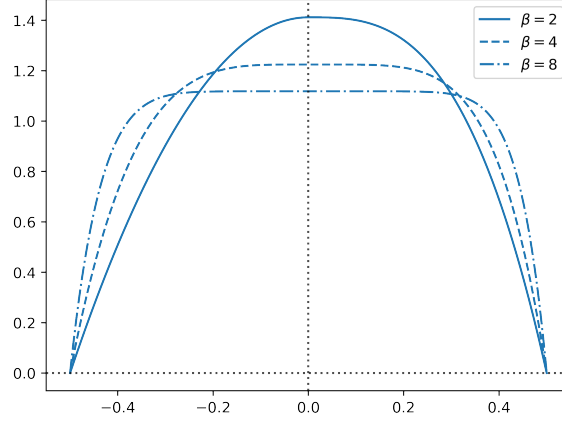
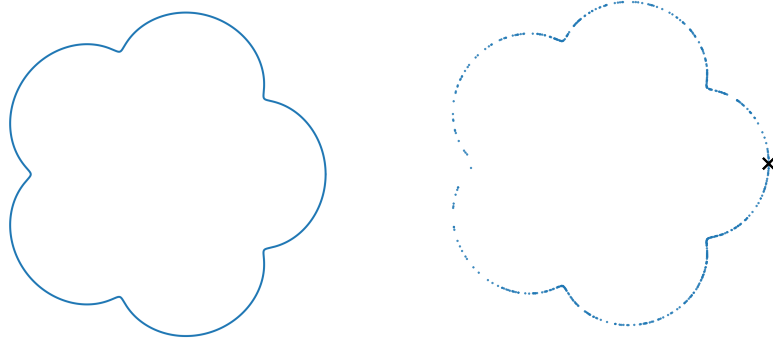
Let  $\beta \in \mathbb{N}^*$  and define the following function for  $v \in [-1/2, 1/2]$

$$g_\beta(v) = C_\beta \times (1 - (-2v)^\beta) \mathbb{1}_{[-1/2, 0)}(v) + C_\beta (1 - (2v)^{\beta+1}) \mathbb{1}_{[0, 1/2]}. \quad (4.17)$$

where  $C_\beta$  is an explicit normalisation constant. The function  $g_\beta$  is positive and  $\int_0^1 g_\beta(v) dt = 1$ ; it defines a probability density over  $[-1/2, 1/2]$ . Also, because the  $(\beta - 1)$ -th derivative of  $g_\beta$  is 1-Lipschitz, but its  $\beta$ -derivative is discontinuous at  $v = 0$ , the function  $g_\beta$  is  $\beta$ -Hölder but *not*  $(\beta + \varepsilon)$ -Hölder for any  $\varepsilon > 0$ . See Figure 4.4.1 for a few plots of the functions  $g_\beta$ . We next consider the parametric curve

$$\Phi : \begin{cases} \mathbb{T}^1 \rightarrow \mathbb{R}^2 \\ t \mapsto (\cos(2\pi v) + a \cos(2\pi \omega v), \sin(2\pi v) + a \sin(2\pi \omega v)) \end{cases}$$

Short computations show that  $\Phi$  is indeed an embedding as soon as  $a\omega < 1$ , in which case  $M = \Phi(\mathbb{T}^1)$  is indeed a smooth compact submanifold of  $\mathbb{R}^2$ . For the rest of this section, we set  $a = 1/8$  and  $\omega = 6$ . See Figure 4.4.2 for a plot of  $M$  with these parameters. We are interested in estimating the density  $f_\beta$  with respect to  $d\mu_M$  of the push-forward


 Figure 4.4.1 – Plots of the densities  $g_\beta$  for  $\beta = 2, 4, 8$ .

 Figure 4.4.2 – Plot of the submanifold  $M$  (Left) for parameters  $a = 1/8$  and  $\omega = 6$ , and 500 points sampled independently from  $\Phi_* g_\beta \cdot \mu_{\mathbb{T}^1}$  for  $\beta = 3$  (Right). The black cross denotes the point  $x = \Phi(0)$ .

measure  $P_\beta = \Phi_* g_\beta \cdot \mu_{\mathbb{T}^1}$ , at point  $x_0 = (1 + a, 0) \in M$ . We use formula (4.15) to compute  $f_\beta(x_0)$ : We have  $\Phi^{-1}(x_0) = 0$  and  $\|\Phi'(0)\| = 2\pi(1 + a\omega)$  hence

$$f_\beta(x_0) = \frac{C_\beta}{2\pi(1 + a\omega)} \text{ at } x_0 = (1 + a, 0).$$

Our aim here is to provide an empirical measure for the convergence of the risk  $n \mapsto \mathbb{E}_{P_\beta^{\otimes n}}[|\hat{f}_h(x_0) - f_\beta(x_0)|^p]^{1/p}$  when  $h$  is tuned optimally (in an oracle way). We pick  $p = 2$ . Our numerical procedure is detailed in Algorithm 1 below, and the numerical results are presented in Figure 4.4.5.

#### 4.4.2 Simulation on two-dimensional submanifolds

We consider a non-isometric embedding of the flat torus  $\mathbb{T}^2$ . We first construct a density function. For an integer  $\beta \geq 1$ , define

$$G_\beta : (v, u) \in [-1/2, 1/2]^2 \mapsto g_\beta(v)g_\beta(u) \quad (4.18)$$

---

**Algorithm 1** MSE rate of convergence estimation

---

- 1: Provide integers  $\beta \geq 1$  and  $\ell \geq \beta$ .
  - 2: Set a grid of increasing number of points  $\mathbf{n} = (n_1, \dots, n_k) \in \mathbb{N}^k$  and a number of repetition  $N$ .
  - 3: **for**  $n_i \in \mathbf{n}$  **do**
  - 4:   Sample  $n_i$  points independently from  $P_\beta$ ,
  - 5:   Compute  $\widehat{f}_{h_i}(x_0)$  with kernel  $K^{(\ell)}$  and bandwidth  $h_i = n_i^{-1/(2\beta+1)}$ ,
  - 6:   Compute the square error  $(\widehat{f}_{h_i}(x) - f_\beta(x))^2$ ,
  - 7:   Repeat the three previous steps  $N$  times,
  - 8:   Average the errors to get a Monte-Carlo approximation  $\widehat{R}(n_i)$  of  $\mathbb{E}_{P_\beta^{\otimes n_i}}[|\widehat{f}_h(x_0) - f_\beta(x_0)|^2]$ .
  - 9: **end for**
  - 10: Perform an Ordinary Least Square Linear Regression on the curve  $\log n_i \mapsto \log \widehat{R}(n_i)$ .
  - 11: **return** The coefficient of the linear regression.
- 

where  $g_\beta$  is defined as in (4.17). Obviously,  $G_\beta$  defines a density function on  $\mathbb{T}^2$  that is  $\beta$ -Hölder (but not  $(\beta + \varepsilon)$ -Hölder for any  $\varepsilon > 0$ ).

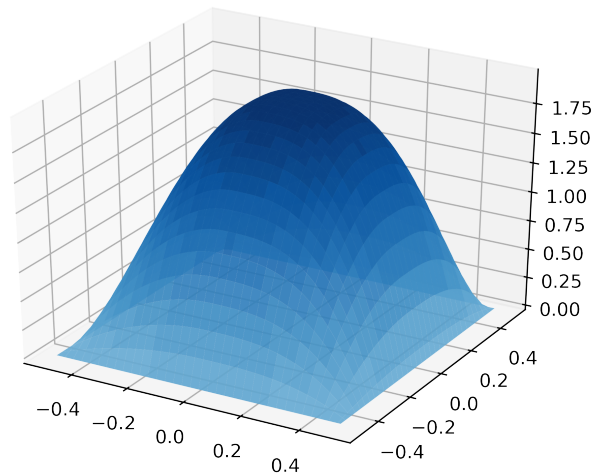


Figure 4.4.3 – Plot of the probability density function  $G_\beta$  for  $\beta = 3$ .

We next consider the parametric surface

$$\Psi : \begin{cases} \mathbb{T}^2 \rightarrow \mathbb{R}^3 \\ (v, u) \mapsto \begin{pmatrix} (b + \cos(2\pi v)) \cos(2\pi u) + a \sin(2\pi \omega v) \\ (b + \cos(2\pi v)) \sin(2\pi u) + a \cos(2\pi \omega v) \\ \sin(2\pi v) + a \sin(2\pi \omega u) \end{pmatrix} \end{cases}$$

for some  $a, b, \omega \in \mathbb{R}$ . In the remaining of the section, we set  $a = 1/8$ ,  $b = 3$  and  $\omega = 5$ . We show that  $\Phi$  indeed defines an embedding. See Figure 4.4.4 for a plot of the submanifold  $M = \Psi(\mathbb{T}^2)$ . For an integer  $\beta \geq 1$ , we denote by  $F_\beta$  the density of the push forward measure  $Q_\beta = \Psi_* G_\beta \cdot \mu_{\mathbb{T}^2}$  with respect to the volume measure  $\mu_M$ . Let  $x = (b + 1, a, 0)$  be the image of 0 by  $\Psi$ . Simple calculations show that the differential of  $\Psi$  at 0 evaluated at

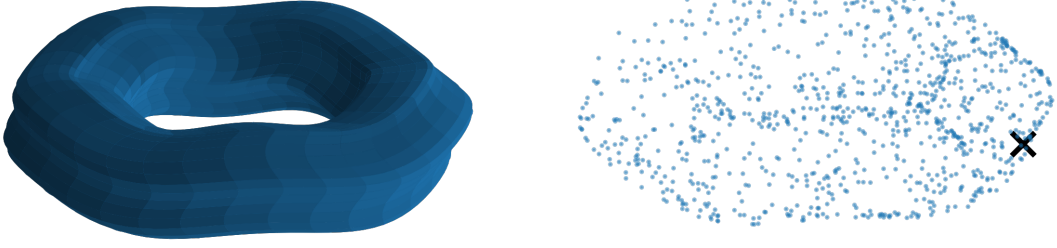


Figure 4.4.4 – Plot of the submanifold  $M$  (Left) for parameters  $a = 1/8$ ,  $b = 3$  and  $\omega = 6$ , and 500 points sampled independently from  $\Psi_* G_\beta \cdot \mu_{\mathbb{T}^2}$  with  $\beta = 3$  (Right). The black cross marks the point  $x$ .

$e_1 = (1, 0) \in T_0 \mathbb{T}^2$  and  $e_2 = (0, 1) \in T_0 \mathbb{T}^2$  is equal to respectively  $d\Psi(0)[e_1] = 2\pi(a\omega, 0, 1)$  and  $d\Psi(0)[e_2] = 2\pi(0, b + 1, a\omega)$ . Hence formula (4.16) yields

$$\det d\Psi(0) = (2\pi)^2 \left( (1 + a^2\omega^2) \left( (b + 1)^2 + a^2\omega^2 \right) - a^2\omega^2 \right)^{1/2} \quad (4.19)$$

and we obtain

$$F_\beta(x_0) = \frac{C_\beta^2}{4\pi^2} \left( (1 + a^2\omega^2) \left( (b + 1)^2 + a^2\omega^2 \right) - a^2\omega^2 \right)^{-\frac{1}{2}}.$$

In the same way as in the previous section, we aim at providing an empirical measure for the rate of convergence of the risk  $\mathbb{E}_{P_\beta^{\otimes n}}[|\hat{f}_h(x_0) - f_\beta(x_0)|^2]$  when  $h$  is suitably tuned with respect to  $n$  and  $\beta$ . This is done using again Algorithm 1. The results are presented in Figure 4.4.5.

### 4.4.3 Adaptation

In this section we estimate a density when its regularity is unknown, contrary to the previous simulation where the regularity parameter  $\beta$  is pugged in the bandwidth choice  $n^{-1/(2\beta+d)}$ . This is performed using Lepski's method presented in Section 4.3.4. The rate is computed using Algorithm 1, for both the one-dimensional and the two-dimensional synthetic datasets.

For the adaptive estimation on the two-dimensional manifold, we observe that the corrective term  $\det d\Psi(0)$  computed in (4.19) results in a density  $F_\beta(x_0)$  that is quite small, while the function  $\psi$  defined at (4.11) and used to tune the bandwidth soars dramatically because of the retained value of  $\omega_d = 4^d \zeta_d \|K^{(d,\ell)}\|_\infty^2 f_{\max}$ , so that the values

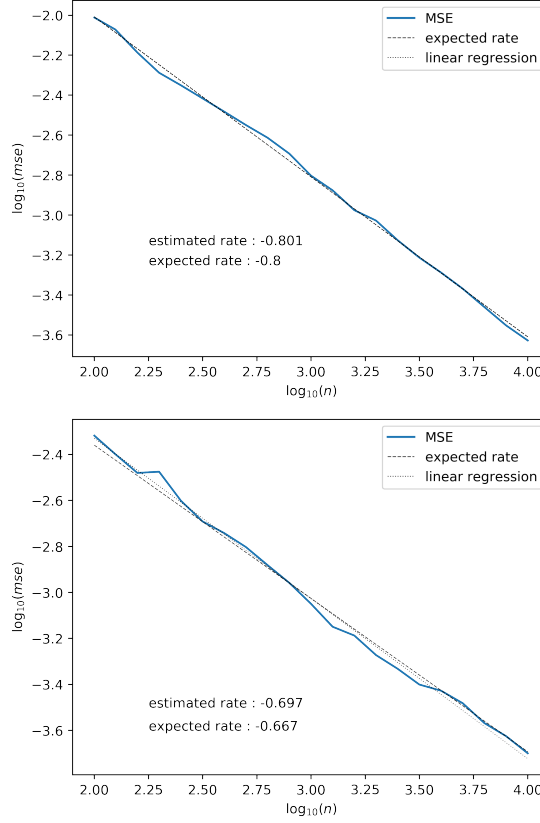


Figure 4.4.5 – Plot of the empirical mean square error (blue) for a density supported by a one-dimensional submanifold (Left) and two-dimensional submanifold (Right) with parameter  $\beta = 2$ . We use a log-regular grid  $\mathbf{n}$  of 21 points ranging from 100 to  $10^4$ . Each experiment is repeated  $N = 500$  times.

of  $\hat{f}_h(x_0)$  and  $\psi(h, \cdot)$  (defined at (4.11)) are not of the same order anymore at this scale (using maximum  $10^6$  observations). To circumvent this effect, we introduce a scaling parameter  $\lambda$  as follows

$$G_{\beta, \lambda} = v, u \mapsto \lambda^2 G_{\beta}(\lambda v, \lambda u).$$

Like before, we consider the push-forward probability measure  $\Psi_* G_{\beta, \lambda} \cdot \mu_{\mathbb{T}^2}$  which has density  $F_{\beta, \lambda}$  with respect to  $\mu_M$ . For  $\lambda = 4$ , we find that  $F_{\beta, \lambda}(x_0)$  is of order 1 for most values of  $\beta$ , and we use the function  $\psi^{\text{num}}(h, \eta) = \Omega^{\text{num}}(h)\lambda(h) + \Omega^{\text{num}}(\eta)\lambda(\eta)$  using simply  $\Omega^{\text{num}}(h) = \sqrt{1/nh^d}$ . We have no theoretical guarantee that such a method will work but we recover nonetheless the right rate in the estimation of the value of the density, see Figure 4.4.6 for a plot of the estimated rate.

We find an empirical error with a relatively high dispersion, hence our choice to represent the median of the squared error instead of the more traditional mean squared error.

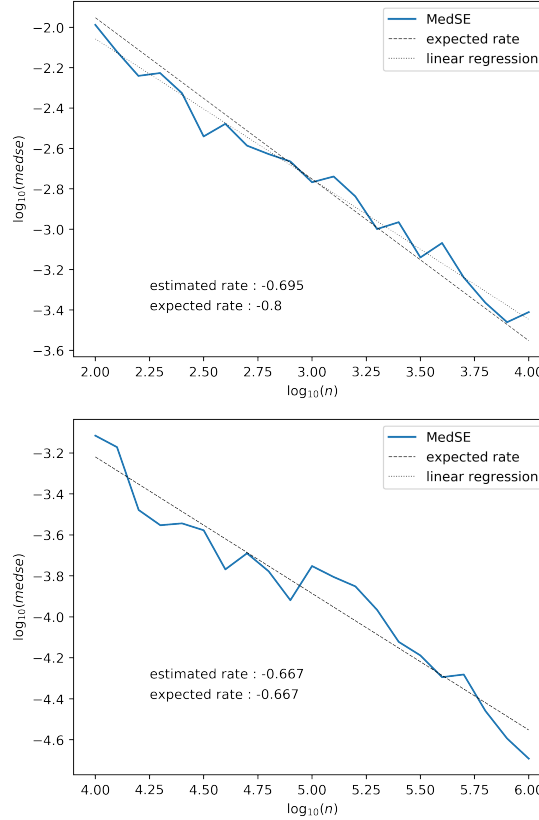


Figure 4.4.6 – (Left) Plot of the empirical median square error for the one-dimensional submanifold with  $\beta = 2$ . The bandwidth  $h$  is chosen adaptively using Lepski's method of order  $\ell = 3$  as in Section 4.3.4. We used a log-regular grid  $\mathbf{n}$  of 11 points ranging from 100 to  $10^4$  and each experience was repeated  $N = 500$  times. (Right) Same experiment but for the two dimensional manifold, with a grid ranging from  $10^4$  to  $10^6$  and  $N = 100$  repetitions.

# Appendix

## Contents

---

<b>4.A Appendix of Chapter 4 . . . . .</b>	<b>133</b>
4.A.1 Additional results of geometry . . . . .	133
4.A.2 Proof of Theorem 4.2.10 . . . . .	134
4.A.3 Proof of Theorem 4.3.2 . . . . .	136
4.A.4 Proofs of Section 4.3.2 . . . . .	137
4.A.5 Proofs of Section 4.3.3 . . . . .	141
4.A.6 Proofs of Section 4.3.4 . . . . .	144
4.A.7 Proofs of Section 4.3.5 . . . . .	146

---

## 4.A Appendix of Chapter 4

### 4.A.1 Additional results of geometry

We first state a few classical results that we will need in the upcoming proofs. We start with a quantitative bound that links the reach to the curvature of a submanifold. We denote by  $\mathbb{I}$  the second fundamental form.

**Proposition 4.A.1** (Prp. 6.1 in [NSW08]). *Let  $M$  be a compact smooth submanifold of  $\mathbb{R}^D$ . Then, for any  $x \in M$ , we have  $\|\mathbb{I}_x\|_{\text{op}} \leq 1/\tau_M$ .*

Since  $\mathbb{I}_x$  is the differential of order two of the mapping  $\exp_x$  at the  $0 \in T_x M$ , Proposition 4.A.1 has several convenient implications. First, it gives a uniform lower bound for the injectivity radii of  $M$  as stated in Proposition 4.2.6. Second, it also yields nice bounds on how well the Euclidean distance on  $\mathbb{R}^D$  approximates the Riemannian distance  $d_M$  on  $M \times M$ .

**Proposition 4.A.2.** [NSW08, Prp. 6.3] *For any compact submanifold  $M$  of  $\mathbb{R}^D$  and any  $x, y \in M$  such that  $\|x - y\| \leq \tau_M/2$ , we have*

$$\|x - y\| \leq d_M(x, y) \leq \tau_M \left( 1 - \sqrt{1 - \frac{2\|x - y\|}{\tau_M}} \right).$$

Proposition 4.A.2 allows in turn to compare the volume measure  $\mu_M$  to the Lebesgue measure on its tangent spaces.

**Lemma 4.A.3.** *For any  $d$ -dimensional compact smooth submanifold  $M$  of  $\mathbb{R}^D$ , for any  $x \in M$  and any  $\eta \leq \tau_M/2$ , we have*

$$(1 - \eta^2/6\tau_M^2)^d \zeta_d \eta^d \leq \mu_M(B(x, \eta)) \leq \left\{ (1 + (\xi(\eta/\tau_M)\eta)^2/\tau_M^2) \xi(\eta/\tau_M) \right\}^d \zeta_d \eta^d$$

where  $\xi(s) = (1 - \sqrt{1 - 2s})/s$  and  $\zeta_d$  is the volume of the unit Euclidean ball in  $\mathbb{R}^d$ .

*Proof.* This result already appears in [Aam17, Lem III.23] but we prove it here to make constants explicit. Let us denote by  $\text{Leb}$  the Lebesgue measure on  $T_x M$ . Using [Aam17, Prp III.22.v], we know that, as long as  $\xi(\eta/\tau_M)\eta \leq \tau_M$  (which holds if  $\eta \leq \tau_M/2$ ),

$$\begin{aligned} (1 - \eta^2/6\tau^2)^d \text{Leb}(B_{T_x M}(0, \eta)) &\leq \mu_M(\exp_x(B_{T_x M}(0, \eta))) \\ &\leq \mu_M(\exp_x(B_{T_x M}(0, \xi(\eta/\tau_M)\eta))) \\ &\leq (1 + (\xi(\eta/\tau_M)\eta)^2/\tau_M^2)^d \text{Leb}(B_{T_x M}(0, \xi(\eta/\tau_M)\eta)). \end{aligned}$$

Thanks to Proposition 4.A.2, if  $\eta \leq \tau_M/2$ , then  $\exp_x(B_{T_x M}(0, \eta)) \subset M \cap B(x, \eta) \subset \exp_x(B_{T_x M}(0, \xi(\eta/\tau_M)\eta))$ . These inclusions combined with the last inequalities yield the result.  $\square$

#### 4.A.2 Proof of Theorem 4.2.10

We go along a classical line of arguments, thanks to a Bayesian two-point inequality by means of Le Cam's lemma [Yu97, Lem. 1], restated here in our context. For two probability measures  $P_1, P_2$ , we write  $\text{TV}(P_1, P_2) = \sup_A |P_1(A) - P_2(A)|$  for their variational distance and  $H^2(P_1, P_2) = \int (\sqrt{dP_1} - \sqrt{dP_2})^2$  for their (squared) Hellinger distance.

**Lemma 4.A.4.** (Le Cam) *For any  $P_1, P_2 \in \Sigma_{\alpha, \beta}^d$ , we have,*

$$\begin{aligned} &\inf_{\hat{f}} \sup_{P \in \Sigma_{\alpha, \beta}^d} \mathbb{E}_{P^{\otimes n}}[|\hat{f}(x_0) - f_P(x_0)|^p]^{1/p} \\ &\geq \frac{1}{2} |f_{P_1}(x_0) - f_{P_2}(x_0)| (1 - \text{TV}(P_1^{\otimes n}, P_2^{\otimes n})) \\ &\geq \frac{1}{2} |f_{P_1}(x_0) - f_{P_2}(x_0)| \left(1 - \sqrt{2 - 2(1 - H^2(P_1, P_2)/2)^n}\right). \end{aligned} \tag{4.20}$$

*Proof.* The proof of (4.20) can be found in [Yu97, Lem. 1]. It only remains to see that  $\text{TV}(P_1^{\otimes n}, P_2^{\otimes n}) \leq \sqrt{2 - 2(1 - H^2(P_1, P_2)/2)^n}$ . This comes from classical inequalities on the Hellinger distance, see [Tsy08, Lem. 2.3 p.86] and [Tsy08, Prp.(i)-(iv) p.83].  $\square$

*Proof of Theorem 4.2.10.* With no loss of generality, we pick  $x_0 = 0$ . We work in  $\mathbb{R}^{d+1} \subset \mathbb{R}^D$ , and denote  $(e_1, \dots, e_{d+1})$  the canonical basis of  $\mathbb{R}^{d+1}$ . We consider a family of submanifolds  $M_\delta \subset \mathbb{R}^{d+1}$  such that

$$M_\delta \cap \{(z, t) \in \mathbb{R}^d \times \mathbb{R} \mid \|z\| \leq 1\} = O(\delta) \cup O(-\delta)$$



where  $O(t) = \{(z, t) \mid z \in \mathbb{R}^d, \|z\| \leq 1\}$ . We do not give the construction explicitly, but refer instead to Figure 4.A.1 for a diagram of such a manifold.

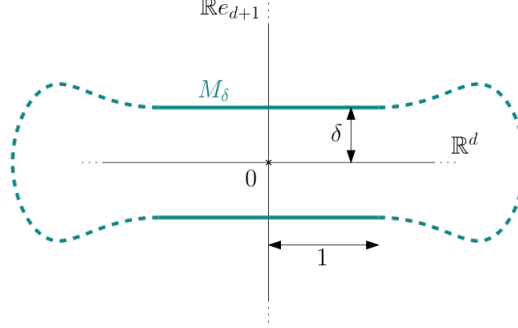


Figure 4.A.1 – Diagram of a candidate for  $M_\delta$ .

We endow each  $M_\delta$  with a density  $f_\delta$  such that

$$\forall z \in O(\delta), f_\delta(z) = f_{\max} \text{ and } \forall z \in O(-\delta), f_\delta(z) = f_{\min}$$

and we denote  $Q_\delta = f_\delta d\mu_{M_\delta}$ . If  $f_{\min}$  is small enough (due to the constraint  $\text{vol supp } P \leq 1/f_{\min}$  for any  $P \in \Sigma_{\alpha, \beta}^d$ ) we can always choose  $M_\delta$  and  $f_\delta$  so that  $|\iota_{M_\delta}|_{\alpha+1} \leq R/2$  and  $|f_\delta|_\beta \leq L/2$ .

Let now  $\Phi : \mathbb{R}^{d+1} \rightarrow \mathbb{R}$  be a smooth, positive, radial function with support in  $B(0, 1)$  with  $\Phi(0) = 1$ . Because the exponential map smoothly depends on the metric, for any  $h < 1$ , there exists  $\delta_h \in (0, h)$  sufficiently small such that the push-forward measures of  $Q_{\delta_h}$  through the mappings

$$\Psi_h^+(z) = \text{Id} - \delta_h \Phi\left(\frac{z - \delta_h e_{d+1}}{h}\right) e_{d+1} \text{ and } \Psi_h^-(z) = \text{Id} + \delta_h \Phi\left(\frac{z + \delta_h e_{d+1}}{h}\right) e_{d+1}$$

are both in  $\Sigma_{\alpha, \beta}^d$ . We write  $N_h^\pm = \Psi_h^\pm(M_{\delta_h})$ ,  $P_h^\pm = (\Psi_h^\pm)_*(Q_{\delta_h})$  and  $g_h^\pm$  for the continuous version of the density  $dP_h^\pm/d\mu_{N_h^\pm}$ . See Figure 4.A.2 for a diagram of  $N_h^+$  and  $N_h^-$ .

Using Lemma 4.A.4, we obtain

$$\inf_{\hat{f}} \sup_{P \in \Sigma_{\alpha, \beta}^d} \mathbb{E}_{P^{\otimes n}}[|\hat{f}(0) - f_P(0)|^p]^{1/p} \geq \frac{1}{2} |g_h^+(0) - g_h^-(0)| (1 - n \text{TV}(P_h^+, P_h^-)).$$

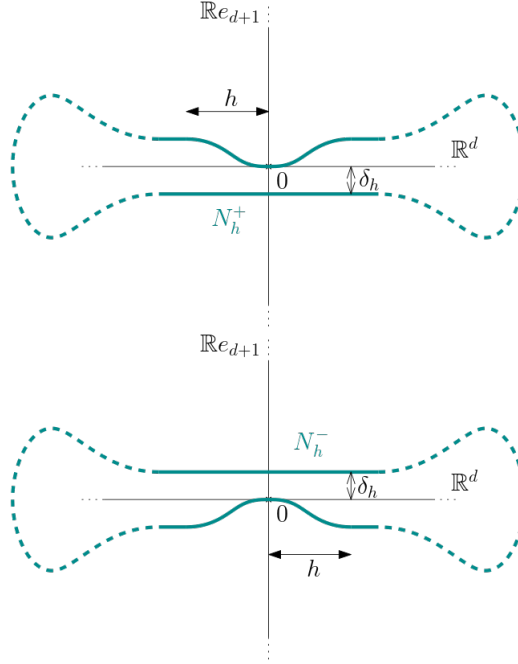
But now  $g_h^+(0) = f_{\delta_h}(\delta_h e_{d+1}) \times |\det d\Psi_h^+(\delta_h e_{d+1})|^{-1} = f_{\max}$  and, likewise,  $g_h^-(0) = f_{\min}$ . As for the total variation distance, we have that  $\text{TV}(P_h^+, P_h^-)$  is equal to

$$\begin{aligned} & P_h^+(\Psi_h^+(hO(\delta_h))) + P_h^+(hO(-\delta_h)) + P_h^-(\Psi_h^-(hO(-\delta_h))) + P_h^-(hO(\delta_h)) \\ &= 2 \text{vol}(hO(\delta_h)) \times (f_{\min} + f_{\max}) = 2\zeta_d h^d (f_{\min} + f_{\max}) \end{aligned}$$

where we recall that  $\zeta_d$  is the volume of the  $d$ -dimensional unit-ball. Putting all the estimates together, we conclude

$$\inf_{\hat{f}} \sup_{P \in \Sigma_{\alpha, \beta}^d} \mathbb{E}_{P^{\otimes n}}[|\hat{f}(0) - f_P(0)|^p]^{1/p} \geq \frac{1}{2} (f_{\max} - f_{\min}) (1 - 2n\zeta_d h^d (f_{\min} + f_{\max})).$$

Letting  $h$  goes to 0 yields the result.  $\square$


 Figure 4.A.2 – Diagram of manifolds  $N_h^+$  (Left) and  $N_h^-$  (right).

### 4.A.3 Proof of Theorem 4.3.2

*Proof of Theorem 4.3.2.* Suppose without loss of generality that  $x = 0$  and consider a smooth submanifold  $M$  of  $\mathbb{R}^{d+1} \subset \mathbb{R}^D$  that contains the disk  $B_{\mathbb{R}^d}(0, 1) \subset \mathbb{R}^d \times \{0_{\mathbb{R}^{D-d}}\}$  with reach greater than  $\tau$ , see Figure 4.A.3 for a diagram of such an  $M$ . By smoothness and compactness of  $M$ , there exists  $L_*$  (depending on  $\tau$ ) such that  $M \in \mathcal{C}_{d,\alpha}(\tau, L_*)$ . Let  $P$  be the uniform probability measure over  $M$ , with density  $f : x \mapsto 1/\text{vol } M$ . We have  $P \in \Sigma_{\alpha,\beta}^d$  as long as  $L^* \leq L$  and  $f_{\min} \leq 1/\text{vol } M \leq f_{\max}$  an assumption we make from now on. For  $0 < \delta \leq 1$ , let  $P_\delta = f_\delta \cdot d\mu_M$  with

$$f_\delta(x) = \begin{cases} f(x) + \delta^\beta G(x/\delta) & \text{if } x \in B_{\mathbb{R}^d}(0, \delta) \\ f(x) & \text{otherwise} \end{cases}$$

with  $G : \mathbb{R}^d \rightarrow \mathbb{R}$  a smooth function with support in  $B_{\mathbb{R}^d}(0, 1)$  and such that  $\int_{\mathbb{R}^d} G(y) dy = 0$ . We pick  $G$  such that  $f_\delta \in \mathcal{F}_\beta$  for small enough  $\delta$ , depending on  $\tau$ . Such a  $G$  can be chosen to depend on  $R$  only.

For  $\delta$  small enough (depending on  $\tau$ ), we thus have  $P_\delta \in \Sigma_{\alpha,\beta}^d$  as well. By Lemma 4.A.4, we infer

$$\inf_{\widehat{f}} \sup_{P \in \Sigma_{\alpha,\beta}^d} \mathbb{E}_{P^{\otimes n}} [|\widehat{f}(x) - f_P(x)|^p]^{1/p} \geq \frac{1}{2} \delta^\beta |G(0)| \left(1 - \sqrt{2 - 2(1 - H^2(P, P_\delta))^n}\right)$$

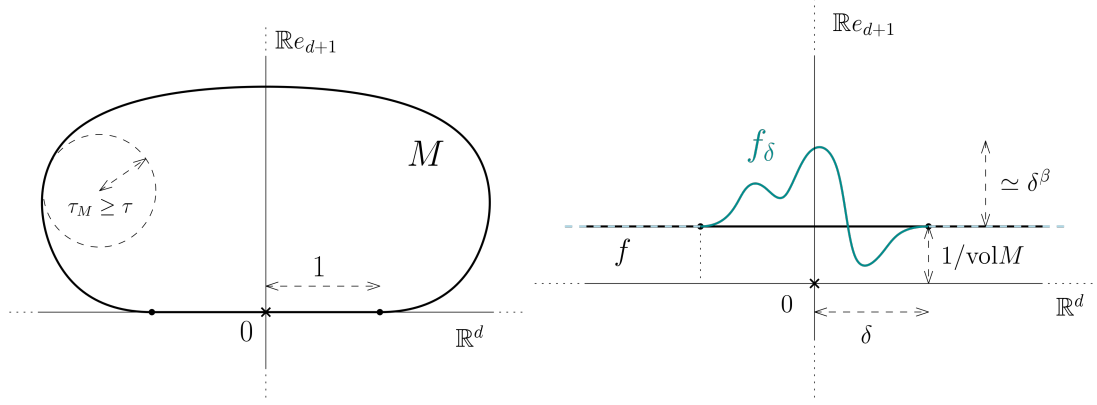


Figure 4.A.3 – Diagram of a candidate for  $M$  (Left) and of the densities  $f$  and  $f_\delta$  around 0 (Right).

so that it remains to compute  $H^2(P, P_\delta)$ . We have the following bound

$$\begin{aligned} H^2(P, P_\delta) &= \int_{B_{\mathbb{R}^d}(0, \delta)} (1 - \sqrt{1 + \text{vol } M \delta^\beta G(x/\delta)})^2 dx \\ &\leq \int_{B_{\mathbb{R}^d}(0, \delta)} (\text{vol } M)^2 \delta^{2\beta} G^2(x/\delta) dx \\ &\leq (C \vee 1) \delta^{2\beta+d} \end{aligned}$$

with  $C = \text{vol } M \times \int_{B_{\mathbb{R}^d}(0, 1)} G(z)^2 dz$  depending on  $\tau$  and  $R$  only. Taking  $\delta = (1/(C \vee 1)n)^{1/(2\beta+d)}$  we obtain, for large enough  $n$  (depending on  $\tau$ )

$$\begin{aligned} &\inf_{\hat{f}} \sup_{P \in \Sigma_{\alpha, \beta}^d} \mathbb{E}_{P \otimes n} [|\hat{f}(x) - f_P(x)|^p]^{1/p} \\ &\geq \frac{1}{2} ((C \vee 1)n)^{-\beta/(2\beta+d)} \sqrt{2 - 2(1 - 1/n)^n} \geq C_* n^{-\beta/(2\beta+d)}, \end{aligned}$$

with  $C_* = (C \vee 1)^{-1/2}$  depending on  $\tau$  and  $R$ . □

#### 4.A.4 Proofs of Section 4.3.2

We set  $K_h(x) = h^{-d} K(x/h)$  and start with bounding the variance of  $K_h(X - x_0)$  when  $X$  is distributed according to  $P \in \Sigma_{\alpha, \beta}^d$ . Let first observe that

$$|K_h(X - x_0)| \leq \frac{\|K\|_\infty}{h^d} \mathbb{1}_{B_D(x_0, h)}(X) \leq \frac{\|K\|_\infty}{h^d} \quad (4.21)$$

**Lemma 4.A.5.** *For any  $P \in \Sigma_{\alpha, \beta}^d$  and for any  $h \leq \tau/2$ ,*

$$\text{Var}_P(K_h(X - x_0)) \leq \frac{\omega}{h^d} \quad \text{where } \omega = 4^d \zeta_d \|K\|_\infty^2 b.$$

with  $\zeta_d$  being the volume of the unit ball in  $\mathbb{R}^d$ .

*Proof.* We have

$$\mathrm{Var}_P(K_h(X - x_0)) \leq \mathbb{E}_P[K_h(X - x_0)^2] \leq \frac{\|K\|_\infty^2}{h^{2d}} P(B(x_0, h)) \leq \frac{4^d \zeta_d f_{\max} \|K\|_\infty^2}{h^d}$$

where we used (4.21) and Lemma 4.A.3 with  $\eta = \tau/2$ .  $\square$

Using Bernstein inequality [BLM13, Thm. 2.10 p.37], for any  $P \in \Sigma_{\alpha, \beta}^d$  and any  $t > 0$ , we infer

$$\mathbb{P}\left(|\widehat{\xi}_h(P, x)| \geq \sqrt{\frac{2\omega t}{nh^d}} + \frac{\|K\|_\infty t}{nh^d}\right) \leq 2e^{-t}, \quad (4.22)$$

where  $\mathbb{P}$  is a short-hand notation for the distribution  $P^{\otimes n}$  of the  $n$ -sample  $X_1, \dots, X_n$  taken under  $P$ . The bound (4.22) is the main ingredient needed to bound the  $L_p$ -norm of the stochastic deviation of  $\widehat{f}_h$ .

*Proof of Proposition 4.3.5.* We denote by  $u_+ = \max\{u, 0\}$  the positive part of a real number  $u$ . We start with

$$\mathbb{E}_{P^{\otimes n}}[|\widehat{\xi}_h(P, x_0)|^p] \leq 2^{p-1} (\Omega(h)^p + \mathbb{E}_{P^{\otimes n}}[(|\widehat{\xi}_h(P, x_0)| - \Omega(h))_+^p]).$$

The first term has the right order. For the second one, we make use of (4.22) to infer

$$\begin{aligned} & \mathbb{E}_{P^{\otimes n}}[(|\widehat{\xi}_h(P, x_0)| - \Omega(h))_+^p] \\ &= \int_0^\infty \mathbb{P}(|\widehat{\xi}_h(P, x_0)| > \Omega(h) + u) p u^{p-1} du \\ &= p \Omega(h)^p \int_0^\infty \mathbb{P}(|\widehat{\xi}_h(P, x_0)| > \Omega(h)(1 + u)) u^{p-1} du \\ &\leq p \Omega(h)^p \left(1 + \int_1^\infty \mathbb{P}\left(|\widehat{\xi}_h(P, x_0)| > \sqrt{\frac{2\omega(1+u)}{nh^d}} + \frac{\|K\|_\infty(1+u)}{nh^d}\right) u^{p-1} du\right) \\ &\leq p \Omega(h)^p \left(1 + \int_1^\infty 2e^{-1-u} u^{p-1} du\right) \\ &\leq p \Omega(h)^p (1 + \Gamma(p)) \end{aligned}$$

which ends the proof.  $\square$

The proof of Lemma 4.3.7 partly relies on the following elementary lemma.

**Lemma 4.A.6.** *Let  $\gamma \geq 0$  be a real number and let  $g : \mathbb{R}^m \rightarrow \mathbb{R}$  for  $m \in \mathbb{N}^*$  satisfying that  $\|g\|_\infty \leq b$  and that the restriction of  $g$  to  $B(0, r)$  (denoting here the open ball in  $\mathbb{R}^m$ ) is  $\beta$ -Hölder, meaning that*

$$\forall v, w \in B(0, r), \quad \|d^k g(v) - d^k g(w)\|_{\mathrm{op}} \leq A \|v - w\|^\delta$$

*for some  $A > 0$  with  $k = \lceil \gamma - 1 \rceil$  and  $\delta = \gamma - k$ . Then there exists a constant  $C$  (depending on  $m, \gamma, r, b$  and  $A$ , and depending on  $m$  and  $\gamma$  when  $r = \infty$ ) such that, for all  $1 \leq j \leq k$ ,*

$$\sup_{v \in B(0, r/2)} \|d^j g(v)\|_{\mathrm{op}} \leq C b^{1-j/\gamma} A^{j/\gamma}.$$

*Proof.* Let  $v \in B(0, r/2)$ . Since  $g$  is  $\gamma$ -Hölder on  $B(0, r)$ , we know that there exists a function  $R_v$  such that, for any  $z$  such that  $v + z \in B(0, r)$ , we have

$$g(v + z) - \sum_{j=0}^k \frac{1}{j!} d^j g(v)[z^{\otimes j}] = R_v(z)$$

with  $|R_v(z)| \leq A\|z\|^\beta/k!$ . Let  $h = (2bk!/A)^{1/\gamma}$ , and  $z_0 \in \mathbb{R}^m$  be unit-norm. Pick  $a_1, \dots, a_k \in (0, 1)$  all distincts and small enough such that  $ha_k z_0 \in B(0, r/2)$  for all  $k$  (if  $r = \infty$ , then we can pick the  $a_i$  independently from  $A, b$  and  $\gamma$ ). Introducing the vectors of  $\mathbb{R}^k$

$$X = \left( hdg(v)[z_0], \dots, \frac{h^k}{k!} d^k g(v)[z_0^{\otimes k}] \right) \quad \text{and}$$

$$Y = (g(v + ha_1 z_0) - g(v) - R_v(ha_1 z_0), \dots, g(v + ha_k z_0) - g(v) - R_v(ha_k z_0))$$

we have  $Y = VX$  with  $V$  being the Vandermonde matrix associated with the real numbers  $(a_1, \dots, a_k)$ . The former being invertible, we have  $\|X\| \leq \|V^{-1}\|_{\text{op}} \|Y\|$  and thus, for any  $1 \leq j \leq k$

$$\left| \frac{h^j}{j!} d^j g(v)[z_0^{\otimes j}] \right| \leq \|V^{-1}\|_{\text{op}} \left( 2b + \frac{A}{k!} h^\gamma \right).$$

Substituting the value of  $h$  and noticing that the former inequality holds for every unit-norm vector  $z_0$ , we can conclude. □

*Proof of Lemma 4.3.7.* We set  $B_h = B(x_0, h)$ . Since  $\tau/2$  is smaller than the injectivity radius of  $\exp_{x_0}$  (see Proposition 4.2.6) we can write

$$\begin{aligned} f_h(P, x_0) &= \int_{B_h} K_h(p - x_0) f(p) d\mu_M(p) \\ &= \int_{\exp_{x_0}^{-1} B_h} K_h(\exp_{x_0} v - x_0) f(\exp_{x_0} v) \zeta(v) dv \end{aligned} \quad (4.23)$$

with  $\zeta(v) = \sqrt{\det g^{x_0}(v)}$ . We set  $\gamma = \alpha \wedge \beta$  and  $k = \lceil \gamma - 1 \rceil$ . Let  $F$  denote the map  $f \circ \exp_{x_0}$ . For  $h$  smaller than  $\tau/2$ , we have  $\exp_{x_0}^{-1} B_h \subset B_{T_{x_0} M}(0, 2h) \subset B_{T_{x_0} M}(0, \tau)$  (see Proposition 4.A.2). We can thus write the following expansion, valid for all  $v \in \exp_{x_0}^{-1} B_h$  and all  $w \in T_{x_0} M$ ,

$$\exp_{x_0}(v) = x + v + \sum_{j=2}^{k+1} \frac{1}{j!} d^j \exp_{x_0}(0)[v^{\otimes j}] + R_1(v) \quad \text{with } \|R_1(v)\| \leq C_1 \|v\|^{\gamma+1}, \quad (4.24)$$

$$F(v) = f(x_0) + \sum_{j=1}^k \frac{1}{j!} d^j F(0)[v^{\otimes j}] + R_2(v) \quad \text{with } |R_2(v)| \leq C_2 \|v\|^\gamma, \quad (4.25)$$

$$K(v + w) = K(v) + \sum_{j=1}^k \frac{1}{j!} d^j K(v)[w^{\otimes j}] + R_3(v, w) \quad \text{with } |R_3(v, w)| \leq C_3 \|w\|^\gamma, \quad (4.26)$$

with  $C_1$  depending on  $\alpha, \tau$  and  $L$ ,  $C_2$  depending on  $\beta, \tau, f_{\max}$  and  $R$  (see Lemma 4.A.6), and  $C_3$  depending on  $K$ . Since now we know that  $g_{ij}^{x_0}(v) = \langle d \exp_{x_0}(v)[e_i], d \exp_{x_0}(v)[e_j] \rangle$ , we have a similar expansion for the mapping  $\zeta(v) = \sqrt{\det g^{x_0}(v)}$

$$\zeta(v) = 1 + \sum_{j=1}^k \frac{1}{j!} d^j \zeta(0)[v^{\otimes j}] + R_4(v) \quad \text{with } |R_4(v)| \leq C_4 \|v\|^\gamma \quad (4.27)$$

with  $C_4$  depending on  $\alpha, \tau$  and  $L$ . Making the change of variable  $v = hw$  in (4.23), we get

$$f_h(P, x_0) = \sum_{k=0}^k G_j(h, P, x_0) + R_h(P, x_0)$$

with  $G_j$  corresponding to the integration of the  $j$ -th order terms in the expansion around 0 of the function  $w \mapsto K\left(\frac{\exp_{x_0}(hw) - x_0}{h}\right) F(hw) \zeta(hw)$ . In particular  $G_j$  can be written as a sum of terms of the type

$$I = h^j \int_{\frac{1}{h} \exp_{x_0}^{-1} B_h} d^m K(w) [\phi(w)^{\otimes m}] \psi(w) dw$$

where  $\psi$  and  $\phi$  are monomials in  $w$  satisfying  $m \deg \phi + \deg \psi = j$ , with coefficients bounded by constants depending on  $\alpha, \tau, L, \beta, f_{\max}$  and  $R$  (again, use Lemma 4.A.6 to bound the derivatives). Since now  $B_{T_{x_0}M}(0, 1) \subset \frac{1}{h} \exp_{x_0}^{-1} B_h$ , and since  $d^j K$  is zero outside of  $B(0, 1)$ , we have that  $G_j$  can actually be written  $G_j(h, P, x) = h^j G_j(P, x_0)$  with  $|G_j(P, x_0)| \leq C$  for some  $C$  depending on  $K, \alpha, \tau, L, \beta, f_{\max}$  and  $R$ . Similar reasoning leads to  $R_h(P, x_0) \leq Ch^\gamma$  with  $C$  depending again on  $K, \alpha, \tau, L, \beta, f_{\max}$  and  $R$ . To conclude, it remains to compute  $G_0(P, x_0)$ . Looking at the zero-th order terms in the expansions (4.24) to (4.27), we find that

$$G_0(P, x_0) = \int_{B_{T_{x_0}M}(0, 1)} K(w) f(x_0) dw = f(x_0)$$

where we used Assumption 4.3.6. The proof of Lemma 4.3.7 is complete.  $\square$

*Proof of Proposition 4.3.8.* For a positive integer  $\ell \geq 1$ , let  $f_h^{(\ell)}(P, x_0)$  be the mean of the estimator  $\widehat{f}_h(x_0)$  computed using  $K^{(\ell)}$ . Let  $\gamma = \alpha \wedge \beta$  and  $k = \lceil \gamma - 1 \rceil$ . We recursively prove on  $1 \leq \ell < \infty$  the following identity

$$\forall h \leq \tau/2, \quad f_h^{(\ell)}(P, x_0) = f(x_0) + \sum_{j=\ell}^k h^j G_j^{(\ell)}(P, x_0) + R_h^{(\ell)}(P, x_0) \quad (4.28)$$

where  $|R_h^{(\ell)}(P, x_0)| \leq C^{(\ell)} h^\gamma$  for some constant  $C^{(\ell)}$  depending on  $\tau, \ell, L, R, f_{\max}$  and  $\beta$ . The initialisation step  $\ell = 1$  has been proven in Lemma 4.3.7. Let now  $1 \leq \ell \leq k$ . By linearity of  $f_h(P, x_0)$  with respect to  $K$ , we have

$$f_h^{(\ell+1)}(P, x_0) = 2f_{2^{-1}\ell h}^{(\ell)}(P, x_0) - f_h^{(\ell)}(P, x_0).$$

Since  $2^{-1/\ell}h \leq h$ , we can use our induction hypothesis (4.28) and find

$$f_h^{(\ell+1)}(P, x_0) = f(x_0) + \sum_{j=\ell}^k (2^{1-j/\ell} - 1) h^j G_j^{(\ell)}(P, x_0) + 2R_{2^{-1/\ell}h}^{(\ell)}(P, x_0) - R_h^{(\ell)}(P, x_0).$$

We conclude noticing that  $2^{1-j/\ell} - 1 = 0$  for  $j = \ell$ , and setting  $G_j^{(\ell+1)}(P, x_0) = (2^{1-j/\ell} - 1)G_j^{(\ell)}(P, x_0)$  and  $R_h^{(\ell+1)}(P, x_0) = 2R_{2^{-1/\ell}h}^{(\ell)}(P, x_0) - R_h^{(\ell)}(P, x_0)$ . The new remainder term verifies

$$|R_h^{(\ell+1)}(P, x_0)| \leq (2^{1-\gamma/\ell} + 1)C^{(\ell)}h^\gamma \leq 3C^{(\ell)}h^\gamma \quad (4.29)$$

ending the induction by setting  $C^{(\ell+1)} = 3C^{(\ell)}$ . When  $\ell \geq k + 1$ , the induction step is trivial.  $\square$

#### 4.A.5 Proofs of Section 4.3.3

*Proof of Lemma 4.3.9.* Let  $\gamma : [0, L_M] \rightarrow M$  be a unit speed parametrisation of  $M$  and extend  $\gamma$  to a smooth function on  $\mathbb{R}$  by  $L_M$ -periodicity. Suppose without loss of generality that  $\gamma(0) = x$ . For any  $t \in \mathbb{R}$ , there is a canonical identification between  $T_{\gamma(t)}M$  and  $\mathbb{R}$  through the map  $v \mapsto \langle \dot{\gamma}(t), v \rangle$ . With such an identification, we can write that for  $s \in \mathbb{R} \simeq T_x M$ ,  $\exp_x(s) = \gamma(s)$  because  $\gamma$  is unit-speed. We thus have  $d\exp_x(s)[h] = h\dot{\gamma}(s)$  for any  $h \in T_{\gamma(s)}M \simeq \mathbb{R}$ . It follows that  $\det g^x(s) = \|d\exp_x(s)[1]\|^2 = \|\dot{\gamma}(s)\|^2 = 1$  and this completes the proof.  $\square$

We write  $V = (x_0, X_1, \dots, X_n)$  for the vertices of  $\mathcal{G}_\varepsilon$  and  $\widehat{\eta} = \sup_{x \in M} d(x, V)$ . For small enough  $\widehat{\eta}$  we have that  $\mathcal{G}_\varepsilon$  is connected, therefore the distance  $\widehat{d}_\varepsilon$  is well-defined on  $V$ . We have in that case a good reverse control of  $d_M$  by  $\widehat{d}_\varepsilon$ , as shown in the next two lemmata.

**Lemma 4.A.7.** *If  $\varepsilon \leq 8\tau$  and  $16\widehat{\eta} \leq \varepsilon$ , then  $\widehat{d}_\varepsilon(x, y) \leq d_M(x, y)$  for any  $x, y \in V$ .*

**Lemma 4.A.8.** *If  $\varepsilon \leq \tau/2$ , then  $d_M(x, y) \leq \left(1 + \frac{\pi^2}{48\tau^2}\varepsilon^2\right)\widehat{d}_\varepsilon(x, y)$  for any  $x, y \in V$ .*

*Proof of Lemma 4.A.7.* We can take the shortest path in  $M$  between  $x$  and  $y$  as a unit-speed path of the form  $\gamma : [0, \ell] \rightarrow \mathbb{R}^D$  with  $\ell = d_M(x, y) \leq L_M/2$ . We let  $\delta = \ell/(4\lfloor \ell/\varepsilon \rfloor)$  and  $N = 4\lfloor \ell/\varepsilon \rfloor$ . Notice that  $\varepsilon/4 \leq \delta \leq \varepsilon/2$ . Let us define  $p_j = \gamma(j\delta)$ , so that  $p_0 = x$  and  $p_N = y$ . Since  $\widehat{\eta} \leq \varepsilon/16$ , for every  $1 \leq j \leq N - 1$ , there exists among our vertices  $V$  a point denoted by  $\widehat{p}_j$  such that  $\|p_j - \widehat{p}_j\| \leq \varepsilon/16$ . We set  $\widehat{t}_j \in [0, L_M]$  for its coordinate, namely  $\widehat{p}_j = \gamma(\widehat{t}_j)$ .

Let us show first that for  $1 \leq j < N$ , we have  $\widehat{t}_j \in [0, \ell]$ . Indeed, thanks to Proposition 4.A.2, since  $\varepsilon/16 \leq \tau/2$ , we have  $|t_j - \widehat{t}_j| \leq 2\|p_j - \widehat{p}_j\| \leq \varepsilon/8$ . Since  $\delta \geq \varepsilon/4$ , we thus have  $0 \leq \widehat{t}_1 \leq \dots \leq \widehat{t}_{N-1} \leq \ell$ . Furthermore, writing  $\widehat{p}_0 = x$  and  $\widehat{p}_N = y$ , we have

$$\|\widehat{p}_j - \widehat{p}_{j+1}\| \leq \|\widehat{p}_j - p_j\| + \|p_j - p_{j+1}\| + \|p_{j+1} - \widehat{p}_{j+1}\| \leq \varepsilon$$

for any  $0 \leq j < \ell$ . The sequence  $s = (\widehat{p}_0, \dots, \widehat{p}_N)$  is thus a path in  $\mathcal{G}_\varepsilon$  and so

$$\widehat{d}_\varepsilon(p, q) \leq L_s = \|\widehat{p}_1 - \widehat{p}_0\| + \dots + \|\widehat{p}_N - \widehat{p}_{N-1}\| \leq |\widehat{t}_1 - \widehat{t}_0| + \dots + |\widehat{t}_N - \widehat{t}_{N-1}| = \widehat{t}_N - \widehat{t}_0$$

where we set  $\widehat{t}_0 = 0$  and  $\widehat{t}_N = \ell = d_M(x, y)$ , ending the proof.  $\square$

*Proof of Lemma 4.A.8.* Following the proof of [ACLG19, Lem. 5] if there exists  $\delta > 0$  such that  $\|x - y\| \leq \delta$  implies  $d_M(x, y) \leq \pi\tau$  for all  $x, y \in M$ , then we must have that for any  $x, y \in M$  satisfying  $\|x - y\| \leq \delta$ ,

$$d_M(x, y) \leq \left(1 + \frac{\pi^2}{48\tau^2} \|x - y\|^2\right) \|x - y\|.$$

Thanks to Proposition 4.A.2, this must hold for  $\delta = \tau/2$ . Now let  $p_0, \dots, p_m$  be one shortest path in  $\mathcal{G}_\varepsilon$  between  $x$  and  $y$ . Since  $\varepsilon \leq \tau/2$ , we have

$$\begin{aligned} d_M(x, y) &\leq \sum_{j=1}^m d_M(p_j, p_{j-1}) \leq \sum_{j=1}^m \left(1 + \frac{\pi^2}{48\tau^2} \|p_j - p_{j-1}\|^2\right) \|p_j - p_{j-1}\| \\ &\leq \left(1 + \frac{\pi^2}{48\tau^2} \varepsilon^2\right) \widehat{d}_\varepsilon(x, y) \end{aligned}$$

which ends the proof.  $\square$

In view of Lemma 4.A.7 and Lemma 4.A.8, we want to tune  $\varepsilon$  so that it is the smallest possible and so that  $16\widehat{\eta} \leq \varepsilon$  holds with high probability. This is achieved for  $\varepsilon$  of order  $\log n/n$ .

**Lemma 4.A.9.** *Setting  $\varepsilon = \frac{32(p+1)\log n}{f_{\min} n}$ , for every  $n \geq 3$ , we have  $\mathbb{P}(16\widehat{\eta} \leq \varepsilon) \geq 1 - 1/n^p$ .*

*Proof.* Let  $\delta > 0$ , and let  $N = \lfloor L_M/\delta \rfloor$ . We split  $[0, L_M]$  into  $N$  intervals  $I_1, \dots, I_N$  of length  $L_M/N$ . We denote  $\mathcal{A}$  the event for which each  $I_j$  contains at least one coordinate among those of the sample of observations  $(X_1, \dots, X_n)$ . On  $\mathcal{A}$ , we have  $\widehat{\eta} \leq L_M/N \leq 2\delta$ . Moreover,

$$\begin{aligned} \mathbb{P}(\mathcal{A}) &= 1 - \mathbb{P}(\exists j, \gamma(I_j) \text{ contains no observation}) \\ &\geq 1 - N \left(1 - \min_{1 \leq j \leq N} P(\gamma(I_j))\right)^n \geq 1 - N \left(1 - \frac{aL_M}{N}\right)^n. \end{aligned}$$

Using that  $N \leq L_M/\delta$  and that  $L_M \leq 1/a$  we infer

$$\mathbb{P}(\widehat{\eta} \leq 2\delta) \geq 1 - \frac{1}{a\delta} (1 - a\delta)^n \geq 1 - \frac{e^{-a\delta n}}{a\delta}.$$

Setting  $\delta = \frac{(p+1)\log n}{an}$  and  $\varepsilon = 32\delta$  yields

$$\mathbb{P}(16\widehat{\eta} \leq \varepsilon) \geq 1 - \frac{n}{(p+1)n^{p+1}\log n} \geq 1 - \frac{1}{n^p}$$

as soon as  $\log n \geq 1$ , i.e. for  $n \geq 3$ .  $\square$



*Proof of Proposition 4.3.10.* Recall that we set  $K^{1D} = K^{(1,\ell)}$  where  $K^{(1,\ell)}$  is defined starting from kernel  $\lambda_1^{-1}\Lambda$ . Let  $\mathcal{A}$  be the event  $\{16\widehat{\eta} \leq \varepsilon\}$ . By triangle inequality,  $\mathbb{E}_{P^{\otimes n}}[|\hat{f}_h^{1D}(x_0) - f_P(x_0)|^p]^{1/p} \leq \mathcal{R}_{\mathcal{A}} + \mathcal{R}_{\mathcal{A}^c}$ , with

$$\mathcal{R}_{\mathcal{A}} = \left( \mathbb{E}_{P^{\otimes n}}[|\hat{f}_h^{1D}(x_0) - f_P(x_0)|^p \mathbb{1}_{\mathcal{A}}] \right)^{1/p}$$

and

$$\mathcal{R}_{\mathcal{A}^c} = \left( \mathbb{E}_{P^{\otimes n}}[|\hat{f}_h^{1D}(x_0) - f_P(x_0)|^p \mathbb{1}_{\mathcal{A}^c}] \right)^{1/p}.$$

On  $\mathcal{A}$ , we have, for  $n$  large enough (depending on  $p, f_{\min}$  and  $\tau$ ) such that  $\varepsilon \leq \tau/2$  holds,  $|\widehat{d}_\varepsilon(X_i, x_0) - d_M(X_i, x_0)| \leq C_1 \varepsilon^2$  with  $C_1$  depending on  $\tau$  only. This is inferred from Lemmas 4.A.7 and 4.A.8. We deduce that, on this event,

$$|\hat{f}_h^{1D}(x_0) - \hat{g}_h^{1D}(x_0)| \leq \frac{C_1 \|K^{1D'}\|_\infty \varepsilon^2}{h^2} \quad \text{with} \quad \hat{g}_h^{1D}(x_0) = \frac{1}{n} \sum_{i=1}^n K_h^{1D}(d_M(X_i, x_0)).$$

It follows that

$$\begin{aligned} \mathcal{R}_{\mathcal{A}} &\leq \frac{C_1 \|K^{1D'}\|_\infty \varepsilon^2}{h^2} + \left( \mathbb{E}_{P^{\otimes n}}[|\hat{g}_h^{1D}(x_0) - f_P(x_0)|^p] \right)^{1/p} \\ &\leq \frac{C_1 \|K^{1D'}\|_\infty \varepsilon^2}{h^2} + \left( \mathbb{E}_{P^{\otimes n}}[|\widehat{\xi}_h^*(P, x_0)|^p] \right)^{1/p} + |\mathcal{B}_h^*(P, x_0)| \end{aligned}$$

with  $\mathcal{B}_h^*$  and  $\widehat{\xi}_h^*$  denoting the bias and stochastic deviation of estimator  $\hat{g}_h^{1D}(x_0)$ . Following the same arguments as in proof of Proposition 4.3.5, we have

$$\mathbb{E}_{P^{\otimes n}}[|\widehat{\xi}_h^*(P, x_0)|^p]^{1/p} \leq c_p \Omega(h)^p,$$

with  $c_p$  depending only on  $p$ . For the bias term, as soon as  $h \leq \pi\tau$ , we have

$$\begin{aligned} \mathcal{B}_h^*(P, x) &= \mathbb{E}_{P^{\otimes n}}[\hat{g}_h^{1D}(x_0)] - f(x_0) = \int_M K_h^{1D}(d(p, x_0)) f(p) d\mu_M(p) - f(x_0) \\ &= \int_{B_{T_{x_0}M}(0,1)} K^{1D}(\|v\|) (f \circ \exp_{x_0}(hv) - f(x_0)) dv. \end{aligned}$$

Since now  $f \circ \exp_x$  is  $\beta$ -Hölder on  $B_{T_{x_0}M}(0, \pi\tau)$ , we know that all the terms in the development of  $\mathcal{B}_h^*(P, x_0)$  up to order  $\lceil \beta - 1 \rceil$  cancels. We deduce  $|\mathcal{B}_h^*(P, x_0)| \leq C_2 h^\beta$  with  $C_2$  depending on  $\ell$  and  $R$  only. For the other term  $\mathcal{R}_{\mathcal{A}^c}$ , we write  $|\hat{f}_h^{1D}(x_0) - f(x_0)| \leq \frac{\|K^{1D}\|_\infty}{h} + f_{\max}$ , so that, according to Lemma 6.2.1,

$$\mathcal{R}_{\mathcal{A}^c} \leq \left( \frac{\|K^{1D}\|_\infty}{h} + f_{\max} \right) \mathbb{P}(\mathcal{A}^c)^{1/p} \leq C_3 \frac{1}{nh}$$

with  $C_3$  depending on  $\ell$  and  $f_{\max}$ . Putting all these estimates together yields the result.  $\square$

#### 4.A.6 Proofs of Section 4.3.4

**Lemma 4.A.10.** *For any  $P \in \Sigma_{\alpha,\beta}^d$ , and  $\Theta > p$ , we have*

$$\mathbb{E}_{P^{\otimes n}}[|\widehat{f}(x_0) - f_P(x_0)|^p]^{1/p} \lesssim \Omega(h^*(P, x_0))\lambda(h^*(P, x_0))$$

up to a constant depending on  $p$  and  $\Theta$ , with

$$h^*(P, x_0) = \max \left\{ h \in \mathbb{H} \mid \forall \eta \in \mathbb{H}(h), |f_\eta(P, x_0) - f(x_0)| \leq \frac{1}{2}\Omega(h)\lambda(h) \right\}.$$

*Proof.* We fix  $P \in \Sigma_{\alpha,\beta}^d$  and write  $\widehat{h}$  and  $h^*$  for  $\widehat{h}(x_0)$  and  $h^*(P, x_0)$  respectively. Let  $\mathcal{A} = \{\widehat{h} \geq h^*\}$ . We can write  $\mathbb{E}_{P^{\otimes n}}[|\widehat{f}(x_0) - f_P(x_0)|^p] = \mathcal{R}_{\mathcal{A}} + \mathcal{R}_{\mathcal{A}^c}$ , where

$$\mathcal{R}_{\mathcal{A}} = \mathbb{E}_{P^{\otimes n}}[|\widehat{f}(x_0) - f_P(x_0)|^p \mathbb{1}_{\mathcal{A}}] \quad \text{and} \quad \mathcal{R}_{\mathcal{A}^c} = \mathbb{E}_{P^{\otimes n}}[|\widehat{f}(x_0) - f_P(x_0)|^p \mathbb{1}_{\mathcal{A}^c}].$$

We start with bounding  $\mathcal{R}_{\mathcal{A}}$ . Firstly,

$$\begin{aligned} \mathcal{R}_{\mathcal{A}} &\leq 3^{p-1} \left( \mathbb{E}_{P^{\otimes n}}[|\widehat{f}_{\widehat{h}}(x_0) - \widehat{f}_{h^*}(x_0)|^p \mathbb{1}_{\mathcal{A}}] + \mathbb{E}_{P^{\otimes n}}[|\widehat{f}_{h^*}(x_0) - f_{h^*}(P, x_0)|^p \mathbb{1}_{\mathcal{A}}] \right. \\ &\quad \left. + \mathbb{E}_{P^{\otimes n}}[|f_{h^*}(P, x_0) - f_P(x_0)|^p \mathbb{1}_{\mathcal{A}}] \right). \end{aligned}$$

Next, by definition of  $\widehat{h}$  and  $\mathcal{A}$ , we have

$$|\widehat{f}_{\widehat{h}}(x_0) - \widehat{f}_{h^*}(x_0)| \mathbb{1}_{\mathcal{A}} \leq \psi(\widehat{h}, h^*) \mathbb{1}_{\mathcal{A}} \leq 2\Omega(h^*)\lambda(h^*).$$

By definition of  $h^*$ , we also have  $|f_{h^*}(P, x_0) - f(x_0)| \leq \frac{1}{2}\Omega(h^*)\lambda(h^*)$ . Finally, using Proposition 4.3.5

$$\mathbb{E}_{P^{\otimes n}}[|f_{h^*}(x_0) - f_{h^*}(P, x_0)|^p \mathbb{1}_{\mathcal{A}}] \leq c_p \Omega(h^*)^p \leq c_p (\Omega(h^*)\lambda(h^*))^p$$

holds as well. Putting all three inequalities together yields

$$\mathcal{R}_{\mathcal{A}} \leq C_{\mathcal{A}} (\Omega(h^*)\lambda(h^*))^p \quad \text{with} \quad C_{\mathcal{A}} = 3^{p-1} (2^p + c_p + 2^{-p}).$$

We now turn to  $\mathcal{R}_{\mathcal{A}^c}$ . Notice that for any  $h \in \mathbb{H}(h^*)$ , we have

$$|f_h(P, x_0) - f(x_0)| \leq \frac{1}{2}\Omega(h^*)\lambda(h^*) \leq \frac{1}{2}\Omega(h)\lambda(h),$$

hence

$$|\widehat{f}_h(x_0) - f(x_0)| \leq \frac{1}{2}\Omega(h^*)\lambda(h^*) + |\widehat{\xi}_h(P, x_0)|.$$

We can thus write

$$\begin{aligned} \mathcal{R}_{\mathcal{A}^c} &= \sum_{h \in \mathbb{H}(h^*/2)} \mathbb{E}_{P^{\otimes n}}[|\widehat{f}_h(x_0) - f(x_0)|^p \mathbb{1}_{\{\widehat{h}=h\}}] \\ &\leq \sum_{h \in \mathbb{H}(h^*/2)} \mathbb{E}_{P^{\otimes n}} \left[ \left( \frac{1}{2}\Omega(h^*)\lambda(h^*) + |\widehat{\xi}_h(P, x_0)| \right)^p \mathbb{1}_{\{\widehat{h}=h\}} \right]. \end{aligned}$$

Now, for any  $h \in \mathbb{H}(h^*/2)$ , we have

$$\begin{aligned} \{\widehat{h} = h\} &\subset \{\exists \eta \in \mathbb{H}(h), |\widehat{f}_{2h}(x_0) - \widehat{f}_\eta(x_0)| > \psi(2h, \eta)\} \\ &\subset \bigcup_{\eta \in \mathbb{H}(h)} \{\Omega(h^*)\lambda(h^*) + |\widehat{\xi}_{2h, \eta}(P, x_0)| > \psi(2h, \eta)\}, \end{aligned}$$

where  $\widehat{\xi}_{2h, \eta}(P, x_0) = \widehat{\xi}_{2h}(P, x_0) - \widehat{\xi}_\eta(P, x_0)$ , and where we used the triangle inequality and the definition of  $h^*$ . Now, we have  $\Omega(h^*)\lambda(h^*) \leq \Omega(2h)\lambda(2h)$  since  $2h \leq h^*$  and by definition of  $\psi(2h, \eta)$ , we infer

$$\{\widehat{h} = h\} \subset \bigcup_{\eta \in \mathbb{H}(h)} \{|\widehat{\xi}_{2h, \eta}(P, x_0)| > \Omega(\eta)\lambda(\eta)\}$$

so that

$$\mathbb{P}(\widehat{h} = h) \leq \sum_{\eta \in \mathbb{H}(h)} \mathbb{P}(|\widehat{\xi}_{2h, \eta}(P, x_0)| > \Omega(\eta)\lambda(\eta)) \quad (4.30)$$

$$\begin{aligned} &\leq \sum_{\eta \in \mathbb{H}(h)} \mathbb{P}\left(|\widehat{\xi}_{2h, \eta}(P, x_0)| > \sqrt{\frac{8\omega\lambda(\eta)}{n\eta^d}} + \frac{2\|K\|_\infty\lambda(\eta)}{n\eta^d}\right) \\ &\leq \sum_{\eta \in \mathbb{H}(h)} 2\exp(-\lambda(\eta)^2). \end{aligned} \quad (4.31)$$

For (4.30) we use the fact that  $\lambda(\eta) \geq 1$  and Bernstein's inequality on the random variable  $\widehat{\xi}_{2h, \eta}(P, x_0)$  for (4.31). Noticing now that  $\lambda(\eta)^2 \geq d\Theta \log(1/\eta)$ , we further obtain

$$\mathbb{P}(\widehat{h} = h) \leq 2h^{\Theta d} \times \sum_{j=0}^{\lfloor \log_2(1/h^-) \rfloor} 2^{-j\Theta d} \leq \frac{2}{1 - 2^{-\Theta d}} h^{\Theta d}.$$

For any  $h \in \mathbb{H}(h^*/2)$ , we thus get the following bound, using Cauchy-Schwarz inequality

$$\begin{aligned} &\mathbb{E}_{P^{\otimes n}} \left[ \left( \frac{1}{2} \Omega(h^*)\lambda(h^*) + |\widehat{\xi}_h(P, x_0)| \right)^p \mathbb{1}_{\{\widehat{h}=h\}} \right] \\ &\leq \mathbb{P}(\widehat{h} = h)^{1/2} E_{P^{\otimes n}} \left[ \left( \frac{1}{2} \Omega(h^*)\lambda(h^*) + |\widehat{\xi}_h(P, x_0)| \right)^{2p} \right]^{1/2} \\ &\leq 2^{(2p-1)/2} \sqrt{\frac{2}{1 - 2^{-\Theta d}}} h^{\Theta d/2} \left( 2^{-p} \Omega(h^*)^p \lambda(h^*)^p + c_{2p}^{1/2} \Omega(h)^p \right). \end{aligned} \quad (4.32)$$

We plan to sum over  $h \in \mathbb{H}(h^*/2)$  the RHS of (4.32). Notice first that

$$\sum_{h < h^*} h^{\Theta d/2} \leq (h^*)^{\Theta d/2} (1 - 2^{-\Theta d/2})^{-1}.$$

Moreover, for any  $h \geq h^-$ , we have  $\Omega(h) \leq 2\sqrt{2\omega/(nh^d)}$  by definition of  $h^-$ . It follows that

$$\Omega(h^*) \leq \Omega(h) \leq 2\Omega(h^*) \left( \frac{h^*}{h} \right)^{d/2}.$$

for any  $h \leq h^*$ . This enables us to bound the following sum

$$\begin{aligned} \sum_{h \in \mathbb{H}(h^*/2)} h^{\Theta d/2} \Omega(h)^p &\leq 2^p \Omega(h^*)^p h^{*pd/2} \sum_{h \in \mathbb{H}(h^*/2)} h^{\Theta d/2 - pd/2} \\ &\leq \frac{2^p}{1 - 2^{(p-\Theta)d/2}} \Omega(h^*)^p h^{*\Theta d/2} \end{aligned}$$

where we used that  $\Theta > p$ . Putting all these estimates together, using that  $h^* \leq 1$  and  $\lambda(h^*) \geq 1$ , we eventually obtain  $\mathcal{R}_{\mathcal{A}^c} \leq C_{\mathcal{A}^c} \Omega(h^*)^p \lambda(h^*)^p$ , with

$$C_{\mathcal{A}^c} = \frac{2^p}{\sqrt{1 - 2^{-\Theta d}}} \left( \frac{2^{-p}}{1 - 2^{-\Theta d/2}} + \frac{\sqrt{c_{2p}} 2^p}{1 - 2^{(p-\Theta)d/2}} \right).$$

In conclusion  $\mathbb{E}_{P^{\otimes n}}[|\widehat{f}(x_0) - f_P(x_0)|^p]^{1/p} \leq (C_{\mathcal{A}} + C_{\mathcal{A}^c}) \Omega(h^*)^p \lambda(h^*)^p$  which completes the proof.  $\square$

*Proof of Theorem 4.3.4.* Let  $P \in \Sigma_{\alpha, \beta}^d$  and let  $\bar{h} = (\rho \log n / n)^{1/(2\gamma+d)}$  with  $\gamma = \alpha \wedge \beta$  and for some constant  $\rho$  to be specified later. By Proposition 4.3.8 we know that for  $n$  large enough (depending on  $\rho, \alpha, \beta, d$ ) such that  $\bar{h} \leq \tau/2$ , we have  $|f_\eta(P, x_0) - f(x_0)| \leq C_1 \eta^\gamma$  for all  $\eta \leq \bar{h}$  with  $C_1$  depending on  $K, \ell, \alpha, \tau, L, \beta, f_{\max}$  and  $R$ . Moreover, we also have

$$\frac{2^{-2} \Omega(\bar{h})^2 \lambda(\bar{h})^2}{C_1^2 \bar{h}^{2\gamma}} \geq \frac{d\Theta 2\omega \log(1/\bar{h})}{4C_1^2 n \bar{h}^{2\gamma+d}} = \frac{d\Theta \omega (2\gamma + d)^{-1} \log n - \log \log n - \log \rho}{2C_1^2 \rho \log n}.$$

Thus, picking  $\rho = d\Theta \omega (2\gamma + d)^{-1} / (2C_1^2)$  yields  $C_1 \bar{h}^\gamma \leq \frac{1}{2} \Omega(\bar{h}) \lambda(\bar{h})$  for  $n$  large enough (depending on  $\rho$ ), and therefore  $\bar{h} \leq h^*(P, x)$ . By Lemma 4.A.10 this implies

$$\mathbb{E}_{P^{\otimes n}}[|\widehat{f}(x_0) - f_P(x_0)|^p]^{1/p} \leq C_2 \Omega(\bar{h}) \lambda(\bar{h})$$

where  $C_2$  depends on  $p$  and  $\Theta$ . But using that both  $\bar{h} \geq h^-$  and  $\lambda(\bar{h})^2 = d\Theta \log(1/\bar{h})$  for  $n$  large enough (depending on  $\rho, d, K$  and  $\Theta$ ), we also obtain

$$\begin{aligned} \Omega(\bar{h})^2 \lambda(\bar{h})^2 &\leq \frac{8\omega d\Theta \log(1/\bar{h})}{n \bar{h}^d} \\ &= \frac{8\omega d\Theta (2\gamma + d)^{-1} \log n - \log \log n - \log \omega}{\rho \log n} \bar{h}^{2\gamma} \leq 16C_1^2 \bar{h}^{2\gamma}. \end{aligned}$$

This last estimate yields

$$\mathbb{E}_{P^{\otimes n}}[|\widehat{f}(x_0) - f_P(x_0)|^p]^{1/p} \leq \left( 4C_1 C_2 \rho^{\gamma/(2\gamma+d)} \right) \left( \frac{\log n}{n} \right)^{\gamma/(2\gamma+d)}$$

for  $n$  large enough depending on  $\rho, \alpha, \beta, d, K$  and  $\Theta$ , which completes the proof.  $\square$

#### 4.A.7 Proofs of Section 4.3.5

*Proof of Proposition 4.3.13.* By the triangle inequality, for any  $P \in \Sigma_{\alpha, \beta}^d$ , we write

$$\begin{aligned} \mathbb{E}_{P^{\otimes n}}[|\widehat{f}_{\bar{h}}(\widehat{d}; x_0) - f_P(x_0)|^p]^{1/p} &\leq \left( \mathbb{E}_{P^{\otimes n}}[|\widehat{f}_{\bar{h}}(\widehat{d}; x_0) - f_P(x_0)|^p \mathbb{1}_{\{\widehat{d}=d\}}] \right)^{1/p} \\ &\quad + \left( \mathbb{E}_{P^{\otimes n}}[|\widehat{f}_{\bar{h}}(\widehat{d}; x_0) - f_P(x_0)|^p \mathbb{1}_{\{\widehat{d} \neq d\}}] \right)^{1/p}. \end{aligned}$$

The first term in the right-hand side has the right order thanks to Theorem 4.3.4. For the second one, using that  $|f(x_0)| \leq f_{\max}$  and

$$|f_{\widehat{h}}(\widehat{d}, x_0)| \leq \sup_{1 \leq d < D} \frac{\|K^{(\ell)}(d, \cdot)\|_{\infty}}{(h_d^-)^d} \lesssim n$$

up to a constant that depend on  $D, K$  and  $\ell$ , we infer

$$\left( \mathbb{E}_P \left[ |\widehat{f}_{\widehat{h}}(\widehat{d}; x_0) - f(x_0)|^p \mathbb{1}_{\widehat{d} \neq d} \right] \right)^{1/p} \lesssim \mathbb{P}(\widehat{d} \neq d)^{1/p} \times n.$$

Finally, since  $\widehat{d}$  satisfies Assumption 4.3.12, we have

$$\mathbb{E}_{P^{\otimes n}} [|\widehat{f}_{\widehat{h}}(\widehat{d}; x_0) - f_P(x_0)|^p]^{1/p} \lesssim \left( \frac{\log n}{n} \right)^{\alpha \wedge \beta / (2\alpha \wedge \beta + d)} + n^{-1/2}$$

for  $n$  large enough depending on  $p, \Theta, K, \ell, \alpha, \tau, L, \beta, f_{\max}, f_{\min}$  and  $R$ , so that the result indeed holds up to a constant depending on the same parameters and  $D$ .  $\square$

*Proof of Proposition 4.3.15.* Let  $P \in \Sigma_{\alpha, \beta}^d$  and  $\eta > 0$ . Assume that  $\widehat{P}_{\eta} > 0$ . We have

$$\begin{aligned} |\widehat{\delta}_{\eta} - d| &\leq |\log_2 \widehat{P}_{2\eta} - \log_2 P_{2\eta}| + |\log_2 \widehat{P}_{\eta} - \log_2 P_{\eta}| + |\log_2 P_{2\eta} - \log_2 P_{\eta} - d| \\ &\leq \frac{1}{\log 2} \left( \frac{|\widehat{P}_{2\eta} - P_{2\eta}|}{\widehat{P}_{2\eta} \wedge P_{2\eta}} + \frac{|\widehat{P}_{\eta} - P_{\eta}|}{\widehat{P}_{\eta} \wedge P_{\eta}} \right) + |\log_2 (P_{2\eta} / (2^d P_{\eta}))| \end{aligned}$$

We first consider the determinist term. For  $\eta \leq \tau/2$ , we have, writing  $r_{\eta} = \xi(\eta/\tau)\eta$  and using Lemma 4.A.3,

$$L_{2\eta}(1 - \eta^2/6\tau^2)(2\eta)^2 \zeta_d \leq P_{2\eta} \leq U_{2\eta}(1 + r_{2\eta}^2/\tau^2)r_{2\eta}^d \zeta_d$$

and

$$L_{\eta}(1 - \eta^2/6\tau^2)\eta^d \zeta_d \leq P_{\eta} \leq U_{\eta}(1 + r_{\eta}^2/\tau^2)r_{\eta}^d \zeta_d,$$

where  $L_{\eta} = \inf_{M \cap B(x_0, \eta)} f$  and  $U_{\eta} = \sup_{M \cap B(x_0, \eta)} f$ . Using again Lemma 4.A.3, we have that for  $\eta \leq \tau/2$ ,  $M \cap B(x_0, \eta) \subset \exp_{x_0} B_{T_{x_0}M}(0, 2\eta)$ , and, since  $2\eta \leq \pi\tau/2$  and that  $f \in \mathcal{F}_{\beta}$ , we know, using Lemma 4.A.6, that there exists  $R_1 > 0$  (depending on  $\beta, \tau, f_{\max}$  and  $R$ ) such that  $f(x) - R_1\eta \leq L_{\eta} \leq U_{\eta} \leq f(x) + R_1\eta$ . If  $\eta < \tau/4$ , the same bounds apply for  $2\eta$  and we thus obtain

$$(f(x) - R_1 2\eta)(1 - \eta^2/6\tau^2)(2\eta)^d \zeta_d \leq P_{2\eta} \leq (f(x) + R_1 2\eta)(1 + r_{2\eta}^2/\tau^2)r_{2\eta}^d \zeta_d$$

and

$$(f(x) - R_1 \eta)(1 - \eta^2/6\tau^2)\eta^d \zeta_d \leq P_{\eta} \leq (f(x) + R_1 \eta)(1 + r_{\eta}^2/\tau^2)r_{\eta}^d \zeta_d. \quad (4.33)$$

Using these two inequalities, and the fact that  $r_{\eta}/\eta \rightarrow 1$  as  $\eta \rightarrow 0$ , we find that  $|P_{2\eta}/(2^d P_{\eta}) - 1| \lesssim \eta$  up to a constant that depends on  $R_1, \tau$  and  $f_{\min}$ , for  $\eta$  small enough (depending

on  $R_1, \tau$  and  $f_{\min}$  as well). For the other terms, a simple use of Hoeffding's inequality yields for any  $\eta, \varepsilon > 0$ ,

$$\mathbb{P}(|\widehat{P}_\eta - P_\eta| > \varepsilon) \leq 2 \exp(-2n\varepsilon^2).$$

On the event  $\mathcal{A}_\eta = \{|\widehat{P}_\eta - P_\eta| \leq \varepsilon\}$ , we have moreover  $\widehat{P}_\eta \wedge P_\eta \geq P_\eta - \varepsilon$ . Setting  $\varepsilon = \eta^{d+1}$ , and using (4.33), we see that  $P_\eta - \varepsilon \gtrsim \eta^d$  for  $\eta$  small enough (depending on  $R_1, \tau$  and  $f_{\min}$ ). Thus, on the event  $\mathcal{A}_\eta \cap \mathcal{A}_{2\eta}$ , with probability at least  $1 - 4 \exp(-2n\eta^{2d+2})$ , we derive

$$|\widehat{\delta}_\eta - d| \lesssim \eta + \frac{\varepsilon}{P_\eta - \varepsilon} + \frac{\varepsilon}{P_{2\eta} - \varepsilon} \lesssim \eta, \quad (4.34)$$

for  $\eta$  small enough (depending on  $R_1, \tau$  and  $f_{\min}$ ), up to a constant that depends on  $R_1, \tau$  and  $f_{\min}$ . Now setting  $\eta = n^{-1/(2D+2)}$ , we have  $\widehat{d} = \widehat{\delta}_\eta = d$  on the event  $\mathcal{A}_\eta \cap \mathcal{A}_{2\eta}$  as soon as  $n$  is large enough so that the LHS of (4.34) is strictly smaller than  $1/2$ , ending the proof. □

# Chapter 5

## Bayesian manifold density estimation

We study the Bayesian density estimation of data living in the offset of an unknown submanifold of the Euclidean space. In this perspective, we introduce a new notion of anisotropic Hölder for the underlying density and obtain posterior rates that are mini-max optimal and adaptive to the regularity of the density, to the intrinsic dimension of the manifold, and to the size of the offset. Our Bayesian procedure appears to be convenient to implement and yields good practical results, even for quite singular data. This chapter is based on [\[BRR22\]](#).

### Contents

---

<b>5.1</b>	<b>Introduction</b>	<b>150</b>
5.1.1	Manifold density estimation	150
5.1.2	Our approach and contributions	151
5.1.3	Organisation of the Chapter	153
5.1.4	Notations	153
<b>5.2</b>	<b>Model : distributions concentrated near manifolds</b>	<b>154</b>
5.2.1	General anisotropic Hölder functions	155
5.2.2	Manifold anisotropic Hölder functions	156
5.2.3	Location-scale mixtures of normal priors	159
<b>5.3</b>	<b>Main results</b>	<b>162</b>
5.3.1	Posterior contraction rates	162
5.3.2	Approximating M-anisotropic densities	165
<b>5.4</b>	<b>Numerical experiments</b>	<b>166</b>
<b>5.5</b>	<b>Proofs of the main results</b>	<b>169</b>
5.5.1	Proof of Theorem 5.3.2	169
5.5.2	Proof of Theorem 5.3.7	172
<b>5.6</b>	<b>Discussion</b>	<b>178</b>

---

## 5.1 Introduction

### 5.1.1 Manifold density estimation

In many high dimensional statistical problems it is common to consider that the data has an intrinsic low dimensional structure. More precisely, statistics and computer sciences have seen a growing interest in the so-called *manifold hypothesis* where the data is believed to be supported (or near supported) on a low dimensional submanifold  $M$  of an ambient space (see [MF12] for an introduction).

There are good intuitive reasons to believe that real world data (such as natural images, sounds, texts, etc . . .) belong to the vicinity of a low dimensional submanifold, often due to physical constraints, see for instance [LV07] or [FMN16b]. Empirical evidence has also been shown in a number of important cases such as texts data sets [BN01], sounds [KPP70, BN01], images and videos [WS06b, KW13] or more recently in Covid data [VSFD<sup>+</sup>22]. Analysing such data sets is often called manifold learning. Manifold learning deal with either nonlinear dimension reduction techniques, manifold estimation or the construction of generative models and the estimation of the distribution on or near an unknown manifold. These problems are strongly connected. Dimension reduction consists in finding low dimensional representations of the data. This is typically done by constructing mappings as in Kernel PCA or graph based methods such as Isomap, Locally Linear Embeddings or Laplacian Eigenmaps.

Instead of estimating an embedding, the problem of reconstructing the manifold is another popular aspect of manifold learning, see [GPPVW12b], [AL19], [Div21a] or [DW21] among others. Finally the estimation of the distributions on or near manifolds and the construction of generative models have received recent wide interests in the statistics and machine learning community, specially with the developments of deep learning algorithms. There is a growing literature on generative models under the manifold hypothesis with many methodological developments around variational autoencoders (see [GBC16, Sec 14.6] or [KW13]), Generative adversarial networks (see [GPAM<sup>+</sup>14, AB17, ACB17] among others) or recent versions of normalizing flows (see [HP21]). The theoretical results associated to these approaches control the error between the true generative process and the estimated generative models typically under adversarial losses such as the Wasserstein distance, as in [TY22], since the focus is more on generating interesting samples than on estimating the distribution per se.

In this Chapter we study the estimation of the density in the vicinity of an unknown



submanifold  $M$  of the ambient space  $\mathbb{R}^D$ . Density estimation is an important class of problems in statistics and machine learning and in addition to being of interest in itself can be used as an intermediate steps in many tasks of unsupervize learning such as clustering, prediction or in ridge estimation ([GPPVW14, CGW15]). Density or distribution estimation under the exact manifold hypothesis (assuming that the data belong to a submanifold) has been studied theoretically for instance in [OG09] or [Div21b] under Wasserstein losses and in [BH21] under the pointwise loss. Assuming that the data belong exactly to a smooth manifold may be too restrictive since signals are often corrupted by noise. Hence in this Chapter we assume that the data belong to a neighbourhood  $M^\delta = \{x \in \mathbb{R}^D; d(x, M) \leq \delta\}$ , neither  $M$  nor  $d$  or the density  $f$  are considered known. This problem is studied in [CKKL21] in the special case of data corrupted with Gaussian noise and [MLD20] proposes a Bayesian nonparametric method to estimate a density on  $M^\delta$  based on mixtures of Fisher - Gaussian distributions for which they prove consistency under the assumption that the width  $\delta$  of the tube  $M^\delta$  is fixed.

### 5.1.2 Our approach and contributions

As far as we are aware there is no theoretical results on convergence rates - either from a frequentist or a Bayesian approach - for estimating a density on tubes  $M^\delta$  when  $M$  and  $\delta$  are unknown and  $\delta$  is possibly small. In this Chapter we bridge this gap and we propose a Bayesian nonparametric method based on specific families of location - scale mixtures of Gaussian distributions. We study the posterior concentration rates associated to these priors, i.e. the smallest possible  $\varepsilon_n$  such that

$$\Pi(d(f_0, f) \leq \varepsilon_n | X_1, \dots, X_n) \rightarrow 1,$$

in probability when the data  $X_1, \dots, X_n$  are a  $n$  sample from  $f_0$  and where  $\Pi(\cdot | X_1, \dots, X_n)$  denotes the posterior distribution, see [GGVDV00]. As is well known, when the distance  $d(\cdot, \cdot)$  is the Hellinger or the  $L_1$  metric, this posterior concentration rates induces also a convergence rate  $\varepsilon_n$  for the posterior mean  $\hat{f}$ , see for instance [GGVDV00]. Typically the rate  $\varepsilon_n$  depends on regularity properties of the density  $f_0$  and on the prior.

To do so we first define a general mathematical framework describing regularity properties of densities defined on possibly small neighbourhoods  $M^\delta$  of submanifolds  $M$ , with the idea that the density has a given smoothness  $\beta_0$  along the manifold  $M$  and another smoothness  $\beta_\perp$  along the normal to the manifold. This manifold driven anisotropic smoothness is defined in Section 5.2.2 and is an extension to anisotropic Hölder functions along coordinate axes. Building on that we show that the posterior concentration rate depends on  $\beta_0, \beta_\perp$  together with the dimension  $d$  of  $M$  and the width  $\delta$  of the tube. Interestingly the prior  $\Pi$  does not need to depend on  $\beta_0, \beta_\perp, d, \delta$  or

$M$  which makes the approach fully adaptive and the rate we obtain, at least when  $\delta$  is not too small is of order

$$n^{-\gamma} \text{ with } \gamma = \frac{\beta_0}{2\beta_0 + d + (D - d)\beta_0/\beta_\perp},$$

up to a  $\log n$  term, which is minimax.

Nonparametric location mixtures of Gaussians are known to be flexible models for densities, and adaptive minimax rates of convergence on Hölder types spaces have been obtained using Bayesian or frequentist estimation procedures based on location mixtures of normals, see [KRVDV10], [STG13], and [GVDV07] for Bayesian methods and [MRM13] for a penalized likelihood approach. However, location mixtures are not versatile enough since the covariance matrix remains fixed across the components, so we instead take advantage of the flexibility of location-scale mixtures of Gaussians. In [CDB17] the authors derive a suboptimal posterior concentration rate for isotropic positive Hölder densities on  $\mathbb{R}^D$ , while [MRM13] obtained minimax convergence rates for penalized maximum likelihood methods based on the same type of location-scale mixtures and [NR17] obtained also minimax posterior concentration rates using a hybrid location-scale mixture prior in the regression model. These results thus indicate that one has to be careful in designing the prior in nonparametric location-scale mixtures of Gaussians. The priors we consider in this Chapter are variants of location-scale mixture priors, see Section 5.2.3, which are flexible enough to adapt to the non linear or manifold driven smoothness of the class of densities studied here. This prior construction can also be seen as tiling the manifold by low-rank *Gaussian pancakes*, a method that is similar to mixtures of factors analyzers [GH96, DCS<sup>+</sup>10] or manifold Parzen windows [VB02] where, however, no theoretical guarantees on the estimation of the density were proven.

Hence our contributions are both methodological and theoretical. From a methodological point of view, we provide with a family of versatile priors (see 5.2) that are shown empirically and theoretically to perform very well in modelling data that are singularly supported near submanifolds. In particular we show empirically that these variants or location scale mixtures of Gaussian priors behave much better than the standard *conjugate* location-scale mixture of Gaussian prior, see Section 5.4.

From a theoretical point of view, we introduce a new notion of Hölder smoothness along a submanifold (see Section 5.2.2) which is proving to be adequate for the study of such almost-degenerated densities, and we derive posterior concentration rates for this new model (Section 5.3.1). The rates are optimal if the data do not collapse too quickly towards the manifold. These results rely on an intermediate result in approximation

theory which has interests in its own right and is provided in Section 5.3.2.

### 5.1.3 Organisation of the Chapter

In Section 5.2, we define manifold-anisotropic Hölder function, together with the families of priors we consider in the Chapter. Section 5.3 contains the main theoretical results and Section 5.4 the empirical results, with a description of the algorithms used to simulate from the posterior distribution. We provide in Section 5.5 proofs of the main results, namely the contraction rate Theorem 5.3.2 and the approximation Theorem 5.3.7. Some useful facts on manifolds are presented in Appendix 5.A. The other Appendices contain additional proofs and lemmata, as well as details on the numerical setting of Section 5.4.

### 5.1.4 Notations

For a multi-index  $k = (k_1, \dots, k_D) \in \mathbb{N}^D$ , we set  $|k| = k_1 + \dots + k_D$  and  $k! = k_1! \dots k_D!$ . For  $x \in \mathbb{R}^D$ , we write  $x^k = x_1^{k_1} \dots x_D^{k_D} \in \mathbb{R}$  and  $x_{\max}$  (resp.  $x_{\min}$ ) to be the maximal (resp. minimal) value of its entries. For any two indices  $i, j \in \{1, \dots, D\}$  with  $i \leq j$ , we set  $x_{i:j} = (x_i, \dots, x_j) \in \mathbb{R}^{j-i+1}$ . Finally, for a sufficiently regular function  $f : \mathbb{R}^D \rightarrow \mathbb{R}$ , we define its  $k$ -th partial derivative as

$$D^k f(x) = \frac{\partial^{|k|} f}{\partial x_1^{k_1} \dots \partial x_D^{k_D}}(x).$$

If  $M \subset \mathbb{R}^D$  is a measurable subset with Hausdorff dimension  $d$ , one denote  $\mu_M$  for the Borel measure  $\mu_M = \mathcal{H}^d(\cdot \cap M)$  where  $\mathcal{H}^d$  is the  $d$ -dimensional Hausdorff measure on  $\mathbb{R}^D$ . For  $r > 0$  and  $x \in \mathbb{R}^D$ , one write  $B_M(x, r) = B(x, r) \cap M$  where  $B(x, r)$  is the usual Euclidean ball of  $\mathbb{R}^D$ . If  $M$  is closed, then  $\text{pr}_M$  defines the (possible multi-valued) orthonormal projection from  $\mathbb{R}^D$  to  $M$ .

We will denote by  $\|\cdot\|$  the usual Euclidean norm of  $\mathbb{R}^k$  for any  $k \in \mathbb{N}^*$ . When  $\mathcal{L}$  is a linear map between such spaces, we write  $\|\mathcal{L}\|_{\text{op}}$  for the operator norm associated with the Euclidean norms. The notation  $\|\cdot\|_1$  (resp.  $\|\cdot\|_\infty$ ) will refer to both the  $L^1$ -norm (resp sup-norm) for vectors of  $\mathbb{R}^k$  for any  $k \in \mathbb{N}^*$ , and to the  $L^1$ -norm (resp sup-norm) for measurable functions from  $\mathbb{R}^k$  to  $\mathbb{R}$  for any  $k \in \mathbb{N}^*$ . The brackets  $\langle \cdot, \cdot \rangle$  will be used to denote the usual Euclidean product in  $\mathbb{R}^k$  for any  $k \in \mathbb{N}^*$ . For any matrix  $A \in \mathbb{R}^{k \times k}$ , the notation  $\|\cdot\|_A$  will refer to the quadratic form over  $\mathbb{R}^k$  defined by  $x \mapsto \langle Ax, x \rangle$ , which is a norm if  $A$  is positive. The set of orthogonal transform of  $\mathbb{R}^D$  will be denoted by  $\mathcal{O}(D, \mathbb{R})$ , or sometimes simply  $\mathcal{O}(D)$ .

For two positive functions  $f, g : \mathbb{R}^D \rightarrow \mathbb{R}$  we write the Hellinger distance as

$$d_H(f, g) = \left\{ \int_{\mathbb{R}^D} (\sqrt{f(x)} - \sqrt{g(x)})^2 dx \right\}^{1/2}.$$

In this Chapter,  $M$  will designate a closed submanifold of  $\mathbb{R}^D$  of dimension  $1 \leq d \leq D - 1$ . For any point  $x \in M$ , the tangent and normal spaces of  $M$  at  $x$  will be denoted  $T_x M$  and  $N_x M$ , and the corresponding bundles  $TM$  and  $NM$ . We write  $\exp_x : (T_x M, 0) \rightarrow (M, x)$  for the exponential map of  $M$  at point  $x$ . We let  $d_M(x, y)$  denote the intrinsic distance between  $x$  and  $y$  in  $M$ .

Finally we will use throughout the symbols  $\simeq$ ,  $\lesssim$  and  $\gtrsim$  to denote equalities or inequalities up to a constant, when the constant is not important.

## 5.2 Model : distributions concentrated near manifolds

We assume that we observe  $X_1, \dots, X_n$  independent and identically distributed from  $P_0$  on  $\mathbb{R}^D$  with density  $f_0$  with respect to Lebesgue measure. We assume that there is a low dimensional structure underlying our observations, i.e. that  $f_0$  has support concentrated near a low dimensional manifold  $M$  which is unknown. More precisely there exists  $\delta > 0$  unknown and typically small such that  $P_0(M^\delta) = 1$ , where  $M^\delta$  is the  $\delta$ -offset of  $M$ : it is the set of points that are at distance less than  $\delta$  from  $M$ ,

$$M^\delta := \bigcup_{x \in M} B(x, \delta) = \{z \in \mathbb{R}^D \mid d(z, M) \leq \delta\}.$$

A typical example is when the observations are noisy versions of data whose support is  $M$ :  $X = Y + Z$  with  $Y \in M$  and  $|Z| \leq \delta$  almost surely. When the noise  $Z$  has a density smoother than the density of  $Y$  on  $M$  (with respect to the Hausdorff measure), the density of  $X$  is anisotropic with a smoothness along the manifold  $M$  smaller than that along the normal directions. In this Chapter we thus aim at constructing priors which are flexible enough to lead to *good* estimation of  $f_0$  in situations where the density has a complex anisotropic structure in that it has an unknown smoothness  $\beta_0$  along an unknown manifold  $M$  and a different (larger) smoothness  $\beta_\perp$ , also unknown, along the normal spaces of the manifold. In this context, since the anisotropy varies spatially, it is therefore important to consider priors which adapt spatially to such *non linear smoothness*. In Section 5.2.3 below we consider certain families of location-scale mixtures with a careful modelling of the prior on the variance of the components and we show in Section 5.3 that these priors allow for manifold driven smoothness.

To begin with, we define what we think is a new notion of anisotropic Hölder spaces on the Euclidean space  $\mathbb{R}^D$ , and which happens to be a natural extension of the

usual notion of (isotropic) Hölder smoothness. We are aware that there exist various notions of anisotropic smoothness, see for instance [KLP01, HL02, CL13, GL11, GL14], with most of them stemming from the anisotropic smoothness as defined in [Nik12]. In all the aforementioned references, the anisotropy was consistently defined as a control of the variations of the partial derivatives along each axis separately, with no control of the cross-derivatives (and no guarantee that they, in fact, exist). While this is enough in a Euclidean framework, we argue that, to the best of our effort, we could not make such assumptions sufficient in our non-linear setting, as the proofs presented in Section 5.5 or in the Appendices might highlight. Instead, we come out with a new notion of Hölder anisotropy, in the footsteps of what [STG13] already sketched in their paper, that handles cross-derivatives in the same way that the usual notion of (isotropic) Hölder smoothness does, and which in fact coincides with the latter when the anisotropy vector is isotropic. This new class is defined in the first subsection below, and its main properties reviewed in Section 5.B.1.

### 5.2.1 General anisotropic Hölder functions

An anisotropic Hölder functions  $f : \mathbb{R}^D \rightarrow \mathbb{R}$  is, informally, a function whose smoothness is different along each axis of  $\mathbb{R}^D$ . Letting  $\beta = (\beta_1, \dots, \beta_D) \in (\mathbb{R}_+^*)^D$ , which will represent the regularity indices along each axis, we define

$$\alpha = (\alpha_1, \dots, \alpha_D) \text{ where } \alpha_i = \beta / \beta_i \in [0, D] \text{ and } \beta^{-1} = \frac{1}{D} \sum_i \beta_i^{-1}.$$

The coefficient  $\beta$  acts as the effective smoothness of the function  $f$ . Notice that  $\alpha_1 + \dots + \alpha_D = D$ . In this section, we define the spaces of anisotropic functions over bounded open subset of  $\mathbb{R}^D$ . We defer to Section 5.B.1 the introduction of the same class over general open subsets. We let  $\mathcal{U} \subset \mathbb{R}^D$  be a bounded open subset and  $L : \mathcal{U} \rightarrow \mathbb{R}_+$  be any non-negative function.

**Definition 5.2.1.** *The anisotropic Hölder spaces  $\mathcal{H}_{\text{an}}^\beta(\mathcal{U}, L)$  is the set of all functions  $f : \mathcal{U} \rightarrow \mathbb{R}^D$  satisfying:*

- i) *For any multi-index  $k \in \mathbb{N}^D$  such that  $\langle k, \alpha \rangle < \beta$ , the partial derivative  $D^k f$  is well defined on  $\mathcal{U}$  and  $|D^k f(x)| \leq L(x)$  for all  $x \in \mathcal{U}$ ;*
- ii) *For any multi-index  $k \in \mathbb{N}^D$  such that  $\beta - \alpha_{\max} \leq \langle k, \alpha \rangle < \beta$ , there holds*

$$|D^k f(y) - D^k f(x)| \leq L(x) \sum_{i=1}^D |y_i - x_i|^{\frac{\beta - \langle k, \alpha \rangle}{\alpha_i} \wedge 1} \quad \forall x, y \in \mathcal{U}. \quad (5.1)$$

See Figure 5.2.1 for a graphical representation of the quantities at stake. The function  $L$  acts as an upper-bound for the localized and anisotropic version of the usual

Hölder-norm:

$$\left\{ \max_{\langle k, \alpha \rangle < \beta} |D^k f(x)| \right\} \vee \max_{\beta - \alpha_{\max} \leq \langle k, \alpha \rangle < \beta} \sup_{y \in \mathcal{U}} \frac{|D^k f(y) - D^k f(x)|}{\sum_{i=1}^D |y_i - x_i|^{\frac{\beta - \langle k, \alpha \rangle}{\alpha_i} \wedge 1}}.$$

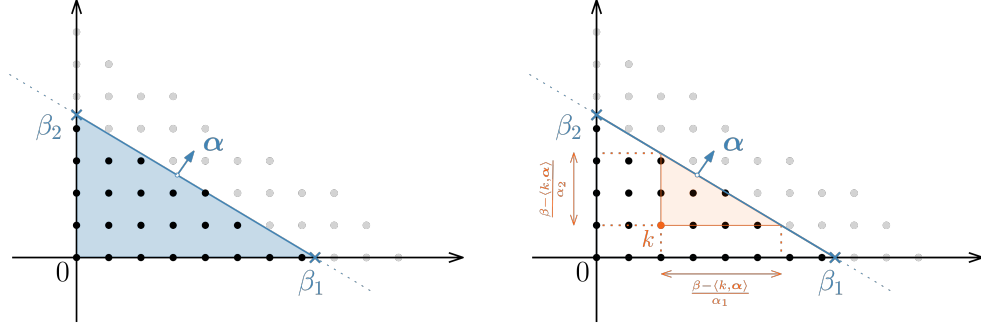


Figure 5.2.1 – An example in dimension  $D = 2$ . The vector  $\alpha$  is the only vector of 1-norm  $D$  which has positive coordinates and which is orthogonal to the simplex of vertices  $\{\beta_i e_i\}_{1 \leq i \leq D}$ . In black are the points  $k$  of  $\mathbb{N}^2$  such that  $\langle k, \alpha \rangle < \beta$ .

Note that the constraint on the intermediate derivatives  $D^k f(x)$  for  $0 < \langle k, \alpha \rangle < \beta$  may seem superfluous since some Kolmogorov-Landau type inequalities would yield some bounds on these derivatives, but we add them nonetheless to our functional class to simplify some notations. We list and prove in Section 5.B.1 various useful properties of functions in the anisotropic Hölder class. Also the function  $L$  in the definition of  $\mathcal{H}_{\text{an}}^\beta(\mathcal{U}, L)$  can be constant, in which case we will typically denote it  $C$ , to make it more explicit (leading to  $\mathcal{H}_{\text{an}}^\beta(\mathcal{U}, C)$ ).

*Remark 5.2.2.* The usual isotropic Hölder spaces are special case of our definition of  $\mathcal{H}_{\text{an}}^\beta(\mathcal{U}, L)$  corresponding to  $\beta = (\beta, \dots, \beta)$  with  $\beta > 0$ . In this case we write

$$\mathcal{H}_{\text{iso}}^\beta(\mathcal{U}, L) := \mathcal{H}_{\text{an}}^\beta(\mathcal{U}, L) \quad \text{for } \beta = (\beta, \dots, \beta).$$

As a final remark, we will use the same notations for the spaces of multivalued functions when their coordinate functions are all in the corresponding space. For instance, if  $\Psi : \mathcal{U} \rightarrow \mathbb{R}^D$ , then

$$\Psi = (\Psi_1, \dots, \Psi_D) \in \mathcal{H}_{\text{an}}^\beta(\mathcal{U}, L) \quad \Leftrightarrow_{\text{def}} \quad \Psi_i \in \mathcal{H}_{\text{an}}^\beta(\mathcal{U}, L) \text{ for all } i \in \{1, \dots, D\},$$

and the same holds for the other spaces defined in this subsection.

## 5.2.2 Manifold anisotropic Hölder functions

We now consider functions whose smoothness directions at point  $x \in \mathbb{R}^D$  are dependant on the position of  $x$  with respect to a given submanifold  $M \subset \mathbb{R}^D$  of dimension  $1 \leq$

$d \leq D - 1$ . More specifically, we extend the above notions of anisotropy to functions with a given regularity in the tangential directions of  $M$ , and of another regularity in the normal directions of  $M$ . We call such functions *manifold-anisotropic Hölder*, or sometimes simply *M-anisotropic*. To define such a class of function, we assume that  $M$  is a closed submanifold with reach bounded from below by  $\tau > 0$  (see Appendix 5.A for definition and properties of the reach) and we consider local parametrizations at any  $x_0 \in M$

$$\Psi_{x_0} : \mathcal{V}_{x_0} \rightarrow M,$$

where  $\mathcal{V}_{x_0}$  is a neighborhood of 0 in  $T_{x_0}M$ . The maps  $\Psi_{x_0}$  can be taken in a wide class of parametrizations of  $M$ . For instance, one could consider taking  $\Psi_{x_0}$  to be (close to) the inverse projection over  $M \rightarrow T_{x_0}M$  where  $T_{x_0}M$  is seen as an affine subspace of  $\mathbb{R}^D$  going through  $x_0$ , see for instance [AL19] or [Div21b]. For purely practical matter, we choose  $\Psi_{x_0}$  to be the exponential map  $\exp_{x_0}$ , although the results in this Chapter could be carried out with other well-behaved parametrizations, such as the one mentioned above. In particular, in the case of the exponential maps, we can define the domain of  $\Psi_{x_0}$  to be  $B_{T_{x_0}M}(0, \pi\tau)$ , see Appendix 5.A. In the rest of this Chapter, we set

$$\mathcal{V}_{x_0} := B_{T_{x_0}M}(0, \tau/8),$$

with factor  $1/8$  being there for technical reasons. If all the maps  $\Psi_{x_0}$  are of regularity  $\beta_M > 1$ , meaning that there exists a constant  $C_M > 0$  such that

$$\Psi_{x_0} \in \mathcal{H}_{\text{iso}}^{\beta_M}(\mathcal{V}_{x_0}, C_M), \quad \forall x_0 \in M \quad (5.2)$$

(in particular  $M$  is at least  $\mathcal{C}^k$  with  $k = \lceil \beta_M - 1 \rceil$ ), then one can construct a map

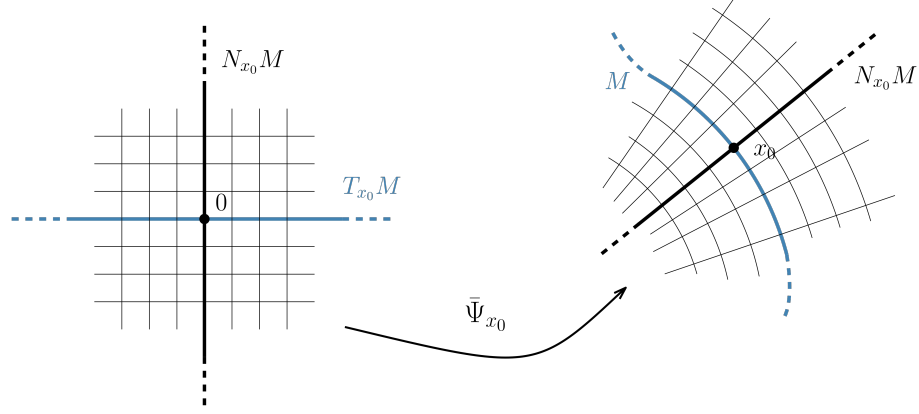
$$\bar{\Psi}_{x_0} : \begin{cases} \mathcal{V}_{x_0} \times N_{x_0}M & \rightarrow \mathbb{R}^D \\ (v, \eta) & \mapsto \Psi_{x_0}(v) + N_{x_0}(v, \eta). \end{cases}$$

where  $N_{x_0}(v, \cdot)$  is an isometry from  $N_{x_0}M$  to  $N_{\Psi_{x_0}(v)}M$  and where  $v \mapsto N_{x_0}(v, \cdot) \in \mathcal{H}_{\text{iso}}^{\beta_M-1}(\mathcal{V}_{x_0}, C_M^\perp)$  for some other constant  $C_M^\perp$  depending on  $C_M$ ,  $\tau$  and  $\beta_M$ . We refer to Appendix 5.A for further details concerning the construction of  $\bar{\Psi}_{x_0}$  and the proof of its regularity. When restricting the latter map, one get a local parametrization of the offset  $M^{\tau/2}$  around  $x_0$

$$\bar{\Psi}_{x_0} : \mathcal{V}_{x_0} \times B_{N_{x_0}M}(0, \tau/2) \rightarrow M^{\tau/2}$$

as shown in Lemma 5.A.2. This parametrization is such that  $\text{pr}_M(\bar{\Psi}_{x_0}(v, \eta)) = \Psi_{x_0}(v)$  for any  $(v, \eta) \in \mathcal{V}_{x_0} \times B_{N_{x_0}M}(0, \tau/2)$  and  $\bar{\Psi}_{x_0}$  is a diffeomorphism from  $\mathcal{V}_{x_0} \times B_{N_{x_0}M}(0, \tau/2)$  to its image which satisfies

$$\bar{\Psi}_{x_0} \in \mathcal{H}_{\text{iso}}^{\beta_M-1}(\mathcal{V}_{x_0} \times B_{N_{x_0}M}(0, \tau/2), C_M^*) \quad (5.3)$$


 Figure 5.2.2 – A visual interpretation of the parametrization  $\bar{\Psi}_{x_0}$ .

for some  $C_M^* > 0$  depending on  $C_M$ ,  $\tau$  and  $\beta_M$ . See Figure 5.2.2 for a visual interpretation of this parametrizations.

For any  $\delta > 0$ , we define  $\bar{\Psi}_{x_0,\delta}(v,\eta) := \bar{\Psi}_{x_0}(v,\delta\eta)$  to be the rescaled version of  $\bar{\Psi}_{x_0}$  in the normal directions. It is a well defined parametrization of  $M^{\tau/2}$  on the set  $\mathcal{W}_{x_0,\delta} := \mathcal{V}_{x_0} \times B_{N_{x_0}M}(0, \tau/2\delta)$ . We let  $\beta_0, \beta_\perp$  be two positive real numbers, and define the vector

$$\beta_{0,\perp} = (\underbrace{\beta_0, \dots, \beta_0}_d, \underbrace{\beta_\perp, \dots, \beta_\perp}_{D-d}) \in \mathbb{R}^D.$$

Now for any function  $L : \mathbb{R}^D \rightarrow \mathbb{R}_+$ , we define:

**Definition 5.2.3.** Let  $L : \mathbb{R}^D \rightarrow \mathbb{R}_+$  be a function; the class  $\mathcal{H}_\delta^{\beta_0, \beta_\perp}(M, L)$  is the set of all functions  $f : \mathbb{R}^D \rightarrow \mathbb{R}$  which satisfy:

- i)  $f$  is supported on  $M^\delta$ ;
- ii) For any  $x_0 \in M$ , set  $\bar{f}_{x_0,\delta} := \delta^{D-d} f \circ \bar{\Psi}_{x_0,\delta}$  and  $L_{x_0,\delta} := \delta^{D-d} L \circ \bar{\Psi}_{x_0,\delta}$ , then

$$\bar{f}_{x_0,\delta} \in \mathcal{H}_{\text{an}}^{\beta_{0,\perp}}(\mathcal{W}_{x_0,\delta}, L_{x_0,\delta}). \quad (5.4)$$

Informally, such a function is  $\beta_0$ -Hölder along the manifold  $M$ , and  $\beta_\perp$ -Hölder normal to the manifold  $M$ . The normalization  $\delta^{D-d}$  accounts for the scaling  $\eta \mapsto \delta\eta$  along the normal spaces (which are of dimension  $D-d$ ) in the definition of  $\bar{\Psi}_{x_0,\delta}$ . Its presence is natural and can be understood as follows: when  $f$  is a density supported on  $M^\delta$ , the typical magnitude of its values is of order  $1/\delta^{D-d}$ , and the absence of normalization would whence make the above functional class irrelevant to describe the regularity of such densities.

M-anisotropic functions happen to be a convenient way to describe the regularity of a number of densities that are naturally supported around  $M$ . To illustrate this, take



$f_* : M \rightarrow \mathbb{R}$  to be a  $\beta_0$ -Hölder density, meaning that there exists  $L_0 : M \rightarrow \mathbb{R}$  such that for any  $x_0 \in M$ ,

$$f_* \circ \Psi_{x_0} \in \mathcal{H}_{\text{iso}}^{\beta_0}(\mathcal{V}_{x_0}, L_0 \circ \Psi_{x_0}).$$

Now take  $K : \mathbb{R}^D \rightarrow \mathbb{R}$  to be a normalized positive smooth isotropic kernel supported on  $B(0, 1)$ . We introduce  $c_\perp^{-1} = \int K(\varepsilon) d\mu_E(\varepsilon)$  where  $E$  is any (through isotropy)  $(D-d)$ -dimensional subspaces of  $\mathbb{R}^D$ . We also assume that  $K \in \mathcal{H}_{\text{iso}}^{\beta_\perp}(\mathbb{R}^D, L_\perp)$  for some function  $L_\perp$  which is also rotationally invariant.

**Proposition 5.2.4.** *Let  $f$  be the density of a random variable  $Z = X + \delta\mathcal{E}$  where  $X \sim f_*(x)\mu_M(dx)$  and  $0 < \delta < \tau$ . Then,*

1. (Orthonormal noise) *If  $\beta_\perp \leq \beta_M - 1$ , and if  $\mathcal{E}|X \sim c_\perp K(\varepsilon)\mu_{N_X M}(d\varepsilon)$ , then*

$$f \in \mathcal{H}_\delta^{\beta_0, \beta_\perp}(M, L), \quad \text{with,} \quad L(x) := C\delta^{-(D-d)} L_0(\text{pr}_M x) \times L_\perp(x - \text{pr}_M x), \quad C > 0.$$

2. (Isotropic noise) *If  $\delta < \tau/32$  and  $\beta_0 \leq \beta_\perp \leq \beta_M - 1$ , and if  $\mathcal{E} \sim K(\varepsilon)d\varepsilon$ , independently of  $X$ , then*

$$f \in \mathcal{H}_\delta^{\beta_0, \beta_\perp}(M, L), \quad \text{with} \quad L(x) := C\delta^{-D} \int_M L_\perp\left(\frac{x-y}{\delta}\right) L_0(y)\mu_M(dy), \quad C > 0.$$

In both cases  $C$  depends on  $C_M, \tau, \beta_M, \beta_0, \beta_\perp$ .

See Section 5.A.2 for a proof of this result.

**Remark 5.2.5.** Throughout the Chapter we assume that the true density  $f_0 \in \mathcal{H}_\delta^{\beta_0, \beta_\perp}(M, L)$ , which implies that  $P_0(M^\delta) = 1$ . However it is enough to assume that  $P_0(M^\delta) \geq 1 - o(1/n)$  where  $n$  is the number of observations. This makes no difference in terms of the results presented in Section 5.3. The weaker assumption  $P_0(M^\delta) \geq 1 - o(1/n)$  is for instance fulfilled in the additive noise model (see Proposition 5.2.4) with  $Z = X + \frac{\delta}{\sqrt{C \log n}} \mathcal{E}$  with the Gaussian kernel  $K(x) := (2\pi)^{D/2} \exp(-\|x\|^2/2)$ .

In the following section we describe the family of priors which we use to estimate the above family of densities.

### 5.2.3 Location-scale mixtures of normal priors

We model the manifold-anisotropic Hölder densities using location-scale mixtures of normals. We parametrize the covariances of the components by  $\Sigma = O^T \Lambda O$  where  $O$  is a unitary matrix and  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_D)$  is diagonal. Location-scale mixtures can then be written as:

$$f_P(x) = \int_{\mathbb{R}^D} \varphi_{O^T \Lambda O}(x - \mu) dP(\mu, O, \Lambda), \quad P = \sum_{k=1}^K p_k \delta_{(\mu_k, O_k, \Lambda_k)}, \quad K \in \mathbb{N} \cup \{+\infty\}, \quad (5.5)$$

where, for any positive definite matrix  $\Sigma$ ,

$$\varphi_{\Sigma}(z) := \frac{1}{\det^{1/2}(2\pi\Sigma)} \exp\left\{-\frac{1}{2}\|z\|_{\Sigma^{-1}}^2\right\},$$

is the density of a centered Gaussian with covariance matrix  $\Sigma$ . The two most well known families of priors on  $P$  are Dirichlet process priors and mixtures with random number of components, also known as mixtures of finite mixtures. Recall that if  $P$  follows a Dirichlet process priors with parameters  $A$  and  $H$  where  $A > 0$  and  $H$  is a probability measure on some measurable space  $\Theta$ , then

$$P = \sum_{k=1}^{\infty} p_k \delta_{\theta_k} \quad \text{with} \quad p_k = V_k \prod_{i < k} (1 - V_i), \quad V_i \stackrel{iid}{\sim} \text{Beta}(1, A) \quad \text{and} \quad \theta_k \stackrel{iid}{\sim} H.$$

If  $P$  follows a mixture of finite mixtures prior of parameters  $\alpha_K$  and  $\pi_K$  where  $\alpha_K > 0$  and  $\pi_K$  is a probability measure on  $\mathbb{N}$ , then

$$P = \sum_{k=1}^K p_k \delta_{\theta_k} \quad \text{with} \quad K \sim \pi_K, \quad (p_1, \dots, p_K) | K \sim \mathcal{D}(\alpha_K, \dots, \alpha_K) \quad \text{and} \quad \theta_k \stackrel{iid}{\sim} H.$$

In both cases (Dirichlet process and mixture of finite mixtures) we call  $H$  the base probability measure. Obviously in the case of mixtures of finite mixtures the conditional prior on  $(p_1, \dots, p_K)$  and  $\theta_1, \dots, \theta_K$  could be different but we consider this setup for the sake of simplicity.

As shown empirically by [MLD20] location-scale Dirichlet process mixtures with base measure constructed from the conjugate prior of the Gaussian model are not well adapted to the problem at hand. We show however that if particular care is given to the choice of  $H$ , the posterior on manifold-anisotropic density is well behaved. In particular we consider the two following types of location-scale mixtures:

- *Partial location-scale mixtures:* The eigenvalues  $\Lambda$  of the covariance of the Gaussians are common accross components,

$$f_{\Lambda, P}(x) = \int_{\mathbb{R}^D} \varphi_{O^T \Lambda O}(x - \mu) dP(\mu, O), \quad P = \sum_{k=1}^K p_k \delta_{(\mu_k, O_k)}, \quad K \in \mathbb{N} \cup \{+\infty\} \quad (5.6)$$

where  $P$  is a probability distribution on  $\mathbb{R}^D \times \mathcal{O}(D)$  (where  $\mathcal{O}(D)$  is the set of unitary matrices in  $\mathbb{R}^D$ ) and is either a Dirichlet process prior or a mixture of finite mixtures.

- *Hybrid location-scale mixtures:* The density  $f_P$  is written as (5.5) where  $P$  conditionally on a probability  $Q_2$  on  $\mathbb{R}_+^D$  follows a Dirichlet process mixture or a mixture of finite mixtures with base measure  $H_0(d\mu, dO, d\lambda) = H_1(d\mu, dO) \otimes Q_2(d\lambda)$ , and  $Q_2$  follows a distribution  $\tilde{\Pi}_{\Lambda}$ .

We denote by  $\Pi$  the prior on the parameter and we consider the following assumptions on  $\Pi$ . These conditions differs whether  $\Pi$  is assumed to come from a partial location-scale mixture prior or a hybrid location-scale mixture prior.

**Conditions on the partial location-scale mixtures.**  $f$  is modelled as in (5.6) and  $P$  follows either a Dirichlet process with base measure  $H$  or a mixture of finite mixtures with base measure  $H$  and prior on  $K$  satisfying

$$-\log \Pi_K(K = x) \simeq x(\log x)^r, \quad r = 0, 1. \quad (5.7)$$

Here  $r = 0$  corresponds to the geometric prior on  $K$  and  $r = 1$  to the Poisson one. The base measure  $H(d\mu, dO) = h(\mu, O)d\mu dO$  where  $d\mu$  designates the Lebesgue measure on  $\mathbb{R}^D$  and  $dO$  the Haar measure on  $\mathcal{O}(D)$ . and we further assume that there exists  $c_1, b_1 > 0$  and  $b_2 > 2D - 1$  such that

$$e^{-c_1 \|\mu\|^{b_1}} \lesssim h(\mu, O) \lesssim (1 + \|\mu\|)^{-b_2} \quad \text{with} \quad \forall \mu, O. \quad (5.8)$$

We also assume that  $\Lambda$  is drawn from a probability measure  $\Pi_\Lambda$  that has a density  $\pi_\Lambda$  with respect to Lebesgue measure on  $\mathbb{R}^D$ , and that this density satisfies: there exists  $c_2, c_3, b_3 > 0$  and  $b_4 > D(D - 1)/2$  such that

$$\begin{aligned} e^{-c_2 \sum_{i=1}^D \lambda_i^{-d/2}} &\lesssim \pi_\Lambda(\lambda_1, \dots, \lambda_D) \quad \text{for small } \lambda_1, \dots, \lambda_D \in (\mathbb{R}_+^*)^D, \\ \Pi_\Lambda \left( \min_{1 \leq i \leq D} \lambda_i < x \right) &\lesssim e^{-c_3 x^{-b_3}} \quad \text{for small } x > 0, \\ \text{and} \quad \Pi_\Lambda \left( \max_{1 \leq i \leq D} \lambda_i > x \right) &\lesssim x^{-b_4} \quad \text{for large } x > 0. \end{aligned} \quad (5.9)$$

Condition (5.8) is weak and is for instance satisfied as soon as  $\mu$  and  $O$  are independent under  $H$  with positive and continuous density for  $O$  and positive density for  $\mu$  with weak tail assumptions. Condition (5.9) is also weak and common in the case of location Gaussian mixtures and is verified in particular if the  $\sqrt{\lambda_i}$ 's are independent inverse Gammas under  $\Pi_\Lambda$ , or if the  $\lambda_i$ 's are independent inverse Gammas and  $d \geq 2$ .

**Conditions on the hybrid location-scale mixtures.**  $H_1$  satisfies (5.8) and  $Q_2$  is random with distribution  $\tilde{\Pi}_\lambda$  which satisfies: for all  $b > 0$  there exists  $B_0, c_2 > 0$  such that for  $2x_1 \leq x_2$  both small,

$$\tilde{\Pi}_\Lambda \left[ Q_2 \left( [x_1, x_1(1 + x_1^b)]^d \times [x_2, x_2(1 + x_1^b)]^{D-d} \right) \geq x_1^{B_0} \right] \gtrsim e^{-c_2 x_1^{-d/2}}. \quad (5.10)$$

Moreover we assume that for some positive constant  $c_3, b_3, c_4, b_4 > 0$  such that for  $x > 0$  small,

$$\begin{aligned} \mathbb{E}_{\tilde{\Pi}_\Lambda} \left[ Q_2 \left( \min_{1 \leq i \leq D} \lambda_i \leq x \right) \right] &\lesssim e^{-c_3 x^{-b_3}} \quad \text{for small } x, \\ \mathbb{E}_{\tilde{\Pi}_\Lambda} \left[ Q_2 \left( \max_{1 \leq i \leq D} \lambda_i > x \right) \right] &\lesssim e^{-c_4 x^{b_4}} \quad \text{for large } x. \end{aligned} \quad (5.11)$$

*Remark 5.2.6.* One can view the partial location-scale mixture as a special instance of the hybrid location-scale mixture defined above: take  $Q_2$  to be the Dirac mass at a value  $\Lambda$  where  $\Lambda \sim \Pi_\Lambda$ .

Conditions (5.10) and (5.11) are in particular satisfied if  $Q_2$  comes from a Dirichlet process. More precisely, if  $Q_2$  is of the form

$$Q_2(d\lambda) = \prod_{i=1}^D Q_0(d\lambda_i) \quad \text{with} \quad Q_0 \sim \text{DP}(BH_\lambda),$$

with  $B > 0$ , then (5.10) and (5.11) are satisfied for reasonable choices of probability distribution  $H_\lambda$  on  $\mathbb{R}_+$ . We show in the next proposition that this is in particular true when  $H_\lambda$  is a square-root- inverse Gamma.

**Proposition 5.2.7.** *Assume that  $Q_2 = Q_0^{\otimes D}$  where  $Q_0 \sim \text{DP}(BH_\lambda)$ , where  $B > 0$  and under  $H_\lambda$ ,  $\sqrt{\lambda_i}$  follow a truncated inverse Gamma with parameters  $a_1, a_2 > 0$ , and truncation parameter  $T \ll n^\omega$  for some  $\omega > 0$ . Then conditions (5.10) and (5.11) are satisfied.*

A proof of Proposition 5.2.7 can be found in Appendix 5.B.3.

*Remark 5.2.8.* Although the conditions on the prior: (5.9) and (5.10) depend on  $d$ , they are satisfied for all  $d$  by setting  $d = 1$  and in particular they are verified for all  $d \geq 1$  if  $\sqrt{\lambda_i}$  follow an inverse Gamma under the base measure, which is agnostic to  $d$ .

	MFM	DPM	Partial	Hybrid
Conditions	(5.7)+(5.8)	(5.8)	(5.9)	(5.10)+(5.11)

Table 5.2.1 – Summary table of the required conditions depending on the type of mixture and the type of scale sampling.

## 5.3 Main results

### 5.3.1 Posterior contraction rates

Recall that  $X_1, \dots, X_n$  is an  $n$  sample drawn from a distribution  $P_0$  with density  $f_0$ . This density is concentrated around a submanifold  $M$ , with a given smoothness  $\beta_0$  along the manifold and a typically much larger smoothness  $\beta_\perp$  along the normal spaces. More precisely, we will assume:

- *Conditions on  $M$ :* the submanifold  $M$  is of dimension  $d$  and has a reach greater than  $\tau > 0$ . Furthermore, there exists  $\beta_M > 2$  and  $C_M > 0$  such that  $\Psi_{x_0} \in \mathcal{H}_{\text{iso}}^{\beta_M}(\mathcal{V}_{x_0}, C_M)$ ,  $\forall x_0 \in M$ . In particular,  $M$  also satisfies (5.3).

- *Conditions on  $f_0$ :* the density  $f_0$  is in  $\mathcal{H}_\delta^{\beta_0, \beta_\perp}(M, L)$ . Furthermore, there exists  $c_5 > 0$  and  $\kappa > 0$ ,

$$f_0(x) \lesssim e^{-c_5 \|x\|^\kappa} \quad \forall x \in \mathbb{R}^D, \quad (5.12)$$

and for some  $\omega > 6\beta$  and  $C_0 < \infty$ ,

$$\int_{\mathcal{W}_{x_0, \delta}} \left| \frac{D^k \bar{f}_{x_0, \delta}}{\bar{f}_{x_0, \delta}} \right|^{\omega / \langle k, \alpha \rangle} \bar{f}_{\delta, x_0} \leq C_0 \quad \text{and} \quad \int_{\mathcal{W}_{x_0, \delta}} \left| \frac{L_{x_0, \delta}}{\bar{f}_{x_0, \delta}} \right|^{\omega / \beta} \bar{f}_{\delta, x_0} \leq C_0. \quad (5.13)$$

for all  $\delta$  small,  $x_0 \in M$  and all  $0 \leq \langle k, \alpha \rangle < \beta$

- *Conditions on  $\Pi$ :* the prior  $\Pi$  is originating from a Partial / Hybrid location-scale mixture of finite mixtures / Dirichlet process mixture satisfying the conditions displayed in the Table 5.2.1.

**Remark 5.3.1.** Note that the conditions regarding the prior  $\Pi$  do not involve  $M$ ,  $\delta$ ,  $L$ ,  $\beta$ , or  $\tau$ : they are regarded as unknown in this framework. In fact, the only feature of  $M$  or  $f_0$  from which  $\Pi$  seems to depend is the intrinsic dimension  $d$ , through (5.9) or (5.10). However, as noted in Remark 5.2.8 we can choose priors which do not depend on  $d$  and such that these assumptions are verified for all  $d \geq 1$ .

In the rest of this Chapter the symbols  $\simeq$ ,  $\lesssim$  and  $\gtrsim$  denote equalities or inequalities up to a constant depending on  $D$ ,  $d$ ,  $\tau$ ,  $\beta_M$ ,  $C_M$ ,  $\beta_0$ ,  $\beta_\perp$  and all the other parameters appearing in conditions (5.7) to (5.13).

**Theorem 5.3.2.** Let  $X_1, \dots, X_n$  be a  $n$ -sample from  $f_0$ . We assume that  $\omega > 6\beta$  is large enough so that  $n^{-\frac{\omega-6\beta}{2\beta+D}} = o(\delta^{D-d})$  and that  $\beta_0 \leq \beta_\perp \leq \beta_M - 3$ . Then, under the conditions stated above,

$$\Pi(d_H(f_P, f_0) \geq \varepsilon_n \mid \mathcal{X}_n) \xrightarrow[n \rightarrow \infty]{} 0 \quad \text{in } \mathbb{P}_0^{\otimes n}\text{-probability}$$

where

$$\varepsilon_n \simeq \log^p n \times \left\{ \frac{1}{\sqrt{n \delta^{\frac{D}{\alpha_0 - \alpha_\perp}}}} \vee n^{-\frac{\beta}{2\beta+D}} \right\},$$

with  $p > 0$  depending on  $D$ ,  $\kappa$  and  $\beta$ .

**Remark 5.3.3.** The case where  $\beta_\perp$  is infinite is particularly of interest. Then,  $\beta \rightarrow \beta_0$ ,  $\alpha_0 \rightarrow D/d$ ,  $\alpha_\perp \rightarrow 0$  and the rate  $\varepsilon_n$  becomes

$$\varepsilon_n^\infty = \log^p(n) \times \left\{ \frac{1}{\sqrt{n \delta^d}} \vee n^{-\frac{\beta_0}{2\beta_0+d}} \right\},$$

which is, when  $\delta$  is not too small (i.e  $\delta \gtrsim n^{-1/(2\beta_0+d)}$ ), the minimax rate for estimating a  $\beta_0$  Hölder density in  $\mathbb{R}^d$ , up to a log term. Here the strength of our result lies in that the manifold (and thus the support of  $f_0$ ) is unknown and the prior depends neither on  $\beta$  nor  $\delta$  or  $d$  (or  $M$ ).

Since the class of densities contains the case where  $M$  is a  $d$  dimensional subspace of  $\mathbb{R}^D$ , when  $\delta \gtrsim n^{-(\alpha_0 - \alpha_\perp)/(2\beta + D)}$  the rate  $\varepsilon_n$  is the minimax estimation rate (up to a  $\log n$  term). Although no proof of a minimax bound exists in our framework, a careful look at the proof of [GL14, Thm 4 (ii)] for  $p = 1$ ,  $r = (\infty, \dots, \infty)$ ,  $s = \infty$  and  $\theta = 1$  (tail dominance from (5.12)) show that the lower-bound translates in our context. Indeed the densities used to derived the lower bound are obtained from a smooth density with additive perturbations of the form  $x \mapsto h^\beta \prod g(x_i/h^{\alpha_i})$  where  $h > 0$  and  $g$  is a smooth and compactly supported function of zero-mean. Such a perturbations belong to anisotropic Hölder classes defined in Section 5.2.1.

*Remark 5.3.4.* A refinement of [STG13] shows in fact that some location mixtures priors are able to achieve the rate of  $n^{-\frac{\beta_0}{2\beta_0 + d + (D-d)\frac{\beta_0}{\beta_\perp}}}$  (up to a  $\ln$  factor) in the models 5.2.4 when the manifold  $M$  is in fact a  $d$ -dimensional subspace of  $\mathbb{R}^d$ , thanks to the simple euclidian anisotropic smoothness property of the underlying density (up to an orthogonal change of basis adapted to the linear subspace). Our result is a nonlinear extension of this work.

*Remark 5.3.5.* Because the approximation results of Subsection 5.3.2 are stable under finite mixtures, so do the results of Theorem 5.3.2. In particular, the support of  $f_0$  can be a finite union of submanifolds  $M_i$  with non trivial intersections and with each  $M_i$  fulfilling (5.2). See Figure 5.3.1 for a diagram of such a situation. Hence the assumption on the lower bound on the reach can be significantly weakened. We consider such an example in the simulations of Section 5.4.

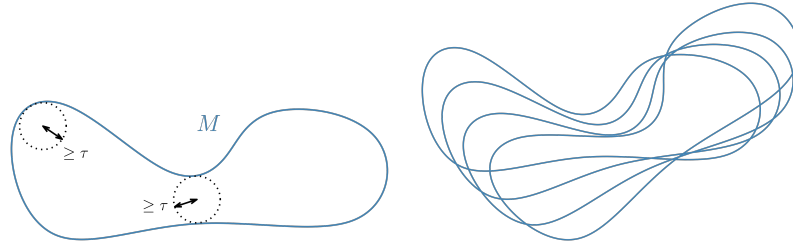


Figure 5.3.1 – (Left) An example of smooth submanifolds with a reach constrained to be greater than  $\tau$  and (Right) a finite union of such manifolds. Both subsets are admissible as the (near) support for a density  $f_0$  satisfying the contraction rates displayed in Theorem 5.3.2, as explained in Remark 5.3.5.

Assumptions (5.12) and (5.13) are common assumptions in density estimation based on mixtures of Gaussians, see for instance [KRVDV10] or [STG13] for the Bayesian approaches and [MRM13] for the frequentist approaches. They are rather weak assumptions. The difficulty with (5.13) is that it is expressed on  $\bar{f}_{x_0, \delta}$ , which is natural in our context since the smoothness assumption on  $f_0$  is expressed in terms of  $\bar{f}_{x_0, \delta}$ , but is not so intuitive. However, a careful examination of (5.13) show that this assumption

is for instance implied by the stronger, but chart-independent assumption:

$$\int_{\mathbb{R}^D} \left\{ \frac{L(x)}{f_0(x)} \right\}^{\omega_*} f_0(x) dx \leq C_0,$$

for some large  $\omega_*$ . Moreover, to understand better what (5.13) means in terms of  $f_0$ , we illustrate it in our archetypal model where  $f_0$  is the density of  $X = Y + \delta\mathcal{E}$  as in Proposition 5.2.4. We then have the following result:

**Lemma 5.3.6.** *Under the conditions of Proposition 5.2.4 and if*

$$\int_M \left\{ \frac{L_0(x)}{f_*(x)} \right\}^{\omega_*} f_*(x) d\mu_M(x) < \infty \quad \text{and} \quad \int_{B(0,1)} \left\{ \frac{L_\perp(\eta)}{K(\eta)} \right\}^{\omega_*} K(\eta) d\eta < \infty,$$

*for some  $\omega_*$  large. Then (5.13) is satisfied in both the orthonormal noise model and in the isotropic noise model.*

The proof of Lemma 5.3.6 can be found in Section 5.B.2. Note that the conditions in Lemma 5.3.6 are fulfilled by a number of natural kernels or densities such as  $K(\eta) \approx (1 - \|\eta\|^2)_+^p$  and  $L_\perp \approx (1 - \|\eta\|^2)_+^{p-\beta_\perp}$  with large  $p$  or  $K(\eta) \approx \exp(-(1 - \|\eta\|^2)_+^{-1})$  and  $L_\perp(\eta) \approx (1 - \|\eta\|^2)^{-\beta_M} K(\eta)$ , see Lemma 5.A.6 for further details.

### 5.3.2 Approximating M-anisotropic densities

Theorem 5.3.2 is proved using the approach of [GVDV07], with a control on the prior mass of Kullback-Leibler neighbourhoods of  $f_0$  and on the entropy of the support of the prior. The main difficulty in our setup is in proving the Kullback-Leibler prior mass condition. To do so, we need to construct an *efficient* approximation of  $f_0$  by mixtures of Gaussian. This construction is of interest in its own as it sheds light on the behaviour of such mixtures and on the geometry of M-anisotropic densities.

To explain the construction, we denote, for any  $x \in M^\tau$ ,  $T_x = T_{\text{pr}_M(x)}M$  and  $N_x = N_{\text{pr}_M(x)}M$ . We also write  $\Sigma(x) = O_x^\top \Delta_{\sigma,\delta}^2 O_x$  where  $O_x$  is the matrix in the canonical basis of  $\mathbb{R}^D$  and in arbitrary orthonormal basis of  $T_x$  and  $N_x$  of the linear map  $z \mapsto (\text{pr}_{T_x} z, \text{pr}_{N_x} z)$  and where

$$\Delta_{\sigma,\delta} = \begin{pmatrix} \sigma^{\alpha_0} \text{Id}_d & 0 \\ 0 & \delta \sigma^{\alpha_\perp} \text{Id}_{D-d} \end{pmatrix}.$$

Note that  $\Sigma_x$  does not depends on the choice of an orthonormal basis of  $T_x$  and  $N_x$  since for any orthonormal base change  $P$  that preserves  $T_x$ , one have  $P^\top \Delta_{\sigma,\delta} P = \Delta_{\sigma,\delta}$ . For any function  $f : \mathbb{R}^D \rightarrow \mathbb{R}$ , we define

$$K_\Sigma f(x) := \int_{M^\tau} \varphi_{\Sigma(y)}(x - y) f(y) dy,$$

where we recall that  $\varphi_{\Sigma(y)}$  is the density of a centered Gaussian with variance  $\Sigma(y)$ . The idea behind the construction of the approximation of  $f_0$  by a mixture of gaussians is to show that  $K_{\Sigma}f_0(x)$  is close to  $f_0$  and then to define a perturbation  $f_1$  of  $f_0$  such that  $K_{\Sigma}f_1(x) - f_0(x) = O(L(x)\sigma^{\beta})$ . Compared to the construction proposed by [KRVDV10] in the univariate case or [STG13] in the multivariate case where  $K_{\Sigma}f = \varphi_{\Sigma} * f$ , in our construction  $\Sigma$  varies with the location  $y$ . Note that in particular  $\int \varphi_{\Sigma(y)}(x - y)dy$  may be different from 1. This dependence in  $y$  is crucial to adapt to the geometry of the manifold but considerably complicates the proof as the underlying kernel integral operator can no longer be written as a convolution. We first show that  $f_0$  can be efficiently approximated pointwise.

**Theorem 5.3.7.** *Assume that  $\sigma, \delta \leq 1$ , that  $f_0 \in \mathcal{H}_{\delta}^{\beta_0, \beta_{\perp}}(M, L)$  satisfies (5.12) and (5.13), that  $\sigma^{\alpha_0 - \alpha_{\perp}} \leq \delta$  and that the manifold satisfies (5.2) with  $\beta_0 \leq \beta_{\perp} \leq \beta_M - 3$ . Then there exists a function  $g : \mathbb{R}^D \rightarrow \mathbb{R}$  such that, for any  $H > 0$ ,*

$$|K_{\Sigma}g(x) - f_0(x)| \lesssim \sigma^{\beta} L(x) \mathbb{1}_{M^{\tau}} + (H \log(1/\sigma))^{D/\kappa} \sigma^H \|L\|_{\infty} \quad \forall x \in \mathbb{R}^D.$$

The function  $g$  has the form

$$g(x) = f_0 + \frac{1}{\delta^{D-d}} \sum_{0 < \langle k, \alpha \rangle < \beta} \sigma^{\langle k, \alpha \rangle} \sum_{j=1}^J d_{j,k}(x, \sigma, \delta) D_z^k (\overline{\chi_j f_0})_{x_j, \delta}(z_{j,x}),$$

where  $z_{j,x} := \Delta_{1,\delta}^{-1} \bar{\Psi}_{x_j, \delta}^{-1}(x)$ , where  $(\chi_j)_{j \leq J}$  is a partition of unity, defined in Section 5.C.1, of the set  $M^{\tau} \cap B(0, R_0(\log(1/\sigma))^{1/\kappa})$  associated with a  $\tau/64$ -packing  $(x_j)_{j \leq J}$  of  $M^{\tau} \cap B(0, R_0(\log(1/\sigma))^{1/\kappa})$  and where  $d_{j,k}(x, \sigma, \delta)$  are smooth and bounded functions depending on  $\chi_j$  and  $M$ .

We then establish that the previous bound translates to a control in terms of Hellinger distance.

**Corollary 5.3.8.** *In the context of Theorem 5.3.7, if  $f_0$  also satisfies (5.13) and if  $\sigma^{\omega - 6\beta} = o(\delta^{D-d})$  and  $\delta \sigma^{\alpha_{\perp}} = o(|\log \sigma|^{-1/2})$ , the probability density  $\tilde{h} \propto g \mathbb{1}_{g \geq f_0/2} + f_0/2 \mathbb{1}_{g < f_0/2}$  verifies*

$$d_H(K_{\Sigma} \tilde{h}, g)^2 \lesssim \sigma^{2\beta} |\log \sigma|^{16\beta + D/\kappa}. \quad (5.14)$$

Theorem 5.3.7 is proven in Section 5.5.2 while the proof of Corollary 5.3.8, which appears to be a non-trivial consequence of Theorem 5.3.7, is delayed to Section 5.C.1.

## 5.4 Numerical experiments

In this section, we provide a few numerical experiments using the Maximum A Posteriori (MAP) to approximate the true distribution, as implemented in the python package `pyro` [BCJ<sup>+</sup>19]. The goal of this section is not to check that the theoretical results of this



Chapter hold numerically, but rather to give some visual examples as how suitable the location-scale mixtures of Subsection 5.2.3 can be to describe in a relevant way data that can be very singular. More precisely, we consider the hierarchical model:

$$\begin{aligned}
 (y_i)_{i=1}^n \mid (\mu_i, O_i)_{i=1}^n, \Lambda &\sim \bigotimes_{i=1}^n \mathcal{N}(\cdot \mid \mu_i, O_i \Lambda O_i^T) \quad \text{with} \quad \Lambda = \text{diag}(\lambda_1, \dots, \lambda_D), \\
 (\mu_i, O_i)_{i=1}^n \mid P &\sim P^{\otimes n}, \\
 P &\sim \text{DP}(\alpha P_0) \quad \text{with} \quad P_0 = \mathcal{N}(\cdot \mid \mu_0, \Sigma_0) \otimes \text{Unif}(\mathcal{O}(D)), \\
 (\lambda_j)_{j=1}^D \mid (b_j)_{j=1}^D &\sim \bigotimes_{j=1}^D \text{Inv}\Gamma(a_j, b_j), \quad \text{with} \quad (b_j)_{j=1}^D \sim \bigotimes_{j=1}^D \text{Exp}(\kappa_j).
 \end{aligned} \tag{5.15}$$

We set in our experiments the value of the hyperparameters as  $\alpha = 1$ ,  $a_j = \kappa_j = 1$  for all  $1 \leq j \leq D$  and  $\mu_0 = 0$ ,  $\Sigma_0 = I_D$ . We will restrict our numerical study to the ambient dimension  $D = 2$  and  $D = 3$ , and use the partial location-scales mixture described in Section 5.2 with exponential hyperpriors on the common eigenvalues  $\lambda_1$  and  $\lambda_2$  of the scales, and with the eigenvalue matrix  $\Lambda$  having the form

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \text{ for } D = 2, \quad \text{or} \quad \Lambda = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_2 \end{pmatrix} \text{ for } D = 3.$$

We refer to Appendix 5.E for further details regarding the design of the prior. The synthetic dataset we use are supported near four geometric shapes: the 2D-spiral ( $D = 2$ ,  $d = 1$ ) the two circles ( $D = 2$ ,  $d = 1$ ); the 3D-spiral ( $D = 3$ ,  $d = 1$ ) and the torus ( $D = 3$ ,  $d = 2$ ). We also refer to Appendix 5.E for the exact parametric equations of these sets. For each shape, we generate between 500 and 10000 points for various values of  $\delta$ , and generate the same amount of data through the estimated posterior obtain with SVI by predictive posterior sampling. The results are presented in Figures 5.4.1 to 5.4.4. As expected, the posterior does visually concentrate around the true distribution, even when  $\delta$  is very small (i.e the true probability measure is very singular) or when the support of the density is based on two crossing manifolds as it is the case for the two circles, see also remark 5.3.5.

One interesting observation is the fact that the prior naturally adapts to the intrinsic dimension  $d$  of the support: for the 3D spiral and the torus one notice that the value of  $\lambda_1$  (which corresponds to the 1-dimensional side of the covariance matrix) is predominant compared to  $\lambda_2$ , while it is automatically the other way around for the torus — see Figure 5.4.5 for a visualisation of this fact.

We conclude this series of numerical experiments with the inspection of the decrease of contraction rate for the 2D spiral and the torus, with  $n$  ranging from 100 to 10000. As in [MLD20] we evaluate the risk using the histogram metric which is computationally much less expensive than the  $L_1$  metric. It is computed as follow: take a new

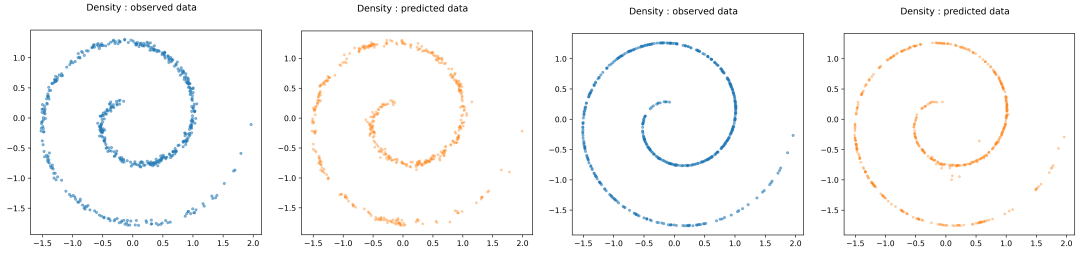


Figure 5.4.1 – The 2D spiral: For  $\delta = 0.1$  (Left), we observed  $n = 500$  data points (in Blue) and predicted the same amount from the posterior distribution (in Orange). In (Right), same experiment but with  $\delta = 0.01$ .

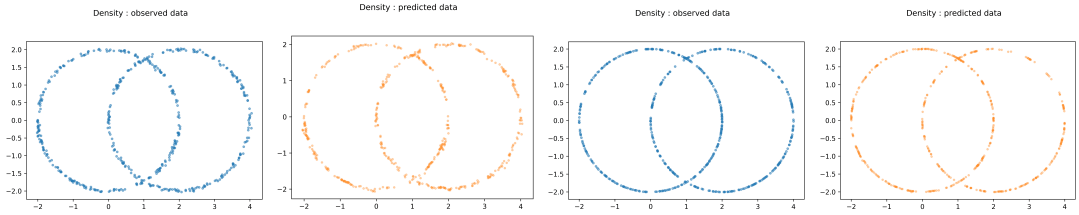


Figure 5.4.2 – The circles: For  $\delta = 0.1$  (Left), we observed  $n = 500$  data points (in Blue) and predicted the same amount from the posterior distribution (in Orange). In (Right), same experiment but with  $\delta = 0.01$ .

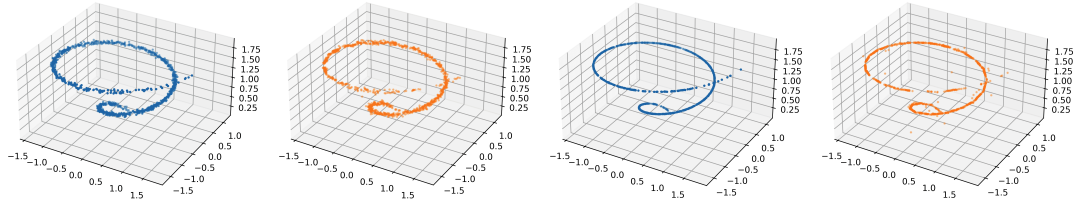


Figure 5.4.3 – The 3D spiral: For  $\delta = 0.1$  (Left), we observed  $n = 1000$  data points (in Blue) and predicted the same amount from the posterior distribution (in Orange). In (Right), same experiment but with  $\delta = 0.01$ .

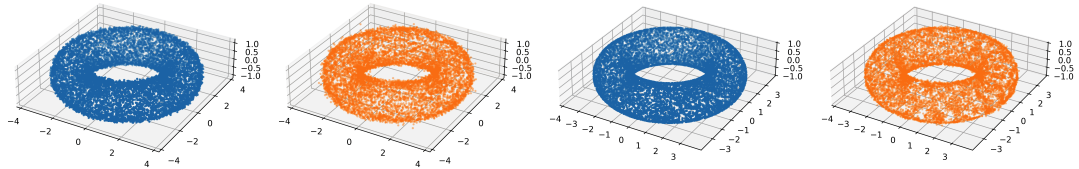


Figure 5.4.4 – The torus: For  $\delta = 0.5$  (Left), we observed  $n = 10000$  data points (in Blue) and predicted the same amount from the posterior distribution (in Orange). In (Right), same experiment but with  $\delta = 0.05$ .

test sample  $\mathcal{X}_N = \{x_1, \dots, x_N\}$  and, for any predictive  $n$ -sample  $\mathcal{Z}_n := \{z_1, \dots, z_n\}$  built

on a training sample, compute

$$\text{hist}_\varepsilon(\mathcal{X}_N, \mathcal{Z}_n) := \frac{1}{N} \sum_{i=1}^N |\mathbb{P}_{\mathcal{X}_N}(\mathcal{B}(x_i, \varepsilon)) - \mathbb{P}_{\mathcal{Z}_n}(\mathcal{B}(x_i, \varepsilon))| \quad (5.16)$$

where  $\mathbb{P}_{\mathcal{X}_N}$  (resp.  $\mathbb{P}_{\mathcal{Z}_n}$ ) is the empirical distribution associated with  $\mathcal{X}_N$  (resp.  $\mathcal{Z}_n$ ). We present the results in Figures 5.4.6 and 5.4.7.

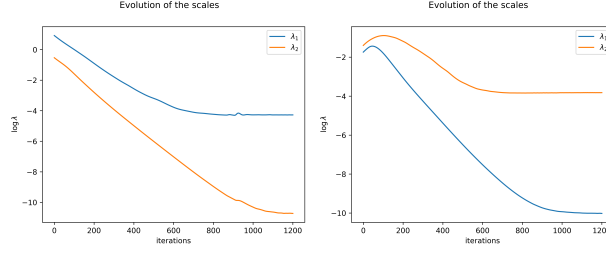


Figure 5.4.5 – The values of  $\lambda_1$  and  $\lambda_2$  as functions of the number of iterations in the SVI process for the 3D spiral with  $n = 1000$  observations and  $\delta = 0.05$  (Left) and the torus with  $n = 10000$  observations and  $\delta = 0.05$  (Right). Each value is initialized in the same way but are separated as soon as the first iteration.

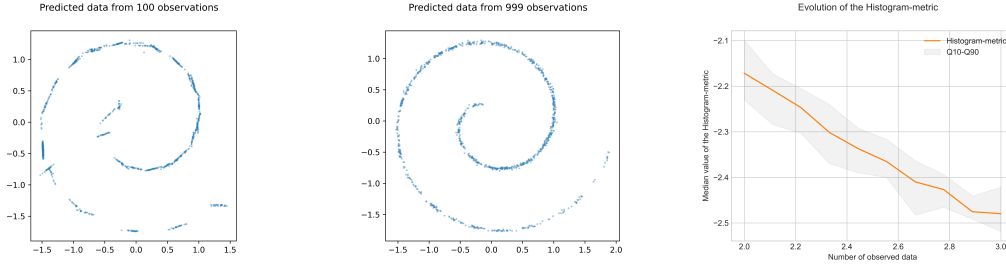


Figure 5.4.6 – For the 2D-spiral with  $\delta = 0.05$  we sampled 1000 points from the predictive distribution after observing (Left) 100 points and (Middle) 999 points. (Right) We computed the histogram metric (5.16) for a new test sample of size  $N = 10000$  and a predictive  $n = 1000$  sample, for  $\varepsilon = 0.05$ . The experiment was repeated 30 times for a training sample of size ranging from 100 to 999 on a log-regular grid of length 10. We show the median value of the metric in orange and its 10 and 90 percentiles in gray. Results are in double  $\log_{10}$ -scale.

## 5.5 Proofs of the main results

### 5.5.1 Proof of Theorem 5.3.2

Theorem 5.3.2 is proven using [GVDV07, Thm 5], which relies on two things: making sure that the prior probability distribution puts enough mass around the true density  $f_0$ , and ensuring that most of its probability mass is concentrated on a subset of manageable entropy.

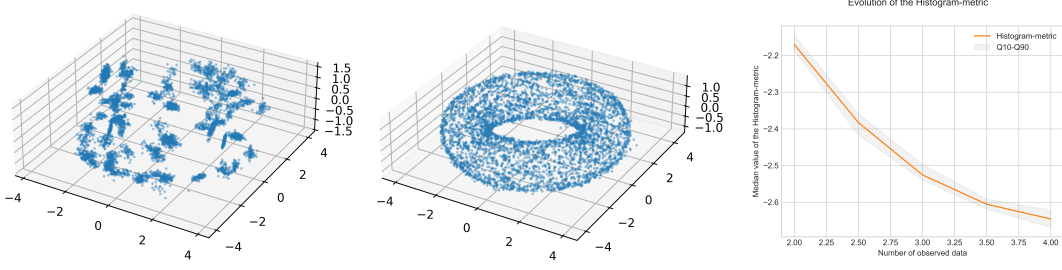


Figure 5.4.7 – Same experiment as in Figure 5.4.6 but for the torus with  $\delta = 0.05$ . Prediction is made using (Left) 100 observed points and (Middle) 10000 points. (Right) We computed the histogram metric (5.16) for a new test sample of size  $N = 10000$  and a predictive  $n = 1000$  sample, for  $\varepsilon = 0.5$ . The experiment was repeated 10 times for a training sample of size ranging from 100 to 10000 on a log-regular grid of length 5.

*Proof of Theorem 5.3.2.* We first check that our prior fulfils the entropic condition of [GVDV07, Thm 5]. To do so we define, for any sequence  $\varepsilon_n$  going to 0,  $\mathcal{F}_n = \mathcal{F}_n(\varepsilon_n, R_0, H_0, \sigma_0, \sigma_1)$  to be the set of all probability density function  $f_P$  with  $P = \sum_{h \geq 1} \pi_h \delta_{\mu_h, U_h, \Lambda_h}$  such that

$$\sum_{h \geq H_n} \pi_h \leq \varepsilon_n, \quad \forall h \leq H_n, \mu_h \in B(0, R_n) \quad \text{and} \quad \forall h \leq H_n, \Lambda_h \in \mathcal{Q}_n = [\underline{\sigma}_n^2, \bar{\sigma}_n^2]^D, U_h \in \mathcal{O}(D)$$

where  $R_n = \exp(R_0 n \varepsilon_n^2)$ ,  $H_n = \lfloor H_0(n \varepsilon_n^2) / \log n \rfloor$ ,  $\underline{\sigma}_n^2 = \sigma_0^2(n \varepsilon_n^2)^{-1/b_3}$  for some positive constant  $R_0, H_0$  and  $\sigma_0$  and where, for some  $\sigma_1 > 0$ ,

- $\bar{\sigma}_n^2 = \exp(\sigma_1^2 n \varepsilon_n^2)$  in the case of the Partial location-scale mixture
- $\bar{\sigma}_n^2 = \sigma_1^2(n \varepsilon_n^2)^{1/b_4}$  in the case of the Hybrid location-scale mixture.

We then show in Lemma 5.5.1 below that  $\Pi(\mathcal{F}_n^c) \lesssim \exp(-c_1 n \varepsilon_n^2)$  as soon as  $n \varepsilon_n^2 \geq n^\omega$  for some  $\omega > 0$ , and that  $\mathcal{F}_n$  satisfies the entropic bound of [GVDV07, Thm 5] with partitions described in the aforementioned lemma.

We finally define for any  $\varepsilon > 0$ ,

$$\mathcal{B}(f_0, \varepsilon) = \left\{ f : \mathbb{R}^D \rightarrow \mathbb{R} \mid \mathbb{P}_0 \left( \log \frac{f_0}{f} \right) \leq \varepsilon^2 \quad \text{and} \quad \mathbb{P}_0 \left( \log^2 \frac{f_0}{f} \right) \leq \varepsilon^2 \right\},$$

and introduce  $p = s \vee t/2$  and

$$\tilde{\varepsilon}_n := \delta^{\frac{\beta}{\alpha_0 - \alpha_1}} \wedge n^{-\frac{\beta}{2\beta + D}} \quad \text{and} \quad \varepsilon_n := \left\{ \frac{C^{1/2}}{\sqrt{n}} \tilde{\varepsilon}_n^{-D/2\beta} \log^{t/2}(1/\tilde{\varepsilon}_n) \right\} \vee \{ \tilde{\varepsilon}_n \log^s(1/\tilde{\varepsilon}_n) \}$$

where  $s, t$  and  $C$  are introduced in Lemma 5.5.2. The sequence  $\tilde{\varepsilon}_n$  goes to 0 and is such that  $\tilde{\varepsilon}_n^{\alpha_0} \leq \delta^\beta \tilde{\varepsilon}_n^{\alpha_1}$  so that Lemma 5.5.2 below applies and

$$\begin{aligned} \Pi(f_P \in \mathcal{B}(f_0, \varepsilon_n)) &\geq \Pi(f_P \in \mathcal{B}(f_0, \tilde{\varepsilon}_n \log^s(1/\tilde{\varepsilon}_n))) \gtrsim \exp\left(-C \tilde{\varepsilon}_n^{-D/\beta} \log^t(1/\tilde{\varepsilon}_n)\right) \\ &\gtrsim \exp(-n \varepsilon_n^2), \end{aligned}$$

so that the posterior contracts at rate  $\varepsilon_n$ , so long as  $n\varepsilon_n^2 \geq n^\omega$  for some  $\omega > 0$ , which allows application of Lemma 5.5.1 and in turn of [GVDV07, Thm 5]. We now distinguish three cases:

1. If first  $\delta^{\frac{D}{\alpha_0 - \alpha_\perp}} \leq n^{-1}$  then the results of Theorem 5.3.2 is trivial and there is nothing to show (because the contraction rate goes to  $\infty$  instead of 0);
2. If  $\delta^{\frac{D}{\alpha_0 - \alpha_\perp}} \geq n^{-\frac{D}{2\beta + D}}$ , then  $\tilde{\varepsilon}_n = n^{-\frac{\beta}{2\beta + D}}$  and one easily get that  $\varepsilon_n \simeq n^{-\frac{\beta}{2\beta + D}} \log^p n$ . In particular,  $n\varepsilon_n^2 \gg n^\omega$  for  $\omega = D/(2\beta + D)$ .
3. If finally  $n^{-1} < \delta^{\frac{D}{\alpha_0 - \alpha_\perp}} < n^{-\frac{D}{2\beta + D}}$ , then  $\tilde{\varepsilon}_n = \delta^{\frac{\beta}{\alpha_0 - \alpha_\perp}}$  and  $\log(1/\delta) \gtrsim \log n$  so that

$$n\varepsilon_n^2 \geq C\tilde{\varepsilon}_n^{-D/\beta} \log^t(1/\tilde{\varepsilon}_n) \gtrsim \delta^{-\frac{D}{\alpha_0 - \alpha_\perp}} \log^t(n) \gg n^{\frac{D}{2\beta + D}}.$$

In particular, Lemma 5.5.1 applies and the conclusion follows again from [GVDV07, Thm 5], to the last detail that we need to understand how  $\varepsilon_n$  depends on  $\delta$ . By assumption, there holds  $\log(1/\tilde{\varepsilon}_n) \lesssim \log(1/\delta) \lesssim \log(n)$  and  $\tilde{\varepsilon}_n < n^{-\beta/(2\beta + D)}$  and thus

$$\varepsilon_n \lesssim \log^p n \times \left\{ \frac{\tilde{\varepsilon}_n^{-D/2\beta}}{\sqrt{n}} \vee \tilde{\varepsilon}_n \right\} = \log^p n \times \frac{1}{\sqrt{n\delta^{D/(\alpha_0 - \alpha_\perp)}}}.$$

□

We now turn on the entropic condition on the sieve  $\mathcal{F}_n$  introduced in the proof below. We define, for  $\mathbf{j} = (j_h, h \leq H_n) \in \mathbb{N}^{H_n}$ ,

$$\mathcal{F}_{n,\mathbf{j}} := \{f_P \in \mathcal{F}_n \mid \forall h \leq h_n, j_h \sqrt{n} < \|\mu_h\| \leq (j_h + 1)\sqrt{n}\}$$

along with the following refinement

$$\mathcal{F}_{n,\mathbf{j},0} := \left\{ f_P \in \mathcal{F}_{n,\mathbf{j}} \mid \frac{\max_i \lambda_i}{\min_i \lambda_i} \leq n \right\}$$

and  $\forall \ell \geq 1, \mathcal{F}_{n,\mathbf{j},\ell} := \left\{ f_P \in \mathcal{F}_{n,\mathbf{j}} \mid n^{2\ell-1} < \frac{\max_i \lambda_i}{\min_i \lambda_i} \leq n^{2\ell} \right\}.$

**Lemma 5.5.1.** *Under assumptions of Table 5.2.1, for any sequence  $\varepsilon_n \rightarrow 0$  such that  $n\varepsilon_n^2 \geq n^\omega$ , for some  $\omega > 0$ , and for all  $c_1 > 0$ , if  $R_0, H_0, \sigma_1^2$  are large enough and  $\sigma_0^2$  is small enough, there exists  $M_0 > 0$  such that,*

$$i) \Pi(\mathcal{F}_n^c) \lesssim \exp(-c_1 n\varepsilon_n^2);$$

ii) *In the case of the partial location-scale prior*

$$\sum_{\mathbf{j}, \ell} \sqrt{\Pi(\mathcal{F}_{n,\mathbf{j},\ell}) N(\varepsilon_n, \mathcal{F}_{n,\mathbf{j},\ell}, \|\cdot\|_1)} e^{-M_0 n\varepsilon_n^2} = o(1);$$

iii) *In the case of the hybrid location-scale prior*

$$\sum_{\mathbf{j}} \sqrt{\Pi(\mathcal{F}_{n,\mathbf{j}}) N(\varepsilon_n, \mathcal{F}_{n,\mathbf{j}}, \|\cdot\|_1)} e^{-M_0 n\varepsilon_n^2} = o(1).$$

The proof of Lemma 5.5.1 can be found in Section 5.D.1. The last elementary brick in the proof of Theorem 5.3.2 is the control of the probability of small balls around  $\mathbb{P}_0$ , which is stated below.

**Lemma 5.5.2.** *Let  $\varepsilon > 0$  and assume that it is small enough so that  $\varepsilon^{\alpha_0} \leq \delta^\beta \varepsilon^{\alpha_1}$  and that  $\varepsilon^{\alpha_1} \ll \log^{-1}(1/\varepsilon)$ . Then, in the context of Theorem 5.3.2, there holds*

$$\Pi(f_P \in B(f_0, \tilde{\varepsilon})) \gtrsim \exp \left\{ -C \varepsilon^{-D/\beta} \log^t(1/\varepsilon) \right\},$$

where the constant  $C$  depends on the parameters, where  $\tilde{\varepsilon} \simeq \varepsilon \log^s(1/\varepsilon)$  and with  $s, t > 0$  depending on  $D, \beta$  and  $\kappa$ .

The complete proof is given in Section 5.D.2 and is sketched as follow: the first step is the Hellinger approximation of  $f_0$  by  $K_\Sigma \tilde{h}$  as expressed in Corollary 5.3.8. Then we exhibit an  $\varepsilon$ -approximation of  $K_\Sigma \tilde{h}$  by a discrete location scale mixture with a controlled number of atoms through the use of Lemma 5.5.3 below. The result then follows from similar arguments as in [STG13, KRVDV10] or [NR17].

**Lemma 5.5.3.** *Let  $\varepsilon > 0$  such that  $\delta \sigma^{\alpha_1} \log(1/\varepsilon) \ll 1$ . For any density  $g$  on  $M^\delta$  satisfying (5.12), there exists a discrete probability measure  $G$  on  $\mathbb{R}^D$  with at most  $N \simeq \sigma^{-D} \log^D(1/\varepsilon)$  atoms such that*

$$\|K_\Sigma G - K_\Sigma g\|_\infty \lesssim \frac{\varepsilon}{\sigma^D \delta^{D-d}} \quad \text{and} \quad \|K_\Sigma G - K_\Sigma g\|_1 \lesssim \varepsilon \log^{D/2}(1/\varepsilon).$$

The atoms of  $G$  are in  $M^\delta$  and are  $\sigma^{2\alpha_0} \varepsilon$ -apart.

The proof of Lemma 5.5.3 can be found in Section 5.D.3. We underline that although it uses similar ideas to [GVDV01, Lem 3.1], it is not a straightforward adaptation of it, since in  $K_\Sigma$  the covariances depend on the locations of the mixture in a complicated way.

## 5.5.2 Proof of Theorem 5.3.7

As explained in Section 5.3.2 a key ingredient of the proof of Theorem 5.3.2 is the pointwise approximation of  $f_0$  by  $K_\Sigma \tilde{g}$  where  $g$  is close to  $f_0$  and is explicited in the proof of Theorem 5.3.7 below.

*Proof of Theorem 5.3.7.* Let  $x_0 \in M$  and define

$$\mathcal{W}_{x_0}^j := B_{T_{x_0}} \left( 0, \frac{2+j}{16} \tau \right) \times B_{N_{x_0}} \left( 0, \frac{6+j}{8} \tau \right) \quad \text{for } j \in \{0, 1, 2\},$$

and  $\mathcal{O}_{x_0}^j = \bar{\Psi}_{x_0}(\mathcal{O}_{x_0}^j)$  for  $j \in \{0, 1, 2\}$ . We have  $\mathcal{W}_{x_0}^0 \subset \mathcal{W}_{x_0}^1 \subset \mathcal{W}_{x_0}^2$  and  $\mathcal{O}_{x_0}^0 \subset \mathcal{O}_{x_0}^1 \subset \mathcal{O}_{x_0}^2$ . Furthermore, the sets  $\mathcal{O}_{x_0}^0$  for  $x_0 \in M$  forms a covering of  $M^{3\tau/4}$ , see Section 5.A.4 for more details.

We now drop the  $x_0$  from the notation. Let  $f : \mathbb{R}^D \rightarrow \mathbb{R}$  be in  $\mathcal{H}_\delta^{\beta_0, \beta_\perp}(M, L)$  and supported on  $\mathcal{O}^0$  — it is to be thought of as  $f_0$  multiplied by a smooth function supported on  $\mathcal{O}^0$ . Take  $x \in \mathcal{O}^1$  and compute

$$K_\Sigma f(x) := \int_{\mathbb{R}^D} \varphi_{\Sigma(u)}(x - u) f(u) du = \int_{M^\delta \cap \mathcal{O}^0} \varphi_{\Sigma(u)}(x - u) f(u) du.$$

We first prove that we can construct a function  $g$  such that  $K_\Sigma g$  is close to  $f$  and we then apply this result to  $f = f_0 \chi_j$  with  $\chi_j$  the partition of unity defined in Lemma 5.A.6.

The idea is to use the fact that  $\Sigma(u) = o(1)$  and the smoothness of  $u \mapsto \Sigma(u)$  so that

$$K_\Sigma f(x) \approx \int_{M^\delta \cap \mathcal{O}^0} \varphi_{\Sigma(x)}(x - u) f(u) du \approx f(x).$$

We now write down the approximation rigorously and quantify the error, taking into account the geometry of the manifold  $M$ . In all that follows, we use the notation  $z = (v, \eta)$  for points belonging to  $B_{T_{x_0}M}(0, \tau/16) \times B_{N_{x_0}M}(0, \tau/2)$ , while throughout  $w = (v, \delta\eta) \in B_{T_{x_0}M}(0, \tau/16) \times B_{N_{x_0}M}(0, \delta\tau/2)$ . We first make the change of variable  $w = \bar{\Psi}^{-1}(u)$ , yielding

$$K_\Sigma f(x) = \int_{\bar{\Psi}^{-1}(M^\delta \cap \mathcal{O}^0)} \varphi_{\Sigma(\bar{\Psi}(w))}(x - \bar{\Psi}(w)) f(\bar{\Psi}(w)) |\det d\bar{\Psi}(w)| dw.$$

Then, denoting by  $w_x = \bar{\Psi}^{-1}(x)$ , we write

$$w = \Delta_{\sigma, \delta} z + w_x = \Delta_{\sigma, \delta} z + \Delta_{1, \delta} z_x, \quad w_x = (v_x, \delta\eta_x), \quad z_x = \Delta_{1, \delta}^{-1} w_x,$$

in the integral above, giving

$$K_\Sigma f(x) = \frac{1}{(2\pi)^{D/2} \delta^{D-d}} \int_{\Delta_{\sigma, 1}^{-1}(\mathcal{W}^0 - z_x)} e^{-B_\sigma(x, z)} \bar{f}_\delta(\Delta_{\sigma, 1} z + z_x) \zeta(\Delta_{\sigma, \delta} z + w_x) dz, \quad (5.17)$$

with

$$B_\sigma(x, z) := \frac{1}{2} \|x - \bar{\Psi}(\Delta_{\sigma, \delta} z + w_x)\|_{\Sigma^{-1}(\bar{\Psi}(\Delta_{\sigma, \delta} z + w_x))}^2$$

$$\text{and } \zeta(\Delta_{\sigma, \delta} z + w_x) := |\det d\bar{\Psi}(\Delta_{\sigma, \delta} z + w_x)|,$$

and where we used the fact that  $|\det \Sigma(u)|$  is constantly  $\sigma^D \delta^{D-d}$  for  $u \in M^\tau$ . Since  $z \mapsto \bar{f}_\delta(\Delta_{\sigma, 1} z + z_x)$  is zero outside of  $\Delta_{\sigma, 1}^{-1}(\mathcal{W}^0 - z_x)$ , we can replace the latter set with  $T_{x_0}M \times N_{x_0}M \approx \mathbb{R}^D$  in the integral above. We now develop each term separately. First

$$\zeta(\Delta_{\sigma, \delta} z + w_x) = |\det d\bar{\Psi}(w_x)| + \sum_{1 \leq |k| < \beta_M - 2} \frac{(\Delta_{\sigma, \delta} z)^k}{k!} D^k \zeta(w_x) + R_\sigma^\zeta(x, z)$$

with  $R_\sigma^\zeta(x, z) \lesssim \|\Delta_{\sigma, \delta} z\|^{\beta_M - 2}$ , up to a constant that depends on  $C_M$ . Secondly, there holds

$$\bar{f}_\delta(\Delta_{\sigma, 1} z + z_x) = \underbrace{\bar{f}_\delta(z_x)}_{=\delta^{D-d} f(x)} + \sum_{0 < \langle k, \alpha \rangle < \beta} \frac{z^k}{k!} \sigma^{\langle k, \alpha \rangle} D^k \bar{f}_\delta(z_x) + R_\sigma^f(x, z) \quad (5.18)$$

with  $R_\sigma^f(x, z) \leq DL_\delta(z_x)\sigma^\beta \|1 \vee z\|_1^{\beta_{\max}} = D\delta^{D-d}L(x)\sigma^\beta \|1 \vee z\|_1^{\beta_{\max}}$  in application of Corollary 5.A.3. It remains to understand  $B_\sigma(x, z)$ . Notice that

$$\bar{\Psi}(\Delta_{\sigma,\delta}z + w_x) = x + d\bar{\Psi}(w_x)[\Delta_{\sigma,\delta}z] + \sum_{2 \leq |k| < \beta_M - 1} \frac{(\Delta_{\sigma,\delta}z)^k}{k!} D^k \bar{\Psi}(w_x) + R_\sigma^{\bar{\Psi}}(x, z). \quad (5.19)$$

with again  $R_\sigma^{\bar{\Psi}}(x, z) \lesssim \|\Delta_{\sigma,\delta}z\|^{\beta_M - 1}$ . Let write  $k = (k_0, k_\perp)$  with  $k_0 \in \mathbb{N}^d$  and  $k_\perp \in \mathbb{N}^{D-d}$ . We know that  $\bar{\Psi}$  is affine with respect to its second variable, so that

$$D^{(k_0, k_\perp)} \bar{\Psi} = 0 \quad \text{for } |k_\perp| \geq 2. \quad (5.20)$$

The sum in the RHS of (5.19) thus rewrites

$$\sum_{\substack{k_\perp=0 \\ 2 \leq |k_0| < \beta_M - 1}} \sigma^{|k_0|\alpha_0} \frac{z^{k_0}}{k_0!} D^{k_0} \bar{\Psi}(w_x) + \sum_{\substack{|k_\perp|=1 \\ \text{ind} \leq |k_0| < \beta_M - 2}} \sigma^{|k_0|\alpha_0} \sigma^{\alpha_\perp} \delta \frac{z^{k_\perp}}{k_\perp!} D^{k_\perp} \bar{\Psi}(w_x) \quad (5.21)$$

and  $R_\sigma^{\bar{\Psi}}$  is bounded by the more precise quantity

$$\|R_\sigma^{\bar{\Psi}}(x, z)\| \lesssim \sum_{\substack{|k|=\lceil \beta_M - 2 \rceil \\ |k_\perp| \leq 1}} |\Delta_{\sigma,\delta}z|^k \|\Delta_{\sigma,\delta}z\|^{\beta_M - 1 - \lceil \beta_M - 2 \rceil}.$$

Furthermore, the first differential of  $\bar{\Psi}$  reads

$$\begin{aligned} d\bar{\Psi}(w_x)[\Delta_{\sigma,\delta}z] &= \underbrace{\sigma^{\alpha_0} d\Psi(v_x)[v] + \sigma^{\alpha_0} \text{pr}_{T_x} d^v N(w_x)[v]}_{\in T_x} \\ &\quad + \underbrace{\sigma^{\alpha_0} \text{pr}_{N_x} d^v N(w_x)[v] + \delta \sigma^{\alpha_\perp} N(v_x, \eta)}_{\in N_x}. \end{aligned} \quad (5.22)$$

Now recall that  $u = \bar{\Psi}(\Delta_{\sigma,\delta}z + w_x)$  and that

$$B_\sigma(x, z) = \frac{1}{2} \|x - u\|_{\Sigma^{-1}(u)}^2 = \frac{1}{2\sigma^{2\alpha_0}} \|\text{pr}_{T_u}(x - u)\|^2 + \frac{1}{2\delta^2 \sigma^{2\alpha_\perp}} \|\text{pr}_{N_u}(x - u)\|^2$$

Using the development of  $u$  that we have in (5.19, 5.21), we find that, noting again  $k = (k_0, k_\perp)$ , for any  $y \in \mathbb{R}^D$ ,

$$\text{pr}_{T_u}(y) = \text{pr}_{T_x}(y) + \sum_{\substack{1 \leq |k| < \beta_M - 1 \\ |k_0| \geq 1}} (\Delta_{\sigma,\delta}z)^k \Phi_k^T(w_x)[y] + R_\sigma^T(x, y, z) \quad (5.23)$$

with

$$|R_\sigma^T(x, y, z)| \lesssim \|y\| \sum_{\substack{|k|=\lceil \beta_M - 2 \rceil \\ |k_0| \geq 1}} |\Delta_{\sigma,\delta}z|^k \|\Delta_{\sigma,\delta}z\|^{\beta_M - 1 - \lceil \beta_M - 2 \rceil},$$

and for some  $(\beta_M - 1 - |k|)$ -Hölder functions  $\Phi_k^T(w_x)$  (which are smooth functions of  $D^\ell \Psi(w_x)$  for  $1 \leq |\ell| \leq |k|$ , see for instance Lemma 5.A.2). Here the important fact is that the sums in both displays above start at  $|k_0| \geq 1$ . This is because:



- i)  $s \mapsto \text{pr}_M(x+s)$  is constant over the fiber  $N_x$ , so there is no sole contribution of the normal part of  $u-x$  (only mixed contribution with the tangential part) in (5.23);
- ii) The tangential part of  $d\bar{\Psi}(w_x)[\Delta_{\sigma,\delta}z]$  does not inherit the contribution of the normal displacement in  $\delta\alpha^\perp$ , as shown in (5.22);
- iii) In (5.21), neither sums contain any term of type  $(0, k_\perp)$ .

Plugging  $y = u - x = \bar{\Psi}(\Delta_{\sigma,\delta}z + w_x) - x$  in (5.23) and using (5.22) yields

$$\begin{aligned} \sigma^{-\alpha_0} \text{pr}_{T_u}(x-u) &= d\Psi(v_x)[v] + \text{pr}_{T_x} d^v N(w_x)[v] \\ &\quad + \sum_{\substack{2 \leq |k| < \beta_M - 1 \\ |k_0| \geq 1}} \sigma^{-\alpha_0} (\Delta_{\sigma,\delta}z)^k \tilde{\Phi}_k^T(w_x) + R_\sigma^T(x, z) \end{aligned}$$

with  $\tilde{\Phi}_k^T$  being again smooth functions of  $D^\ell \bar{\Psi}(w_x)$  for  $|\ell| \leq |k|$ , and where

$$\|R_\sigma^T(x, z)\| \lesssim \sigma^{-\alpha_0} \sum_{\substack{|k| = \lceil \beta_M - 2 \rceil \\ |k_0| \geq 1}} |\Delta_{\sigma,\delta}z|^k \|\Delta_{\sigma,\delta}z\|^{\beta_M - 1 - \lceil \beta_M - 2 \rceil} \lesssim \delta^{\beta_M - 2} \sigma^{(\beta_M - 2)\alpha_\perp} \|z\|^{\beta_M - 1}$$

up to a constant that depends on  $C_M$ . Now notice that

$$\text{pr}_{N_u}(z) = z - \text{pr}_{T_u} z = \text{pr}_{N_x} z - \sum_{\substack{1 \leq |k| < \beta_M - 1 \\ |k_0| \geq 1}} (\Delta_{\sigma,\delta}z)^k \Phi_k^T(w_x)[z] - R_\sigma^T(x, y, z)$$

whence again plugging  $z = u - x$  and using (5.22)

$$\begin{aligned} \delta^{-1} \sigma^{-\alpha_\perp} \text{pr}_{N_u}(x-u) &= N(v_x, \eta) + \sigma^{\alpha_0 - \alpha_\perp} \delta^{-1} \text{pr}_{N_x} d^v N(w_x)[v] \\ &\quad + \sum_{\substack{2 \leq |k| < \beta_M - 1 \\ |k_0| \geq 1}} \sigma^{-\alpha_\perp} \delta^{-1} (\Delta_{\sigma,\delta}z)^k \tilde{\Phi}_k^N(w_x) + R_\sigma^N(x, z) \end{aligned}$$

where again  $\tilde{\Phi}_k^N(w_x)$  is polynomial in  $D^\ell \bar{\Psi}(x)$  for  $\ell \leq k$  and

$$\begin{aligned} \|R_\sigma^N(x, z)\| &\lesssim \sigma^{-\alpha_\perp} \delta^{-1} \sum_{\substack{|k| = \lceil \beta_M - 2 \rceil \\ |k_0| \geq 1}} |\Delta_{\sigma,\delta}z|^k \|\Delta_{\sigma,\delta}z\|^{\beta_M - 1 - \lceil \beta_M - 2 \rceil} \\ &\lesssim \delta^{\beta_M - 3} \sigma^{\alpha_\perp (\beta_M - 3) + \alpha_0} \|z\|^{\beta_M - 1}. \end{aligned}$$

Recall that by assumption  $\sigma^{\alpha_0} \leq \delta \sigma^{\alpha_\perp}$  so that  $R_N$  is of greater order than  $R_T$ . There thus exists functions  $\Phi_k$  similar to  $\tilde{\Phi}_k^T$  and  $\tilde{\Phi}_k^N$  such that

$$B_\sigma(x, z) = \frac{1}{2} \|A(w_x)[z]\|^2 + \sum_{\substack{1 \leq |k| < \beta_M - 1 \\ |k_0| \geq 1}} \sigma^{-\alpha_\perp} \delta^{-1} (\Delta_{\sigma,\delta}z)^k \Phi_k(w_x) + R_\sigma^B(x, z)$$

with  $R_\sigma^B(x, z) \lesssim \delta^{\beta_M - 2} \sigma^{(\beta_M - 2)\alpha_\perp} \|z\|^{\beta_M - 1}$  and where

$$A(w_x)[v, \eta] := d\Psi(v_x)[v] + \text{pr}_{T_x} d^v N(w_x)[v] + N(v_x, \eta).$$

We can rewrite the development of  $B_\sigma$ , up to a slight modification of the  $\Phi_k$  and of  $R^B$ , which we write again  $\Phi_k$  and  $R^B$  with a slight abuse of notation, in the following form

$$\begin{aligned} & -B_\sigma(x, z) + \frac{1}{2} \|A(w_x)[z]\|^2 \\ & = \log \left\{ 1 + \sum_{\substack{1 \leq |k| < \beta_M - 1 \\ |k_0| \geq 1}} \sigma^{-\alpha_\perp} \delta^{-1} (\Delta_{\sigma, \delta} z)^k \Phi_k(w_x) + R_\sigma^B(x, z) \right\} \end{aligned} \quad (5.24)$$

or again

$$e^{-B_\sigma(x, z)} = e^{-\frac{1}{2} \|A(w_x)[z]\|^2} \times \left\{ 1 + \sum_{\substack{1 \leq |k| < \beta_M - 1 \\ |k_0| \geq 1}} \sigma^{-\alpha_\perp} \delta^{-1} (\Delta_{\sigma, \delta} z)^k \Phi_k(w_x) + R_\sigma^B(x, z) \right\}$$

All in all, we obtain a development of  $K_\Sigma f(x)$  of the form

$$\begin{aligned} K_\Sigma f(x) &= \sum_{0 \leq \langle k, \alpha \rangle < \beta} \sum_{0 \leq |m| < \beta_M - 2} \sigma^{\langle k+m, \alpha \rangle} \delta^{|m_\perp|} \xi_{k,0,m}(x) \\ &+ \sum_{0 \leq \langle k, \alpha \rangle < \beta} \sum_{\substack{1 \leq |\ell| < \beta_M - 1 \\ |\ell_0| \geq 1}} \sum_{0 \leq |m| < \beta_M - 2} \sigma^{\langle k+\ell+m, \alpha \rangle - \alpha_\perp} \delta^{|\ell_\perp| + |m_\perp| - 1} \xi_{k,\ell,m}(x) + R_\sigma(x) \end{aligned} \quad (5.25)$$

with  $|R_\sigma(x)| \lesssim \{\sigma^\beta + \sigma^{\beta(\beta_M - 2)/\beta_\perp} \delta^{\beta_M - 2}\} L(x) \lesssim \sigma^\beta L(x)$  as soon as  $\beta_M - 2 \geq \beta_\perp$ . The term  $L(x)$  appears in the control of  $R_\sigma$  because every term in the remainder is multiplied by one term of the development (5.18) of  $\bar{f}_\delta$ , and each one of this term is upper bounded by  $L(x)$ . The term  $\xi_{k,\ell,m}(x)$  (for both  $\ell = 0$  and  $\ell > 0$  with the convention  $\Phi_0 = 1$ ) is exactly

$$D^k \bar{f}_\delta(z_x) \times \Phi_\ell(w_x) \times D^m \zeta(w_x) \times \frac{1}{(2\pi)^{D/2} \delta^{D-d}} \int_{\mathbb{R}^D} e^{-\frac{1}{2} \|A(w_x)[z]\|^2} z^{(k+\ell+m)} dz.$$

The zero-th order term is equal to

$$\bar{f}_\delta(z_x) \times \zeta(w_x) \times \frac{1}{(2\pi)^{D/2} \delta^{D-d}} \int_{\mathbb{R}^D} e^{-\frac{1}{2} \|A(w_x)[z]\|^2} dz = f(x) \times |\det \bar{\Psi}(w_x)| \times \frac{1}{|\det A(w_x)|}$$

But recall that

$$d\bar{\Psi}(w_x)[z] = d\Psi(v_x)[v] + \text{pr}_{T_x} d^v N(w_x)[v] + \text{pr}_{N_x} d^v N(w_x)[v] + N(v_x, \eta)$$

which, written from an orthonormal basis concatenated from orthonormal bases of  $T_{x_0}$  and  $N_{x_0}$  to an orthonormal basis concatenated from orthonormal bases of  $T_x$  and  $N_x$  leads to a matrix which is block triangular inferior with diagonal blocks corresponding to  $d\Psi(v_x)[v] + \text{pr}_{T_x} d^v N(w_x)[v]$  and  $N(v_x, \eta)$  so that  $|\det d\bar{\Psi}(w_x)| = |\det A(w_x)|$ . Now for the higher-order terms, notice that the map

$$w \mapsto D^m \zeta(w) \times \Phi_\ell(w) \times \frac{1}{(2\pi)^{D/2}} \int_{\mathbb{R}^D} e^{-\frac{1}{2} \|A(w)[\eta]\|^2} \eta^{(k+\ell+m)} d\eta$$

belongs, in application of Proposition 5.B.6 and Proposition 5.B.1, to  $\mathcal{H}_{\text{iso}}^{\beta_M-2-|m|\vee|\ell|}(\mathcal{W}_{x_0,\delta}, C)$  for some  $C$  depending on  $C_M$  and  $\tau$ . Likewise,  $z \mapsto D^k \bar{f}_\delta(z)$  belongs to  $\mathcal{H}^{\beta^{(k)}}(\mathcal{W}_{x_0,\delta}, L)$  according to Proposition 5.B.1. Using Proposition 5.B.6 once again and the definition of manifold-driven Hölder spaces, one get that

$$\xi_{k,\ell,m} \in \mathcal{H}_{\delta}^{\tilde{\beta}_0, \tilde{\beta}_\perp}(M, CL) \quad \text{with} \quad \begin{cases} \tilde{\beta}_0 = \{\beta_0 - \langle k, \alpha \rangle / \alpha_0\} \wedge \{\beta_M - 2 - |m| \vee |\ell|\} \\ \tilde{\beta}_\perp = \{\beta_\perp - \langle k, \alpha \rangle / \alpha_\perp\} \wedge \{\beta_M - 2 - |m| \vee |\ell|\}. \end{cases}$$

Looking at (5.25), if we were to prove that

$$\begin{cases} \langle k+m, \alpha \rangle + \tilde{\beta} \geq \beta & \text{if } \ell = 0 \\ \langle k+\ell+m, \alpha \rangle + \tilde{\beta} \geq \alpha_0 + \beta & \text{if } |\ell_0| \geq 1 \end{cases} \quad \text{where } \tilde{\beta} := D \times \left\{ \frac{d}{\tilde{\beta}_0} + \frac{D-d}{\tilde{\beta}_\perp} \right\}^{-1}, \quad (5.26)$$

then we would get, by induction, that there exists a function  $g$ , supported on  $\mathcal{O}^0$ , of the form

$$g(x) = f(x) + \frac{1}{\delta^{D-d}} \sum_{0 < \langle k, \alpha \rangle < \beta} \sigma^{\langle k, \alpha \rangle} d_k(x, \sigma, \delta) D^k \bar{f}_\delta(z_x)$$

with  $d_k$  uniformly bounded, such that

$$|K_\Sigma g(x) - f(x)| \lesssim L(x) \sigma^\beta \quad \forall x \in \mathcal{O}^1.$$

It only remains to prove (5.26). We start with  $\ell = 0$ . Then  $|m| \vee |\ell| = |m|$  and since  $\beta_0 - \langle k, \alpha \rangle / \alpha_0 \leq \beta_\perp - \langle k, \alpha \rangle / \alpha_\perp$ , there are three cases to cover: case 1.:  $\beta_M - 2 - |m| \geq \beta_\perp - \langle k, \alpha \rangle / \alpha_\perp$ ; case 2.:  $\beta_M - 2 - |m| \leq \beta_0 - \langle k, \alpha \rangle / \alpha_0$  and case 3:  $\beta_0 - \langle k, \alpha \rangle / \alpha_0 \leq \beta_M - 2 - |m| \leq \beta_\perp - \langle k, \alpha \rangle / \alpha_\perp$ .

We start with case 1: in this case,  $\tilde{\beta} = \beta - \langle k, \alpha \rangle$  and (5.26) follows immediately. In both case 2 and case 3, notice that denoting  $\tilde{\alpha}_\perp := \tilde{\beta} / \tilde{\beta}_\perp$ , we either have  $\tilde{\alpha}_\perp = 1 \geq \alpha_\perp$  (case 2) or  $\tilde{\alpha}_\perp = \tilde{\beta} / \tilde{\beta}_\perp \geq \alpha_\perp$  because  $\tilde{\beta}_\perp / \tilde{\beta}_0 \leq \beta_\perp / \beta_0$  by assumption (case 3). This yields that

$$\begin{aligned} \tilde{\beta} &= (\beta_M - 2 - |m| \vee |\ell|) \tilde{\alpha}_\perp \geq (\beta_M - 2 - |m|) \alpha_\perp \geq (\beta_M - 2) \alpha_\perp - \langle m, \alpha \rangle \\ &\geq \frac{\beta_M - 2}{\beta_\perp} \beta - \langle m, \alpha \rangle \end{aligned}$$

ending the proof of (5.26) for  $\ell = 0$  since  $\beta_M - 2 \geq \beta_\perp$ . Now for the case  $|\ell_0| \geq 1$ , if  $|\ell| \leq |m|$ , the previous reasoning still holds and we get

$$\langle k+\ell+m, \alpha \rangle + \tilde{\beta} \geq \langle \ell, \alpha \rangle + \langle k+m, \alpha \rangle + \tilde{\beta} \geq \alpha_0 + \beta.$$

If finally  $|\ell| > |m|$ , then we can write  $\ell = \tilde{\ell} + e_i$  for some  $i \in \{1, \dots, d\}$  so that  $|\ell| = |\tilde{\ell}| + 1$  and

$$\langle k+\ell+m, \alpha \rangle + \tilde{\beta} \geq \alpha_0 + \langle k+\tilde{\ell}, \alpha \rangle + \tilde{\beta}.$$

Now noticing that  $\beta_M - 2 - |\ell| = \beta_M - 3 - |\tilde{\ell}|$ , we obtain, using the same reasoning as above, that  $\langle k+\tilde{\ell}, \alpha \rangle + \tilde{\beta} \geq \beta$  as soon as  $\beta_M - 3 \geq \beta_\perp$ , ending the proof of (5.26).

We are now ready to prove Theorem 5.3.7. We let

$$R = \{H \log(1/\sigma)\}^{1/\kappa}$$

and  $\chi_1, \dots, \chi_J$  be the functions defined at Lemma 5.A.6 from a  $\tau/64$ -packing of  $M \cap B(0, R)$ . Recall that we can always chose  $J$  of order less than  $R^D$ . In light of the point (iv) of Lemma 5.A.6, the function  $f_j := \chi_j f_0$  is still in  $\mathcal{H}_\delta^{\beta_0, \beta_\perp}(M, CL)$  for some constant  $C$  depending on  $\tau$  and  $\beta_\perp$ . Since  $\text{supp } f_j \subset \mathcal{O}_{x_j}^0$  (point (i) of Lemma 5.A.6), the first part of this proof yields that there exists some functions  $g_j$  supported on  $\mathcal{O}_{x_j}^0$ , such that

$$|K_\Sigma g_j(x) - f_j(x)| \lesssim L(x) \sigma^\beta \quad (5.27)$$

uniformly on  $\mathcal{O}_{x_j}^1$ . Now notice that for  $x$  outside of  $\mathcal{O}_{x_j}^1$ , we have  $d(x, \mathcal{O}_{x_j}^0) > (\sigma^{\alpha_0} \vee \delta \sigma^{\alpha_\perp}) \sqrt{(H+D) \log(\sigma)}$  so that

$$|K_\Sigma g_j(x)| \leq \int_{\mathcal{O}_{x_j}^0} |\varphi_\Sigma(u)(x-u)g_j(u)| du \leq \frac{\sigma^H}{(2\pi)^{D/2} \delta^{D-d}} \int_{\mathcal{O}_{x_j}^0} |g_j(u)| du \lesssim \sigma^H \sup_{\mathcal{O}_{x_j}^0} L$$

and the equality (5.27) extends to the whole set  $\mathbb{R}^D$  with the bound  $\sigma^H \|L\|_\infty$  on  $\mathbb{R}^D \setminus \mathcal{O}_{x_j}^1$ . Using the linearity of  $K_\Sigma$ , we thus find that for  $g = \sum_{j=1}^J g_j$ , and for any  $x \in \mathbb{R}^D$ , there holds,

$$\begin{aligned} |K_\Sigma g(x) - f_0(x)| &\leq \sum_{j=1}^J |K_\Sigma g_j(x) - f_j(x)| \leq \sum_{j \in J(x)} |K_\Sigma g_j(x) - f_j(x)| + \sum_{j \notin J(x)} |K_\Sigma g_j(x)| \\ &\lesssim |J(x)| \times L(x) \sigma^\beta + (J - |J(x)|) \times \sigma^H \|L\|_\infty \\ &\lesssim \sigma^\beta L(x) + \{H \log(1/\sigma)\}^{D/\kappa} \sigma^H \|L\|_\infty \end{aligned}$$

where we denoted  $J(x) = \{1 \leq j \leq J \mid x \in \mathcal{O}_{x_j}^1\}$ , and used the fact that  $J(x)$  is bounded from above by something depending on  $D$  and  $\tau$  only, ending the proof.  $\square$

## 5.6 Discussion

With the aim of developing Bayesian procedures in the framework of manifold learning, we exhibited a new family of priors based on location-scales of Dirichlet mixture of Gaussians, and described a general setting for studying density supported near a submanifold. The latter relies on two things: first, a parametrization of the offset of the manifold and second, an anisotropic class of Hölder functions. In this model, we obtained concentration rates in Theorem 5.3.2 for the associated posterior distribution that are adaptive to the regularity of the underlying density while being totally agnostic of the underlying submanifolds and their main features. Our procedure is therefore fully adaptive.

An interesting feature of our theoretical framework is that it allows to express the rate in terms of the smoothnesses of the density and the manifold together with the thickness  $\delta$  of the support around the manifold. When  $\delta$  is fixed, our results can be viewed as an extension of minimax rates for regular anisotropic densities to manifold driven anisotropic densities. But we are also considering the regime where  $\delta = o(1)$  which corresponds to the manifold learning problem. The rates obtained in Theorem 5.3.2 have two regimes: one when  $\delta$  is not too small with rate  $n^{-\beta/(2\beta+D)}$  (up to  $\log n$  terms) and the concentrated regime where  $\delta$  is very small ( $\delta = o(n^{-1/(2\beta+D)})$ ) where we obtain the rate  $(n\delta^{D/(\alpha_0-\alpha_1)})^{-1/2}$ . It is not clear if this latter rate is optimal or not. The case  $\delta = 0$  would correspond to the observations belonging to the manifold  $M$  (and for which we would expect the rate  $n^{-\beta_0/(2\beta_0+d)}$ ) but cannot be thought as a limiting case of our problem since then the distribution has density with respect to the Hausdorff measure on  $M$  and not with respect to the Lebesgue measure on  $\mathbf{R}^D$ . When  $M$  is unknown the model is not dominated and our approach is not applicable. The problem of posterior contraction rates when the distribution lives on an unknown manifold remains open, although some interesting ideas in [TY22] or [CDFM22] could be used to address it.

Another interesting output of our results is that if nonparametric mixtures of normal densities define a versatile and flexible model for smooth densities, the structure of the prior on the mixing distribution is crucial. In this Chapter we propose two classes of priors which we believe enjoy many strong theoretical properties while remaining reasonably simple to implement. Moreover variational Bayes algorithms using pyro can be easily implemented and for which the same theoretical guarantees hold. It is quite possible that other nonparametric mixture models such as the Fisher-Gaussian kernels of [MLD20] would enjoy the same theoretical guarantees and we believe that our approximation result can be useful to study the theoretical properties of mixtures of Fisher-Gaussian kernels which are strongly related to Gaussian kernels.



# Appendix

## Contents

---

<b>5.A</b>	<b>Some facts on submanifolds with bounded reach . . . . .</b>	<b>181</b>
5.A.1	The geometry of submanifolds with bounded reach . . . . .	181
5.A.2	Taylor expansion of M-anisotropic Hölder functions . . . . .	185
5.A.3	Stability by a change of basis . . . . .	185
5.A.4	Partitions of unity and packings . . . . .	186
<b>5.B</b>	<b>Appendix to Section 5.2 . . . . .</b>	<b>189</b>
5.B.1	Auxiliary results on general anisotropic Hölder functions . . . . .	189
5.B.2	Proofs associated to the examples of Proposition 5.2.4 . . . . .	195
5.B.3	Proof of Proposition 5.2.7 . . . . .	198
<b>5.C</b>	<b>Appendix to Section 5.5.2: proof of Theorem 5.3.7 . . . . .</b>	<b>199</b>
5.C.1	Technical Lemmata . . . . .	199
5.C.2	Proof of Corollary 5.3.8 . . . . .	201
<b>5.D</b>	<b>Appendix to Section 5.5.1: proof of Theorem 5.3.2 . . . . .</b>	<b>205</b>
5.D.1	Proof of Lemma 5.5.1 . . . . .	206
5.D.2	Proof of Lemma 5.5.2 . . . . .	208
5.D.3	Proof of Lemma 5.5.3 . . . . .	211
<b>5.E</b>	<b>Appendix to Section 5.4 . . . . .</b>	<b>216</b>

---

## 5.A Some facts on submanifolds with bounded reach

### 5.A.1 The geometry of submanifolds with bounded reach

The reach  $\tau_K$  of a closed subset  $K \subset \mathbb{R}^D$ , initially introduced by [Fed59, Def 4.1 p.432], is defined the supremum of all the  $r \geq 0$  such that the orthogonal projection from the  $r$ -offset  $K^r = \{x \in \mathbb{R}^D \mid d(x, K) \leq r\}$  to  $K$  is well-defined, namely

$$\tau_K := \sup \{r \geq 0 \mid \forall x \in K^r, \forall y, z \in K, d(x, K) = \|x - y\| = \|x - z\| \Rightarrow y = z\}.$$

When the reach of a closed submanifold  $M \subset \mathbb{R}^D$  is bounded away from zero,  $M$  enjoys a number useful properties, that we list and prove below. In all the results stated hereafter, the reach of  $M$  is bounded from below by some  $\tau > 0$ . Lemma 5.A.1 provide

already existing results from the literature, sometimes slightly rephrased to better suit our needs. We start off with a control of the exponential map over submanifold with bounded curvature together with a comparison between the intrinsic distance on  $M$ , denoted by  $d_M(\cdot, \cdot)$ , and the ambient Euclidean distance. Recall that for any  $x, y \in M$ ,  $d_M(x, y)$  is the infimum of the length of all continuous path between  $x$  and  $y$  in  $M$  and if  $x$  and  $y$  are in two separate path-connected components, then  $d_M(x, y) = \infty$ .

**Lemma 5.A.1.** *The following facts hold true*

i) *For any  $x \in M$ , the exponential map  $\exp_x$  is a diffeomorphism from  $B_{T_x M}(0, \pi\tau)$  to  $\exp_x\{B_{T_{x_0} M}(0, \pi\tau)\}$ .*

ii) *It is double Lipschitz from  $B_{T_x M}(0, \tau/4)$  to its image with*

$$\forall v, w \in B_{T_x M}(0, \tau/4), \quad \frac{11}{16}\|v - w\| \leq \|\exp_x(v) - \exp_x(w)\| \leq \frac{21}{16}\|v - w\|. \quad (5.28)$$

iii) *Letting  $\kappa : \gamma \mapsto 2(1 - \sqrt{1 - \gamma})/\gamma$ . If  $\|x - y\| \leq \gamma\tau/2$  with  $\gamma \leq 1$ , then*

$$\|x - y\| \leq d_M(x, y) \leq \kappa(\gamma)\|x - y\|. \quad (5.29)$$

iv) *Finally, if  $\|x - y\| \leq \tau/2$ , there holds*

$$\|\text{pr}_{T_x M} - \text{pr}_{T_y M}\|_{\text{op}} \leq \frac{d_M(x, y)}{\tau} \leq \frac{2}{\tau}\|x - y\|. \quad (5.30)$$

*Proof.* The first result on  $\exp_x$  is an application of [AB06, Thm 1.3]. For ii), denoting  $R_x(v) = \exp_x(v) - x - v$ , there holds that,

$$\|\exp_x(v) - \exp_x(w)\| - \|v - w\| \leq \|R_x(v) - R_x(w)\| \leq \frac{5}{16}\|v - w\|$$

where we used [AL19, Lem 1]. Finally iii) comes from the monotonicity of  $\kappa$  and [NSW08, Prp 6.3], and using [BLW19, Lem 6], there holds  $\|\text{pr}_{T_x M} - \text{pr}_{T_y M}\|_{\text{op}} \leq d_M(x, y)/\tau$ , which, together with iii) for  $\gamma = 1$ , leads to iv).  $\square$

We now wish to define the parametrization of the  $\tau/2$ -offset of  $M$  that we introduced in Subsection 5.2.2. This requires to identify in a non-ambiguous way every normal fiber  $N_x M$  to the base fiber  $N_{x_0} M$  for  $x$  in the vicinity of  $x_0$ . A natural way to do that would be to use parallel transport, and define

$$N_{x_0}(v, \eta) := t_\gamma(\eta)$$

where  $t_\gamma : N_{x_0} M \rightarrow N_{\exp_{x_0}(v)} M$  is the parallel transport along the path  $\gamma(s) := \exp_{x_0}(sv)$ . We refer to [Lee06, Sec 4] for a formal introduction to parallel transport.

In order to make things more comprehensible for the reader who is unfamiliar with parallel transports, and in order to have clear and quantitative controls and the



quantity at stake, we suggest another, more elementary approach. We assert that the two approaches yields similar regularity classes as introduced in Subsection 5.2.2. We start off with a few notations. For a matrix  $A \in \mathbb{R}^{D \times D}$  and  $1 \leq k \leq D$ , we let  $V_k(A)$  be the vector space spanned by the first  $k$  columns of  $A$ . We denote by  $\text{Norm} : x \mapsto x/\|x\|$ . We let

$$\mathcal{G} : GL(D, \mathbb{R}) \rightarrow O(D, \mathbb{R})$$

be the Gram-Schmidt process, defined recursively on the columns of any invertible matrix  $A = (A_1, \dots, A_D)$  as  $\mathcal{G}(A) = (\mathcal{G}_1(A), \dots, \mathcal{G}_D(A))$  with

$$\begin{aligned} \mathcal{G}_1(A) &:= \text{Norm}(A_1) \quad \text{and} \quad \forall 1 \leq j \leq D-1, \quad \mathcal{G}_{j+1}(A) := \text{Norm}(\bar{\mathcal{G}}_{j+1}(A)) \\ \text{where } \bar{\mathcal{G}}_{j+1}(A) &:= A_{j+1} - \sum_{1 \leq i \leq j} \langle A_{j+1}, \mathcal{G}_i(A) \rangle \mathcal{G}_i(A). \end{aligned}$$

Because  $\mathcal{G}$  is such that  $V_k(A) = V_k(\mathcal{G}(A))$  for every  $1 \leq k \leq D$ , there holds that  $\bar{\mathcal{G}}_{k+1}(A) = \text{pr}_{V_k(A)^\perp}(A_{k+1})$ , and that  $\bar{\mathcal{G}}_j(A)$  is thus non zero everywhere, so that  $\mathcal{G}$  is a well-defined, smooth application. In order to bound its derivatives, we need to control how  $\bar{\mathcal{G}}_k$  is far away from zero. We introduce

$$GL_\varepsilon(D, \mathbb{R}) := \{A \in GL(D, \mathbb{R}) \mid d(A_{k+1}, V_k(A)) \geq \varepsilon \quad \forall 1 \leq k \leq D\},$$

so that  $\|\bar{\mathcal{G}}_k(A)\| \geq \varepsilon$  for every  $k$  and any  $A \in GL_\varepsilon(D, \mathbb{R})$ , and thus straightforwardly all the derivatives, up to any order, of  $\mathcal{G}$  are bounded on  $GL_\varepsilon(D, \mathbb{R})$ . We let  $B_0$  be an arbitrary basis of  $T_{x_0}M$ ,  $B_\perp$  be an arbitrary basis of  $N_{x_0}M$  and let  $B = (B_0, B_\perp)$ . We define, for  $v \in B_{T_{x_0}M}(0, \tau/4)$ ,  $A_{x_0}(B, v) := (d\Psi_{x_0}(v)[B_0], B_\perp)$ . Note that since  $\Psi_{x_0}$  is a diffeomorphism, there holds

$$V_d(A_{x_0}(B, v)) = \text{Vect}(d\Psi_{x_0}(v)[B_0]) = T_{\Psi_{x_0}(v)}M. \quad (5.31)$$

Set

$$N_{x_0}^B(v, \cdot) := \mathcal{G}(A_{x_0}(B, v))[\cdot], \quad \bar{\Psi}_{x_0}^B(v, \eta) := \Psi_{x_0}(v) + N_{x_0}^B(v, \eta).$$

We show in Lemma 5.A.2 that  $N_{x_0}^B$  and  $\bar{\Psi}_{x_0}^B$  are well defined and smooth, which combined with (5.31) yields in particular that  $N_{x_0}^B(v, \cdot)$  is an isometry between  $N_{x_0}M$  and  $N_{\Psi_{x_0}(v)}M$ .

**Lemma 5.A.2.** *For any  $v \in B_{T_{x_0}M}(0, \tau/4)$ , there holds that  $A_{x_0}(B, v) \in GL_\varepsilon(D, \mathbb{R})$  with  $\varepsilon = 1/2^d$ . Consequently, the map  $v \mapsto N_{x_0}^B(v, \cdot)$  is in  $\mathcal{H}_{\text{iso}}^{\beta_M-1}(B_{T_{x_0}M}(0, \tau), C)$  for some constant  $C_M, \tau, \beta_M$  and  $D$  (and not on  $B$ ). Moreover for any basis  $B$ , it holds that*

$$|\det d\bar{\Psi}_{x_0}^B(v, \eta)| \geq (3/16)^d,$$

for any  $v \in B_{T_{x_0}M}(0, \tau/4)$  and  $\eta \in B_{N_{x_0}M}(0, \tau/2)$ .

*Proof.* First notice that for  $d \leq k \leq D-1$  and  $v \in B_{T_{x_0}M}(0, \tau/4)$ , there holds, letting  $A = A_{x_0}(B, v)$ , and because of (5.31) and the fact that  $B_\perp$  is an orthonormal frame of  $N_{x_0}M$ ,

$$d(A_{k+1}, V_k(A)) = d(A_{k+1}, T_{\Psi_{x_0}(v)}M) \geq 1 - \|\text{pr}_{T_{x_0}M} - \text{pr}_{T_{\Psi_{x_0}(v)}M}\|_{\text{op}} \geq 3/4$$

where we used (5.30). Now we let  $1 \leq k \leq d-1$  and let  $V_0(A) = \{0\}$ . Letting  $Q = (A_1, \dots, A_d)$ , here holds that

$$\prod_{k=0}^{d-1} d(A_{k+1}, V_k(A))^2 = \det Q^\top Q = \det d\Psi_{x_0}(v)^\top d\Psi_{x_0}(v).$$

Using [AL19, Lem 1], it holds  $\|d\Psi_{x_0}(v) - \iota\|_{\text{op}} \leq 5/16$  for  $v \in B_{T_{x_0}M}(0, \tau/4)$  where  $\iota : T_{x_0}M \rightarrow \mathbb{R}^D$  is the inclusion so that for any  $h \in T_{x_0}M$ ,  $\|d\Psi_{x_0}(v)[h]\| \geq (1 - 5/16)\|h\| = 11/16\|h\|$ . In particular  $\det \Psi_{x_0}(v)^\top d\Psi_{x_0}(v) \geq (11/16)^{2d}$ . Since now

$$d(A_{k+1}, V_k(A))^2 \leq \|A_{k+1}\|^2 \leq \|d\Psi_{x_0}(v)\|_{\text{op}}^2 \leq (21/16)^2$$

there holds that for any  $1 \leq k \leq d$ ,

$$d(A_{k+1}, V_k(A))^2 \geq (16/21)^{2(d-1)}(11/16)^d \geq (11/21)^{2d} \geq 1/2^{2d}.$$

Using that the Gram-Schmidt transform is smooth on  $GL_\varepsilon(D, \mathbb{R})$  with  $\varepsilon = 1/2^d$  and Proposition 5.B.7, we obtain that the map  $v \mapsto N_{x_0}^B(v, \cdot)$  is in  $\mathcal{H}_{\text{iso}}^{\beta_M-1}(B_{T_{x_0}M}(0, \tau), C)$ .

Also, in  $B$  and in orthonormal bases of  $T_{\Psi_{x_0}(v)}M$  and  $N_{\Psi_{x_0}(v)}$ , the Jacobian of  $\bar{\Psi}_{x_0}^B(v, \eta)$  write

$$\begin{pmatrix} d\Psi_{x_0}(v) + \text{pr}_{T_{\Psi_{x_0}(v)}M} \circ d_v N_{x_0}^B(v, \eta) & 0 \\ \text{pr}_{N_{\Psi_{x_0}(v)}M} \circ d_v N_{x_0}^B(v, \eta) & N_{x_0}^B(v, \cdot) \end{pmatrix}$$

so that  $|\det d\bar{\Psi}_{x_0}^B(v, \eta)| = \left| \det \left\{ d\Psi_{x_0}(v) + \text{pr}_{T_{\Psi_{x_0}(v)}M} \circ d_v N_{x_0}^B(v, \eta) \right\} \right|$ . We saw earlier in the proof that  $\|d\Psi_{x_0}(v)\|_{\text{op}} \geq 11/16$ . Furthermore, using (5.30), we find that for any small  $w \in T_{x_0}M$ ,

$$\begin{aligned} \|\text{pr}_{T_{\Psi_{x_0}(v)}} \{N_{x_0}^B(v+w, \eta) - N_{x_0}^B(v, \eta)\}\| &= \|(\text{pr}_{T_{\Psi_{x_0}(v+w)}} - \text{pr}_{T_{\Psi_{x_0}(v)}})N_{x_0}^B(v+w, \eta)\| \\ &\leq \frac{\|w\|}{\tau} \|\eta\|, \end{aligned}$$

and consequently  $\|\text{pr}_{T_{\Psi_{x_0}(v)}} \circ d_v N_{x_0}^B(v, \eta)\|_{\text{op}} \leq \|\eta\|/\tau \leq 1/2$ . Thus, for any  $h \in T_{x_0}M$ ,  $\|d\Psi_{x_0}(v)[h] + \text{pr}_{T_{\Psi_{x_0}(v)}M} \circ d_v N_{x_0}^B(v, \eta)[h]\|_{\text{op}} \geq (11/16 - 1/2)\|h\| \geq 3/16\|h\|$  and thus

$$|\det d\bar{\Psi}_{x_0}^B(v, \eta)| \geq (3/16)^d.$$

□

An important feature of the parametrizations  $\bar{\Psi}_{x_0}^B$  is that the subsequent Hölder classes as defined in Definition 5.2.3 do not depend, up to a universal constant, to the choice of a collection of basis  $(B_{x_0})_{x_0 \in M}$ . This is shown in Section 5.A.3.

### 5.A.2 Taylor expansion of M-anisotropic Hölder functions

In this section, we derive a Taylor expansion for manifold-anisotropic Hölder functions. Recall from Section 5.3 that for any  $\sigma, \delta > 0$ ,

$$\Delta_{\sigma,\delta} = \begin{pmatrix} \sigma^{\alpha_0} \text{Id}_d & 0 \\ 0 & \delta \sigma^{\alpha_\perp} \text{Id}_{D-d} \end{pmatrix}.$$

**Corollary 5.A.3.** *Let  $f \in \mathcal{H}_\delta^{\beta_0, \beta_\perp}(M, L)$ . Then, for any  $x_0 \in M$ , any  $w \in \mathcal{W}_{x_0, \delta}$ , and any  $z \in T_{x_0}M \times N_{x_0}M$  such that  $w + \Delta_{\sigma,1}z \in \mathcal{W}_{x_0, \delta}$ , there holds*

$$\bar{f}_{x_0, \delta}(w + \Delta_{\sigma,1}z) = \bar{f}_{x_0, \delta}(w) + \sum_{0 < \langle k, \alpha \rangle < \beta} \sigma^{\langle k, \alpha \rangle} \frac{z^k}{k!} D^k \bar{f}_{\delta, x_0}(w) + R(w, z),$$

where the remainder  $R$  satisfies the following bound

$$|R(w, z)| \leq D \sigma^\beta \|1 \vee z\|_1^{\beta_{\max}} L_{x_0, \delta}(w).$$

*Proof of Corollary 5.A.3.* Simply applying Proposition 5.B.2 to  $\bar{f}_{x_0, \delta}$  yields

$$\bar{f}_{x_0, \delta}(w + \Delta_{\sigma,1}z) = \bar{f}_{x_0, \delta}(w) + \sum_{0 < \langle k, \alpha \rangle < \beta} \frac{(\Delta_{\sigma,1}z)^k}{k!} D^k \bar{f}_{\delta, x_0}(w) + R_{x_0, \delta}(w, w + \Delta_{\sigma,1}z)$$

where  $R_{x_0, \delta}(w, \Delta_{\sigma,1}z)$  satisfies

$$\begin{aligned} |R_{x_0, \delta}(w, w + \Delta_{\sigma,1}z)| &\leq L_{x_0, \delta}(w) \sum_{\beta - \alpha_{\max} \leq \langle k, \alpha \rangle < \beta} \frac{|\Delta_{\sigma,1}z|^k}{k!} \sum_{j=1}^D |(\Delta_{\sigma,1}z)_j|^{\frac{\beta - \langle k, \alpha \rangle}{\alpha_j}} \\ &= L_{x_0, \delta}(w) \sum_{\beta - \alpha_{\max} \leq \langle k, \alpha \rangle < \beta} \sigma^{\langle k, \alpha \rangle} \frac{|z|^k}{k!} \sum_{j=1}^D \sigma^{\beta - \langle k, \alpha \rangle} |z_j|^{\frac{\beta - \langle k, \alpha \rangle}{\alpha_j}} \\ &\leq D \sigma^\beta L_{x_0, \delta}(w) \|1 \vee z\|_1^{\beta_{\max}} \end{aligned}$$

where we used the fact that  $k_j + \frac{\beta - \langle k, \alpha \rangle}{\alpha_j} \leq \beta_j \leq \beta_{\max}$  for any  $1 \leq j \leq D$ .  $\square$

### 5.A.3 Stability by a change of basis

We now show that a change of basis does not interfere with the anisotropic regularity of a map seen through  $\bar{\Psi}_{x_0}^B$ .

**Lemma 5.A.4.** *For any orthonormal basis  $B' = (B'_0, B'_\perp)$  subordinated to  $T_{x_0}M$  and  $N_{x_0}M$ , and for any  $\delta > 0$ , it holds that*

$$(\bar{\Psi}_{x_0, \delta}^B)^{-1} \circ \bar{\Psi}_{x_0, \delta}^{B'}(v, \eta) = (v, C_{B, B'}(v) \eta) \quad (5.32)$$

where  $C_{B, B'}$  is independent of  $\delta$  and is in  $\mathcal{H}_{\text{iso}}^{\beta_M - 1}(\mathcal{B}_{T_{x_0}M}(0, \tau), C)$  for some constant  $C$  depending on  $C_M, \tau, \beta_M$  and  $D$  (and not on  $B$  and  $B'$ ).

*Proof.* Short and simple computations shows that  $C_{B,B'}(v) := N_{x_0}^B(v, \cdot)^\top N_{x_0}^{B'}(v, \cdot)$  so that an application of Lemma 5.A.2 with Proposition 5.B.6 immediately yields the result.  $\square$

**Corollary 5.A.5.** *In the context of Subsection 5.2.2. Assume that  $\beta_0 \leq \beta_\perp \leq \beta_M - 1$ . Then, there exists a constant  $C$  depending on  $C_M, \tau, \beta_M$  and  $D$  such that, if there exists a basis  $B$  such that  $f \circ \bar{\Psi}_{x_0, \delta}^B \in \mathcal{H}_{\text{an}}^\beta(\mathcal{W}_{x_0, \delta}, L_{x_0, \delta}^B)$ , then, for any other orthonormal basis  $B'$ ,  $f \circ \bar{\Psi}_{x_0, \delta}^{B'} \in \mathcal{H}_{\text{an}}^\beta(\mathcal{W}_{x_0, \delta}, CL_{x_0, \delta}^{B'})$ .*

*Proof.* Using the lemma above, there holds

$$f \circ \bar{\Psi}_{x_0, \delta}^{B'} = f \circ \bar{\Psi}_{x_0, \delta}^B(v, C_{B,B'}(v)\eta) = f \circ \bar{\Psi}_{x_0, \delta}^B \circ J(v, \eta)$$

where  $J$  is defined through (5.32). We denote for short  $f^{B'} = f \circ \bar{\Psi}_{x_0, \delta}^{B'}$ ,  $f^B = f \circ \bar{\Psi}_{x_0, \delta}^B$  so that  $f^{B'} = f^B \circ J$ . Taking  $\langle k, \alpha \rangle < \beta$  and using the multivariate Faa di Bruno formula [CS96], we find that  $D^k f^{B'}$  is a sum of product of the form

$$D^\ell(f^B \circ J) \times \prod (D^{k^{(j)}} J)^{\ell^{(j)}}$$

subject to  $|\ell| \leq |k|$ ,  $\sum_j \ell^{(j)} = \ell$  and  $\sum_j |\ell^{(j)}| k^{(j)} = k$  with  $k^{(j)} \neq 0$ . Now notice that  $(D^{k^{(j)}} J)_i = 0$  for  $1 \leq i \leq d$  as soon as  $k_\perp^{(j)} \neq 0$ . For a configuration of  $\ell, \ell^{(j)}$  and  $k^{(j)}$  such that the above product is not zero, there thus holds

$$|\ell_0| = \sum_j |\ell_0^{(j)}| = \sum_{k_\perp^{(j)}=0} |\ell_0^{(j)}| \leq \sum_{k_\perp^{(j)}=0} |\ell_0^{(j)}| |k_0^{(j)}| \leq |k_0|$$

which, together with  $|\ell| \leq |k|$  and  $\alpha_0 \geq \alpha_\perp$ , yields that  $\langle \ell, \alpha \rangle \leq \langle k, \alpha \rangle$  whenever the above product is non zero. We conclude with a telescopic argument with Lemma 5.A.4.  $\square$

#### 5.A.4 Partitions of unity and packings

The approximation result uses a particular covering of an offset of the manifold  $M$ , which we describe here. Take  $x_0 \in M$  and define

$$\mathcal{W}_{x_0}^j := B_{T_{x_0}}\left(0, \frac{2+j}{16}\tau\right) \times B_{N_{x_0}}\left(0, \frac{6+j}{8}\tau\right) \quad \text{for } j \in \{0, 1, 2\},$$

and  $\mathcal{O}_{x_0}^j = \bar{\Psi}_{x_0}(\mathcal{O}_{x_0}^j)$  for  $j \in \{0, 1, 2\}$ . We have

$$\mathcal{W}_{x_0}^0 \subset \mathcal{W}_{x_0}^1 \subset \mathcal{W}_{x_0}^2 \quad \text{and} \quad \mathcal{O}_{x_0}^0 \subset \mathcal{O}_{x_0}^1 \subset \mathcal{O}_{x_0}^2.$$

Furthermore, the sets  $\mathcal{O}_{x_0}^0$  for  $x_0 \in M$  forms a covering of  $M^{3\tau/4}$ . See Figure 5.A.1 for an illustration of these open sets.

In what follows, we will need the notion of packing. An  $\varepsilon$ -packing of a subset  $\mathcal{A} \subset \mathbb{R}^D$  is a set  $\{y_1, \dots, y_J\}$  of points of  $\mathcal{A}$  such that  $\|x_j - x_k\| > \varepsilon$  for any  $1 \leq i \neq j \leq J$  and such that no set of  $J+1$  points has this property. We denote by  $\text{pk}(\mathcal{A}, \varepsilon) := J$  the  $\varepsilon$ -packing

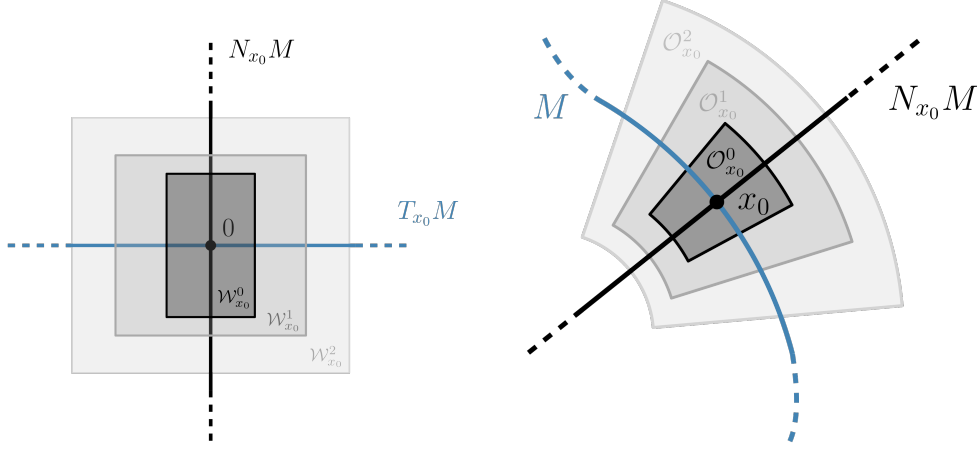


Figure 5.A.1 – A visual representation of the sets used in the proof of Section 5.3.

number of  $\mathcal{A}$ . By maximality of  $J$ , it is straightforward to see that  $\mathcal{A}$  is covered by the union of the balls  $B(x_j, \varepsilon)$ . Furthermore, the balls  $B(x_j, \varepsilon/2)$  must be disjoint by definition of a packing so that

$$\text{pk}(\mathcal{A}, \varepsilon) \times \min_{x \in \mathcal{A}} \text{vol}(\mathcal{A} \cap B(x, \varepsilon/2)) \leq \text{vol} \left\{ \mathcal{A} \cap \bigcup_{1 \leq j \leq J} B(x_j, \varepsilon/2) \right\} \leq \text{vol} \mathcal{A}.$$

In the case where  $\mathcal{A}$  is a ball of radius  $R$  with  $R$  large before  $\varepsilon$ , it is straightforward to see that  $\text{vol}(\mathcal{A} \cap B(x, \varepsilon/2)) \gtrsim \varepsilon^D$  for any  $x \in \mathcal{A}$  so that  $\text{pk}(\mathcal{A}, \varepsilon) \lesssim (R/\varepsilon)^D$ . We now let

$$\rho(x) := \exp \left\{ -\frac{1}{(1 - \|x\|^2)_+} \right\}$$

which is an infinitely differentiable, radially symmetric function from  $\mathbb{R}^D$  to  $[0, 1]$  supported on  $B(0, 1)$ . For any  $x_0 \in M$ , we define

$$\rho_{x_0}(x) := \rho \left( 32 \frac{x - x_0}{\tau} \right)$$

For any large  $R > 0$ , one can take a  $\tau/64$ -packing of  $M \cap B(0, R)$ , say  $\{x_1, \dots, x_J\}$  with  $J$  of order less than  $\text{pk}(B(0, R), \tau/64) \lesssim R^D$ . We can define

$$\chi_j(x) := \frac{\rho_{x_j}(x)}{\sum_{i=1}^J \rho_{x_i}(x)}.$$

In a similar fashion as what is done in [Div21b], we first review a few properties satisfied by the maps  $\{\chi_j\}_{1 \leq j \leq J}$ , which forms a partition of unity associated with the covering  $\{B(x_j, \tau/32)\}_{1 \leq j \leq J}$  of  $M \cap B(0, R)$ . See Figure 5.A.2 for a geometric interpretation of the situation.

**Lemma 5.A.6.** *The following assertions hold true:*

- i) For any  $1 \leq j \leq J$ ,  $\text{supp } \chi_j \subset \mathcal{O}_{x_j}^0$ ;

- ii) There exists a numeric constant  $\gamma > 0$  such that  $M^{\gamma\tau} \cap B(0, R) \subset \text{supp} \sum_j \chi_j$ ;
- iii) There exists a numeric constant  $\nu > 0$ , such that for any  $x \in M^{\gamma\tau} \cap B(0, R)$ , there holds  $\sum_{j=1}^J \rho_{x_j}(x) \geq \nu$ ;
- iv) For any  $|k| \leq K$ , there holds that  $\|D^k \chi_j\|_\infty \leq C < \infty$  with  $C$  depending on  $K$  and  $\tau$ ;
- v) For any  $|k| \leq K$ , there exists a non-negative function  $I_K$  such that, for any  $1 \leq j \leq J$ ,

$$|D^k \chi_j(x)| \leq I_K(x - x_j) \chi_j(x), \quad \forall x \in M^{\gamma\tau} \cap B(0, R)$$

with  $I_K$  being such that for any  $\omega > 0$

$$\sup_{x \in M^{\gamma\tau}} (I_K(x - x_j))^\omega \chi_j(x) \leq C < \infty$$

where  $C$  depends on  $K, \tau$  and  $\omega$ .

We stress out that the constants appearing in Lemma 5.A.6 do not depend on  $R$  or  $J$  at all, so that it can be applied for a family of covering  $\{x_1, \dots, x_J\}$  indexed by  $R \rightarrow \infty$ , as it will be the case in the proofs below.

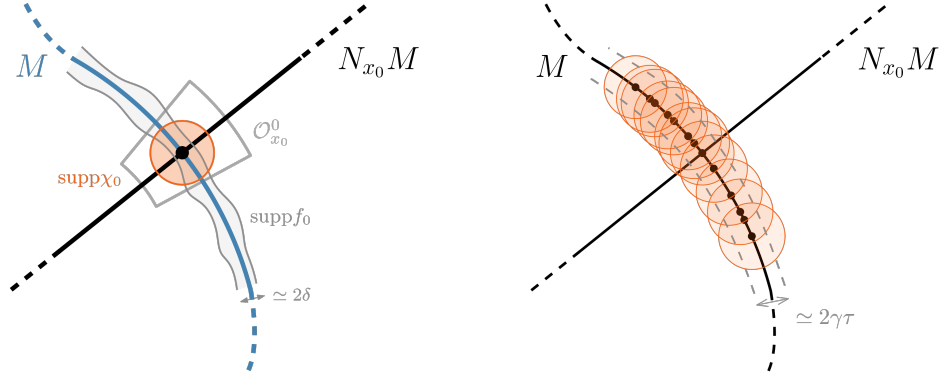


Figure 5.A.2 – A visual interpretation of the framework and results of Lemma 5.A.6.

*Proof of Lemma 5.A.6.* We start with proving i). Let  $x \in B(x_j, \tau/32)$  and let  $z = \text{pr}_M x$ . There holds that

$$\|z - x_j\|^2 = \|z - x\|^2 + \|x - x_j\|^2 + 2\langle z - x, x - x_j \rangle = \|z - x\|^2 - \|x - x_j\|^2 + 2\langle z - x_j, x - x_j \rangle.$$

But now using [Fed59, Thm 4.7 (8)], there holds that  $2\langle z - x_j, x - x_j \rangle \leq \|z - x_j\|^2 \|z - x\|/\tau$  so that in the end

$$\|z - x_j\|^2 \leq \frac{\|z - x\|^2}{1 - \|z - x\|/\tau} \leq \left\{ \sqrt{\frac{32}{31}} \frac{\tau}{32} \right\}^2,$$

where we used the fact that  $\|z - x\| \leq \|x_j - x\| \leq \tau/32$ . In particular, using Lemma 5.29, one get

$$d_M(z, x_j) \leq \kappa \left( \sqrt{\frac{32}{31}} \frac{1}{16} \right) \|z - x_j\| \leq \kappa \left( \sqrt{\frac{32}{31}} \frac{1}{16} \right) \sqrt{\frac{32}{31}} \frac{\tau}{32} < \tau/16$$

where the last inequality was checked with a calculator, so that  $z = \Psi_{x_j}(v)$  with  $v \in B_{T_{x_j}M}(0, \tau/16)$ . Finally  $\|x - z\| \leq \tau/32$  so that  $x - z = N_{x_j}(v, \eta)$  for some  $\eta \in B_{N_{x_j}M}(0, \tau/32)$ . In the end  $x = \bar{\Psi}_{x_j}(v, \eta)$  with  $(v, \eta) \in \mathcal{W}_{x_j}^0$  so  $x \in \mathcal{O}_{x_j}^0$ , ending the proof of the assertion i).

We continue with proving ii) and iii) with  $\gamma = 1/128$  and  $\nu = \exp\{-16/7\}$ . Take  $x \in M^{\gamma\tau} \cap B(0, R)$  and  $z = \text{pr}_M(x)$ . There exists  $x_j$  such that  $\|z - x_j\| \leq \tau/64$ , leading to  $\|x - x_j\| \leq 3\tau/128$  and thus  $x \in \text{supp } \chi_{x_j}$ . Furthermore,

$$\rho_{x_j}(x) = \exp\left\{-\frac{1}{(1 - (32\|x - x_j\|/\tau)^2)_+}\right\} \geq \exp\left\{-\frac{1}{(1 - (3/4)^2)_+}\right\} = \nu,$$

so that points ii) and iii) are proven.

We finish with the proof of iv) and v). We let  $R(x) = (\sum_{i=1}^J \rho_{x_i}(x))^{-1}$ , which is well defined on  $M^{\gamma\tau} \cap B(0, R)$  and bounded from above by  $\nu^{-1}$  so that  $\chi_j(x) = R(x)\rho_{x_j}(x)$  and

$$D^k \chi_j(x) = \sum_{|\ell| \leq k} \binom{k}{\ell} D^{k-\ell} R(x) D^\ell \rho_{x_j}(x).$$

Outside of  $B(x_j, \tau/32)$  one can take  $I_K = 0$ . Now if  $x \in B(x_j, \tau/32)$ , one get that for any  $|\ell| < k$ , using the Faa Di Bruno formula yields that  $D^\ell \rho_{x_j}(x)$  is a sum of term of the form

$$(-1)^s \prod_{i=1}^s D^{r_i} \Upsilon(x) \times \rho_{x_j}(x) \quad \text{with} \quad \Upsilon(x) := \frac{1}{1 - \frac{32^2}{\tau^2} \|x - x_j\|^2} \quad \text{and} \quad \sum_{i=1}^s r_i = \ell.$$

In particular, we get that  $D^\ell \rho_{x_j}(x)$  is bounded from above by some constant depending on  $K$  and  $\tau$  on  $B(x_j, \tau/32)$ . Furthermore, for any  $|\ell| < k$ , the derivative  $D^\ell R(x)$  is a sum of terms of the form

$$R(x)^r \prod_{i=1}^s D^{r_i} R^{-1}(x) \quad \text{with} \quad \sum_{i=1}^s r_i \leq \ell \quad \text{and} \quad r + \sum_{i=1}^s |r_i| = |\ell| + 2,$$

and  $D^{r_i} R^{-1}(x) = \sum_q D^{r_i} \rho_{x_q}(x)$  which is bounded from above by something depending on  $\tau$  and  $K$ . All in all, we get the uniform boundedness of  $D^k \chi_j$  and that

$$|D^k \chi_j(x)| \leq I_K(x - x_j) \chi_j(x)$$

$$\text{with} \quad I_K(x) := \begin{cases} C_{K,\tau} \left\{1 - \frac{32^2}{\tau^2} \|x\|^2\right\}^{-(K+1)} & \text{if } x \in B(0, \tau/32), \\ 0 & \text{otherwise,} \end{cases}$$

where  $C_{K,\tau}$  depends on  $K$  and  $\tau$ , ending the proof.  $\square$

## 5.B Appendix to Section 5.2

### 5.B.1 Auxiliary results on general anisotropic Hölder functions

We first extend Definition 5.2.1 to functions defined on general open sets of  $\mathbb{R}^D$ . For any open set  $\mathcal{U} \subset \mathbb{R}^D$ , any function  $L : \mathcal{U} \rightarrow \mathbb{R}_+$  and any positive real number  $\zeta > 0$ , we

define the anisotropic Hölder spaces  $\mathcal{H}_{\text{an}}^\beta(\mathcal{U}, L, \zeta)$  as the set of functions  $f : \mathcal{U} \rightarrow \mathbb{R}^D$  satisfying that

- i) For any multi-index  $k \in \mathbb{N}^D$  such that  $\langle k, \alpha \rangle < \beta$  the partial derivative  $D^k f$  is well defined on  $\mathcal{U}$  and  $|D^k f(x)| \leq L(x)$  for all  $x \in \mathcal{U}$ ;
- ii) For any multi-index  $k \in \mathbb{N}^D$  such that  $\beta - \alpha_{\max} \leq \langle k, \alpha \rangle < \beta$ , there holds

$$|D^k f(y) - D^k f(x)| \leq L(x) \sum_{i=1}^D |y_i - x_i|^{\frac{\beta - \langle k, \alpha \rangle}{\alpha_i} \wedge 1} \quad \forall x, y \in \mathcal{U}, \|x - y\| \leq \zeta. \quad (5.33)$$

Like in Subsection 5.2.1, we define in a similar fashion  $\mathcal{H}_{\text{iso}}^\beta(\mathcal{U}, L, \zeta)$ ,  $\mathcal{H}_{\text{an}}^\beta(\mathcal{U}, C, \zeta)$  and  $\mathcal{H}_{\text{iso}}^\beta(\mathcal{U}, C, \zeta)$  for some constant  $C > 0$ . We now list a number of useful results which hold for our definition of anisotropy. The proofs of these results are provided below.

**Proposition 5.B.1.** *Let  $f \in \mathcal{H}_{\text{an}}^\beta(\mathcal{U}, L, \zeta)$ . Then for any  $k \in \mathbb{N}^D$  such that  $\langle k, \alpha \rangle < \beta$ , the partially differentiated function  $D^k f$  is in  $\mathcal{H}_{\text{an}}^{\beta^{(k)}}(\mathcal{U}, L, \zeta)$  with*

$$\beta^{(k)} = \left\{ 1 - \frac{\langle k, \alpha \rangle}{\beta} \right\} \beta.$$

Anisotropic Hölder functions enjoy the same convenient Taylor expansion as usual Hölder function.

**Proposition 5.B.2.** *Let  $f \in \mathcal{H}_{\text{an}}^\beta(\mathcal{U}, L, \zeta)$ . Then, for any  $x, y \in \mathcal{U}$  with  $\|x - y\| \leq \zeta$ ,*

$$f(y) = f(x) + \sum_{0 < \langle k, \alpha \rangle < \beta} \frac{(y - x)^k}{k!} D^k f(x) + R(x, y),$$

where the remainder  $R$  satisfies the following bound

$$|R(x, y)| \leq L(x) \sum_{\beta - \alpha_{\max} \leq \langle k, \alpha \rangle < \beta} \frac{|y - x|^k}{k!} \sum_{i=1}^D |y_i - x_i|^{\frac{\beta - \langle k, \alpha \rangle}{\alpha_i}}.$$

**Proposition 5.B.3.** *Let  $f \in \mathcal{H}_{\text{an}}^\beta(\mathcal{U}, L, \zeta)$ . Then, for any  $k \in \mathbb{N}^D$  such that  $\langle k, \alpha \rangle < \beta$ , and any  $x, y \in \mathcal{U}$  with  $\|x - y\| \leq \zeta$ ,*

$$|D^k f(y) - D^k f(x)| \leq CL(x) \left\{ \sum_{i=1}^D |y_i - x_i|^{\frac{\beta - \langle k, \alpha \rangle}{\alpha_i} \wedge 1} \right\}.$$

with  $C$  depending on  $k, \zeta$  and  $D$ .

**Proposition 5.B.4.** *Let  $f \in \mathcal{H}_{\text{an}}^\beta(\mathcal{U}, L, \zeta)$ . Then, for any  $\beta' \leq \beta$ ,  $f \in \mathcal{H}_{\text{an}}^{\beta'}(\mathcal{U}, CL, \zeta)$  with  $C$  depending on  $\beta, \zeta$  and  $D$ .*

**Proposition 5.B.5.** *Let  $f \in \mathcal{H}_{\text{an}}^{\beta_1}(\mathcal{U}_1, L_1, \zeta_1)$  and  $g \in \mathcal{H}_{\text{an}}^{\beta_2}(\mathcal{U}_2, L_2, \zeta_2)$  with  $\mathcal{U}_1 \subset \mathbb{R}^{D_1}$  and  $\mathcal{U}_2 \subset \mathbb{R}^{D_2}$ . Then*

$$f \otimes g \in \mathcal{H}_{\text{an}}^{(\beta_1, \beta_2)}(\mathcal{U}_1 \times \mathcal{U}_2, CL_1 \otimes L_2, \zeta)$$

where  $f \otimes g(x, y) = f(x)g(y)$  and  $\zeta = \zeta_1 \wedge \zeta_2$ , and where  $C$  depends on  $D_1, D_2, \zeta, \beta_1$  and  $\beta_2$ .



**Proposition 5.B.6.** Let  $f \in \mathcal{H}_{\text{an}}^{\beta_1}(\mathcal{U}, L_1, \zeta)$  and  $g \in \mathcal{H}_{\text{an}}^{\beta_2}(\mathcal{U}, L_2, \zeta)$ . Then  $f \times g \in \mathcal{H}_{\text{an}}^{\beta}(\mathcal{U}, L, \zeta)$  with  $\beta = \beta_1 \wedge \beta_2$ , and where, for some constant  $C$  depending on  $\zeta, \beta$  and  $D$ ,

$$L(x) = CL_1(x)L_2(x).$$

**Proposition 5.B.7.** Let  $f \in \mathcal{H}_{\text{iso}}^{\beta_1}(\mathcal{U}_1, L_1)$  and  $g \in \mathcal{H}_{\text{iso}}^{\beta_0}(\mathcal{U}_0, C_0)$  where  $g$  takes its value in  $\mathcal{U}_1$ , and where  $\mathcal{U}_0$  is bounded. Assume furthermore that  $\beta_0 \geq 1$ . Then  $f \circ g \in \mathcal{H}_{\text{iso}}^{\beta}(\mathcal{U}_0, L_1 \circ g)$  where  $\beta = \beta_0 \wedge \beta_1$  and where  $C$  depends on  $C_0, C_1$  and  $\text{diam } \mathcal{U}_0$ .

*Proof of Proposition 5.B.1.* Let  $k \in \mathbb{N}^D$  such that  $\langle k, \alpha \rangle < \beta$ . Note that

$$\alpha = \beta \frac{1}{\beta} = \beta^{(k)} \frac{1}{\beta^{(k)}}.$$

Let  $\ell \in \mathbb{N}^D$  such that  $\langle \ell, \alpha \rangle < \beta^{(k)}$ . Then  $\langle k + \ell, \alpha \rangle < \beta$  so that by definition,  $D^\ell D^k f = D^{k+\ell} f$  exists and is bounded from above by  $L$ . If now  $\ell$  is such that  $\beta^{(k)} - \alpha_{\max} \leq \langle \ell, \alpha \rangle < \beta^{(k)}$ , then  $\beta - \alpha_{\max} \leq \langle k + \ell, \alpha \rangle < \beta$  and for any  $x, y \in \mathcal{U}$  that are at most  $\zeta$ -apart,

$$|D^\ell D^k f(y) - D^\ell D^k f(x)| \leq L(x) \sum_{i=1}^D |y_i - x_i|^{\frac{\beta - \langle k + \ell, \alpha \rangle}{\alpha_i} \wedge 1} = L(x) \sum_{i=1}^D |y_i - x_i|^{\frac{\beta^{(k)} - \langle \ell, \alpha \rangle}{\alpha_i} \wedge 1},$$

ending the proof.  $\square$

*Proof of Proposition 5.B.2.* We let for any  $0 \leq i \leq D$ ,

$$z^{(i)} = (y_1, \dots, y_i, x_{i+1}, \dots, x_D) \in \mathbb{R}^D$$

so that  $z^{(0)} = x$  and  $z^{(D)} = y$ . We prove the results recursively on the integer  $N = \lceil \beta - 1 \rceil$ . If  $N = 0$ , then every coefficient  $\beta_i$  are strictly less than 1, and the results follow immediately from the definition of being  $\beta$ -Hölder (there are no  $k$  that satisfies  $\langle k, \alpha \rangle < \beta$  except  $k = 0$ ). If  $N \geq 1$ , we can order without loss of generality and for ease of notations

$$\beta_1 \leq \dots \leq \beta_k < 1 \leq \beta_{k+1} \leq \dots \leq \beta_D$$

with  $k \leq D - 1$  because  $N \geq 1$ . We write

$$f(y) - f(x) = f(z^{(k)}) - f(x) + \sum_{i=k}^{D-1} f(z^{(i+1)}) - f(z^{(i)}).$$

The first term is simply bounded from above by

$$|f(z^{(k)}) - f(x)| \leq L(x) \sum_{i=1}^k |y_i - x_i|^{\beta/\alpha_i}.$$

Recall we write  $e_i = (0, \dots, 0, 1, 0, \dots, 0)$  with 1 at the  $i$ -th position. For the other terms, we do the Taylor expansion with integral remaining term:

$$\begin{aligned} f(z^{(i)}) - f(z^{(i-1)}) &= \sum_{\ell=1}^{L_i-1} \frac{(y_i - x_i)^\ell}{\ell!} D^{\ell e_i} f(z^{(i-1)}) + R_i(y, x), \quad L_i = \lceil \beta_i - 1 \rceil \geq 1 \\ R_i(y, x) &= \frac{(y_i - x_i)^{L_i}}{(L_i - 1)!} \int_0^1 (1-t)^{L_i-1} D^{L_i e_i} f(z^{(i)}(t)) dt, \\ &\text{with } z^{(i)}(t) = tz_{(i)} + (1-t)z_{(i-1)}. \end{aligned} \tag{5.34}$$

Note that  $L_i \alpha_i = \frac{[\beta_i - 1]}{\beta_i} \beta$  and for all  $k = (k_{1:i}, 0) \neq 0$ ,  $L_i \alpha_i + \langle k, \alpha \rangle \geq \beta$ . Indeed, because the coefficient of  $\beta$  are ordered, there holds

$$L_i \alpha_i + \langle k, \alpha \rangle = \frac{\beta}{\beta_i} \left[ [\beta_i - 1] + \sum_{j \leq i} k_j \beta_i / \beta_j \right] \geq \frac{\beta}{\beta_i} ([\beta_i - 1] + 1) \geq \beta.$$

Therefore

$$\left| D^{L_i e_i} f(z^{(i)}(t)) - D^{L_i e_i} f(x) \right| \leq L(x) \sum_{j=1}^i |y_j - x_j|^{\frac{\beta - L_i \alpha_i}{\alpha_j}} \quad (5.35)$$

and in particular

$$R_i(x, y) = \frac{(y_i - x_i)^{L_i}}{L_i!} D^{L_i e_i} f(x) + \tilde{R}_i(x, y)$$

where  $|\tilde{R}_i(x, y)| \leq L(x) \frac{|y_i - x_i|^{L_i}}{L_i!} \sum_{j=1}^i |y_j - x_j|^{\frac{\beta - L_i \alpha_i}{\alpha_j}}.$

Now we use the induction hypothesis on  $f_{i,\ell} = D^{\ell e_i} f$  which belongs to  $\mathcal{H}^{\beta^{(\ell e_i)}}(\mathcal{U}, L, \zeta)$  (according to Proposition 5.B.1):

$$f_{i,\ell}(z^{(i-1)}) - f_{i,\ell}(x) = \sum_{\substack{0 < \langle k, \alpha \rangle < \beta - \ell \alpha_i \\ k_{i:D} = 0}} \frac{(y - x)^k}{k!} D^{\ell e_i + k} f(x) + R_{i,\ell}(x, y),$$

where the remainder  $R$  satisfies the following bound

$$|R_{i,\ell}(x, y)| \leq L(x) \sum_{\substack{\beta - \alpha_{\max} \leq \langle k + \ell e_i, \alpha \rangle < \beta \\ k_{i:D} = 0}} \frac{|y - x|^k}{k!} \sum_{i=1}^D |y_j - x_j|^{\frac{\beta - \langle k + \ell e_i, \alpha \rangle}{\alpha_j}}.$$

All in all, gathering all the developments yields

$$\begin{aligned} & f(z^{(i)}) - f(z^{(i-1)}) \\ &= \sum_{\ell=1}^{L_i-1} \frac{(y_i - x_i)^\ell}{\ell!} \left\{ D^{\ell e_i} f(x) + \sum_{\substack{0 < \langle k, \alpha \rangle < \beta - \ell \alpha_i \\ k_{i:D} = 0}} \frac{(y - x)^k}{k!} D^{\ell e_i + k} f(x) + R_{i,\ell}(x, y) \right\} + R_i(y, x) \\ &= \sum_{\ell=1}^{L_i-1} \left\{ \sum_{\substack{0 < \langle k + \ell e_i, \alpha \rangle < \beta \\ k_{i:D} = 0}} \frac{(y - x)^{k + \ell e_i}}{(k + \ell e_i)!} D^{\ell e_i + k} f(x) + \frac{(y_i - x_i)^\ell}{\ell!} R_{i,\ell}(y, x) \right\} + R_i(y, x) \\ &= \sum_{\substack{0 < \langle k, \alpha \rangle < \beta \\ k_i \neq 0, k_{(i+1):D} = 0}} \frac{(y - x)^k}{k!} D^k f(x) + \left\{ \tilde{R}_i(y, x) + \sum_{\ell=1}^{L_i-1} \frac{(y_i - x_i)^\ell}{\ell!} R_{i,\ell}(y, x) \right\} \end{aligned}$$

so that

$$f(y) - f(x) = \sum_{0 < \langle k, \alpha \rangle < \beta} \frac{(y-x)^k}{k!} D^k f(x) + \underbrace{\sum_{i \geq k+1} \left\{ \tilde{R}_i(y, x) + \sum_{\ell=1}^{L_i-1} \frac{(y_i - x_i)^\ell}{\ell!} R_{i,\ell}(y, x) \right\}}_{:= R(y, x)} + R_0(y, x)$$

with  $R_0(y, x) = f(z^{(k)}) - f(x)$ , and with  $R(y, x)$  being exactly bounded from above by

$$|R(y, x)| \leq L(x) \sum_{\beta - \alpha_{\max} \leq \langle k, \alpha \rangle < \beta} \frac{|y-x|^k}{k!} \sum_{j=1}^D |y_j - x_j|^{\frac{\beta - \langle k, \alpha \rangle}{\alpha_j}}$$

and Proposition 5.B.2 is proved.  $\square$

*Proof of Proposition 5.B.3.* Either  $\langle k, \alpha \rangle \geq \beta - \alpha_{\max}$  and then there is nothing to show, or  $\langle k, \alpha \rangle < \beta - \alpha_{\max}$ , in which case  $\langle k + e_i, \alpha \rangle < \beta$  for any  $i$  and thus  $d D^k f$  is well defined. There thus exist  $z \in [x, y]$  such that  $D^k f(y) - D^k f(x) = d D^k f(z)[y - x]$  leading to

$$|D^k f(y) - D^k f(x)| \leq \sum_{i=1}^D |D^{k+e_i} f(z)| \times |y_i - x_i|.$$

Using an induction argument, there exists  $C_i > 0$  such that

$$|D^{k+e_i} f(z) - D^{k+e_i} f(x)| \leq C_i L(x) \left\{ \sum_{i=1}^D |y_i - x_i|^{\frac{\beta - \langle k+e_i, \alpha \rangle}{\alpha_i} \wedge 1} \right\} \leq C_i D(1 \vee \zeta) L(x)$$

leading the the right results with  $C = 1 + D(1 \vee \zeta) \max_{1 \leq i \leq D} C_i$ .  $\square$

*Proof of Proposition 5.B.4.* Let  $\beta'$  be the harmonic mean of  $\beta'$  and  $\alpha' = \beta'/\beta'$ . For any  $k \in \mathbb{N}^D$  such that  $\langle k, \alpha' \rangle < \beta'$ , we have  $\langle k, \alpha \rangle < \beta$  so that  $D^k f$  is well defined and bounded from above by  $L$ . What's more, using Proposition 5.B.3, we have

$$|D^k f(y) - D^k f(x)| \leq CL(x) \left\{ \sum_{i=1}^D |y_i - x_i|^{\frac{\beta - \langle k, \alpha \rangle}{\alpha_i} \wedge 1} \right\}.$$

Now notice that

$$\frac{\beta - \langle k, \alpha \rangle}{\alpha_i} = \beta_i \{1 - \langle k, 1/\beta \rangle\} \geq \beta'_i \{1 - \langle k, 1/\beta' \rangle\} = \frac{\beta' - \langle k, \alpha' \rangle}{\alpha'_i}$$

so that

$$|y_i - x_i|^{\frac{\beta - \langle k, \alpha \rangle}{\alpha_i} \wedge 1} \leq (1 \vee \zeta) \times |y_i - x_i|^{\frac{\beta' - \langle k, \alpha' \rangle}{\alpha'_i} \wedge 1}$$

yielding the result.  $\square$

*Proof of Proposition 5.B.5.* Let  $k = (k_1, k_2) \in \mathbb{N}^{D_1+D_2}$ . Let  $\beta = (\beta_1, \beta_2)$  and  $\beta, \beta_1, \beta_2$  be the harmonic means of  $\beta, \beta_1$  and  $\beta_2$ . Let also  $\alpha = \beta/\beta$  and  $\alpha_i = \beta_i/\beta_i$  for  $i \in \{1, 2\}$ . If  $\langle k, \alpha \rangle < \beta$ , then

$$\langle k_1, 1/\beta_1 \rangle + \langle k_2, 1/\beta_2 \rangle < 1$$

so that both term is stricly less that one  $D^{k_1}f$  and  $D^{k_2}g$  are well defined and  $D^k(f \otimes g) = D^{k_1}f \otimes D^{k_2}g$  is well defined as well. Furthermore, if  $x = (x_1, x_2)$  and  $y = (y_1, y_2)$ , is such that  $\|x - y\| \leq \zeta$ , then  $\|x_1 - y_1\| \leq \zeta \leq \zeta_1$  and  $\|x_2 - y_2\| \leq \zeta \leq \zeta_2$ . Using Proposition 5.B.3, one find that

$$\begin{aligned} & |D^k(f \otimes g)(x) - D^k(f \otimes g)(y)| \\ & \leq |D^{k_2}g(x_2)| \times |D^{k_1}f(x_1) - D^{k_1}f(y_1)| + |D^{k_1}f(y_1)| \times |D^{k_2}g(x_2) - D^{k_2}g(y_2)| \\ & \leq CL_2(x_2)L_1(x_1) \left\{ \sum_{i=1}^{D_1} |y_{1,i} - x_{1,i}|^{\frac{\beta_1 - \langle k_1, \alpha_1 \rangle}{\alpha_{1,i}} \wedge 1} \right\} \\ & + C|D^{k_1}f(y_1)|L_2(x_2) \left\{ \sum_{i=1}^{D_2} |y_{2,i} - x_{2,i}|^{\frac{\beta_2 - \langle k_2, \alpha_2 \rangle}{\alpha_{2,i}} \wedge 1} \right\}. \end{aligned}$$

for some constant  $C > 0$  depending on  $D_1, D_2, \zeta$  and  $\beta$ . Now notice first that

$$|D^{k_1}f(y_1)| \leq |D^{k_1}f(y_1) - D^{k_1}f(x_1)| + |D^{k_1}f(x_1)| \leq (CD(1 \vee \zeta) + 1)L_1(x_1),$$

and notice furthermore that of either  $j \in \{1, 2\}$ ,

$$\frac{\beta_j - \langle k_j, \alpha_j \rangle}{\alpha_{j,i}} = \beta_{j,i} \{1 - \langle k_j, 1/\beta_j \rangle\} \geq \beta_{j,i} \{1 - \langle k, 1/\beta \rangle\} = \frac{\beta - \langle k, \alpha \rangle}{\alpha_i} \left( = \frac{\beta - \langle k, \alpha \rangle}{\alpha_{i+D_1}} \text{ if } j = 2 \right)$$

so that

$$\begin{aligned} & |D^k(f \otimes g)(x) - D^k(f \otimes g)(y)| \\ & \leq (1 \vee \zeta)CL_2(x_2)L_1(x_1) \left\{ \sum_{i=1}^{D_1} |y_{1,i} - x_{1,i}|^{\frac{\beta - \langle k, \alpha \rangle}{\alpha_i} \wedge 1} \right\} \\ & + (1 \vee \zeta)C(CD(1 \vee \zeta) + 1)L_1(x_1)L_2(x_2) \left\{ \sum_{i=1}^{D_2} |y_{2,i} - x_{2,i}|^{\frac{\beta - \langle k, \alpha \rangle}{\alpha_{i+D_1}} \wedge 1} \right\} \\ & \leq (1 \vee \zeta)C(CD(1 \vee \zeta) + 1)(L_1 \otimes L_2)(x) \left\{ \sum_{i=1}^{D_1+D_2} |y_i - x_i|^{\frac{\beta - \langle k, \alpha \rangle}{\alpha_i} \wedge 1} \right\} \end{aligned}$$

ending the proof. □

*Proof of Proposition 5.B.6.* We let  $\alpha = \beta/\beta$  with  $\beta$  the harmonic mean of  $\beta$ . Now for any multi-index  $k$  such that  $\langle k, \alpha \rangle < \beta$ , we have  $\langle \ell, \alpha \rangle < \beta$  for any  $\ell \leq k$  so that  $D^kfg$  is indeed well-defined and

$$D^kfg(x) = \sum_{\ell \leq k} \binom{k}{\ell} D^\ell f(x) D^{k-\ell}g(x)$$

so that  $|D^k f g(x)| \leq 2^{|k|} L_1(x) L_2(x)$  and

$$|D^k f g(y) - D^k f g(x)| \leq L_2(x) \sum_{\ell \leq k} \binom{k}{\ell} |D^\ell f(y) - D^\ell f(x)| + L_1(x) \sum_{\ell \leq k} \binom{k}{\ell} |D^\ell g(y) - D^\ell g(x)|.$$

Using now Proposition 5.B.3, we find that there exists  $C > 0$  depending on  $\beta$ ,  $D$  and  $\zeta$  such that

$$\begin{aligned} |D^k f g(y) - D^k f g(x)| &\leq 2C L_1(x) L_2(x) \sum_{\ell \leq k} \binom{k}{\ell} \sum_{i=1}^D |y_i - x_i|^{\frac{\beta - \langle \ell, \alpha \rangle}{\alpha_i} \wedge 1} \\ &\leq 2^{|k|+1} (1 \vee \zeta) C L_1(x) L_2(x) \sum_{i=1}^D |y_i - x_i|^{\frac{\beta - \langle k, \alpha \rangle}{\alpha_i} \wedge 1} \end{aligned}$$

where we used again that

$$|y_i - x_i|^{\frac{\beta - \langle \ell, \alpha \rangle}{\alpha_i} \wedge 1} \leq (1 \vee \zeta) \times |y_i - x_i|^{\frac{\beta - \langle k, \alpha \rangle}{\alpha_i} \wedge 1},$$

for  $\ell \leq k$ , ending the proof.  $\square$

*Proof of Proposition 5.B.7.* A simple use of the multivariate Faa Di Bruno formula [CS96] yields that, for any  $k \in \mathbb{N}^D$  with  $|k| \leq \beta$ ,  $D^k(f \circ g)(x)$  for  $x \in \mathcal{U}_0$  is a sum a term of the form

$$D^\ell f(g(x)) \prod_{j=1}^s (D^{k^{(j)}} g(x))^{\ell^{(j)}}$$

with  $|\ell| \leq |k|$ ,  $s \leq |k|$ ,  $\sum_j \ell^{(j)} = \ell$ , and  $\sum_j |\ell^{(j)}| |k^{(j)}| = k$ . In particular, it is bounded by  $C L_1(g(x))$  where  $C$  depends on  $C_0$  and  $\beta$ . Likewise, a telescopic argument would yield that

$$|D^k(f \circ g)(y) - D^k(f \circ g)(x)| \leq C L_1(g(x)) \prod_{i=1}^D |x_i - y_i|^{(\beta - |k|) \wedge 1}$$

where  $C$  depends on  $C_0$ ,  $\beta$  and  $\text{diam } \mathcal{U}_0$ , ending the proof.  $\square$

## 5.B.2 Proofs associated to the examples of Proposition 5.2.4

*Proof of Proposition 5.2.4.* Take  $x_0 \in M$ . In the orthonormal noise model, the density  $f$  takes the very simple form

$$\begin{aligned} \bar{f}_{x_0, \delta}(v, \eta) &= \delta^{D-d} f \circ \bar{\Psi}_{x_0, \delta}(v, \eta) = \delta^{D-d} \times f_0(\Psi_{x_0}(v)) \times \delta^{-(D-d)} c_\perp K\left(\frac{1}{\delta} N_{x_0}(v, \delta \eta)\right) \\ &= f_{x_0}(v) \times c_\perp K(\eta) \end{aligned}$$

where we used the isotropy of  $K$ . Thus,  $\bar{f}_{x_0, \delta}$  lies in  $\mathcal{H}_{\text{an}}^\beta(\mathcal{W}_{x_0, \delta}, L_{x_0, \delta})$  where  $L$  was defined in the statement of Proposition 5.2.4, through Proposition 5.B.5 and the isotropy of  $L_\perp$ .

In the isotropic noise model, one can write

$$\bar{f}_{x_0, \delta}(v, \eta) = \delta^{D-d} \int_M \delta^{-D} K\left(\frac{\bar{\Psi}_{x_0, \delta}(v, \eta) - x}{\delta}\right) f_0(x) d\mu(x).$$

If  $\|\eta\| \geq 1$ , the integrand is trivially 0, so one may focus on  $\|\eta\| \leq 1$ . Now if  $x \in M$  is at least  $2\delta$  apart from  $\Psi_{x_0}(v)$ , there holds

$$\|x - \bar{\Psi}_{x_0,\delta}(v, \eta)\| \geq \|x - \Psi_{x_0}(v)\| - \delta \geq \delta$$

so that the integrand in the integral above is zero for  $x$  outside of  $B(\Psi_{x_0}(v), 2\delta)$ . Doing the variable change  $x = \Psi_{x_0}(w)$ , we can write

$$\begin{aligned} \bar{f}_{x_0,\delta}(v, \eta) &= \frac{1}{\delta^d} \int_{\exp_{x_0}^{-1} B(\Psi_{x_0}(v), 2\delta)} K\left(\frac{\bar{\Psi}_{x_0,\delta}(v, \eta) - \Psi_{x_0}(w)}{\delta}\right) f_{x_0}(w) |\det d\Psi_{x_0}(w)| dw \\ &= \int_{\mathcal{Z}_\delta} \underbrace{K\left(\frac{\Psi_{x_0}(v) - \Psi_{x_0}(v + \delta s)}{\delta} + N_{x_0}(v, \eta)\right)}_{=: K \circ g_{\delta,s}(v, \eta)} \underbrace{f_{x_0}(v + \delta s) |\det d\Psi_{x_0}(v + \delta s)|}_{=: h_{\delta,s}(v)} ds \end{aligned}$$

where

$$\mathcal{Z}_\delta = \frac{1}{\delta} \left\{ \exp_{x_0}^{-1} B(\Psi_{x_0}(v), 2\delta) - v \right\}.$$

Now notice that, using the first inequality of (5.29), we have  $B(\Psi_{x_0}(v), 2\delta) \subset B(x_0, \tau/8 + 2\delta) \subset B(x_0, 3\tau/16)$ , so that, using the second inequality of (5.29) with  $\gamma = 3/8$ , we find

$$\exp_{x_0}^{-1} B(\Psi_{x_0}(v), 2\delta) \subset B_{T_{x_0}M}(0, \kappa(3/8)3\tau/16) \subset B_{T_{x_0}M}(0, \tau/4)$$

so that in particular, according to (5.28),  $\exp_{x_0}^{-1}$  is  $16/11$ -Lipschitz on  $B(\Psi_{x_0}(v), 2\delta)$  and consequently,  $\mathcal{Z}_\delta \subset B_{T_{x_0}M}(0, 32/11)$ . Furthermore, notice that

$$v + \delta B_{T_{x_0}M}(0, 32/11) \subset B_{T_{x_0}M}(0, 19\tau/88)$$

with  $19\tau/88$  being smaller than the injectivity radius, so that finally

$$\bar{f}_{x_0,\delta}(v, \eta) = \int_{B_{T_{x_0}M}(0, 32/11)} K \circ g_{\delta,s}(v, \eta) \times h_{\delta,s}(v) ds \quad (5.36)$$

It is straightforward to see that

$$v \mapsto \frac{\Psi_{x_0}(v) - \Psi_{x_0}(v + \delta s)}{\delta}$$

has derivatives bounded from above, in virtue of Proposition 5.B.3, by something depending on  $C_M$  and  $\beta_M$  only (and not on  $\delta$ ) and up to order  $\lceil \beta_M - 1 \rceil \geq \beta_\perp$ . Furthermore,  $(v, \eta) \mapsto N_{x_0}(v, \eta)$  is  $(\beta_M - 1)$ -Hölder by construction. It so happens that  $g_{\delta,s} \in \mathcal{H}_{\text{iso}}^{\beta_\perp}(\mathcal{W}_{x_0,\delta}, C_g)$  for some  $g$  depending on  $C_M$  and  $\beta_M$ . Using Proposition 5.B.7, we get that  $K \circ g_{\delta,s} \in \mathcal{H}_{\text{iso}}^{\beta_\perp}(\mathcal{W}_{x_0,\delta}, C_1 L_\perp \circ g_{\delta,s})$  for some constant  $C_1$  depending on  $C_g$ ,  $C_K$  and  $\beta_\perp$ . Similar reasoning and the use of Proposition 5.B.6 would lead to  $h_{\delta,s} \in \mathcal{H}_{\text{iso}}^{\beta_0}(\mathcal{V}_{x_0}, C_2 L_0 \circ \Psi_{x_0} \circ \tau_{\delta s})$  where  $\tau_{\delta s}(v) = v + \delta s$  and for some constant  $C_2$  depending on  $C_M$ ,  $\tau$  and  $\beta_0$ . Now seeing  $h_{\delta,s}$  as a function of  $(v, \eta)$  would trivially lead to

$h_{\delta,s} \in \mathcal{H}_{\text{an}}^\beta(\mathcal{W}_{x_0,\delta}, C_2 L_0 \circ \Psi_{x_0} \circ \tau_{\delta s} \circ \pi_1)$  with  $\beta = (\beta_0, \dots, \beta_0, \beta_\perp, \dots, \beta_\perp)$  and  $\pi_1(v, \eta) = v$ , and an application of Proposition 5.B.6 together with the observation that  $\beta_0 \leq \beta_\perp$  yields

$$K \circ g_{\delta,s} \times h_{\delta,s} \in \mathcal{H}_{\text{an}}^\beta(\mathcal{W}_{x_0,\delta}, C L_\perp \circ g_{\delta,s} \times L_0 \circ \Psi_{x_0} \circ \tau_{\delta s} \circ \pi_1)$$

for some constant  $C$  depending on  $C_2$ ,  $C_1$  and  $\beta_0, \beta_\perp$ . We conclude by using the linearity of the integral, so that integrating (5.36) gives  $\tilde{f}_{x_0,\delta} \in \mathcal{H}_{\text{an}}^\beta(\mathcal{W}_{x_0,\delta}, \tilde{L})$ , where

$$\begin{aligned} \tilde{L}(v, \eta) &= C \int_{\mathcal{B}_{T_{x_0}M}(0, 32/11)} L_\perp \circ g_{\delta,s}(v, s) \times L_0 \circ \Psi_{x_0}(v + \delta s) ds \\ &\lesssim \int_{\mathcal{B}_{T_{x_0}M}(0, 32/11)} L_\perp \circ g_{\delta,s}(v, s) \times L_0 \circ \Psi_{x_0}(v + \delta s) \times |\det d\Psi_{x_0}(v + \delta s)| ds \\ &= \delta^{-d} \int_M L_\perp \left( \frac{\bar{\Psi}_{x_0,\delta}(v, \eta) - x}{\delta} \right) \times L_0(x) d\mu_M(x) \end{aligned}$$

so that

$$L(y) = C \delta^{-D} \int_M L_\perp \left( \frac{y - x}{\delta} \right) \times L_0(x) d\mu_M(x)$$

where  $L$  was defined in the statement of Proposition 5.2.4, and where we used Lemma 5.A.2 to justify the introduction of  $|\det d\Psi_{x_0}(v + \delta s)|$ . This ends the proof.  $\square$

*Proof of Lemma 5.3.6.* In the orthogonal noise model, recall (see proof of Proposition 5.2.4) that

$$\begin{cases} f_0(x) = f_*(\text{pr}_M(x)) \times c_\perp \delta^{-(D-d)} K((x - \text{pr}_M x)/\delta) \\ L(x) = C \delta^{-(D-d)} L_0(\text{pr}_M(x)) \times L_\perp((x - \text{pr}_M x)/\delta), \end{cases}$$

for some  $C > 0$ . There thus holds, letting  $\mathcal{O}_{x_0} = \bar{\Psi}_{x_0,1}(\mathcal{W}_{x_0,1})$ , we a simple use of Cauchy-Schwartz inequality

$$\begin{aligned} &\left\{ \int_{\mathcal{O}_{x_0}} \left\{ \frac{L(x)}{f_0(x)} \right\}^{\omega^*/2} f_0(x) dx \right\}^2 \\ &\lesssim \int_{\mathcal{O}_{x_0}} \left\{ \frac{L_0(\text{pr}_M x)}{f_*(\text{pr}_M x)} \right\}^{\omega^*} f_0(x) dx \int_{\mathcal{O}_{x_0}} \left\{ \frac{L_\perp((x - \text{pr}_M x)/\delta)}{K((x - \text{pr}_M x)/\delta)} \right\}^{\omega^*} f_0(x) dx. \end{aligned}$$

The first integral in the RHS is simply, up to the constant  $c_\perp$ ,

$$\int_{\Psi_{x_0}(\mathcal{V}_{x_0})} \left\{ \frac{L_0(z)}{f_*(z)} \right\}^{\omega^*} f_*(z) d\mu_M(z),$$

which is bounded by assumption. The second integral is exactly

$$\int_{\Psi_{x_0}(\mathcal{V}_{x_0})} f_*(z) d\mu_M(z) \times \int_{\mathcal{B}_{\mathbb{R}^{D-d}}(0,1)} \left\{ \frac{L_\perp(\eta)}{K(\eta)} \right\}^{\omega^*} K(\eta) dx.$$

which is also bounded by assumption. Hence (5.13) holds true in this case.

In the isotropic noise model, there holds this time

$$\begin{cases} f_0(x) = \delta^{-D} \int_{M \cap B(x, \delta)} K((y-x)/\delta) f_*(y) d\mu_M(y) \\ L(x) = C \delta^{-D} \int_{M \cap B(x, \delta)} L_\perp((y-x)/\delta) L_0(y) d\mu_M(y), \end{cases}$$

for some  $C > 0$ . Using the fact that for any integrable function  $\phi, \psi \geq 0$  and any  $\omega \geq 1$ , there holds

$$\left\{ \int \phi \right\}^\omega \leq \left\{ \int \frac{\phi^\omega}{\psi^{\omega-1}} \right\} \times \left\{ \int \psi \right\}^{\omega-1},$$

(which is simple consequence of the Hölder inequality), one find that

$$\begin{aligned} & \int_{\mathcal{O}_{x_0}} \left\{ \frac{L(x)}{f_0(x)} \right\}^{\omega^*/2} f_0(x) dx \\ & \lesssim \int_{\mathcal{O}_{x_0}} \int_M \left\{ \frac{L_\perp((y-x)/\delta) L_0(y)}{K((y-x)/\delta) f_*(y)} \right\}^{\omega^*/2} \delta^{-D} K((y-x)/\delta) f_*(y) d\mu_M(y) dx \\ & \leq \delta^{-D} I_1^{1/2} I_2^{1/2}, \end{aligned}$$

where

$$\begin{aligned} I_1 &:= \int_{\mathcal{O}_{x_0}} \int_M \left\{ \frac{L_0(y)}{f_*(y)} \right\}^{\omega^*} K((y-x)/\delta) f_*(y) d\mu_M(y) dx \\ &\lesssim \delta^D \int_M \left\{ \frac{L_0(y)}{f_*(y)} \right\}^{\omega^*} f_*(y) d\mu_M(y) \lesssim \delta^D, \end{aligned}$$

by assumption, and where

$$\begin{aligned} I_2 &:= \int_{\mathcal{O}_{x_0}} \int_M \left\{ \frac{L_\perp((y-x)/\delta)}{K((y-x)/\delta)} \right\}^{\omega^*} K((y-x)/\delta) f_*(y) d\mu_M(y) dx \\ &\lesssim \delta^D \int_M f d\mu_M \int_{B(0,1)} \left\{ \frac{L_\perp(\eta)}{K(\eta)} \right\}^{\omega^*} K(\eta) d\eta, \end{aligned}$$

so that indeed  $\delta^{-D} I_1^{1/2} I_2^{1/2} \lesssim 1$ , which ends the proof.  $\square$

### 5.B.3 Proof of Proposition 5.2.7

If  $Q_2 = Q_0^{\otimes D}$  with  $Q_0 \sim \text{DP}(BH_\lambda) := \Pi_{\text{DP}}$  then, for  $0 < x_1 \leq x_2/2$  and  $x_1 < 1$ , letting  $A_1 = [x_1, x_1(1+x_1^b)]$ ,  $A_2 = [x_2, x_2(1+x_1^b)]$  and  $A_3 = (A_1 \cup A_2)^c$ , we have

$$(Q_0(A_1), Q_0(A_2), Q_0(A_3)) \sim \mathcal{D}(BH_\lambda(A_1), BH_\lambda(A_2), BH_\lambda(A_3)),$$

and

$$Q_2(A_1^d \times A_2^{D-d}) = Q_0(A_1)^d Q_0(A_2)^{D-d}.$$



We let  $\gamma = \Gamma(B) / (\Gamma(BH_\lambda(A_1))\Gamma(BH_\lambda(A_2))\Gamma(BH_\lambda(A_3)))$ . Noticing that when  $x_1, x_2$  are small,  $H_\lambda(A_3) = 1 - o(1)$ , there holds

$$\begin{aligned} \tilde{\Pi}_\Lambda[Q_2(A_1^d A_2^{D-d}) \geq x_1^{B_0}] &\geq \Pi_{DP}(Q_0(A_1) > x_1^{B_0/D}, Q_0(A_2) > x_1^{B_0/D}) \\ &\geq \frac{\gamma}{2^{(B-1)_+}} \int_{x_1^{B_0/D}}^{1/4} x^{BH_\lambda(A_1)-1} dx \int_{x_1^{B_0/D}}^{1/4} x^{BH_\lambda(A_2)-1} dx \\ &\geq \frac{\Gamma(B)}{(1+o(1))2^{(B-1)_+}} \left[ 4^{-BH_\lambda(A_1)} - x_1^{\frac{BH_\lambda(A_1)B_0}{D}} \right] \left[ 4^{-BH_\lambda(A_2)} - x_1^{\frac{BH_\lambda(A_2)B_0}{D}} \right] \\ &\geq \frac{\Gamma(B)4^{-2BH_\lambda(A_1)}}{(1+o(1))2^{(B-1)_+}} \left[ 1 - e^{-\frac{BH_\lambda(A_1)B_0}{D} \log(4^{D/B_0} x_1)} \right]^2, \end{aligned}$$

where we used the fact that for  $x_1$  and  $x_2$  small,  $H_\lambda(A_1) \leq H_\lambda(A_2)$  and were both small. Under the inverse Gamma assumption on  $H_\lambda$ ,

$$e^{-a_2 x_1^{-1/2}} x_1^{-\frac{a_1+1}{2}} x_1^{1+b} \lesssim H_\lambda(A_1) \lesssim e^{-a_2 x_1^{-1/2}/\sqrt{2}} x_1^{-\frac{a_1+1}{2}} x_1^{1+b} \lesssim e^{-a_2 x_1^{-1/2}/2}$$

so that for small  $x_1$ ,

$$1 - e^{-\frac{BH_\lambda(A_1)B_0}{D} \log(4^{D/B_0} x_1)} \gtrsim e^{-a_2 x_1^{-1/2}} x_1^{-\frac{a_1+1}{2}} x_1^{1+b} |\log(x_1)| \gtrsim e^{-2a_2 x_1^{-1/2}}$$

so that (5.10) holds. Condition (5.11) is verified similarly. First note that  $Q_2(\min_{i \leq D} \lambda_i \leq x) \leq DQ_0((0, x])$  and using the fact that  $Q_0((0, x])$  follows a Beta variable with parameters  $BH_\lambda((0, x]), B(1+o(1))$  we obtain that

$$\mathbb{E}_{\tilde{\Pi}_\Lambda} \left[ Q_2 \left( \min_{i \leq D} \lambda_i \leq x \right) \right] \lesssim H_\lambda((0, x]) \lesssim e^{-a_2 x^{-1/2}/2},$$

for small  $x$ . Similar computation terminates the proof of (5.11)

## 5.C Appendix to Section 5.5.2: proof of Theorem 5.3.7

### 5.C.1 Technical Lemmata

We let  $\chi_1, \dots, \chi_J$  be the partition of unity defined in Section 5.A.4 associated with a  $\tau/64$ -packing of  $\{x_1, \dots, x_J\}$  of  $M \cap B(0, R)$ . We recall that  $J \lesssim R^D$  and we take  $R = (H \log(1/\sigma))^{1/\kappa}$ . We define

$$f_j := c_j^{-1} \chi_j \times f_0 \quad \text{where} \quad c_j = \int_{\mathbb{R}^D} \chi_j(x) f_0(x) dx.$$

Before we give the proof of Corollary 5.3.8 we need a few technical Lemmata.

**Lemma 5.C.1.** *For all  $1 \leq j \leq J$ ,  $f_j \in \mathcal{H}_\delta^{\beta_0, \beta_1}(M, L_j)$  with  $L_j(x) := C c_j^{-1} I_K(x - x_j) \chi_j(x) L(x)$  where  $C > 0$  is a constant depending on  $\tau$  and where we took  $K = \lceil \beta_1 \rceil$ .*

*Proof.* This is an easy corollary of Proposition 5.B.6 and Proposition 5.B.7 along with the assertion iv) from Lemma 5.A.6.  $\square$

We let  $L_{j,\delta} := \delta^{D-d} L_j \circ \bar{\Psi}_{x_j,\delta}$  and

$$\bar{f}_{j,\delta} := f_j \circ \bar{\Psi}_{x_j,\delta} : \mathcal{W}_{x_j,\delta} \rightarrow \mathbb{R}^D.$$

and also introduce the functions  $g_j : \mathbb{R}^D \rightarrow \mathbb{R}$  from the proof of Theorem 5.3.7. Recall that these functions are of the form

$$g_j(x) := f_j(x) + \frac{1}{\delta^{D-d}} \sum_{0 < \langle k, \alpha \rangle < \beta} \sigma^{\langle k, \alpha \rangle} d_{j,k}(x, \sigma, \delta) D^k \bar{f}_{j,\delta}(z_x) \quad \text{where } z_x := \Delta_{1,\delta}^{-1} \bar{\Psi}_{x_j,\delta}^{-1}(x)$$

and satisfy

1. They are supported on  $\mathcal{O}_{x_j}^0$ ;
2. The functions  $d_{j,k}$  are uniformly bounded by a constant  $C$  depending on  $C_M$ ;
3.  $|K_\Sigma g_j(x) - f_j(x)| \lesssim \sigma^\beta L_j(x)$  on  $\mathcal{O}_{x_j}^1$ ;
4.  $|K_\Sigma g_j(x) - f_j(x)| \lesssim \sigma^H \|L_j\|_\infty$  outside of  $\mathcal{O}_{x_j}^1$ .

**Lemma 5.C.2.** *Under (5.13), there holds, for any  $\varepsilon \in (0, \omega - 2\beta)$ , any  $\langle k, \alpha \rangle < \beta$  and any  $1 \leq j \leq J$ ,*

$$\int \left( \frac{|D^k \bar{f}_{j,\delta}(z)|}{\bar{f}_{j,\delta}(z)} \right)^{\frac{2\beta+\varepsilon}{\langle k, \alpha \rangle}} \bar{f}_{j,\delta}(z) dz \lesssim c_j^{-1} \quad \text{and} \quad \int \left( \frac{L_{j,\delta}(z)}{\bar{f}_{j,\delta}(z)} \right)^{\frac{2\beta+\varepsilon}{\beta}} \bar{f}_{j,\delta}(z) dz \lesssim c_j^{-1},$$

up to a constant depending on the parameters.

*Proof.* We denote by  $\bar{\chi}_{j,\delta} = \chi_j \circ \bar{\Psi}_{x_j,\delta}$  and  $\bar{I}_{j,\delta} := I_K(\bar{\Psi}_{x_j,\delta}(\cdot) - x_j)$ . Writing that

$$\left| \frac{D^k \bar{f}_{j,\delta}}{\bar{f}_{j,\delta}} \right| \leq \sum_{\ell \leq k} \binom{k}{\ell} \left| \frac{D^{k-\ell} \bar{\chi}_{j,\delta}}{\bar{\chi}_{j,\delta}} \right| \times \left| \frac{D^\ell \bar{f}_{x_j,\delta}}{\bar{f}_{x_j,\delta}} \right| \lesssim \bar{I}_{j,\delta} \sum_{\ell \leq k} \left| \frac{D^\ell \bar{f}_{x_j,\delta}}{\bar{f}_{x_j,\delta}} \right|$$

we easily see that

$$\begin{aligned} \int \left| \frac{D^k \bar{f}_{j,\delta}}{\bar{f}_{j,\delta}} \right|^{\frac{2\beta+\varepsilon}{\langle k, \alpha \rangle}} \bar{f}_{j,\delta} &\lesssim c_j^{-1} \sum_{\ell \leq k} \int \left\{ \bar{I}_{j,\delta} \left| \frac{D^\ell \bar{f}_{x_j,\delta}}{\bar{f}_{x_j,\delta}} \right| \right\}^{\frac{2\beta+\varepsilon}{\langle k, \alpha \rangle}} \bar{\chi}_{j,\delta} \bar{f}_{x_j,\delta} \\ &\lesssim c_j^{-1} \sum_{\ell \leq k} \int \left| \frac{D^\ell \bar{f}_{x_j,\delta}}{\bar{f}_{x_j,\delta}} \right|^{\frac{2\beta+\varepsilon}{\langle k, \alpha \rangle}} \bar{f}_{x_j,\delta} \lesssim c_j^{-1} \end{aligned}$$

where we used both Lemma 5.A.6 iv) and (5.13) with the fact that  $\langle \ell, \alpha \rangle \leq \langle k, \alpha \rangle$  for  $\ell \leq k$ . The bound on the second integral follows from the same line of reasoning.  $\square$

**Lemma 5.C.3.** *We let  $A_{j,\sigma}$  be the event of all  $x \in \mathcal{O}_{x_j}^2$  such that*

$$\forall 0 < \langle k, \alpha \rangle < \beta, \frac{|D^k \bar{f}_{j,\delta}(z_x)|}{\bar{f}_{j,\delta}(z_x)} \leq \sigma^{-\langle k, \alpha \rangle} \log(1/\sigma)^{-\langle k, \alpha \rangle/2} \quad \text{and} \quad \frac{L_{j,\delta}(z_x)}{\bar{f}_{j,\delta}(z_x)} \leq \sigma^{-\beta} \log(1/\sigma)^{-\beta/2}.$$

*Then, the following assertions hold true for  $\sigma$  small enough*

- i)  $g_j(x) \geq f_j(x)/2$  for all  $x \in \mathcal{A}_{j,\sigma}$ ;
- ii)  $P_{f_j}(\mathcal{A}_{j,\sigma}^c) \lesssim c_j^{-1} \{\sigma \log(1/\sigma)\}^{2\beta+\varepsilon}$ ;
- iii) For any  $\langle k, \alpha \rangle < \beta$ ,  $\int_{\mathcal{A}_{j,\sigma}^c} |D^k \bar{f}_{j,\delta}(z_x)| dx \lesssim c_j^{-1} \delta^{D-d} \{\sigma \log(1/\sigma)\}^{2\beta+\varepsilon-\langle k, \alpha \rangle}$ .

*Proof.* We start with proving i). For any  $x \in \mathcal{A}_{j,\sigma}$ , there holds, using the fact that the functions  $d_{j,k}$  are uniformly bounded,

$$\begin{aligned} |g_j(x) - f_j(x)| &= \left| \frac{1}{\delta^{D-d}} \sum_{0 < \langle k, \alpha \rangle < \beta} \sigma^{\langle k, \alpha \rangle} d_k(y, \sigma, \delta) D^k \bar{f}_{j,\delta}(z_y) \right| \\ &\lesssim \frac{\bar{f}_{j,\delta}(z_x)}{\delta^{D-d}} \sum_{0 < \langle k, \alpha \rangle < \beta} \sigma^{\langle k, \alpha \rangle} \left| \frac{D^k \bar{f}_{j,\delta}(z_x)}{\bar{f}_{j,\delta}(z_x)} \right| \\ &\lesssim f_j(x) \sum_{0 < \langle k, \alpha \rangle < \beta} \log(1/\sigma)^{-\langle k, \alpha \rangle/2} \lesssim \log^{-\alpha_1/2}(1/\sigma) f_j(x) < \frac{1}{2} f_j(x) \end{aligned}$$

provided that  $\sigma$  is chosen small enough. For ii), notice that

$$\begin{aligned} P_{f_j}(\mathcal{A}_{j,\sigma}^c) &\leq \sum_{0 < \langle k, \alpha \rangle < \beta} P_{f_j} \left( \left| \frac{D^k \bar{f}_{j,\delta}(z_x)}{\bar{f}_{j,\delta}(z_x)} \right| > \sigma^{-\langle k, \alpha \rangle} \log(1/\sigma)^{-\langle k, \alpha \rangle/2} \right) \\ &\leq \sum_{0 < \langle k, \alpha \rangle < \beta} \sigma^{2\beta+\varepsilon} \log(1/\sigma)^{2\beta+\varepsilon} \int_{\mathcal{O}_{x_j}^0} \left( \frac{|D^k \bar{f}_{j,\delta}(z_x)|}{\bar{f}_{j,\delta}(z_x)} \right)^{\frac{2\beta+\varepsilon}{\langle k, \alpha \rangle}} f_j(x) dx \\ &\lesssim c_j^{-1} \sigma^{2\beta+\varepsilon} \log(1/\sigma)^{2\beta+\varepsilon} \end{aligned}$$

where we used Lemma 5.C.2 after a variable change  $z = z_x$ . Finally, for iii), there holds

$$\begin{aligned} \int_{\mathcal{A}_{j,\sigma}^c} |D^k \bar{f}_{j,\delta}(z_x)| dx &= \int \frac{|D^k \bar{f}_{j,\delta}(z_x)|}{\bar{f}_{j,\delta}(z_x)} \bar{f}_{j,\delta}(z_x) \mathbb{1}_{\mathcal{A}_{j,\sigma}^c}(x) dx \\ &\leq \left\{ \int \left\{ \frac{|D^k \bar{f}_{j,\delta}(z_x)|}{\bar{f}_{j,\delta}(z_x)} \right\}^{\frac{2\beta+\varepsilon}{\langle k, \alpha \rangle}} \bar{f}_{j,\delta}(z_x) \mathbb{1}_{\mathcal{A}_{j,\sigma}^c}(x) dx \right\}^{\frac{\langle k, \alpha \rangle}{2\beta+\varepsilon}} \times \left\{ \delta^{D-d} P_{f_j}(\mathcal{A}_{j,\sigma}^c) \right\}^{1-\frac{\langle k, \alpha \rangle}{2\beta+\varepsilon}} \\ &\lesssim \left\{ c_j^{-1} \delta^{D-d} \right\}^{\frac{\langle k, \alpha \rangle}{2\beta+\varepsilon}} \times \left\{ c_j^{-1} \delta^{D-d} \{\sigma \log(1/\sigma)\}^{2\beta+\varepsilon} \right\}^{1-\frac{\langle k, \alpha \rangle}{2\beta+\varepsilon}} \\ &\lesssim c_j^{-1} \delta^{D-d} \{\sigma \log(1/\sigma)\}^{2\beta+\varepsilon-\langle k, \alpha \rangle}. \end{aligned}$$

ending the proof. □

### 5.C.2 Proof of Corollary 5.3.8

Let define

$$\tilde{h}_j := g_j \mathbb{1}_{g_j > f_j/2} + \frac{1}{2} f_j \mathbb{1}_{g_j \leq f_j/2}$$

and  $h_j = \tilde{h}_j / I_j$  where  $I_j = \int \tilde{h}_j$ . The function  $h_j$  is a probability measure on  $\mathcal{O}_{x_j}^0$  and furthermore, using the convexity of  $f \rightarrow d_H^2(f, f_0)$ , there holds

$$d_H^2(K_\Sigma(h), f_0) \leq \sum_{j=1}^J c_j d_H^2(K_\Sigma(h_j), f_j) \quad \text{where } h = \sum_{j=1}^J c_j h_j.$$

We will control each term separately. First notice that

$$\sum_{j=1}^J c_j d_H^2(K_\Sigma(h_j), f_j) \leq J\sigma^{2\beta} + \sum_{c_j \geq \sigma^{2\beta}} c_j d_H^2(K_\Sigma(h_j), f_j).$$

Now take  $1 \leq j \leq J$  such that  $c_j \geq \sigma^{2\beta}$  and define  $\mathcal{U}_{j,\sigma} = \{x \in \mathbb{R}^D \mid f_j(x) \geq \sigma^{H_1}\}$  for some  $H_1 > 0$  to be specified later. Then there holds

$$d_H^2(K_\Sigma h_j, f_j) \leq \int_{\mathcal{U}_{j,\sigma} \cap \mathcal{O}_{x_j}^1} \left( \sqrt{K_\Sigma(h_j)} - \sqrt{f_j} \right)^2 + \int_{(\mathcal{O}_{x_j}^1)^c} K_\Sigma(h_j) + \int_{\mathcal{U}_{j,\sigma}^c} [K_\Sigma(h_j) + f_j]$$

and

$$\begin{aligned} \int_{\mathcal{U}_{j,\sigma} \cap \mathcal{O}_{x_j}^1} \left( \sqrt{K_\Sigma(h_j)} - \sqrt{f_j} \right)^2 &\leq \int_{\mathcal{U}_{j,\sigma} \cap \mathcal{O}_{x_j}^1} \frac{(K_\Sigma(h_j) - f_j)^2}{K_\Sigma(h_j) + f_j} \\ &\leq \int_{\mathcal{U}_{j,\sigma} \cap \mathcal{O}_{x_j}^1} \frac{K_\Sigma(h_j - \tilde{h}_j)^2}{K_\Sigma(h_j) + f_j} + \int_{\mathcal{U}_{j,\sigma} \cap \mathcal{O}_{x_j}^1} \frac{K_\Sigma(\tilde{h}_j - g_j)^2}{K_\Sigma(h_j) + f_j} + \int_{\mathcal{U}_{j,\sigma} \cap \mathcal{O}_{x_j}^1} \frac{(K_\Sigma(g_j) - f_j)^2}{K_\Sigma(h_j) + f_j} \end{aligned}$$

and therefore

$$d_H^2(f_j, K_\Sigma(h_j)) \leq \int_{(\mathcal{O}_{x_j}^1)^c} K_\Sigma(h_j) \tag{5.37}$$

$$+ \int_{\mathcal{U}_{j,\sigma}^c} [K_\Sigma(h_j) + f_j] \tag{5.38}$$

$$+ (1 - I_j)^2 \tag{5.39}$$

$$+ \int_{\mathcal{U}_{j,\sigma} \cap \mathcal{O}_{x_j}^1} \frac{K_\Sigma((f_j/2 - g_j) \mathbb{1}_{g_j < f_j/2})^2}{K_\Sigma(h_j) + f_j} \tag{5.40}$$

$$+ \int_{\mathcal{U}_{j,\sigma} \cap \mathcal{O}_{x_j}^1} \left( \frac{K_\Sigma(g_j) - f_j}{f_j} \right)^2 f_j \tag{5.41}$$

where each term will be bounded independently.

1. For (5.37), notice that for any  $y \in \mathcal{O}_{x_j}^0$  and any  $x \notin \mathcal{O}_{x_j}^1$ , there holds that  $\|x - y\|_{\Sigma^{-1}}^2 \geq$

$C/(\delta^2 \sigma^{2\alpha_1})$  for some  $C$  depending on  $\tau$ . We can thus write

$$\begin{aligned}
 \int_{(\mathcal{O}_{x_j}^1)^c} K_\Sigma(h_j) &= \int_{(\mathcal{O}_{x_j}^1)^c} \int h_j(y) \varphi_{\Sigma(y)}(x-y) dy dx \\
 &= \int_{(\mathcal{O}_{x_j}^1)^c} \int_{\mathcal{O}_{x_j}^0} h_j(y) \frac{\exp\left(-\frac{1}{2}\|x-y\|_{\Sigma^{-1}(y)}^2\right)}{(2\pi)^{D/2} \sigma^D \delta^{D-d}} dy dx \\
 &\leq \exp\left\{-\frac{C^2}{4\delta^2 \sigma^{2\alpha_1}}\right\} \int_{(\mathcal{O}_{x_j}^1)^c} \int_{\mathcal{O}_{x_j}^0} h_j(y) \frac{\exp\left(-\frac{1}{4}\|x-y\|_{\Sigma^{-1}(y)}^2\right)}{(2\pi)^{D/2} \sigma^D \delta^{D-d}} dy dx \\
 &\lesssim \exp\left\{-\frac{C^2}{4\delta^2 \sigma^{2\alpha_1}}\right\} \int_{(\mathcal{O}_{x_j}^1)^c} K_{2\Sigma}(h_j) \lesssim \exp\left\{-\frac{C^2}{4\delta^2 \sigma^{2\alpha_1}}\right\} \lesssim \sigma^{2\beta},
 \end{aligned}$$

because  $\delta \sigma^\perp = o(|\log(\sigma)|^{1/2})$  by assumption.

2. For (5.41), there holds

$$\begin{aligned}
 \int_{\mathcal{U}_{j,\sigma} \cap \mathcal{O}_{x_j}^1} \left( \frac{K_\Sigma(g_j) - f_j}{f_j} \right)^2 f_j &\leq \int_{\mathcal{O}_{x_j}^1} \left( \frac{L_j \sigma^\beta}{f_j} \right)^2 f_j \leq \sigma^{2\beta} \left\{ \int_{\mathcal{O}_{x_j}^1} \left( \frac{L_j}{f_j} \right)^{\frac{2\beta+\varepsilon}{\beta}} f_j \right\}^{\frac{2\beta}{2\beta+\varepsilon}} \\
 &\leq \sigma^{2\beta} \left\{ \int_{\mathcal{W}_{x_j}^1} \left( \frac{L_{j,\delta}}{f_{j,\delta}} \right)^{\frac{2\beta+\varepsilon}{\beta}} \bar{f}_{j,\delta} \frac{|\det d\bar{\Psi}_{x_j,\delta}|}{\delta^{D-d}} \right\}^{\frac{2\beta}{2\beta+\varepsilon}} \\
 &\lesssim \sigma^{2\beta} \left\{ \int_{\mathcal{W}_{x_j}^1} \left( \frac{L_{j,\delta}}{f_{j,\delta}} \right)^{\frac{2\beta+\varepsilon}{\beta}} \bar{f}_{j,\delta} \right\}^{\frac{2\beta}{2\beta+\varepsilon}} \lesssim c_j^{-\frac{2\beta}{2\beta+\varepsilon}} \sigma^{2\beta} \lesssim c_j^{-1} \sigma^{2\beta},
 \end{aligned}$$

where we used that  $|\det d\bar{\Psi}_{x_j,\delta}| \lesssim \delta^{D-d}$  and Lemma 5.C.2.

3. For (5.39), notice that

$$I_j = \int g_j \mathbb{1}_{\{f_j < 2g_j\}} + \int \frac{f_j}{2} \mathbb{1}_{\{f_j \geq 2g_j\}} = \int g_j + \int \frac{(f_j - 2g_j)}{2} \mathbb{1}_{\{f_j \geq 2g_j\}}$$

and that  $\int g_j = \int K_\Sigma(g_j)$ . Moreover

$$\int_{\mathcal{O}_{x_j}^1} K_\Sigma(g_j) = \int_{\mathcal{O}_{x_j}^1} f_j + \int_{\mathcal{O}_{x_j}^1} (K_\Sigma(g_j) - f_j) = 1 + \int_{\mathcal{O}_{x_j}^1} (K_\Sigma(g_j) - f_j)$$

and since

$$\begin{aligned}
 \left| \int_{\mathcal{O}_{x_j}^1} K_\Sigma(g_j) - f_j \right| &\leq \int_{\mathcal{O}_{x_j}^1} L_j \sigma^\beta = \sigma^\beta \int_{\mathcal{W}_{x_j}^1} L_{j,\delta} \times \frac{|\det d\bar{\Psi}_{x_j,\delta}|}{\delta^{D-d}} \\
 &\lesssim \sigma^\beta \left\{ \int_{\mathcal{W}_{x_j}^1} \left( \frac{L_{j,\delta}}{f_{j,\delta}} \right)^{\frac{2\beta+\varepsilon}{\beta}} \bar{f}_{j,\delta} \right\}^{\frac{\beta}{2\beta+\varepsilon}} \lesssim c_j^{-\frac{\beta}{2\beta+\varepsilon}} \sigma^\beta \lesssim c_j^{-1/2} \sigma^\beta.
 \end{aligned}$$

where we again used that  $|\det d\bar{\Psi}_{x_j,\delta}| \lesssim \delta^{D-d}$  and Lemma 5.C.2, we have that

$$\int_{\mathcal{O}_{x_j}^1} K_\Sigma(g_j) = 1 + O(c_j^{-1/2} \sigma^\beta).$$

We also have, as in (5.37) that

$$\begin{aligned} \int_{(\mathcal{O}_{x_j}^1)^c} K_\Sigma(|g_j|) &\lesssim e^{-\frac{C^2}{4\delta^2\sigma^{2\alpha_1}}} \int_{\mathcal{O}_{x_j}^1} |g_j|(y) dy \\ &\lesssim e^{-\frac{C^2}{4\delta^2\sigma^{2\alpha_1}}} \|L_j\|_\infty = o(c_j^{-1}\sigma^{2\beta}) = o(c_j^{-1/2}\sigma^\beta), \end{aligned}$$

where the two last inequality comes from  $\delta^2\sigma^{2\alpha_1} = o(1/\log(1/\sigma))$  and  $c_j \geq \sigma^{2\beta}$ . Therefore  $|1 - \int g_j| \lesssim c_j^{-1/2}\sigma^\beta$ . Moreover, using this time Lemma 5.C.3

$$\begin{aligned} \int_{f_j > 2g_j} (f_j - 2g_j) &\lesssim P_{f_j}(f_j > 2g_j) + \sum_{0 < \langle k, \alpha \rangle < \beta} \frac{\sigma^{\langle k, \alpha \rangle}}{\delta^{D-d}} \int_{f_j > 2g_j} |D^k \bar{f}_{j,\delta}(z_x)| dx \\ &\lesssim c_j^{-1} \sigma^{2\beta+\varepsilon}. \end{aligned}$$

All in all, we find that

$$\begin{aligned} (1 - I_j)^2 &\lesssim \exp\{-C^2/(4\delta^2\sigma^{2\alpha_1})\} + \{c_j^{-1}\sigma^{2\beta}\} \vee \{c_j^{-2}\{\sigma \log(1/\sigma)\}^{4\beta+2\varepsilon}\} \\ &\lesssim \exp\{-C^2/(4\delta^2\sigma^{2\alpha_1})\} + c_j^{-1}\sigma^{2\beta} \end{aligned}$$

where the last inequality again holds because  $c_j \geq \sigma^{2\beta}$ .

4. For (5.38), we start with taking  $\langle k, \alpha \rangle < \beta$ . Notice that, thanks to Proposition 5.B.3, there exists a constant  $C > 0$  such that for any  $x, y \in \mathcal{O}_{x_j}^0$ ,  $|D^k \bar{f}_{j,\delta}(z_x) - D^k \bar{f}_{j,\delta}(z_y)| \leq C\delta^{D-d}L_j(x)$ . Then, using (5.17), together with  $|\det d\bar{\Psi}(w)| \leq C$ ,

$$\begin{aligned} |K_\Sigma(D^k \bar{f}_{j,\delta}(z_\square))(x)| &\lesssim \int_{\Delta_{\sigma,1}^{-1}(\mathcal{W}^0 - z_x)} e^{-B_\sigma(x,z)} |D^k \bar{f}_{j,\delta}(\Delta_{\sigma,1}z + z_x)| dz \\ &\lesssim \delta^{D-d}L_j(x) \int_{\Delta_{\sigma,1}^{-1}(\mathcal{W}^0 - z_x)} e^{-B_\sigma(x,z)} dz \lesssim \delta^{D-d}L_j(x). \end{aligned}$$

Now notice that

$$\begin{aligned} &\int_{\mathcal{U}_{j,\sigma}^c} |D^k \bar{f}_{j,\delta}(z_x)| dx \\ &\leq \left\{ \int_{\mathcal{U}_{j,\sigma}^c} \left\{ \frac{|D^k \bar{f}_{j,\delta}(z_x)|}{\bar{f}_{j,\delta}(z_x)} \right\}^{\frac{2\beta+\varepsilon}{\langle k, \alpha \rangle}} \bar{f}_{j,\delta}(z_x) dx \right\}^{\frac{\langle k, \alpha \rangle}{2\beta+\varepsilon}} \times \left\{ \int_{\mathcal{U}_{j,\sigma}^c} \bar{f}_{j,\delta}(z_x) dx \right\}^{1-\frac{\langle k, \alpha \rangle}{2\beta+\varepsilon}} \\ &\lesssim c_j^{-1} \delta^{D-d} \sigma^{H_1 \times \{1-\langle k, \alpha \rangle/(2\beta+\varepsilon)\}} \lesssim c_j^{-1} \delta^{D-d} \sigma^{H_1/2} \end{aligned}$$

and likewise,  $\int_{\mathcal{U}_{j,\sigma}^c} L_j(x) dx \lesssim c_j^{-1} \delta^{D-d} \sigma^{H_1 \times \{1-\beta/(2\beta+\varepsilon)\}} \lesssim \delta^{D-d} \sigma^{H_1/2}$ . We thus have shown that

$$\int_{\mathcal{U}_{j,\sigma}^c} |K_\Sigma(D^k \bar{f}_{j,\delta}(z_\square))(x)| \lesssim c_j^{-1} \delta^{D-d} \sigma^{H_1/2}. \quad (5.42)$$

Coming back to (5.38), there immediately holds that  $\int_{\mathcal{U}_{j,\sigma}^c} f_j \lesssim \sigma^{H_1}$ , and furthermore, noticing that

$$h_j \lesssim \tilde{h}_j \lesssim f_j + \frac{1}{\delta^{D-d}} \sum_{0 \leq \langle k, \alpha \rangle < \beta} \sigma^{\langle k, \alpha \rangle} |D^k \bar{f}_{j,\delta}(z_\square)|$$

and using (5.42) and the monotonicity of  $K_\Sigma$ , we find that  $\int_{\mathcal{U}_{j,\sigma}^c} K_\Sigma(h_j) \lesssim c_j^{-1} \sigma^{H_1/2}$ .

5. Finally, for (5.40), notice that  $f_j/2 - g_j$  is a sum of terms of the form  $\sigma^{\langle k, \alpha \rangle} \delta^{-(D-d)} D^k \bar{f}_{j, \delta}(z_\square)$  for  $0 \leq \langle k, \alpha \rangle < \beta$ . The term  $K_\Sigma((f_j/2 - g_j) \mathbb{1}_{g_j \leq f_j/2})^2$  is thus upper bounded by a sum of terms of the form  $\sigma^{2\langle k, \alpha \rangle} \delta^{-2(D-d)} K_\Sigma(D^k \bar{f}_{j, \delta}(z_\square) \mathbb{1}_{g_j \leq f_j/2})^2$ . Now there holds

$$\begin{aligned} & \int_{\mathcal{U}_{j, \sigma} \cap \mathcal{O}_{x_j}^1} \frac{K_\Sigma(D^k \bar{f}_{j, \delta}(z_\square) \mathbb{1}_{g_j \leq f_j/2})^2}{K_\Sigma(h_j) + f_j} \\ & \lesssim \sigma^{-H_1} \|K_\Sigma(D^k \bar{f}_{j, \delta}(z_\square) \mathbb{1}_{g_j \leq f_j/2})\|_\infty \times \int K_\Sigma(|D^k \bar{f}_{j, \delta}(z_\square)| \mathbb{1}_{g_j \leq f_j/2}) \\ & \lesssim \frac{\sigma^{-H_1}}{c_j} \int_{\mathcal{O}_{x_j}^1} |D^k \bar{f}_{j, \delta}(z_y)| \mathbb{1}_{g_j \leq f_j/2} dy \lesssim \frac{\sigma^{-H_1}}{c_j^2} \delta^{D-d} \{\sigma \log(1/\sigma)\}^{2\beta + \varepsilon - \langle k, \alpha \rangle} \end{aligned}$$

where we used Lemma 5.C.3. Summing all the bounds in  $k$ , we obtain that (5.40) is bounded from above, up to constant, by

$$\max_k c_j^{-2} \sigma^{2\beta + \varepsilon - H_1 + 2\langle k, \alpha \rangle} \delta^{-(D-d)} \log(1/\sigma)^{2\beta + \varepsilon} \lesssim c_j^{-1} \delta^{-(D-d)} \sigma^{\varepsilon - H_1} \log(1/\sigma)^{2\beta + \varepsilon}$$

where we used that  $c_j \geq \sigma^{2\beta}$ .

Collecting all the bounds on (5.37-5.41) together with  $\delta^2 \sigma^{2\alpha_1} = o(1/\log(1/\sigma))$ , we get that

$$c_j d_H^2(K_\Sigma(h_j), f_j) \lesssim c_j \left[ \sigma^{2\beta} \log(1/\sigma)^{4\beta + 2\varepsilon} + \sigma^{H_1/2} + \sigma^{\varepsilon - H_1} \delta^{-(D-d)} \log(1/\sigma)^{2\beta + \varepsilon} \right].$$

Choosing  $H_1 = 4\beta$  and  $\varepsilon \geq 6\beta + \varepsilon_1$  where  $\sigma^{\varepsilon_1} \leq \delta^{D-d}$ , we obtain the result.

## 5.D Appendix to Section 5.5.1: proof of Theorem 5.3.2

For any probability distribution  $P$  on  $\mathbb{R}^D \times \mathcal{S}^{++}(D, \mathbb{R})$ , one defines the probability density function on  $\mathbb{R}^D$

$$f_P(x) := \int \varphi_\Sigma(x - y) dP(y, \Sigma).$$

Note that when  $g$  is a probability distribution on  $\mathbb{R}^D$ , then

$$K_\Sigma g(x) = f_P(x) \quad \text{with} \quad dP(y, \Sigma) = \delta_{\Sigma(y)}(\Sigma) g(y) dy.$$

**Lemma 5.D.1.** *Let  $V_0, \dots, V_N$  be a partition of  $\mathbb{R}^D$  and let  $P = \sum_{j=1}^N \pi_j \delta_{z_j, \Sigma_j}$  with  $z_j \in V_j$ . Then, for any probability measure  $Q$  on  $\mathbb{R}^D \times \mathcal{S}^{++}(D, \mathbb{R})$ ,*

$$\|f_Q - f_P\|_1 \leq 2 \sum_{j=1}^N |Q_1(V_j) - \pi_j| + \frac{1}{2} \sup_{1 \leq j \leq N} \|\Sigma_j^{-1/2}\|_{\text{op}} \text{diam } V_j + \frac{3}{2} \sup_{1 \leq j \leq N} \sup_{\Sigma \in \mathcal{S}_j} \sqrt{\text{tr}(\Sigma_j^{-1} \Sigma - \text{Id})^2},$$

where  $Q_1$  is the first marginal of  $Q$  and  $\mathcal{S}_j$  is the support of the second marginal of  $\mathbb{1}_{V_j}(y) dQ(y, \Sigma)$ .

*Proof.* We write that  $f_Q(x) - f_P(x)$  is

$$\begin{aligned} \int_{V_0} \varphi_\Sigma(x-y) dQ(y, \Sigma) + \sum_{j=1}^N \int_{V_j} \{ \varphi_\Sigma(x-y) - \varphi_{\Sigma_j}(x-z_j) \} dQ(y, \Sigma) \\ + \sum_{j=1}^N (Q_1(V_j) - \pi_j) \varphi_{\Sigma_j}(x-z_j), \end{aligned}$$

The last term can be readily bounded in  $L^1$ -norm by

$$\sum_{j=1}^N |Q_1(V_j) - \pi_j| \int \varphi_{\Sigma_j}(x-z_j) dx = \sum_{j=1}^N |Q_1(V_j) - \pi_j|,$$

and the first one by  $Q_1(V_0) = 1 - \sum_{j=1}^N Q_j \leq \sum_{j=1}^N |Q_1(V_j) - \pi_j|$ . For the second term, notice that each term of the sum is upper-bounded in  $L^1$ -norm by

$$\begin{aligned} \int_{V_j} \|\varphi_\Sigma(\cdot-y) - \varphi_{\Sigma_j}(\cdot-z_j)\|_1 dQ(y, \Sigma) \\ \leq \int_{V_j} \|\varphi_{\Sigma_j}(\cdot-y) - \varphi_{\Sigma_j}(\cdot-z_j)\|_1 dQ(y, \Sigma) + \int_{V_j} \|\varphi_\Sigma(\cdot-y) - \varphi_{\Sigma_j}(\cdot-y)\|_1 dQ(y, \Sigma) \\ \leq \frac{1}{2} \int_{V_j} \|\Sigma_j^{-1/2}(y-z_j)\| dQ_1(y) + \frac{3}{2} \int_{V_j} \sqrt{\text{tr}(\Sigma_j^{-1}\Sigma - \text{Id})^2} dQ(y, \Sigma) \\ \leq \frac{1}{2} Q_1(V_j) \|\Sigma_j^{-1/2}\|_{\text{op}} \text{diam } V_j + \frac{3}{2} Q_1(V_j) \sup_{\Sigma \in \mathcal{S}_j} \sqrt{\text{tr}(\Sigma_j^{-1}\Sigma - \text{Id})^2}, \end{aligned}$$

where we used Prp 2.1 and Thm 1.1 of [DMR18]. □

### 5.D.1 Proof of Lemma 5.5.1

Throughout the proof,  $C$  denotes a generic constant whose value depends only on  $D$ . We start with bounding the probability measure of  $\mathcal{F}_n^c$ . There holds

$$\Pi(\mathcal{F}_n^c) \leq \Pi\{\exists h \leq H_n, \mu_h \notin B(0, R_n)\} + \Pi\left\{\sum_{h>H_n} \pi_h \geq \varepsilon_n\right\} + \Pi(\exists h \leq H_n, \Lambda_h \notin [\underline{\sigma}_n^2, \bar{\sigma}_n^2]^D).$$

The first mass is bounded using (5.8):

$$\Pi\{\exists h \leq H_n, \mu_h \notin B(0, R_n)\} \lesssim H_n \int \|\mu\|^{-b_2} \mathbb{1}_{\|\mu\| \geq R_n} d\mu \lesssim H_n R_n^{-b_2} \lesssim e^{-c_1 n \varepsilon_n^2},$$

as soon as  $b_2 R_0 \geq 2c_1$ . The second term is bounded in [STG13, p. 15] by

$$\Pi\left\{\sum_{h>H_n} \pi_h \geq \varepsilon_n\right\} \leq \left\{\frac{eB}{H_n} \log(1/\varepsilon)\right\}^{H_n} \lesssim e^{-c_1 n \varepsilon_n^2},$$

as soon as  $H_0$  is large enough. Finally, to bound the last probability, we consider separately the partial and hybrid location-scale priors.



- Partial location-scale:  $\Lambda_h = \Lambda_1$  for all  $h$  and using (5.9),

$$\Pi(\Lambda_1 \notin \mathcal{Q}_n) \leq \Pi_\Lambda \left\{ \min_i \Lambda_i \leq \underline{\sigma}_n^2 \right\} + \Pi_\Lambda \left\{ \max_i \Lambda_i \geq \bar{\sigma}_n^2 \right\} \lesssim e^{-c_3 \sigma_0^{-2b_3} n \varepsilon_n^2} + \bar{\sigma}_n^{-2b_4} \leq e^{-c_1 n \varepsilon_n^2},$$

as soon as  $\sigma_1^2$  is large enough and  $\sigma_0^2$  is small enough.

- Hybrid location-scale prior

$$\begin{aligned} \Pi(\exists h \leq H_n, \Lambda_h \notin \mathcal{Q}_n) &\leq H_n \mathbb{E}_{\tilde{\Pi}_\Lambda} \left[ Q_2(\min_i \Lambda_i \leq \underline{\sigma}_n^2) \right] + H_n \mathbb{E}_{\tilde{\Pi}_\Lambda} \left[ Q_2(\max_i \Lambda_i > \bar{\sigma}_n^2) \right] \\ &\leq H_n e^{-c_3 \sigma_0^{-2b_3} n \varepsilon_n^2} + H_n e^{-c_4 \sigma_1^{2b_4} n \varepsilon_n^2} \lesssim e^{-c_1 n \varepsilon_n^2}. \end{aligned}$$

We then turn on bounding the entropy of the partitions of  $\mathcal{F}_n$ . Note that on  $\mathcal{F}_n$   $\min_i \lambda_i \geq \underline{\sigma}_n^2$  so that

$$\frac{\max_i \lambda_i}{\min_i \lambda_i} \leq \max_i \lambda_i \underline{\sigma}_n^{-2} \leq \sigma_0^{-2} n^{2\omega/b_3} \times \max_i \lambda_i.$$

From [CDB17], the covering number of  $\mathcal{F}_{n,j,\ell}$  is bounded by

$$\begin{aligned} N(\varepsilon_n, \mathcal{F}_{n,j,\ell}, \|\cdot\|_1) \\ \leq \exp \left( CH_n [\log n + \log(1/\varepsilon_n)] + (D-1) \sum_{h \leq H_n} \log(j_h + 1) + D(D-1) 2^{\ell-2} H_n \log n \right), \end{aligned}$$

and the covering number of  $\mathcal{F}_{n,j}$  is bounded by

$$\begin{aligned} N(\varepsilon_n, \mathcal{F}_{n,j}, \|\cdot\|_1) &\leq \exp \left( CH_n [\log n + \log(1/\varepsilon_n)] + \sum_{h \leq H_n} \log(j_h + 1) + H_n \log \bar{\sigma}_n^2 \right) \\ &\lesssim \exp \left( CH_n [\log n + \log(1/\varepsilon_n)] + \sum_{h \leq H_n} \log(j_h + 1) \right). \end{aligned}$$

We bound  $\Pi(\mathcal{F}_{n,j,\ell})$  in the case of the partial location-scale prior, for  $\ell \geq 1$ :

$$\begin{aligned} \Pi(\mathcal{F}_{n,j,\ell}) &\lesssim \prod_{h \leq H_n} (j_h \sqrt{n})^{-(b_2-D) \mathbb{1}_{j_h \geq 1}} \Pi_\Lambda \left( \max_i \lambda_i > \sigma_0^2 n^{2^{\ell-1}-2\omega/b_3} \right) \\ &\lesssim \prod_{h \leq H_n} (j_h \sqrt{n})^{-(b_2-D) \mathbb{1}_{j_h \geq 1}} n^{-b_4(2^{\ell-1}-2\omega/b_3)} \\ &= \exp \left\{ -(b_2-D) \sum_{h \leq H_n} \mathbb{1}_{j_h \geq 1} \log(j_h) - \frac{1}{2}(b_2-D)nH_n - b_4(2^{\ell-1}-2\omega/b_3)H_n \log n \right\}. \end{aligned}$$

We bound  $\Pi(\mathcal{F}_{n,j})$  in the case of the hybrid location-scale prior:

$$\begin{aligned} \Pi(\mathcal{F}_{n,j}) &\lesssim \prod_{h \leq H_n} (j_h \sqrt{n})^{-(b_2-D) \mathbb{1}_{j_h \geq 1}} \\ &= \exp \left\{ -(b_2-D) \sum_{h \leq H_n} \mathbb{1}_{j_h \geq 1} \log(j_h) - \frac{1}{2}(b_2-D)nH_n \right\}. \end{aligned}$$

This implies in particular that for the partial location-scale prior

$$\begin{aligned} & \sum_{\mathbf{j}, \ell} \sqrt{\Pi(\mathcal{F}_{n, \mathbf{j}, \ell}) N(\varepsilon_n, \mathcal{F}_{n, \mathbf{j}, \ell}, \|\cdot\|_1)} \lesssim \exp\left(\frac{1}{2} C H_n [\log n + \log(1/\varepsilon_n)]\right) \\ & \times \sum_{\mathbf{j}, \ell} \exp\left(\frac{1}{2} \sum_h [(D-1) \log(j_h + 1) - (b_2 - D)(\log(j_h) \mathbb{1}_{j_h \geq 1} - 1)]\right. \\ & \quad \left. + \mathbb{1}_{\ell \geq 2} 2^{\ell-1} \log n [D(D-1)/2 - b_4]\right) \\ & \lesssim \exp(C H_n [\log n + \log(1/\varepsilon_n)]/2), \end{aligned}$$

since  $b_4 > D(D-1)/2$  and  $b_2 > 2D-1$ . Therefore, by choosing  $M_0 > 0$  large enough

$$\sum_{\mathbf{j}, \ell} \sqrt{\Pi(\mathcal{F}_{n, \mathbf{j}, \ell}) N(\varepsilon_n, \mathcal{F}_{n, \mathbf{j}, \ell}, \|\cdot\|_1)} e^{-M_0 n \varepsilon_n^2} = o(1).$$

We obtain a similar result for the hybrid location-scale prior.

### 5.D.2 Proof of Lemma 5.5.2

We let again  $R = (H \log(1/\varepsilon)/C_2)^{1/\kappa}$  and define  $\sigma := \varepsilon^{1/\beta}$ . Thanks to Corollary 5.3.8, we know there exists a density function  $g$  supported on  $M^\delta$  such that  $d_H^2(K_\Sigma g, f_0) \lesssim \sigma^{2\beta} \log^q(1/\sigma)$  for some  $a > 0$ . We can in turn, thanks to Lemma 5.5.3, find a discrete probability measure  $G$  on  $M^\delta \cap B(0, R)$  with  $N$  atoms at least  $\sigma^{\alpha_1} \varepsilon^2$ -apart such that

$$\|K_\Sigma g - K_\Sigma G\|_1 \lesssim \varepsilon^2 \log^{D/2}(1/\varepsilon) \quad \text{and} \quad N \lesssim \sigma^{-D} \log^D(1/\varepsilon).$$

We thus have  $d_H^2(K_\Sigma G, f_0) \lesssim \sigma^{2\beta} \log^q(1/\sigma) + \varepsilon^2 \log^{D/2}(1/\varepsilon) \lesssim \varepsilon^2 \log^{2r}(1/\varepsilon)$  with  $r := q/2 \vee D/4$ . We let  $z_1, \dots, z_N$  be the atoms of  $G$  and denote by  $p_j = G(z_j)$ . We let  $V_j$  be the ball centered around  $z_j$  with radius  $\sigma^{2\alpha_0} \varepsilon^2/2$ . We complete  $V_1, \dots, V_N$  with sets  $V_{N+1}, \dots, V_J$  that forms a partition of  $M^\tau \cap B(0, R)$  with  $V_j$  included in balls of the form  $\{x \in \mathbb{R}^D \mid \|x - z_j\|_{\Sigma^{-1}(z_j)} \leq 1\}$  for some  $z_j \in M^\tau \cap B(0, R)$ , so that we can take  $J \lesssim N + (R/\sigma)^D \lesssim \sigma^{-D} \log^{D/\kappa}(1/\varepsilon)$ . We then set  $V_0$  to be the complementary set of the reunion of the  $V_j$  and set further  $p_j = 0$  for  $j$  greater than  $N+1$ .

We write under the prior  $\Pi$ ,  $P = \sum_{h=1}^\infty \pi_h \delta_{\mu_h, U_h, \Lambda_h}$  and  $\Sigma_h = U_h^\top \Lambda_h U_h$ . We use the convention that  $\Lambda_h = \Lambda$  for all  $h$  in the case of the Partial location scale prior and  $\pi_h = 0$  for  $h \geq K$  for the mixture of finite mixtures prior. Set  $\tilde{N} = \log(1/\varepsilon) \times N$ , we consider the following events

$$\begin{aligned} \mathcal{P}_J &= \left\{ \sum_{j=1}^J |p_j - P_\mu(V_j)| \leq \varepsilon^2 \quad \text{and} \quad \min_{1 \leq j \leq J} P_\mu(V_j) \geq \varepsilon^4 \right\}, \\ \mathcal{F}_{\tilde{N}} &= \left\{ \sum_{h \leq \tilde{N}} \pi_h \geq 1 - \varepsilon^8 \right\}, \\ \mathcal{O}_{\tilde{N}} &= \left\{ \forall 1 \leq h \leq \tilde{N}, \|U_h O_{\mu_h}^\top - \text{Id}\|_{\text{op}} \leq \sigma^{2\alpha_0} \varepsilon^2 \right\}, \\ \text{and } \mathcal{L}_{\tilde{N}} &= \left\{ \forall 1 \leq h \leq \tilde{N}, \Lambda_h \in \mathcal{S}(\sigma^{\alpha_0})^d \times \mathcal{S}(\delta \sigma^{\alpha_1})^{D-d} \right\} \\ \text{where } \mathcal{S}(t) &= \{s \mid t^2 \leq s \leq t^2(1 + \sigma^{2\beta})\}. \end{aligned}$$

We first show that if  $P \in \mathcal{P}_J \cap \mathcal{F}_{\tilde{N}} \cap \mathcal{O}_{\tilde{N}} \cap \mathcal{L}_{\tilde{N}}$ , then

$$d_H(f_P, f_0) \lesssim \varepsilon \log^r(1/\varepsilon).$$

Indeed, we have

$$d_H(f_P, f_0) \leq d_H(f_P, K_\Sigma G) + d_H(K_\Sigma G, f_0) \leq d_H(f_{\hat{P}}, f_{\hat{G}}) + d_H(f_{\hat{P}}, f_P) + \varepsilon \log^r(1/\varepsilon),$$

where  $d_{\hat{G}}(y, \Sigma) = \delta_{\Sigma(y)} dG(y)$  and  $\hat{P} = \sum_{h \geq 1} \pi_h \delta_{\mu_h, \hat{\Sigma}_h}$  with

$$\hat{\Sigma}_h = \begin{cases} \Sigma_h & \text{if } h \leq \tilde{N} \\ \Sigma(z_h) & \text{if } h > \tilde{N} \text{ and } \mu_h \in V_j \text{ with } j \leq N; \\ 0 & \text{otherwise,} \end{cases}$$

where, conventionally,  $\varphi_\Sigma(\cdot - z)dz = \delta_z$  when  $\Sigma = 0$ . Since  $P \in \mathcal{F}_{\tilde{N}}$ , there easily holds  $\|f_P - f_{\hat{P}}\|_1 \leq 2\varepsilon^2$  and, using Lemma 5.D.1, we find that

$$\begin{aligned} & \|f_{\hat{P}} - f_{\hat{G}}\|_1 \\ & \leq 2 \sum_{j=1}^N |P_\mu(V_j) - p_j| + \frac{1}{2} \sigma^{-\alpha_0} \sup_{1 \leq j \leq N} \text{diam } V_j + \frac{3}{2} \sup_{1 \leq j \leq N} \sup_{\mu_h \in V_j} \sqrt{\text{tr}(\Sigma(z_j)^{-1} \hat{\Sigma}_h - \text{Id})^2}, \\ & \lesssim \varepsilon^2 + \sup_{1 \leq j \leq N} \sup_{\mu_h \in V_j} \|\Sigma(z_j)^{-1} \hat{\Sigma}_h - \text{Id}\|_{\text{op}}. \end{aligned}$$

where we used that  $P \in \mathcal{P}_J$  and that  $\text{diam } V_j \lesssim \sigma^{\alpha_0} \varepsilon^2$  for  $j \leq N$ . In the last supremum, if  $h > \tilde{N}$  and  $\mu_h \in V_j$ , then  $\Sigma_h = \Sigma(z_j)$  so we only need to handle the case when  $h \leq \tilde{N}$ . Moreover,

$$\begin{aligned} \Sigma(z_j)^{-1} \hat{\Sigma}_h - \text{Id} &= O_{z_j}^\top \Delta_{\sigma, \delta}^{-2} O_{z_j} U_h^\top \Lambda_h U_h - \text{Id} \\ &= O_{z_j}^\top \Delta_{\sigma, \delta}^{-2} (O_{z_j} U_h^\top - \text{Id}) \Lambda_h U_h + O_{z_j}^\top \Delta_{\sigma, \delta}^{-2} \Lambda_h U_h - \text{Id} \\ &= O_{z_j}^\top \Delta_{\sigma, \delta}^{-2} (O_{z_j} U_h^\top - \text{Id}) \Lambda_h U_h + O_{z_j}^\top (\Delta_{\sigma, \delta}^{-2} \Lambda_h - \text{Id}) U_h + O_{z_j}^\top U_h - \text{Id}, \end{aligned}$$

so that

$$\begin{aligned} \|\Sigma(z_j)^{-1} \hat{\Sigma}_h - \text{Id}\|_{\text{op}} &\lesssim \|\Delta_{\sigma, \delta}^{-2}\|_{\text{op}} \|\Lambda_h\|_{\text{op}} \|O_{z_j} U_h^\top - \text{Id}\|_{\text{op}} + \|\Delta_{\sigma, \delta}^{-2} \Lambda_h - \text{Id}\|_{\text{op}} + \|O_{z_j}^\top U_h - \text{Id}\|_{\text{op}} \\ &\lesssim \varepsilon^2 + \sigma^{2\beta} \simeq \varepsilon^2, \end{aligned}$$

where we used both that  $h \leq \tilde{N}$  and  $P \in \mathcal{O}_{\tilde{N}} \cap \mathcal{L}_{\tilde{N}}$ .

Using [STG13, Lem B2], we find that for  $\lambda > 0$  small enough

$$\begin{aligned} P_0 \log \frac{f_0}{f_P} &\lesssim d_H^2(f_0, f_P) (1 + \log(1/\lambda)) + P_0 \left\{ \log \frac{f_0}{f_P} \mathbb{1}_{f_P/f_0 < \lambda} \right\} \\ P_0 \left( \log \frac{f_0}{f_P} \right)^2 &\lesssim d_H^2(f_0, f_P) (1 + \log^2(1/\lambda)) + P_0 \left\{ \left( \log \frac{f_0}{f_P} \right)^2 \mathbb{1}_{f_P/f_0 < \lambda} \right\}, \end{aligned}$$

up to numeric constant. Notice that by assumption,  $f_0$  is bounded from above by  $\|L\|_\infty$ . Let  $x \in M^\tau \cap B(0, R)$ , and let  $1 \leq j \leq J$  be such that  $x \in V_j$ . Notice that, since  $P \in \mathcal{F}_{\tilde{N}} \cap \mathcal{P}_J$ , there holds

$$\varepsilon^4 \leq P_\mu(V_j) = \sum_{\mu_h \in V_j} \pi_h \leq \sum_{\substack{\mu_h \in V_j \\ h \leq \tilde{N}}} \pi_h + \varepsilon^8 \quad \text{so that} \quad \sum_{\substack{\mu_h \in V_j \\ h \leq \tilde{N}}} \pi_h \geq \varepsilon^4/2.$$

Now we can write

$$f_P(x) \geq \sum_{\substack{\mu_h \in V_j \\ h \leq \tilde{N}}} \pi_h \varphi_{\Sigma_h}(x - \mu_h) \gtrsim \frac{\varepsilon^4}{\delta^{D-d} \sigma^D},$$

where we used that  $|\det \Sigma_h|^{1/2} \leq (1 + \sigma^{2\beta})^{D/2} \delta^{D-d} \sigma^D \lesssim \delta^{D-d} \sigma^D$  and that  $\|x - y\|_{\Sigma^{-1}(y)} \lesssim 1$  for any  $x, y \in V_j$  with a similar line of reasoning as in the proof of Lemma 5.5.3, relying on  $P \in \mathcal{O}_{\tilde{N}} \cap \mathcal{L}_{\tilde{N}}$ . Furthermore, since  $f_P(x) \leq \varepsilon^8 + \sum_{h \leq \tilde{N}} \pi_h \varphi_{\Sigma_h}(x - \mu_h)$ , there must be some  $h \leq \tilde{N}$  such that both  $\pi_h \geq \varepsilon^2$  and  $\mu_h \in B(0, R)$  holds, otherwise one would get a contradiction looking at the mass of  $f_0$  since one would get

$$\begin{aligned} \|f_0\|_1 &= \|f_0 \mathbb{1}_{B(0, R/2)}\|_1 + \|f_0 \mathbb{1}_{B(0, R/2)^c}\|_1 \\ &\leq \underbrace{\|f_P \mathbb{1}_{B(0, R/2)}\|_1}_{\lesssim \varepsilon^2 + \varepsilon^H} + \underbrace{\|(f_P - f_0) \mathbb{1}_{B(0, R/2)}\|_1}_{\lesssim \varepsilon + \sigma^\beta} + \underbrace{\|f_0 \mathbb{1}_{B(0, R/2)^c}\|_1}_{\lesssim \varepsilon^H}, \end{aligned}$$

where the inequalities occurs up to log-term. For  $x \notin B(0, R)$ , notice that, for this particular  $h \leq \tilde{N}$ ,  $f_P(x) \gtrsim \varepsilon^2 \varphi_{\Sigma_h}(x - \mu_h)$ . Taking  $\lambda = C\varepsilon^4/(\delta^{D-d} \sigma^D)$  for some small constant  $C$ , one get for any  $\ell \geq 1$ ,

$$\begin{aligned} P_0 \left\{ \left( \log \frac{f_0}{f_P} \right)^\ell \mathbb{1}_{f_P/f_0 < \lambda} \right\} &\leq P_0 \left\{ \left( \log \frac{f_0}{f_P} \right)^\ell \mathbb{1}_{B(0, R)^c} \right\} \\ &\lesssim \log^\ell \frac{\varepsilon^2}{\sigma^D \delta^{D-d}} \times P_0(B(0, R)^c) + \int_{B(0, R)^c} \|x - \mu_h\|_{\Sigma_h^{-1}}^{2\ell} f_0(x) dx \\ &\lesssim \varepsilon^H \log^\ell \frac{\varepsilon^2}{\sigma^D \delta^{D-d}} + \sigma^{-2\ell\alpha_0} \left\{ \int \|x\|^{4\ell} f_0(x) dx \right\}^{1/2} P_0(B(0, R)^c)^{1/2} \\ &\lesssim \varepsilon^2 \log^2(1/\varepsilon), \end{aligned}$$

where the last inequality holds for  $\ell \in \{1, 2\}$ , provided that we chose  $H \geq 8\alpha_0 + 4\beta$ . This shows that  $f_P \in B(f_0, \tilde{\varepsilon})$  for  $\tilde{\varepsilon} \approx \varepsilon \log^s(1/\varepsilon)$  with  $s = r \vee 1$ . It only remains to lower bound the prior mass of the event  $\mathcal{P}_J \cap \mathcal{F}_{\tilde{N}} \cap \mathcal{O}_{\tilde{N}} \cap \mathcal{L}_{\tilde{N}}$ .

For this, one can use (5.8) and the fact that the scales are drawn independently to rest to find that

$$\Pi(\mathcal{P}_J \cap \mathcal{F}_{\tilde{N}} \cap \mathcal{O}_{\tilde{N}} \cap \mathcal{L}_{\tilde{N}}) \geq \Pi(\mathcal{P}_J \cap \mathcal{F}_{\tilde{N}}) \times c_o^{\tilde{N}} \prod_{h \leq \tilde{N}} O\{\|UO_{\mu_h} - \text{Id}\|_{\text{op}} \leq \sigma^{2\alpha_0} \varepsilon^2\} \times \Pi(\mathcal{L}_{\tilde{N}}).$$

We easily get that  $O\{\|UO_{\mu_h} - \text{Id}\|_{\text{op}} \leq \sigma^{2\alpha_0}\varepsilon^2\} \gtrsim (\sigma^{2\alpha_0}\varepsilon^2)^{D(D-1)/2}$ . For  $\Pi(\mathcal{L}_{\tilde{N}})$ , we use (5.9) or (5.10) along with a simple Markov inequality to find that

$$\begin{aligned} \Pi(\mathcal{L}_{\tilde{N}}) &= \mathbb{E}_{\Pi} \left\{ Q_2 \left( [\sigma^{2\alpha_0}, (1 + \sigma^{2\beta})\sigma^{2\alpha_0}]^d \times [\delta^2 \sigma^{2\alpha_\perp}, (1 + \sigma^{2\beta})\delta^2 \sigma^{2\alpha_\perp}]^{D-d} \right)^{\tilde{N}} \right\} \\ &\gtrsim \exp(-2\alpha_0 B_0 \log(1/\sigma) \tilde{N}) \exp\{-c_2 D \sigma^{-D}\}. \end{aligned}$$

For  $\Pi(\mathcal{P}_J \cap \mathcal{F}_{\tilde{N}})$ , we write

$$\Pi(\mathcal{P}_J \cap \mathcal{F}_{\tilde{N}}) \geq \Pi(\mathcal{P}_J) - \Pi(\mathcal{F}_{\tilde{N}}^c) \gtrsim e^{-CJ \log(1/\varepsilon)} - e^{-\tilde{N} \log(\tilde{N})},$$

where we used [GVDV07, Lem 10] and the bound [STG13, p. 15]. We conclude by noticing that  $e^{-\tilde{N} \log(\tilde{N})} \ll e^{-N \log(1/\varepsilon)}$  so that in the end

$$\Pi(\mathcal{P}_J \cap \mathcal{F}_{\tilde{N}} \cap \mathcal{O}_{\tilde{N}} \cap \mathcal{L}_{\tilde{N}}) \gtrsim \exp(-CN \log^t(1/\varepsilon)),$$

for some  $C$  depending on the parameters and  $t = D/\kappa + 2$ , ending the proof.

### 5.D.3 Proof of Lemma 5.5.3

*Proof of Lemma 5.5.3.* Let  $g : \mathbb{R}^D \rightarrow \mathbb{R}$  be a density supported on  $M^\delta$  and satisfying (5.12) and (5.13). Let  $R = (H \log(1/\varepsilon)/C_2)^{1/\kappa}$ , so that  $g$  is less than  $C_1 \varepsilon^H$  outside of  $B(0, R)$ . We consider the functions  $\{\chi_i\}_{i \in \mathcal{I}}$  introduced in Lemma 5.A.6 associated with a  $\tau/64$ -packing  $\{x_i\}_{i \in \mathcal{I}}$  of  $M \cap B(0, R)$ . Recall that the number  $|\mathcal{I}|$  of functions is less than of order  $R^D$ . We can write

$$K_\Sigma g(x) = K_\Sigma(g \mathbb{1}_{B^c(0, R)})(x) + \sum_{i \in \mathcal{I}} K_\Sigma(\chi_i g \mathbb{1}_{B(0, R)})(x) = K_\Sigma(g \mathbb{1}_{B^c(0, R)})(x) + \sum_{i \in \mathcal{I}} c_i K_\Sigma g_i(x)$$

where  $c_i = \|\chi_i g \mathbb{1}_{B(0, R)}\|_1$  and  $g_i = \chi_i g \mathbb{1}_{B(0, R)}/c_i$  is a density supported on  $\mathcal{O}_{x_i}^0$ . Notice that for any  $x \in \mathbb{R}^D$

$$K_\Sigma(g \mathbb{1}_{B^c(0, R)})(x) \leq C_1 \varepsilon^H K_\Sigma(\mathbb{1}_{B^c(0, R)})(x) \leq C_1 \varepsilon^H$$

and that

$$\begin{aligned} \|K_\Sigma(g \mathbb{1}_{B^c(0, R)})\|_1 &= \|g \mathbb{1}_{B^c(0, R)}\|_1 \leq \int_{\|x\| \geq R} C_1 e^{-C_2 \|x\|^\kappa} \lesssim \int_{r \geq R} e^{-C_2 r^\kappa} r^{D-1} dr \\ &\lesssim \log^{(D-2)/\kappa+1}(1/\varepsilon) \times \varepsilon^H, \end{aligned}$$

where we used a NP-bound on the incomplete Gamma function in the last inequality. Consequently, the term  $K_\Sigma(g \mathbb{1}_{B^c(0, R)})$  need not be discretized. Take now  $i \in \mathcal{I}$ . As in the proof of Theorem 5.3.7, there holds that

$$\varphi_{\Sigma(y)}(x - y) \lesssim \frac{\varepsilon^H}{\sigma^D \delta^{D-d}} \quad \forall x \notin \mathcal{O}_{x_i}^1, y \in \mathcal{O}_{x_i}^0,$$

provided that  $\tau \gtrsim R\delta\sigma^{\alpha_1}$ . This means that  $|K_{\Sigma}g_i(x)| \lesssim \frac{\varepsilon^H}{\sigma^D\delta^{D-d}}$  for all  $x \notin \mathcal{O}_{x_i}^1$ . If now  $x \in \mathcal{O}_{x_i}^1$ , there holds

$$\begin{aligned} K_{\Sigma}g_i(x) &= \int_{\mathcal{W}_{x_i}^0} \varphi_{\Sigma}(\bar{\Psi}_{x_i}(w))(x - \bar{\Psi}_{x_i}(w))\bar{g}_i(w)dw \\ &= \sum_{j \in \mathcal{J}_i} \bar{c}_{i,j} \int_{\mathcal{W}_{i,j}} \varphi_{\Sigma}(\bar{\Psi}_{x_i}(w))(x - \bar{\Psi}_{x_i}(w))\bar{g}_{i,j}(w)dw \end{aligned}$$

where  $\bar{g}_i(w) = g_i(\bar{\Psi}_{x_i}(w)) \times |\det d\bar{\Psi}_{x_i}(w)|$ ,  $\bar{c}_{i,j} = \int_{\mathcal{W}_{i,j}} \bar{g}_i$  and  $\bar{g}_{i,j} = \bar{g}_i/\bar{c}_{i,j}$ . The sets  $\mathcal{W}_{i,j}$  form a partition of  $\mathcal{W}_{x_i}^0$  that are included in

$$\prod_{\ell=1}^d [w_{t_\ell}^0, w_{t_{\ell+1}}^0) \times \prod_{\ell=d+1}^D [w_{t_\ell}^\perp, w_{t_{\ell+1}}^\perp) \cap \mathcal{W}_{x_i}^0,$$

for  $(w_{t_1}^0, \dots, w_{t_d}^0, w_{t_{d+1}}^\perp, \dots, w_{t_D}^\perp)$  vary on a grid of size  $(\sigma^{\alpha_0}, \dots, \sigma^{\alpha_0}, \delta\sigma^{\alpha_1}, \dots, \delta\sigma^{\alpha_1})$ . Notice that

$$\text{Card } \mathcal{J}_i \simeq \frac{1}{\sigma^{\alpha_0 d}} \times \left( \frac{\delta}{\delta\sigma^{\alpha_1}} \right)^{D-d} = \sigma^{-D}.$$

Let  $i \in \mathcal{I}$  and  $j \in \mathcal{J}_i$  be fixed and denote for short  $\bar{\Psi} = \bar{\Psi}_{x_i}$  and  $\bar{\Sigma} = \Sigma \circ \bar{\Psi}$ . We let  $\Gamma > 0$  and we distinguish two cases:  $\inf_{w \in \mathcal{W}_{i,j}} \|x - \bar{\Psi}(w)\|_{\bar{\Sigma}^{-1}(w)}^2 \geq \Gamma \log(1/\varepsilon)$  and  $\inf_{w \in \mathcal{W}_{i,j}} \|x - \bar{\Psi}(w)\|_{\bar{\Sigma}^{-1}(w)}^2 < \Gamma \log(1/\varepsilon)$ . In the former,

$$\int_{\mathcal{W}_{i,j}} \varphi_{\bar{\Sigma}(w)}(x - \bar{\Psi}(w))\bar{g}_{i,j}(w)dw \lesssim \frac{\varepsilon^{\Gamma/2}}{\sigma^D\delta^{D-d}}.$$

While if  $\inf_{w \in \mathcal{W}_{i,j}} \|x - \bar{\Psi}(w)\|_{\bar{\Sigma}^{-1}(w)}^2 \leq \Gamma \log(1/\varepsilon)$ , we first show that  $\sup_{w \in \mathcal{W}_{i,j}} \|x - \bar{\Psi}(w)\|_{\bar{\Sigma}^{-1}(w)}^2 \leq \Gamma' \log(1/\varepsilon)$  for some  $\Gamma' > \Gamma$ . We let  $w_0 \in \mathcal{W}_{i,j}$  such that  $\|x - \bar{\Psi}(w_0)\|_{\bar{\Sigma}^{-1}(w_0)}^2 \leq 2\Gamma \log(1/\varepsilon)$  and take  $w \in \mathcal{W}_{i,j}$ . Using (5.30), we get that there  $\|\text{pr}_{T_{\bar{\Psi}(w)}} - \text{pr}_{T_{\bar{\Psi}(w_0)}}\|_{\text{op}} \leq 2/\tau\sigma^{\alpha_0}$  and the same holds for  $\text{pr}_{N_{\bar{\Psi}(w)}} - \text{pr}_{N_{\bar{\Psi}(w_0)}}$ . Furthermore, using (5.19) and (5.22) with the same set of notations yields

$$\|\text{pr}_{T_{\bar{\Psi}(w_0)}}(\bar{\Psi}(w) - \bar{\Psi}(w_0))\| \leq C'\sigma^{\alpha_0} \text{ and } \|\text{pr}_{T_{\bar{\Psi}(w_0)}}(\bar{\Psi}(w) - \bar{\Psi}(w_0))\| \leq C'\delta\sigma^{\alpha_1},$$

for some other constant  $C' > 0$ . All in all, there holds

$$\begin{aligned} \|x - \bar{\Psi}(w)\|_{\bar{\Sigma}^{-1}(w)}^2 &= \frac{1}{\sigma^{2\alpha_0}} \|\text{pr}_{T_{\bar{\Psi}(w)}}(x - \bar{\Psi}(w))\|^2 + \frac{1}{\delta^2\sigma^{2\alpha_1}} \|\text{pr}_{N_{\bar{\Psi}(w)}}(x - \bar{\Psi}(w))\|^2 \\ &\leq 4C^2 + \frac{2}{\sigma^{2\alpha_0}} \|\text{pr}_{T_{\bar{\Psi}(w_0)}}(x - \bar{\Psi}(w))\|^2 + \frac{2}{\delta^2\sigma^{2\alpha_1}} \|\text{pr}_{N_{\bar{\Psi}(w_0)}}(x - \bar{\Psi}(w))\|^2 \\ &\leq 4C^2 + 8C'2 + \underbrace{\frac{4}{\sigma^{2\alpha_0}} \|\text{pr}_{T_{\bar{\Psi}(w_0)}}(x - \bar{\Psi}(w_0))\|^2 + \frac{4}{\delta^2\sigma^{2\alpha_1}} \|\text{pr}_{N_{\bar{\Psi}(w_0)}}(x - \bar{\Psi}(w_0))\|^2}_{=4\|x - \bar{\Psi}(w_0)\|_{\bar{\Sigma}^{-1}(w_0)}^2} \\ &\leq 4C^2 + 8C'2 + 8\Gamma \log(1/\varepsilon) \leq 9\Gamma \log(1/\varepsilon), \end{aligned}$$

for  $\varepsilon$  small enough. Denote  $R_T(u) = \exp(u) - \sum_{t=0}^{T-1} u^t/t!$ , then  $|R_T(u)| \leq e^u|u|^T/T!$  and

$$\exp\left\{-\frac{1}{2}\|x - \bar{\Psi}(w)\|_{\bar{\Sigma}^{-1}(w)}^2\right\} = \sum_{t=0}^{T-1} \frac{(-1)^t}{2^t t!} \|x - \bar{\Psi}(w)\|_{\bar{\Sigma}^{-1}(w)}^{2t} + \underbrace{R_T\left(-\|x - \bar{\Psi}(w)\|_{\bar{\Sigma}^{-1}(w)}^2/2\right)}_{:= R_T(x,w)}.$$

Note that  $R_T(x, w)$  is uniformly bounded by

$$|R_T(x, w)| \leq e^{5\Gamma \log(1/\varepsilon)} \frac{(5\Gamma \log(1/\varepsilon))^T}{T!} \approx \frac{\varepsilon^{-5\Gamma \log^T(1/\varepsilon)}}{T!}.$$

Set,

$$A_{i,j}(w) := \langle e_i, \bar{\Sigma}^{-1}(w) e_j \rangle, \quad B_i(w) := \langle e_i, \bar{\Sigma}^{-1}(w) \bar{\Psi}(w) \rangle, \quad \text{and} \quad C(w) := \|\bar{\Psi}(w)\|_{\bar{\Sigma}^{-1}(w)}^2,$$

so that all functions  $A_{i,j}$ ,  $B_i$  and  $C$  are continuous functions of  $w$  and if  $x = (x_1, \dots, x_D)$ ,

$$\begin{aligned} \|x - \bar{\Psi}(w)\|_{\bar{\Sigma}^{-1}(w)}^{2t} &= \left\{ \|x\|_{\bar{\Sigma}^{-1}(w)}^2 - 2\langle x, \bar{\Sigma}^{-1}(w) \bar{\Psi}(w) \rangle + \|\bar{\Psi}(w)\|_{\bar{\Sigma}^{-1}(w)}^2 \right\}^t \\ &= \sum_{|k|=t} \binom{t}{k} (-2)^{k_2} \left\{ \sum_{1 \leq i, j \leq D} x_i x_j A_{i,j}(w) \right\}^{k_1} \left\{ \sum_{i=1}^D x_i B_i(w) \right\}^{k_2} C(w)^{k_3} \\ &= \sum_{|k|=t} \binom{t}{k} (-2)^{k_2} C(w)^{k_3} \sum_{|\ell|=k_1} \binom{k_1}{\ell} \prod_{1 \leq i, j \leq D} (x_i x_j)^{\ell_{i,j}} A_{i,j}^{\ell_{i,j}}(w) \sum_{|m|=k_2} \binom{k_2}{m} \prod_{i=1}^D x_i^{m_i} B_i^{m_i}(w) \\ &= \sum_{(p, \ell, m) \in \mathcal{G}_t} P_{p, \ell, m}(x) \times C(w)^p \underbrace{\prod_{1 \leq i, j \leq D} A_{i,j}^{\ell_{i,j}}(w) \prod_{1 \leq i \leq D} B_i^{m_i}(w)}_{:= Q_{p, \ell, m}(w)} \end{aligned}$$

where  $P_{p, \ell, m}(x)$  are polynomial functions of  $x$ ,  $Q_{p, \ell, m}(w)$  are continuous functions of  $w$ , and where  $\mathcal{G}_t$  is the set  $\{(p, \ell, m) \mid p + |\ell| + |m| = t\} \subset \mathbb{N}^{D^2+D+1}$ . According to [GVDV01, Lem 3.1], one can always find an atomic probability measure  $G_{i,j}$  such that

$$\int Q_{p, \ell, m}(w) \bar{g}_{i,j}(w) dw = \int Q_{p, \ell, m}(w) G_{i,j}(dw)$$

for all  $p, \ell, m$  such that  $p + |\ell| + |m| \leq T-1$ . Since there are less than  $T^{D^2+D+1}$  such triplets, the probability measure  $G_{i,j}$  can be taken to have less than  $T^{D^2+D+1}$  atoms. Note that then, this measure satisfies that

$$\begin{aligned} &\left| \int_{\mathcal{W}_{i,j}} \varphi_{\bar{\Sigma}(w)}(x - \bar{\Psi}(w)) (\bar{g}_{i,j}(w) dw - G_{i,j}(dw)) \right| \\ &= \frac{1}{(2\pi)^{D/2} \delta^{D-d} \sigma^D} \left| \int_{\mathcal{W}_{i,j}} R_T(x, w) (\bar{g}_{i,j}(w) dw - G_{i,j}(dw)) \right| \lesssim \frac{\varepsilon^{-5\Gamma \log^T(1/\varepsilon)}}{\delta^{D-d} \sigma^D T!}. \end{aligned}$$

All in all,  $G_{i,j}$  is such that

$$\begin{aligned} &\left| \int_{\mathcal{W}_{i,j}} \varphi_{\bar{\Sigma}(w)}(x - \bar{\Psi}(w)) \bar{g}_{i,j}(w) dw - \int_{\mathcal{W}_{i,j}} \varphi_{\bar{\Sigma}(w)}(x - \bar{\Psi}(w)) G_{i,j}(dw) \right| \\ &\lesssim \begin{cases} \frac{1}{\sigma^D \delta^{D-d}} \varepsilon^H & \text{if } x \notin \mathcal{O}_{x_i}^1; \\ \frac{1}{\sigma^D \delta^{D-d}} \varepsilon^{\Gamma/2} & \text{if } x \in \mathcal{O}_{x_i}^1 \text{ and } \inf_{w \in J} \|x - \bar{\Psi}(w)\|_{\bar{\Sigma}^{-1}(w)}^2 \geq \Gamma \log(1/\varepsilon); \\ \frac{1}{\sigma^D \delta^{D-d}} \frac{1}{T!} \varepsilon^{-5\Gamma \log^T(1/\varepsilon)} & \text{if } x \in \mathcal{O}_{x_i}^1 \text{ and } \inf_{w \in J} \|x - \bar{\Psi}(w)\|_{\bar{\Sigma}^{-1}(w)}^2 \leq \Gamma \log(1/\varepsilon). \end{cases} \end{aligned}$$

Taking  $H \geq 1$ ,  $\Gamma \geq 2$  and  $T \geq 5\Gamma \log(1/\varepsilon)$  yields a bound of order  $\varepsilon/\sigma^D \delta^{D-d}$  in every case. Then the probability measure

$$G = \sum_{i \in \mathcal{I}} \alpha_i \sum_{j \in \mathcal{J}_i} \bar{c}_{i,j}(\bar{\Psi}_{x_i})_{\#} G_{i,j},$$

is a discrete measure on  $M^\delta$  with at most

$$N_\varepsilon = T^{D^2+D+1} \sum_{i \in \mathcal{I}} \text{Card } \mathcal{J}_i \approx \log^{D^2+D+1}(1/\varepsilon) R^D \sigma^{-D} \leq \sigma^{-D} \log^{D^2+D/(\kappa \wedge 2)}(1/\varepsilon)$$

atoms and such that  $\|K_\Sigma G - K_\Sigma g\|_\infty \lesssim \varepsilon/(\sigma^D \delta^{D-d})$ .

We now turn to bounding the  $L^1$  norm. Let again pick  $i \in \mathcal{I}$  and  $j \in \mathcal{J}_i$ . We let  $x_{i,j} = \bar{\Psi}_{x_i}(w_{i,j})$  for some  $w_{i,j} \in \mathcal{W}_{i,j}$  and for  $T > 0$  we define

$$\mathcal{H}_A = \{x \in \mathbb{R}^D \mid \|x - x_{i,j}\|_{\Sigma^{-1}(x_{i,j})} \leq A\}.$$

There holds that  $\text{Leb } \mathcal{H}_A \simeq \delta^{D-d} \sigma^D T^D$ . Furthermore, we have, denoting respectively  $P_{i,j}$  and  $Q_{i,j}$  the push-forwards of  $\bar{g}_{i,j}$  and  $G_{i,j}$  through  $\bar{\Psi}_{x_i}$

$$\begin{aligned} \|K_\Sigma P_{i,j} - K_\Sigma Q_{i,j}\|_1 &= \int_{\mathcal{H}_A} |K_\Sigma P_{i,j} - K_\Sigma Q_{i,j}| + \int_{\mathcal{H}_A^c} |K_\Sigma P_{i,j} - K_\Sigma Q_{i,j}| \\ &\lesssim \delta^{D-d} \sigma^D A^D \|K_\Sigma P_{i,j} - K_\Sigma Q_{i,j}\|_\infty \int_{\mathcal{H}_A^c} [K_\Sigma Q_{i,j} + K_\Sigma P_{i,j}] \\ &\lesssim A^D \varepsilon + \int_{\mathcal{H}_A^c} [K_\Sigma P_{i,j} + K_\Sigma Q_{i,j}]. \end{aligned} \quad (5.43)$$

Recall that the support of  $P_{i,j}$  and  $Q_{i,j}$  are in  $\bar{\Psi}_{x_i}(\mathcal{W}_{i,j})$ . Furthermore, since

$$\text{diam } \text{pr}_M \bar{\Psi}_{x_i}(\mathcal{W}_{i,j}) \lesssim \sigma^{\alpha_0},$$

there holds, using (5.30), that  $\|\text{pr}_{T_y} - \text{pr}_{T_{y'}}\|_{\text{op}} \lesssim \sigma^{\alpha_0}$  for any two  $y, y' \in \bar{\Psi}_{x_i}(\mathcal{W}_{i,j})$ . Let  $y \in \bar{\Psi}_{x_i}(\mathcal{W}_{i,j})$  and  $x \in \mathcal{H}_A^c$ , we now show that  $\|x - y\|_{\Sigma^{-1}(y)} \geq \zeta A$  is  $A$  is chosen large enough, where  $\zeta > 0$  is a fixed constant. We have

$$\begin{aligned} \|x - y\|_{\Sigma^{-1}(y)} &= \left\| \frac{1}{\sigma^{\alpha_0}} \text{pr}_{T_y}(x - y) + \frac{1}{\sigma^{\alpha_\perp} \delta} \text{pr}_{N_y}(x - y) \right\| \\ &= \left\| \frac{1}{\sigma^{\alpha_0}} \text{pr}_{T_{x_{i,j}}}(x - y) + \frac{1}{\sigma^{\alpha_\perp} \delta} \text{pr}_{N_{x_{i,j}}}(x - y) \right\| + O(\|x - y\|) \\ &= \left\| \frac{1}{\sigma^{\alpha_0}} \text{pr}_{T_{x_{i,j}}}(x - x_{i,j}) + \frac{1}{\sigma^{\alpha_\perp} \delta} \text{pr}_{N_{x_{i,j}}}(x - x_{i,j}) \right\| + O(\|x - y\|) + O(1), \\ &= \|x - x_{i,j}\|_{\Sigma^{-1}(x_{i,j})} + O(\|x - y\|) + O(1). \end{aligned} \quad (5.44)$$

where the term  $O(1)$  comes from noticing that

$$\begin{aligned} \|\text{pr}_{T_{x_{i,j}}}(y - x_{i,j})\| &\leq \|\text{pr}_{T_{x_{i,j}}}(\Psi_{x_i}(v) - \Psi_{x_i}(v'))\| + \|\text{pr}_{T_{x_{i,j}}} N_{x_i}(v, \eta)\| \\ &\lesssim |v - v_{i,j}| + \|\text{pr}_{T_{x_{i,j}}} - \text{pr}_{T_y}\|_{\text{op}} \times \|\eta\| \lesssim \sigma^{\alpha_0}, \end{aligned}$$



and

$$\begin{aligned} \|\text{pr}_{N_{x_{i,j}}}(y - x_{i,j})\| &\leq \|\text{pr}_{N_{x_{i,j}}}(\Psi_{x_i}(v) - \Psi_{x_i}(v'))\| + \|\text{pr}_{N_{x_{i,j}}}(N_{x_i}(v, \eta) - N_{x_i}(v_{i,j}, \eta_{i,j}))\| \\ &\lesssim |v - v_{i,j}| + |\eta - \eta_{i,j}| \lesssim \sigma^{\alpha_1} \delta. \end{aligned}$$

The above inequalities also imply that  $\|y - x_{i,j}\| = O(1)$  so that  $\|x - y\| = \|x - x_{i,j}\| + O(1)$ . Decomposing  $x - x_{i,j}$  into  $\text{pr}_{T_{x_{i,j}}}$  and  $\text{pr}_{N_{x_{i,j}}}$  show that  $\|x - x_{i,j}\| = o(\|x - x_{i,j}\|_{\Sigma^{-1}(x_{i,j})})$  so that finally

$$\|x - y\|_{\Sigma^{-1}(y)} = \|x - x_{i,j}\|_{\Sigma^{-1}(x_{i,j})}(1 + o(1)) + O(1)$$

uniformly for  $x \in \mathcal{H}_A^c$  and  $y \in \mathcal{W}_{i,j}$ . By choosing  $A$  larger than a fixed constant we then obtain that

$$\|x - y\|_{\Sigma^{-1}(y)} \geq A/2 \quad \text{uniformly on } x \in \mathbb{R}^D \setminus \mathcal{H}_A \text{ and } y \in \bar{\Psi}_{x_i}(\mathcal{W}_{i,j}).$$

This yields that

$$\begin{aligned} \int_{\mathcal{H}_A^c} K_{\Sigma} P_{i,j} &= \int_{\bar{\Psi}_{x_i}(\mathcal{W}_{i,j})} \int_{\mathcal{H}_A^c} \varphi_{\Sigma(y)}(x - y) dx P_{i,j}(dy) \\ &= \frac{1}{(2\pi)^{D/2} \sigma^D \delta^{D-d}} \int_{\bar{\Psi}_{x_i}(\mathcal{W}_{i,j})} \int_{\mathcal{H}_A^c} \exp\left\{-\frac{1}{2}\|x - y\|_{\Sigma^{-1}(y)}^2\right\} dx P(dy) \\ &\leq \frac{1}{(2\pi)^{D/2}} \int_{\|z\| \geq A/2} \exp\left\{-\frac{1}{2}\|z\|^2\right\} dz \lesssim \exp(-A^2/8), \end{aligned}$$

where we made the variable change  $z = \Sigma^{-1/2}(y)(x - y)$  in the second to last inequality. The same holds for  $K_{\Sigma} Q_{i,j}$ . Setting  $A = (8 \log(1/\varepsilon))^{1/2}$  and combining with (5.43),  $\|K_{\Sigma} P_{i,j} - K_{\Sigma} Q_{i,j}\|_1 \lesssim \log^{D/2}(1/\varepsilon)\varepsilon$ . We finally obtain that

$$\|K_{\Sigma} g - K_{\Sigma} G\|_1 \lesssim \sum_i \alpha_i \sum_{j \in \mathcal{J}_i} \bar{c}_{i,j} \log^{D/2}(1/\varepsilon)\varepsilon \lesssim \log^{D/2}(1/\varepsilon)\varepsilon.$$

Also we can choose the atoms of  $G$  to be  $\sigma^{\alpha_0}\varepsilon$  apart, thanks to Lemma 5.D.1 together with the following bound on  $\Sigma(z)^{-1}\Sigma(y) - \text{Id}$  when  $\|z - y\| \leq \sigma^{\alpha_0}\varepsilon$ :

$$\begin{aligned} \Sigma(z)^{-1}\Sigma(y) - \text{Id} &= O_z^\top \Delta_{\sigma,\delta}^{-2} O_z O_y^\top \Delta_{\sigma,\delta}^2 O_y - \text{Id} \\ &= O_z^\top \Delta_{\sigma,\delta}^{-2} (O_z O_y^\top - \text{Id}) \Delta_{\sigma,\delta}^2 O_y + O_z O_y^\top - \text{Id}, \end{aligned}$$

so that

$$\begin{aligned} &\sqrt{\text{tr}(\Sigma(z)^{-1}\Sigma(y) - \text{Id})^2} \\ &\leq \sqrt{D} \|\Sigma(z)^{-1}\Sigma(y) - \text{Id}\|_{\text{op}} \lesssim \|\Delta_{\sigma,\delta}^{-2}\|_{\text{op}} \|O_z O_y^\top - \text{Id}\|_{\text{op}} \|\Delta_{\sigma,\delta}^2\|_{\text{op}} + \|O_z O_y^\top - \text{Id}\|_{\text{op}} \\ &\lesssim \sigma^{-2\alpha_0} \|O_z - O_y\|_{\text{op}} \lesssim \sigma^{-2\alpha_0} \|\text{pr}_{T_z} - \text{pr}_{T_y}\|_{\text{op}} \lesssim \sigma^{-2\alpha_0} \|z - y\| \lesssim \varepsilon, \end{aligned}$$

where we used (5.30) in the second to last inequality, ending the proof.  $\square$

## 5.E Appendix to Section 5.4

In this section, we give some details on the procedure we used in Section 5.4. The prior we implemented is a version of the partial location-scale mixture described in Section 5.2: for  $D \in \{2, 3\}$ , and  $n \in \mathbb{N}$ , we set

$$\begin{aligned}
 (x_i)_{1 \leq i \leq n} \mid (\mu_i)_{1 \leq i \leq n}, (\Sigma_i)_{1 \leq i \leq n} &\sim \bigotimes_{i=1}^n \mathcal{N}(\mu_i, \Sigma_i), \\
 (\mu_i)_{1 \leq i \leq n} \mid P &\sim P^{\otimes n}, \\
 (\Sigma_i)_{1 \leq i \leq n} \mid \lambda_1, \lambda_2 &\sim O_{D, \lambda_1, \lambda_2}^{\otimes n}, \\
 P &\sim \text{DP}(\alpha \text{Unif}([-c, c]^D)), \\
 \lambda_1, \lambda_2 \mid b_1, b_2 &\sim \text{Inv}\Gamma(a_1, b_1) \otimes \text{Inv}\Gamma(a_2, b_2), \\
 b_1, b_2 &\sim \text{Exp}(\kappa_1) \otimes \text{Exp}(\kappa_2),
 \end{aligned} \tag{5.45}$$

where  $\kappa_1, \kappa_2, a_1, a_2, \alpha$  and  $c$  are positive real number provided by the user. In all the numerical experiments we did, we chose  $\kappa_1 = \kappa_2 = 1$  and  $a_1 = a_2 = \alpha = c = 10$ . It only remains to describe the scale process (5.45). For this, we take advantage of the low-dimensionality of  $D$ , although other solutions exists for high-dimensional data [JHD21]. For  $D = 2$ ,  $O_{2, \lambda_1, \lambda_2}$  is the law of  $O^\top \Lambda O$  where

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \quad \text{and} \quad O = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \quad \text{with } \theta \sim \text{Unif}([0, 2\pi]).$$

For dimension  $D = 3$ ,  $O_{2, \lambda_1, \lambda_2}$  is the law of  $\lambda_2 \text{Id} + (\lambda_1 - \lambda_2)uu^\top$  where

$$u = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \cos(\phi) \\ \sin(\theta) \sin(\phi) \end{pmatrix} \quad \text{with } \theta \sim \text{Unif}([0, \pi]) \quad \text{and } \phi \sim \text{Unif}([0, 2\pi]).$$

The matrix  $\lambda_2 \text{Id} + (\lambda_1 - \lambda_2)uu^\top$  can actually be put in the form  $O^\top \Lambda O$  with

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_2 \end{pmatrix}$$

and with  $O$  being a matrix of the transform  $x \mapsto (\text{pr}_{\text{Vect } u} x, \text{pr}_{u^\perp} x)$  in orthonormal bases, so that  $\lambda_2$  is associated with the 2-dimensional subspace  $u^\perp$  and  $\lambda_1$  with the one dimensional subspace  $\text{Vect } u$ . Notice that we do not provide the algorithm with any prior knowledge of the intrinsic dimension of the data, but rather let it decide whether density is going to be 1-dimensional ( $\lambda_1 \gg \lambda_2$ ), 2-dimensional ( $\lambda_1 \ll \lambda_2$ ) or 3-dimensional ( $\lambda_1 \approx \lambda_2$ ).

We now described analytically the shapes we use for the experiments.

- *The two circles*: it is the union of  $\mathcal{C}_1$  and  $\mathcal{C}_2$  with equation  $(x - x_i)^2 + (y - y_i)^2 = r_i^2$  for  $i \in \{1, 2\}$ . In the experiments, we chose  $(x_1, y_1) = (0, 0)$ ,  $(x_2, y_2) = (2, 0)$  and  $r_1 = r_2 = 2$ .

- *The 2D-spiral*: it is given by the parametric embedding

$$\varphi_2 : t \in [0, 1] \mapsto \begin{cases} R(\omega t + \theta_0) \cos(\omega t + \theta_0) \\ R(\omega t + \theta_0) \sin(\omega t + \theta_0). \end{cases}$$

In the numerical studies, we chose  $R = 1/2\pi$ ,  $\omega = 7\pi/2$  and  $\theta_0 = \pi/2$ .

- *The 3D-spiral*: it is given by the parametric embedding

$$\varphi_3 : t \in [0, 1] \mapsto \begin{cases} R(\omega t + \theta_0) \cos(\alpha t + \theta_0) \\ R(\omega t + \theta_0) \sin(\alpha t + \theta_0) \\ \nu t. \end{cases}$$

In the numerical studies, we chose  $R = 1/2\pi$ ,  $\omega = 7\pi/2$ ,  $\theta_0 = \pi/2$  and  $\nu = 2$ .

- *The torus*: it is the set  $\mathcal{T}$  given by the equation  $(\sqrt{(x - x_0)^2 + (y - y_0)^2} - R)^2 + (z - z_0)^2 = r^2$ . In the experiments, we chose  $(x_0, y_0, z_0) = (0, 0, 0)$ ,  $R = 3$  and  $r = 1$ .

We finally briefly review how the data was generated around each of this shape. The model we used was the additive isotropic noise model (as described in Proposition 5.2.4). The noise kernel we chose was

$$K_{\beta_\perp}(x) \propto (1 - \|x\|^2)_+^{\beta_\perp},$$

and the base densities  $f_*$  on the shape were as follows: for the two-circles, we set  $f_* = 1/2f_1 + 1/2f_2$  where  $f_i$  for  $i \in \{1, 2\}$  is the uniform distribution on  $\mathcal{C}_i$ . For the torus, we chose a uniform distribution on  $\mathcal{T}$  as well. For the 2D-spiral (resp. the 3D-spiral),  $f_*$  was simply the push-forward through  $\varphi_2$  (resp.  $\varphi_3$ ) of the probability distribution  $g_{\beta_0}$  on  $[0, 1]$  given by

$$g_{\beta_0}(t) \propto (1 - (1 - 2t)^{\beta_0})\mathbb{1}_{t \in [0, 1/2]} + (1 - (2t - 1)^{\beta_0+1})\mathbb{1}_{t \in [1/2, 1]}.$$

In the experiments, the regularities were set to  $\beta_0 = 2$  and  $\beta_\perp = 6$ .



## Chapter 6

# From graph centrality to data depth

Given a sample of points in a Euclidean space, we can define a notion of depth by forming a neighborhood graph and applying a notion of node centrality. In the present chapter, we focus on the degree, iterates of the H-index, and the coreness, which are all well-known measures of centrality of a node in a graph. We study their behaviors when applied to a sample of points drawn i.i.d. from an underlying density and with a connectivity radius properly chosen. Equivalently, we study these notions of centrality in the context of random neighborhood graphs. We show that, in the large-sample limit and under some standard condition on the connectivity radius, the degree converges to the likelihood depth (unsurprisingly), while iterates of the H-index and the coreness converge to new notions of depth. This chapter is the subject of the prepublication [AACB21], in revision.

### Contents

<b>6.1</b>	<b>Introduction</b>	<b>220</b>
6.1.1	From Node Centrality to Data Depth	220
6.1.2	Setting and Notation	221
6.1.3	Contribution and Outline	222
<b>6.2</b>	<b>Degree</b>	<b>222</b>
<b>6.3</b>	<b>H-Index</b>	<b>225</b>
6.3.1	Definition of the H-Index	225
6.3.2	Iterated H-Index	226
<b>6.4</b>	<b>Coreness</b>	<b>229</b>
<b>6.5</b>	<b>Numerical Simulations</b>	<b>240</b>
6.5.1	Illustrative Examples	240
6.5.2	Convergence Rates	240
6.5.3	Iterations of the H-Index	242
<b>6.6</b>	<b>Concluding Remarks and Open Questions</b>	<b>244</b>

## 6.1 Introduction

### 6.1.1 From Node Centrality to Data Depth

In the context of multivariate analysis, a notion of *depth* is meant to provide an ordering of the space. While in dimension one there is a natural order (the one inherited by the usual order on the real line), in higher dimensions this is lacking, and impedes the definition of such foundational objects as a median or other quantiles, for example. By now, many notions of data depth have been proposed and the corresponding literature is quite extensive. Most of the notions are geometrical in nature, as perhaps they should be. Among these, for example, we find the half-space depth [Tuk75, DG92], various notions of simplicial depth [Oja83, Liu90], or the convex hull peeling [Bar76, Edd82]. Other notions of depth are not motivated by geometry, in particular the likelihood depth [FLM97, FM99], which is simply given by the values taken by the density (or an estimate when it is unknown). Notions of depth are surveyed in [LPS99, LSS06, Mos13].

While the focus in multivariate analysis is on point clouds, in graph and network analysis the concern is on relationships between some items represented as nodes in a graph. There, the corresponding notion is that of node *centrality*. (There are notions of centrality that apply to edges, but we will not consider these here.) Quite a few notions have been proposed, including the degree, the H-index [Hir05], the coreness [Sei83], and other notions including some based on graph distances [Fre78] or on (shortest-)path counting [Fre77], and still other ones that rely on some spectral properties of the graph [Kat53, Bon72, PBMW99, Kle99]. Notions of centrality are surveyed in [Kol09, BE06, Fre78].

Thus, on the one hand, notions of depth have been introduced in the context of point clouds, while on the other hand, notions of centrality have been proposed in the context of graphs and networks, and these two lines of work seem to have evolved completely separately, with no cross-pollination whatsoever, at least to our knowledge. The only place where we found a hint of that is in the discussion of [Alo06], who mentions a couple of “graph-based approach[es]” which seem to have been developed for the context of point clouds, although one of them — the method of [TF79] based on pruning the minimum spanning tree — applies to graphs as well. We can also mention the recent work of [CS20], who study the large-sample limit of the convex hull peeling, relating it to a motion by (Gaussian) curvature. This lack of interaction may appear surprising, particularly in view of the important role that neighborhood graphs have played in multivariate analysis, for example, in areas like manifold learning [TSL00, WPS05, BN03], topological data analysis [Was18, CCSM11], and clustering [NJW02, AC11, MHvL09, BCQY97]. The consideration of neighborhood graphs has

also led to the definition of geometrical quantities for graphs inspired by Euclidean or Riemannian geometry, such as the volume, the perimeter, and the conductance [TSVB<sup>+</sup>16, ACPP12, MP20], and to the development of an entire spectral theory, in particular the study of the Laplacian [Chu97, BN08, GK06, Sin06].

Inspired by this movement, we draw a bridge between notions of depth for point clouds and notions of centrality for nodes in a graph. In a nutshell, we consider a multivariate analysis setting where the data consist of a set of points in the Euclidean space. The bridge is, as usual, a neighborhood graph built on this point set, which effectively enables the use of centrality measures, whose large sample limit we examine in a standard asymptotic framework where the number of points increases, while the connectivity radius remains fixed or converges to zero slowly enough. In so doing, we draw a correspondence between some well-known measures of centrality and depth, while some notions of centrality are found to lead to new notions of depth. (A bridge going in the other direction, namely from depth to centrality, can be built by first embedding the nodes of a graph as points in a Euclidean space, thus making depth measures applicable. We do not explore this route in the present paper.)

### 6.1.2 Setting and Notation

We consider a multivariate setting where

$$\begin{aligned} \mathcal{X}_n = \{X_1, \dots, X_n\} \text{ is an i.i.d. sample from} \\ \text{a uniformly continuous density } f \text{ on } \mathbb{R}^d. \end{aligned} \tag{6.1}$$

The bridge between point clouds and graphs is the construction of a neighborhood graph. More specifically, for an arbitrary set of distinct points,  $x_1, \dots, x_k \in \mathbb{R}^d$  and a radius  $r > 0$ , let  $\mathcal{G}_r(\{x_1, \dots, x_k\})$  denote the graph with node set  $V = \{1, \dots, k\}$  and edge set  $E = \{(i, j) : \|x_i - x_j\| \leq r\}$ , where  $\|\cdot\|$  denotes the Euclidean norm. Note that the resulting graph is undirected. Although it is customary to weigh the edges by the corresponding pairwise Euclidean distances — meaning that an edge  $(i, j)$  has weight  $\|x_i - x_j\|$  — we choose to focus on purely combinatorial degree-based properties of the graph, so that it is sufficient to work with the unweighted graph.

In what follows, we fix a point  $x \in \mathbb{R}^d$  and study its centrality  $\mathcal{C}(x; \mathcal{G}_r(x, \mathcal{X}_n))$  in the graph  $\mathcal{G}_r(x, \mathcal{X}_n) := \mathcal{G}_r(\{x\} \cup \mathcal{X}_n)$  as  $n \rightarrow \infty$ . This graph is random and sometimes called a *random geometric graph* [Pen03]. The connectivity radius may depend on  $n$  (i.e.,  $r = r_n$ ), although this dependency will be left implicit for the most part.

Everywhere,  $B(x, r)$  will denote the closed ball centered at  $x$  and of radius  $r$ . For a measurable set  $A$ ,  $|A|$  will denote its volume. In particular, we will let  $\omega$  denote the volume of the unit ball, so that  $|B(x, r)| = \omega r^d$  for all  $x \in \mathbb{R}^d$  and  $r \geq 0$ . We will let

$$N := n\omega r^d, \tag{6.2}$$

which, as we shall see, will arise multiple times as a renormalization factor.

### 6.1.3 Contribution and Outline

We study the large-sample ( $n \rightarrow \infty$ ) limit of the centrality of  $x$  in the random neighborhood graph  $\mathcal{G}_r(x, \mathcal{X}_n)$ , where the sample  $\mathcal{X}_n$  is generated as in (6.1). More specifically, we focus on the degree  $\deg_r(x, \mathcal{X}_n)$ ; on the  $k$ th iterate of the H-index  $H_r^k(x, \mathcal{X}_n)$ ; and on the coreness  $C_r(x, \mathcal{X}_n)$ . As will be made clear, these notions of centrality can all be seen as iterates of the H-index, since  $\deg_r(x, \mathcal{X}_n) = H_r^0(x, \mathcal{X}_n)$  and  $C_r(x, \mathcal{X}_n) = H_r^\infty(x, \mathcal{X}_n)$ . Given their prominence in the literature [MGPV20], the degree and the coreness are examined separately. The main limits are taken as the sample size  $n$  goes to infinity while the neighborhood radius  $r$  remains fixed or converges to zero slowly enough. See Figure 6.1.1 for a compact summary of the main results that we derive.

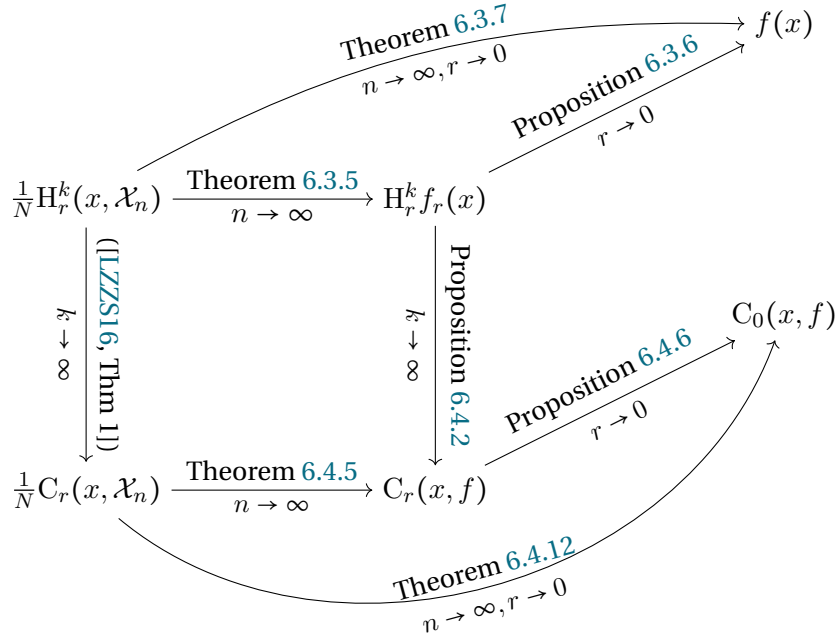


Figure 6.1.1 – These are the main relationships that we establish between notions of centrality and notions of depth.

Section 6.2 is dedicated to the degree, Section 6.3 to the  $k$ th iterate of the H-index for  $1 \leq k < \infty$ , and Section 6.4 to the coreness. In Section 6.5 we report on some numerical simulations.

## 6.2 Degree

The *degree* is arguably the most basic measure of centrality, and also one of the earliest to have been proposed [Fre78]. In our context, the point set  $\mathcal{X}_n$  is an i.i.d sample with common density  $f$  on  $\mathbb{R}^d$ , so that it is composed of  $n$  distinct points almost surely. The



degree of  $x \in \mathbb{R}^d \setminus \mathcal{X}_n$  in the graph  $\mathcal{G}_r(x, \mathcal{X}_n)$  is<sup>1</sup>

$$\deg_r(x, \mathcal{X}_n) := \sum_{i=1}^n \mathbb{1}_{\|x - X_i\| \leq r}. \quad (6.3)$$

Dealing with the degree centrality is rather standard (and very straightforward), but as we will consider more complex notions of centrality below, it helps to draw intuition from the continuum model where we effectively let the sample size diverge ( $n \rightarrow \infty$ ).

**A Stochastic Convergence Result** The analysis of the degree centrality — and of the other centrality measures hereafter — relies on the following elementary lemma, which will be used throughout to control stochastic terms. It involves the Vapnik-Chervonenkis dimension of classes of subsets. Recall that given a class  $\mathcal{S}$  of subsets of  $\mathbb{R}^d$  and an integer  $m \geq 1$ , the *scattering coefficient* of  $\mathcal{S}$  for  $m$  point is defined as

$$\Delta_{\mathcal{S}}(m) := \max_{x_1, \dots, x_m \in \mathbb{R}^d} \# \{(\mathbb{1}_{x_1 \in S}, \dots, \mathbb{1}_{x_m \in S}), S \in \mathcal{S}\},$$

which is the maximum number of different labelings of  $m$  points that  $\mathcal{S}$  can produce. The *Vapnik-Chervonenkis (VC) dimension* of  $\mathcal{S}$  is then defined as the maximum number of points that can be arbitrary labeled with  $\mathcal{S}$ , that is,

$$\text{VC}(\mathcal{S}) := \sup \{m \geq 1, \Delta_{\mathcal{S}}(m) = 2^m\} \in \mathbb{N} \cup \{\infty\}.$$

**Lemma 6.2.1.** *Let  $(\mathcal{S}_r)_{r>0}$  be a family of classes of subsets of  $\mathbb{R}^d$  such that:*

- (i) *The VC-dimension of  $\mathcal{S}_r$  is bounded from above by some  $v \in \mathbb{N}$  uniformly for all  $r > 0$ ;*
- (ii) *For all  $r > 0$  and  $S \in \mathcal{S}_r$ , we have  $\text{diam}(S) \leq 2r$ .*

*Then, for any sequence  $r = r_n$  such that  $nr^d \gg \log n$ , we have*

$$\eta := \sup_{S \in \mathcal{S}_r} \frac{1}{\omega r^d} |P_n(S) - P(S)| \xrightarrow[n \rightarrow \infty]{} 0 \quad \text{a.s.}$$

*Proof.* We use [AST93, Thm 2.1] to get that for any  $r > 0$  and any  $\varepsilon > 0$

$$\mathbb{P} \left( \sup_{S \in \mathcal{S}_r} \frac{1}{\sqrt{P(S)}} |P_n(S) - P(S)| \geq \varepsilon \right) \leq 4\Delta_{\mathcal{S}_r}(2n) \exp \left( -\frac{1}{4} \varepsilon^2 n \right),$$

where  $\Delta_{\mathcal{S}_r}(2n)$  is the scattering number of  $\mathcal{S}_r$  on  $2n$ -points. Using Sauer's lemma, we find that as soon as  $2n \geq v$ , we have  $\Delta_{\mathcal{S}_r}(2n) \leq (2en/v)^v$ . Furthermore, since  $\text{diam}(S) \leq$

<sup>1</sup>If  $x = X_{i_0} \in \mathcal{X}_n$ , the degree of  $x$  in the graph  $\mathcal{G}_r(x, \mathcal{X}_n)$  writes as  $\sum_{i \neq i_0} \mathbb{1}_{\|x - X_i\| \leq r} = (\sum_{i=1}^n \mathbb{1}_{\|x - X_i\| \leq r}) - 1$ , and therefore only differs by 1 from the formula of (6.3). As this difference will be negligible after renormalization by  $1/N$ , we will only consider the sum of indicators of (6.3) for simplicity.

$2r$  for all  $S \in \mathcal{S}_r$ , we have  $\sqrt{P(S)} \leq \sqrt{\|f\|_\infty \omega(2r)^d}$  and setting  $\varepsilon = \kappa \sqrt{\omega(2r)^d / \|f\|_\infty}$  yields

$$\begin{aligned} \mathbb{P}\left(\frac{1}{\omega r^d} \sup_{S \in \mathcal{S}_r} |P_n(S) - P(S)| \geq \kappa\right) &\leq 4(2en/v)^v \exp\left(-\frac{1}{4\|f\|_\infty} \kappa^2 n \omega(2r)^d\right) \\ &\ll 4(2en/v)^v \exp\left(-\frac{2^d}{4\|f\|_\infty} \kappa^2 \log n\right), \end{aligned}$$

which yields the results when taking  $\kappa$  of the form  $c \log(n)^{-1/4}$  for  $c$  large enough and using Borel-Cantelli lemma.  $\square$

Note that in particular the result holds if  $r = r_0$  is constant and if  $\mathcal{S}_{r_0}$  has finite VC-dimension.

**Continuum degree:  $r > 0$  fixed** The continuous analog to the degree is naturally obtained by replacing quantities that depend on  $\mathcal{X}_n$  by their large-sample limit, after being properly normalized. As we consider  $r$ -neighborhood geometric graphs, the degree of  $x$  hence transforms into the convoluted density

$$f_r(x) := \frac{1}{|\mathcal{B}(x, r)|} \int_{\mathcal{B}(x, r)} f(z) dz. \quad (6.4)$$

More formally, we have the following well-known asymptotic behavior.

**Theorem 6.2.2.** *If  $r > 0$  is fixed, then almost surely,*

$$\frac{1}{N} \deg_r(x, \mathcal{X}_n) \xrightarrow[n \rightarrow \infty]{} f_r(x) \text{ uniformly in } x \in \mathbb{R}^d.$$

*Proof.* This comes from a direct application of Lemma 6.2.1 to  $\mathcal{S}_r = \{\mathcal{B}(x, r) \mid x \in \mathbb{R}^d\}$ .  $\square$

*Remark 6.2.3.* We recover that for a neighborhood graph, the counterpart of the degree is the convoluted density  $f_r$ . This object satisfies some of the desirable properties for a depth function identified by [ZS00]. It clearly satisfies P1 [‘affine invariance’] for rigid transformations and also P4 [‘vanishing at infinity’] due to  $f_r$  being a uniformly continuous density. It also satisfies P2 [‘maximality at the center’] and P3 [‘monotonicity relative to deepest point’], at least when the density is unimodal and rotationally invariant with respect to its mode — the latter being a direct consequence of [And55, Thm 1].

**Continuum degree:  $r \rightarrow 0$**  Now letting  $r = r_n$  go to zero slowly enough naturally leads us to recover the actual density.

**Theorem 6.2.4.** *If  $r = r_n$  is such that  $r \rightarrow 0$  and  $nr^d \gg \log n$ , then almost surely,*

$$\frac{1}{N} \deg_r(x, \mathcal{X}_n) \xrightarrow[n \rightarrow \infty]{} f(x) \text{ uniformly in } x \in \mathbb{R}^d.$$

Thus, as a measure of depth, the degree is asymptotically (as the connectivity radius tends to zero slowly enough) equivalent to the likelihood depth of [FM99].

*Proof.* This comes from a simple application of Lemma 6.2.1 to the collection of sets  $\{\mathcal{S}_r\}_{r>0}$  with  $\mathcal{S}_r := \{B(x, r) \mid x \in \mathbb{R}^d\}$ , and of the fact that  $f_r$  converges uniformly to  $f$  since  $f$  is assumed to be uniformly continuous on  $\mathbb{R}^d$ .  $\square$

**Remark 6.2.5 (Kernel Density Estimator).** Defining the kernel density estimator as

$$\hat{f}(x) = \frac{1}{N} \# \{X_i \in B(x, r)\} = \frac{1}{N} \deg_r(x, \mathcal{X}_n), \quad (6.5)$$

Theorem 6.2.4 simply restates the well-known fact that this estimator is uniformly consistent<sup>2</sup> over  $\mathbb{R}^d$  when  $r \rightarrow 0$  slowly enough that  $nr^d \rightarrow \infty$ .

**Remark 6.2.6 (Eigenvector Centrality).** Among spectral notions of centrality, *PageRank* is particularly famous for being at the origin of the Google search engine [PBMW99]. This notion of centrality was first suggested for measuring the ‘importance’ of webpages in the World Wide Web, seen as an oriented graph with nodes representing pages (URLs specifically) and a directed edge from page  $i$  to page  $j$  representing a hyperlink on page  $i$  pointing to page  $j$ . For an undirected graph, like the random geometric graphs that concern us here, the method amounts to using the stationary distribution of the random walk on the graph as a measure of node centrality. This is the walk where, at a given node, we choose one of its neighbor uniformly at random. (The edge weights play no role.) However, it is well-known that the stationary distribution is proportional to the vector of degrees, so that in this particular case, PageRank as a measure of centrality is equivalent to the degree. (Again, this is not true in general for directed graphs.)

## 6.3 H-Index

### 6.3.1 Definition of the H-Index

The *H-index* is named after [Hir05], who introduced this centrality measure in the context of citation networks of scientific publications. For a given node in a graph, it is defined as the maximum integer  $h$  such that the node has at least  $h$  neighbors with degree at least  $h$ . That is, in our context, the H-index of  $x$  in  $\mathcal{G}_r(x, \mathcal{X}_n)$  writes as

$$H_r(x, \mathcal{X}_n) := \text{largest } h \text{ such that } \# \{X_i \in B(x, r) : \deg_r(X_i, \mathcal{X}_n) \geq h\} \geq h.$$

The H-index was put forth as an improvement on the total number of citations as a measure of productivity, which in a citation graph corresponds to the degree. We show below that in the latent random geometric graph model of (6.1), the H-index can be asymptotically equivalent to the degree<sup>3</sup> (see Theorems 6.3.5 and 6.3.7).

<sup>2</sup>Recall that throughout, as defined in (6.1),  $f$  is assumed to be uniformly continuous over  $\mathbb{R}^d$ .

<sup>3</sup>Of course, there is no reason why the underlying geometry of a citation graph ought to be Euclidean.

### 6.3.2 Iterated H-Index

[LZZS16] consider iterates of the mechanism that defines the H-indices as a function of the degrees: The second iterate at a given node is the maximum  $h$  such that the node has at least  $h$  neighbors with H-index at least  $h$ , and so on. More generally, given any (possibly random) bounded measurable function  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ , we define the (random) bounded measurable function  $H_{n,r}\phi : \mathbb{R}^d \rightarrow \mathbb{R}$  as

$$\begin{aligned} H_{n,r}\phi(x) &:= \text{largest } h \text{ such that } \# \{X_i \in B(x, r) : \phi(X_i) \geq h\} \geq h \\ &= N \max \left\{ h \left| \frac{1}{N} \sum_{i=1}^n \mathbb{1}_{\|x-X_i\| \leq r} \mathbb{1}_{\phi(X_i)/N \geq h} \geq h \right. \right\}. \end{aligned} \quad (6.6)$$

The H-index  $H_r(x, \mathcal{X}_n)$  can be simply written  $H_{n,r}\deg_r(x, \mathcal{X}_n)$ , where  $\deg_r(x, \mathcal{X}_n)$  was defined in the previous section. The successive iterations of the H-index  $H_r^k(x, \mathcal{X}_n)$  are simply  $H_{n,r}^k \deg_r(x, \mathcal{X}_n)$ .

Given the variational formula (6.6), a natural continuous equivalent of the H-index is the  $H_r$  transform of the density  $f$ , where  $H_r$  is defined for any non-negative bounded measurable function  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$  as

$$H_r\phi(x) = \sup \left\{ t \geq 0 \left| \frac{1}{\omega r^d} \int_{B(x,r)} \mathbb{1}_{\phi(z) \geq t} f(z) dz \geq t \right. \right\}. \quad (6.7)$$

See Figure 6.3.1 for an illustration of this transform. The  $k$ -th iteration of  $H_r$  applied to  $\phi$  is simply denoted by  $H_r^k\phi$ . The  $H_r$  transform enjoys a few elementary properties, such as monotonicity, Lipschitzness and modulus of continuity preservation, which we now present. We recall that the modulus of continuity of a function  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  is defined by  $\omega_g(u) := \sup \{|g(x) - g(y)| : \|x - y\| \leq u\}$ , for all  $u \geq 0$ . As in our framework (see (6.1)),  $f$  is assumed to be uniformly continuous,  $\lim_{u \rightarrow 0} \omega_f(u) = 0$ . In what follows,  $\ell^\infty(\mathbb{R}^d)$  denotes the class of bounded measurable maps  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ .

**Lemma 6.3.1.**  *$H_r$  is monotonous, meaning that for any two functions  $\phi, \psi \in \ell^\infty(\mathbb{R}^d)$  such that  $\phi \leq \psi$ , we have  $H_r\phi \leq H_r\psi$ .*

*Proof.* This result is trivial once noted that the functional

$$\phi \mapsto \frac{1}{\omega r^d} \int_{B(x,r)} \mathbb{1}_{\phi(z) \geq t} f(z) dz$$

that appears in the definition of  $H_r$  is non-decreasing in  $\phi$ . □

**Lemma 6.3.2.**  *$H_r$  is 1-Lipschitz, meaning that for two functions  $\phi, \psi \in \ell^\infty(\mathbb{R}^d)$  we have  $\|H_r\phi - H_r\psi\|_\infty \leq \|\phi - \psi\|_\infty$ .*

*Proof.* Let  $\varepsilon = \|\phi - \psi\|_\infty$ . We have

$$\frac{1}{\omega r^d} \int_{B(x,r)} \mathbb{1}_{\phi(z) \geq t} f(z) dz \leq \frac{1}{\omega r^d} \int_{B(x,r)} \mathbb{1}_{\psi(z) \geq t - \varepsilon} f(z) dz$$

so that  $H_r\phi(x) \leq H_r\psi(x) + \varepsilon$ , and the proof follows. □

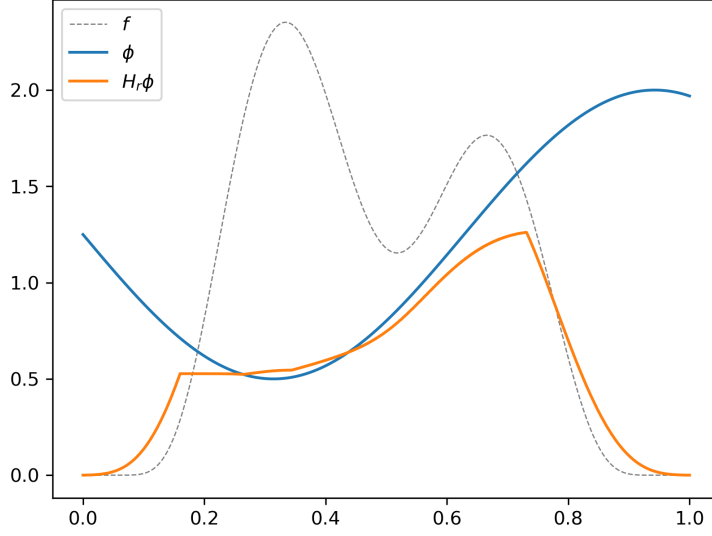


Figure 6.3.1 – A density  $f$ , a function  $\phi$ , and its transform  $H_r\phi$  for  $r = 0.1$ . Both  $f$  and  $\phi$  are smooth.  $H_r\phi$  does not appear to be continuously differentiable everywhere but is nonetheless Lipschitz, with Lipschitz constant no bigger than that of  $f$  and  $\phi$  (see Lemma 6.3.3).

**Lemma 6.3.3.** *If  $\phi \in \ell^\infty(\mathbb{R}^d)$  is uniformly continuous, then so is  $H_r\phi$  and we have  $\omega_{H_r\phi} \leq \omega_\phi \wedge \omega_f$ . In particular, since  $\omega_{f_r} \leq \omega_f$ , we have  $\omega_{H_r^k f_r} \leq \omega_f$  for all  $k \geq 1$ .*

*Proof.* Let  $x, y \in \mathbb{R}^d$ , and denote  $u = y - x$  and  $\varepsilon = \omega_f \vee \omega_\phi(\|x - y\|)$ . We have

$$\frac{1}{\omega r^d} \int_{B(x,r)} \mathbb{1}_{\phi(z) \geq t} f(z) dz = \frac{1}{\omega r^d} \int_{B(y,r)} \mathbb{1}_{\phi(z+u) \geq t} f(z+u) dz \leq \int_{B(y,r)} \mathbb{1}_{\phi(z) \geq t-\varepsilon} f(z) dz + \varepsilon$$

so that we immediately find that  $H_r\phi(x) \leq H_r\phi(y) + \varepsilon$ , and the proof follows.  $\square$

**Continuum H-indices:  $r > 0$  fixed** As intuited above, we have the following general convergence result of the random discrete transform  $H_{n,r}$  towards the continuum one  $H_r$ .

**Lemma 6.3.4.** *Let  $\phi_n, \phi \in \ell^\infty(\mathbb{R}^d)$  be random variables such that almost surely,  $\frac{1}{N}\phi_n \xrightarrow[n \rightarrow \infty]{} \phi$  uniformly. Then almost surely,  $\frac{1}{N}H_{n,r}\phi_n \xrightarrow[n \rightarrow \infty]{} H_r\phi$  uniformly.*

*Proof.* Notice that

$$\begin{aligned} H_{n,r}\phi_n(x) &= \sup \{h \geq 0 \mid \text{Card} \{X_i \in B(x,r), \phi(X_i) \geq h\} \geq h\} \\ &= N \times \sup \{t \geq 0 \mid \text{Card} \{X_i \in B(x,r), \phi(X_i) \geq Nt\} \geq Nt\} \\ &= N \times \sup \left\{ t \geq 0 \mid \frac{1}{N} \sum_{i=1}^n \mathbb{1}_{X_i \in B(x,r)} \mathbb{1}_{\frac{1}{N}\phi_n(X_i) \geq t} \geq t \right\}. \end{aligned}$$

Let  $\varepsilon = \|\frac{1}{N}\phi_n - \phi\|_\infty$ . We have

$$\frac{1}{N} \sum_{i=1}^n \mathbb{1}_{X_i \in B(x,r)} \mathbb{1}_{\frac{1}{N}\phi_n(X_i) \geq t} \leq \frac{1}{N} \sum_{i=1}^n \mathbb{1}_{X_i \in B(x,r)} \mathbb{1}_{\phi(X_i) \geq t-\varepsilon}.$$

Note that the class of balls of  $\mathbb{R}^d$  is a VC-class, and so is the set of super-level sets of  $\phi$ . As a result, the class

$$\mathcal{S}_r = \{B(y, r) \cap \{\phi \geq s\}, y \in \mathbb{R}^d, s \geq 0\}$$

thus satisfies the assumptions of Lemma 6.2.1. Furthermore, using notation  $\eta$  from Lemma 6.2.1, we get

$$\frac{1}{N} \sum_{i=1}^n \mathbb{1}_{X_i \in B(x,r)} \mathbb{1}_{\phi(X_i) \geq t-\varepsilon} \leq \frac{1}{\omega r^d} \int_{B(x,r)} \mathbb{1}_{\phi(z) \geq t-\varepsilon} f(z) dz + \eta$$

uniformly in  $x$  and  $t$ . We thus have

$$\frac{1}{N} H_{n,r} \phi_n(x) \leq \sup \left\{ t \geq 0 \mid \frac{1}{\omega r^d} \int_{B(x,r)} \mathbb{1}_{\phi(z) \geq t-\varepsilon} f(z) dz \geq t - \eta \right\},$$

yielding  $\frac{1}{N} H_{n,r} \phi_n(x) \leq H_r \phi(x) + \varepsilon \vee \eta$ . The lower bound can be obtained in the same fashion. We conclude by letting  $n \rightarrow \infty$ , so that  $\eta$  goes to 0 a.s. (Lemma 6.2.1) and  $\varepsilon$  as well by assumption.  $\square$

When applied iteratively to the sequence of degree functions of  $\mathcal{G}_r(x, \mathcal{X}_n)$ , Lemma 6.3.4 yields the following result.

**Theorem 6.3.5.** *If  $r > 0$  and  $k \in \mathbb{N}^*$  are fixed, then almost surely,*

$$\frac{1}{N} H_r^k(x, \mathcal{X}_n) \xrightarrow{n \rightarrow \infty} H_r^k f_r(x) \text{ uniformly in } x \in \mathbb{R}^d.$$

*Proof.* Apply Lemma 6.3.4 recursively to find that  $\frac{1}{N} H_{n,r}^k \phi_n \rightarrow H_r^k \phi$  for all  $k \geq 1$ . The stated result follows readily starting from  $\phi_n = \deg_r(\cdot, \mathcal{X}_n)$  and  $\phi = f_r$ .  $\square$

We note that  $H_r^k f_r$  satisfies Properties P1 for rigid transformations and P4 of [ZS00], and also P2 and P3 under the same conditions as in Remark 6.2.3. That being said, the iterated continuum H-indices  $H_r^k f_r$  behave very differently from the likelihood depth, as shown in Figure 6.4.1. Note also that for  $k \geq 1$ ,  $H_r^k f_r(x)$  depends on  $f$  in an even less local way than  $f_r$ , since it depends on the values of  $f$  on  $B(x, (k+1)r)$ .

**Continuum H-indices:**  $r \rightarrow 0$  To gain insights on what the discrete H-indices converge to as  $r = r_n \rightarrow 0$ , let us first examine how their fixed- $r$  continuous counterparts  $H_r^k f_r$  behave in the same regime.

**Proposition 6.3.6.** *For all  $k \geq 1$ ,  $H_r^k f_r(x) \xrightarrow{r \rightarrow 0} f(x)$  uniformly in  $x \in \mathbb{R}^d$ .*

*Proof.* On one hand, we have  $H_r^k f_r(x) \leq f_r(x) \leq f(x) + \omega_f(r)$ . On the other hand, we get from the definition of  $H_r f_r$  that  $H_r f_r \geq f - \omega_f(r)$ . Using this bound recursively together with Lemma 6.3.3, we find that  $H_r^k f_r \geq f - k\omega_f(r)$ . At the end of the day, we have proven that  $\|H_r^k f_r - f\|_\infty \leq k\omega_f(r)$ , which concludes the proof.  $\square$

Coming back to the discrete H-indices, we naturally get that the  $k$ -th iteration of the H-index converges to  $f(x)$  as  $r = r_n$  converges to 0 slowly enough, thus coinciding with the likelihood depth.

**Theorem 6.3.7.** *If  $r = r_n$  is such that  $r \rightarrow 0$  and  $nr^d \gg \log n$ , then for all  $k \in \mathbb{N}$ , almost surely,*

$$\frac{1}{N} H_r^k(x, \mathcal{X}_n) \xrightarrow{n \rightarrow \infty} f(x) \text{ uniformly in } x \in \mathbb{R}^d.$$

Hence, as for the degree (Section 6.2), we see that the iterated H-indices are asymptotically equivalent to the likelihood depth when  $r \rightarrow 0$  slowly enough.

*Proof.* First, decompose

$$\left| \frac{1}{N} H_r^k(x, \mathcal{X}_n) - f(x) \right| \leq \left| \frac{1}{N} H_r^k(x, \mathcal{X}_n) - H_r^k f_r(x) \right| + |H_r^k f_r(x) - f(x)|.$$

Proposition 6.3.6 asserts that the second (deterministic) term converges uniformly to zero as  $r \rightarrow 0$ . For the first (stochastic) one, we use expressions (6.6) and (6.7) of  $H_{n,r}$  and  $H_r$  respectively, and the proof of Theorem 6.3.5, to get that

$$\left| \frac{1}{N} H_r^k(x, \mathcal{X}_n) - H_r^k f_r(x) \right| \leq \eta := \sup_{S \in \mathcal{S}_r} \frac{1}{\omega r^d} |P_n(S) - P(S)|,$$

where  $P_n(dz) = n^{-1} \sum_{i=1}^n \delta_{X_i}(dz)$ ,  $P(dz) = f(z)dz$ , and

$$\mathcal{S}_r = \{B(y, r) \cap \{\phi \geq s\} \mid y \in \mathbb{R}^d, s \geq 0, \phi \in \{f_r, \dots, H_r^k f_r\}\}.$$

As an intersection class of two VC classes,  $\mathcal{S}_r$  is also VC, with dimension uniformly bounded in  $r$ . It is composed of sets of radii at most  $r$ , so that Lemma 6.2.1 applies and yields  $\eta \rightarrow 0$  almost surely as  $n \rightarrow \infty$ .  $\square$

## 6.4 Coreness

The notion of *coreness* is based on the concept of *core* as introduced by [Sei83]. (Seidman does not mention ‘coreness’ and only introduces cores, and we are uncertain as to the origin of the coreness.) For an integer  $\ell \geq 0$ , an  $\ell$ -core of a given graph is a maximal induced subgraph which has minimum degree  $\ell$ . To be sure, this means that any node in an  $\ell$ -core is neighbor to at least  $\ell$  nodes in that core. In a given graph, the coreness of a node is the largest integer  $\ell$  such that the node belongs to an  $\ell$ -core. For a recent paper focusing on the computation of the  $\ell$ -cores, see [MGPV20].

The coreness is closely related to the degree and H-index. In fact, [LZZS16, Thm 1] shows that it arises when iterating the definition of the H-index ad infinitum, when starting with the degree function. That is, in our context, we will study the random coreness

$$C_r(x, \mathcal{X}_n) := H_r^\infty(x, \mathcal{X}_n). \quad (6.8)$$

In particular, the coreness satisfies the following fixed-point property: The coreness of node  $i$  is the maximum  $\ell$  such that at least  $\ell$  of its neighbors have coreness at least  $\ell$ . Said otherwise, it is the maximal minimal degree of a subgraph  $H$  that contains  $x$ :

$$C_r(x, \mathcal{X}_n) = \max \left\{ \ell \mid \text{there is a subgraph } H \text{ of } \mathcal{G}_r(x, \mathcal{X}_n) \text{ with } x \in H \text{ and } \min_{i \in H} \deg_H(i) \geq \ell \right\}. \quad (6.9)$$

The coreness was analyzed in the context of an Erdős–Rényi–Gilbert random graph in a number of papers, for example, in [Luc91, JL08, PSW96, Rio08, JL07], and also in the context of other graph models, for example, in [FKRD09]. We are not aware of any work that analyzes the coreness in the context of a random geometric graph.

*Remark 6.4.1.* As the non-negative integer sequence  $(H_r^k(x, \mathcal{X}_n))_{k \geq 0}$  is non-increasing, it becomes stationary after some index  $k^\infty < \infty$ . Said otherwise, the naive algorithm computing  $H_r^\infty(x, \mathcal{X}_n)$  by iterating the H-index terminates after a finite number of iterations, so that bounding  $k^\infty$  is of particular computational interest. Such a bound, depending on the geometric structure of the graph, is discussed in Section 6.5.3.

**Continuum coreness:  $r > 0$  fixed** As defined above in (6.8), the discrete coreness is obtained by applying the H-index operator to the degree infinitely many times. Having in mind Theorem 6.3.5, we naturally define the notion of continuum  $r$ -coreness by taking the limit of the iterated continuum H-index  $H_r^k f_r(x)$  as the number of iteration  $k$  goes to  $\infty$ .

**Proposition 6.4.2.**  $H_r^k f_r(x)$  converges uniformly in  $x$  as  $k \rightarrow \infty$ . Its limit, denoted by  $C_r(x, f)$ , is called the continuum  $r$ -coreness at  $x$ .

*Remark 6.4.3.* Note that since the convergence is uniform,  $C_r(\cdot, f)$  is uniformly continuous and its modulus of continuity is bounded from above by  $\omega_f$  (Lemma 6.3.3). See Figure 6.4.1 for an illustration of the convergence of the iterations  $H_r^k f_r$  towards  $C_r(\cdot, f)$ .

*Proof.* Since for all  $t \geq 0$  and  $x \in \mathbb{R}^d$ ,

$$\frac{1}{\omega r^d} \int_{B(x, r)} \mathbb{1}_{f_r(z) \geq t} f(z) dz \leq f_r(x),$$



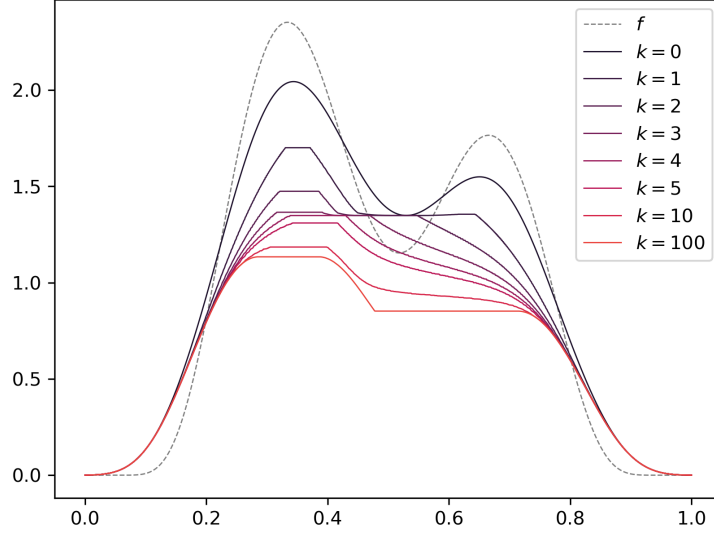


Figure 6.4.1 – The successive iterations of  $H_r^k f_r$  (solid) for a given density  $f$  (dashed), for  $k$  ranging from 0 to 100 with  $r = 0.1$ . The hundredth iteration is very close to its limit  $C_r(x, f)$ .

so that  $H_r f_r \leq f_r$ . Using monotonicity of the operator  $H_r$  (Lemma 6.3.1) we find that  $(H_r^k f_r)_{k \in \mathbb{N}}$  is a non-increasing sequence of functions, bounded from above by  $f_r$  and from below by 0. In particular, it converges towards a function  $C_r(\cdot, f)$  pointwise. Since  $|f_r(x)| \leq \sup_{B(x,r)} |f|$  and that the latter goes to 0 when  $x$  goes to  $\infty$  (since  $f$  is integrable and uniformly continuous over  $\mathbb{R}^d$ ), we can focus on establishing the uniform convergence of  $H_r^k f_r$  on a ball  $B(0, R)$  for an arbitrary large radius  $R$ . Having done so, the sequence  $H_r^k f_r$  is equicontinuous (from Lemma 6.3.3), and the Arzelà–Ascoli theorem insures that the convergence towards  $C_r(\cdot, f)$  is uniform over  $B(0, R)$ .  $\square$

By analogy with (6.9), we may also seek a variational characterization of  $C_r(x, f)$  in terms of subsets of  $\mathbb{R}^d$ , which are the natural continuous counterparts of subgraphs. This formulation, besides offering additional geometrical insights, will help with proving convergence from discrete to continuous  $r$ -coreness (see the proof of Theorem 6.4.5).

**Lemma 6.4.4.** *Let  $\Omega(x)$  be the class of measurable sets  $S \subset \mathbb{R}^d$  that contain  $x$ . Then for  $r > 0$ , the continuum  $r$ -coreness admits the following expression*

$$C_r(x, f) = \sup \left\{ t \mid \exists S \in \Omega(x) \text{ such that } \inf_{y \in S} \frac{1}{\omega r^d} \int_{B(y,r) \cap S} f(z) dz \geq t \right\}. \quad (6.10)$$

*Proof.* Let us write  $F(x)$  for the supremum on the right-hand side, and show that  $C_r(\cdot, f) = F$  by considering their super-level sets. Let  $t \geq 0$ , and  $S = \{F \geq t\}$ . For all  $y \in \mathbb{R}^d$ , we define

$$g(y) := \frac{1}{\omega r^d} \int_{S \cap B(y, r)} f(z) dz,$$

which, by definition of  $S$ , satisfies  $g(y) \geq t$  for all  $y \in S$ . In particular, we get that for all  $y \in S$ ,

$$\frac{1}{\omega r^d} \int_{B(y, r)} \mathbb{1}_{g(z) \geq t} f(z) dz \geq \frac{1}{\omega r^d} \int_{B(y, r)} \mathbb{1}_{z \in S} f(z) dz = g(y) \geq t,$$

so that  $H_r g(y) \geq t$ . By induction on  $k \geq 1$ , we find that  $H_r^k g(y) \geq t$  for all  $y \in S$ , and letting  $k \rightarrow \infty$ , that  $C_r(y, f) \geq t$  for all  $y \in S$ , so that  $S \subset \{C_r(\cdot, f) \geq t\}$ .

For the converse inclusion, notice that since the operator  $H_r$  is 1-Lipschitz (thanks to Lemma 6.3.2) and that  $H_r^k f_r$  converges uniformly towards  $C_r(\cdot, f)$  (Proposition 6.4.2), we have that  $H_r C_r(\cdot, f) = C_r(\cdot, f)$ . Therefore, if  $y \in \{C_r(\cdot, f) \geq t\}$ , meaning  $C_r(y, f) \geq t$ , by definition of  $H_r$ , we get

$$\frac{1}{\omega r^d} \int_{B(y, r)} \mathbb{1}_{C_r(z, f) \geq t} f(z) dz \geq t$$

yielding, by maximality of  $S$ , that  $\{C_r(\cdot, f) \geq t\} \subset S$ , ending the proof.  $\square$

By definition, the continuum  $r$ -coreness  $C_r(\cdot, f)$  behaves roughly like  $H_r^k f_r$  for  $k$  large enough, as shown in Figure 6.4.1. The variational formulation of Lemma 6.4.4 also highlights the fact that  $C_r(\cdot, f)$  depends on  $f$  globally, as it depends on values it takes in the entire space, at least in principle. That is, perturbing  $f$  very far away from  $x$  may change  $C_r(x, f)$  drastically. In Figure 6.4.1, this phenomenon translates into the wider and wider plateaus that  $H_r^k(\cdot, f)$  exhibits as  $k$  grows, which eventually approaches  $C_r(\cdot, f)$ . It is no surprise, then, that  $C_r(\cdot, f)$  satisfies Properties P1 for rigid transformations and P4 of [ZS00], and also P2 and P3 under the same conditions as in Remark 6.2.3 — just like  $H_r^k f_r$ .

We are now in position to prove the convergence of the renormalized discrete coreness towards the  $r$ -continuum coreness, for a bandwidth parameter  $r > 0$  being fixed.

**Theorem 6.4.5.** *If  $r > 0$  is fixed, then almost surely,*

$$\frac{1}{N} C_r(x, \mathcal{X}_n) \xrightarrow[n \rightarrow \infty]{} C_r(x, f) \text{ uniformly in } x \in \mathbb{R}^d. \quad (6.11)$$

*Proof.* Let  $k \geq 1$ . By the decreasingness of the iterations of the H-index  $H_r^k(x, \mathcal{X}_n)$  and their convergence towards  $C_r(x, \mathcal{X}_n)$  [LZZS16, Thm 1], we have that  $C_r(x, \mathcal{X}_n) \leq H_r^k(x, \mathcal{X}_n)$ . Taking  $n$  to  $\infty$  and using Theorem 6.3.5, we find that almost surely,

$$\limsup_{n \rightarrow \infty} \frac{1}{N} C_r(x, \mathcal{X}_n) \leq H_r^k f_r(x)$$

uniformly in  $x$ , so that letting  $k \rightarrow \infty$  and using Proposition 6.4.2, we have

$$\limsup_{n \rightarrow \infty} \frac{1}{N} C_r(x, \mathcal{X}_n) \leq C_r(x, f).$$

For the converse inequality, we will use the variational formulation of  $C_r(x, f)$  given by Lemma 6.4.4. Let  $t < C_r(x, f)$  and  $S \subset \mathbb{R}^d$  be such that  $x \in S$  and

$$\frac{1}{\omega r^d} \int_{B(y, r) \cap S} f(z) dz \geq t \quad \forall y \in S.$$

Let  $H$  denote the subgraph of  $\mathcal{G}_r(x, \mathcal{X}_n)$  with vertices in  $S$ , and  $\deg_H$  the degree of the vertices in this subgraph. We have, for all vertex  $s$  in  $S$ ,

$$\deg_H(s) = n \times P_n(B(s, r) \cap S) - 1 \geq N \times (P(B(s, r) \cap S) - \eta) - 1 \geq N \times (t - \eta) - 1,$$

where

$$\eta := \sup_{A \in \mathcal{S}_r} \frac{1}{\omega r^d} |P_n(A) - P(A)|, \quad \text{with } \mathcal{S}_r := \{S \cap B(y, r) \mid y \in \mathbb{R}^d\},$$

so that  $C_r(x, \mathcal{X}_n) \geq N(t - \eta) - 1$ . The class  $\mathcal{S}_r$  satisfies the assumptions of Lemma 6.2.1, and applying that lemma with  $r > 0$  fixed yields that, almost surely,

$$\liminf_{n \rightarrow \infty} \frac{1}{N} C_r(x, \mathcal{X}_n) \geq t$$

uniformly in  $x \in \mathbb{R}^d$ . Letting  $t \nearrow C_r(x, f)$  establishes

$$\liminf_{n \rightarrow \infty} \frac{1}{N} C_r(x, \mathcal{X}_n) \geq C_r(x, f),$$

which concludes the proof.  $\square$

**Continuum coreness:**  $r \rightarrow 0$  Seeking to complete the construction above to include asymptotic regimes where  $r \rightarrow 0$ , we first opt for a purely functional approach. That is, taking the limit of the continuum  $r$ -coreness as  $r$  goes to zero.

**Proposition 6.4.6.**  $C_r(x, f)$  converges uniformly in  $x \in \mathbb{R}^d$  as  $r \rightarrow 0$ . Its limit, denoted by  $C_0(x, f)$ , is called the continuum coreness at  $x$ .

The proof of this result relies on an intermediary notion of coreness at scale  $\alpha > 0$ . Given  $K \subset \mathbb{R}^d$  and  $y \in \mathbb{R}^d$ , we write  $d(y, K) := \inf_{z \in K} \|y - z\|$  for the distance from  $y$  to  $K$ . We let  $\mathcal{B}_\alpha := \{K^\alpha \mid K \subset \mathbb{R}^d\}$ , where  $K^\alpha := \{y \in \mathbb{R}^d \mid d(y, K) \leq \alpha\}$  and define

$$C^\alpha(x, f) := \sup \{t \geq 0 \mid \exists S \in \mathcal{B}_\alpha \text{ with } x \in S, S \subset \{f \geq t\} \text{ and } \partial S \subset \{f \geq 2t\}\}.$$

Since  $(K^\alpha)^\beta = K^{\alpha+\beta}$  for all  $\alpha, \beta \geq 0$ , the class  $\mathcal{B}_\alpha$  is increasing as  $\alpha \rightarrow 0^+$ , so is  $C^\alpha(x, f)$ , and since the latter is bounded from above by  $\|f\|_\infty$ , it converges to a finite limit. The following lemma asserts that this limit actually coincides with the limit of  $C_r(x, f)$  as  $r \rightarrow 0^+$ .

**Lemma 6.4.7.** *We have  $\lim_{r \rightarrow 0} C_r(x, f) = \lim_{\alpha \rightarrow 0} C^\alpha(x, f)$ .*

This result thus asserts the existence of  $C_0(x, f)$  pointwise, as used in the proof of Proposition 6.4.6. To show Lemma 6.4.7, we first need the following volume estimate.

**Lemma 6.4.8.** *For all  $r \in (0, \alpha/2]$ ,  $x \in \mathbb{R}^d$  and  $y \in B(x, \alpha)$ , we have*

$$|B(y, r) \cap B(x, \alpha)| \geq \frac{1}{2} \omega r^d \left(1 - \frac{Cr}{\alpha}\right),$$

where  $C$  is a positive constant depending on  $d$  only.

*Proof.* The quantity  $|B(y, r) \cap B(x, \alpha)|$  is a decreasing function of  $\|y - x\|$ , so we can only consider the case where  $\|x - y\| = \alpha$ . Let now

$$\rho = r \left(1 - \frac{r}{2\alpha}\right) \quad \text{and} \quad x_0 = x + (\alpha - r + \rho)(y - x).$$

Easy computations show that the half ball

$$B^+ = B(x_0, \rho) \cap \{z \in \mathbb{R}^d \mid \langle z - x_0, x - x_0 \rangle \geq 0\}$$

is a subset of  $B(y, r) \cap B(x, \alpha)$  so that

$$|B(y, r) \cap B(x, \alpha)| \geq |B^+| = \frac{1}{2} \omega r^d (1 - r/2\alpha)^d \geq \frac{1}{2} \omega r^d (1 - Cr/\alpha),$$

with  $C = d/2$ . □

*Proof of Lemma 6.4.7.* Let  $0 < r \leq \alpha$  and let  $t = C^\alpha(x, f)$ . Let  $K \subset \mathbb{R}^d$  be such that  $K^\alpha \subset \{f \geq t - \varepsilon\}$  and  $\partial K^\alpha \subset \{f \geq 2t - 2\varepsilon\}$  for some arbitrarily small  $\varepsilon > 0$ . For all  $y \in K^\alpha$  at distance at least  $r$  from  $\partial K^\alpha$ , we have  $B(y, r) \subset K^\alpha$ , so that

$$\frac{1}{\omega r^d} \int \mathbb{1}_{z \in K^\alpha} \mathbb{1}_{z \in B(y, r)} f(z) dz = \frac{1}{\omega r^d} \int_{B(y, r)} f(z) dz \geq t - \varepsilon - \omega_f(r),$$

where we recall that  $\omega_f$  denotes the modulus of continuity of  $f$ . Otherwise if  $d(y, \partial K^\alpha) \leq r$ , we have for any  $v \in B(y, r)$  that  $f(v) \geq 2t - 2\varepsilon - \omega_f(2r)$ . We then have, thanks to Lemma 6.4.8,

$$|B(y, r) \cap K^\alpha| \geq |B(y, r) \cap B(z_0, \alpha)| \geq \omega r^d (1/2 - O(r/\alpha)),$$

where  $z_0 \in K$  is such that  $y \in B(z_0, \alpha)$ . We hence deduce that

$$\frac{1}{\omega r^d} \int \mathbb{1}_{K^\alpha} \mathbb{1}_{B(y, r)} f \geq t - \varepsilon - O(r/\alpha) - \omega_f(2r),$$

so that  $C_r(x, f) \geq t - \varepsilon - O(r/\alpha) - \omega_f(2r)$ . Taking  $r \rightarrow 0$  and  $\varepsilon \rightarrow 0$ , we obtain

$$\liminf_r C_r(x, f) \geq C^\alpha(x, f),$$

for any  $\alpha > 0$ .

Conversely, let  $S$  be a set containing  $x$  such that

$$\forall y \in S, \frac{1}{\omega r^d} \int \mathbb{1}_{z \in S} \mathbb{1}_{z \in B(y, r)} f(z) dz \geq t.$$

In particular, we have for any  $y \in S$ ,  $f(y) \geq t - \omega_f(r)$ , so that for any  $y \in S^\alpha$ , we have  $f(y) \geq t - \omega_f(r) - \omega_f(\alpha)$ . Let now take  $y \in \partial S^\alpha$ , and let  $z_0 \in S$  be a point at distance at most  $\alpha$  from  $y$ . We have

$$\begin{aligned} f(y) &\geq f(z_0) - \omega_f(\alpha) \geq \frac{1}{|S \cap B(z_0, r)|} \int \mathbb{1}_{z \in S} \mathbb{1}_{z \in B(z_0, r)} f(z) dz - \omega_f(\alpha) - \omega_f(r) \\ &\geq \frac{\omega r^d}{|S \cap B(z_0, r)|} t - \omega_f(\alpha) - \omega_f(r). \end{aligned}$$

But now, Lemma 6.4.8 again yields

$$|S \cap B(z_0, r)| \leq |B(z_0, r) \setminus B(y, \alpha)| = \omega r^d - |B(z_0, r) \cap B(y, \alpha)| \leq \omega r^d (1/2 + O(r/\alpha)),$$

which gives

$$C^\alpha(x, f) \geq t - O(r/\alpha) - \omega_f(r) - \omega_f(\alpha)$$

and hence  $C^\alpha(x, f) + \omega_f(\alpha) \geq \limsup_r C_r(x, f)$ . We thus proved that

$$C^\alpha(x, f) \leq \liminf_r C_r(x, f) \leq \limsup_r C_r(x, f) \leq C^\alpha(x, f) + \omega_f(\alpha), \quad \forall \alpha > 0,$$

which allows to conclude.  $\square$

*Proof of Proposition 6.4.6.* From Lemma 6.4.7, we get that  $C_r(\cdot, f)$  converges pointwise towards a limit  $C_0(\cdot, f)$ . Since  $C_r(x, f) \leq f_r(x) \leq f(x) + \omega_f(r)$ , and since  $f \rightarrow 0$  at  $\infty$  (because  $f$  is integrable and is uniformly continuous), we can focus on the uniform convergence of  $C_r(\cdot, f)$  on a ball  $B(0, R)$  for some arbitrarily large  $R > 0$ . But now, the uniform convergence on  $B(0, R)$  is only a consequence of the Arzelà–Ascoli theorem and the equicontinuity of  $C_r(\cdot, f)$  (Remark 6.4.3).  $\square$

As was shown to be the case for  $C_r(\cdot, f)$  in Lemma 6.4.4, we also give a geometric variational formulation of  $C_0(\cdot, f)$ , which is illustrated in Figure 6.4.2.

**Lemma 6.4.9.** *Let  $\Sigma(x)$  be the class of open sets  $S \subset \mathbb{R}^d$  with smooth boundaries<sup>4</sup> that contain  $x$ . Then the continuum coreness admits the following expression*

$$C_0(x, f) = \sup \{t \geq 0 \mid \exists S \in \Sigma(x) \text{ such that } S \subset \{f \geq t\} \text{ and } \partial S \subset \{f \geq 2t\}\}.$$

<sup>4</sup>That is,  $\partial S$  is a disjoint union of smooth  $(d-1)$ -dimensional submanifolds of  $\mathbb{R}^d$ .

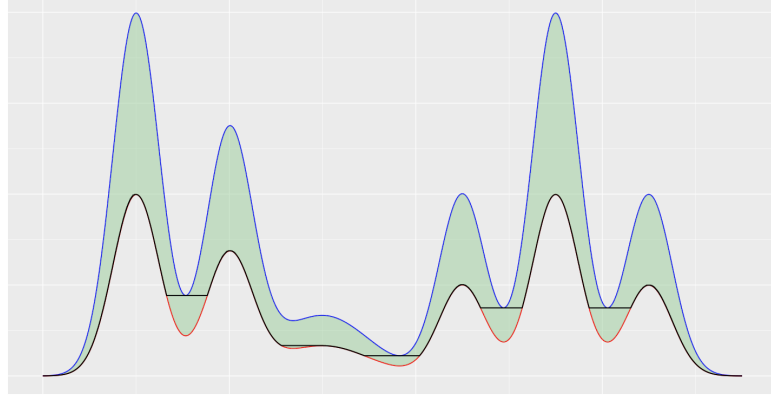


Figure 6.4.2 – An illustration of  $f$  (blue),  $f/2$  (red) and  $C_0(\cdot, f)$  (black) for a mixture of 6 Gaussians in dimension  $d = 1$ . In the zones where  $C_0(\cdot, f)$  does not coincide with  $f/2$ , it exhibits plateaus over intervals  $[x_{\min}, x_{\max}]$ . For  $x \in (x_{\min}, x_{\max})$ , the supremum of Lemma 6.4.9 is attained for  $S = (x_{\min}, x_{\max})$ . Otherwise, this supremum is asymptotically attained for  $S = \{x\}$ .

*Proof.* Write  $C^*$  for the supremum of the right hand side. We want to show that  $C^* = C_0(x, f)$ . For this, take  $t > 0$  such that there exists  $S$  containing  $x$ , with smooth boundary, and such that  $S \subset \{f \geq t\}$  and  $\partial S \subset \{f \geq 2t\}$ . Then, for any  $\alpha > 0$ ,  $S^\alpha$  satisfies

$$\forall y \in S^\alpha, f(y) \geq t - \omega_f(\alpha) \quad \text{and} \quad \forall y \in \partial S^\alpha, f(y) \geq 2t - \omega_f(\alpha).$$

As a result,  $C^\alpha(x, f) \geq t - \omega_f(\alpha)$  and thus, letting  $\alpha \rightarrow 0$ , we have  $C_0(x, f) \geq t$ , and thus  $C_0(x, f) \geq C^*$ .

Conversely, denote  $t = C_0(x, f)$  and let  $\varepsilon > 0$  and  $\alpha > 0$  such that  $C_\alpha(x, f) \geq t - \varepsilon$ . There exists  $K \subset \mathbb{R}^d$  containing  $x$  such that  $K^\alpha$  satisfies  $K^\alpha \subset \{f \geq t - 2\varepsilon\}$  and  $\partial K^\alpha \subset \{f \geq 2t - 4\varepsilon\}$ . For  $\delta > 0$ , let us define

$$\Psi_\delta(y) := \frac{1}{\delta^d} \int_{\mathbb{R}^d} \kappa\left(\frac{y-v}{\delta}\right) \mathbb{1}_{K^{\alpha+\delta}}(v) dv,$$

where  $\kappa$  is a smooth positive normalized kernel supported in  $B(0, 1)$ . The function  $\Psi_\delta : \mathbb{R}^d \rightarrow \mathbb{R}$  is a smooth function with values in  $[0, 1]$ , with  $\Psi_\delta = 1$  on  $K^\alpha$  and  $\Psi_\delta = 0$  outside of  $K^{\alpha+2\delta}$ . Using Sard's lemma, we can find a regular value of  $\Psi_\delta$  in  $[1/4, 3/4]$ , say  $\lambda$ . The set  $S = \{\Psi_\delta > \lambda\}$  is then an open set of  $\mathbb{R}^d$  with smooth boundary  $\partial S = \{\Psi_\delta = \lambda\}$ , which contains  $K$ , so in particular, it contains  $x$ . Furthermore, any point of  $S$  (resp.  $\partial S$ ) is at distance at most  $2\delta$  from  $K^\alpha$  (resp.  $\partial K^\alpha$ ). We thus have

$$\forall y \in S, f(y) \geq t - 2\varepsilon - \omega_f(2\delta) \quad \text{and} \quad \forall y \in \partial S^\alpha, f(y) \geq 2t - 4\varepsilon - \omega_f(2\delta),$$

so that  $C^* \geq t - 2\varepsilon - \omega_f(2\delta)$ . Letting  $\varepsilon, \delta \rightarrow 0$ , we find that  $C^* \geq C_0(x, f)$ , ending the proof.  $\square$

The above formulation clearly establishes that  $C_0(x, f) \leq f(x)$ . On the other hand, taking for  $S$  a ball centered around  $x$  with an arbitrary small radius, we find that  $C_0(x, f) \geq f(x)/2$ . The equality actually occurs whenever the homology of the super-level sets of  $f$  is simple enough, as shown in Proposition 6.4.10. In particular, this is the case when the super-level sets are contractible sets (such as star-shaped ones), or the union of contractible sets.

**Proposition 6.4.10.** *If all the super-level sets of  $f$  have a trivial  $(d - 1)$ -th homology group over  $\mathbb{Z}$ , then  $C_0(x, f) = f(x)/2$  for all  $x \in \mathbb{R}^d$ . This is the case, for example, if  $f$  is a mixture of symmetric unimodal densities with disjoint supports.*

The proof of this proposition relies on the following topological result.

**Lemma 6.4.11.** *Let  $X \subset \mathbb{R}^d$  be a compact subset with  $H_{d-1}(X; \mathbb{Z}) = \{0\}$ . Then  $\mathbb{R}^d \setminus X$  is path-connected.*

*Proof.* We introduce the Alexandrov compactification  $Y = \mathbb{R}^d \cup \{\infty\}$  of  $\mathbb{R}^d$ , which is homeomorphic to the sphere  $S^d$ . Using Alexander's duality theorem [Hat02, Cor 3.45 p.255], we find that  $\tilde{H}^0(Y \setminus X; \mathbb{Z}) = \tilde{H}_{d-1}(X; \mathbb{Z}) = H_{d-1}(X; \mathbb{Z}) = \{0\}$  where  $\tilde{H}_\bullet$  and  $\tilde{H}^\bullet$  denote respectively the reduced homology and cohomology groups. As pointed out in [Hat02, Paragraph 2, p.199], the group  $\tilde{H}^0(Y \setminus X; \mathbb{Z})$  is identified to the group of functions  $Y \setminus X \rightarrow \mathbb{Z}$  that are constant on the path-connected component of  $Y \setminus X$ , quotiented by the group of constant functions. We conclude that  $Y \setminus X$ , and hence  $\mathbb{R}^d \setminus X$  by boundedness of  $X$ , has only one path-connected component.  $\square$

*Proof of Proposition 6.4.10.* From the formulation of Lemma 6.4.9 applied with  $S$  ranging within open balls centered at  $x$  and radius  $\delta \rightarrow 0$ , we see that we always have  $C_0(x, f) \geq f(x)/2$ .

Conversely, if  $t < C_0(x, f)$ , there exists a smooth set  $S \subset \{f \geq t\}$  with  $\partial S \subset \{f \geq 2t\}$  that contains  $x$ . Assume for a moment that  $S \setminus \{f \geq 2t\}$  is non-empty, and take a point  $y$  in it. Since  $\{f \geq 2t\}$  is compact with a trivial  $(d - 1)$ -th homology group, we have that  $\mathbb{R}^d \setminus \{f \geq 2t\}$  is path-connected thanks to Lemma 6.4.11, so that there exists a continuous path from  $y$  to any point  $z \in \mathbb{R}^d \setminus S$  that stays in  $\mathbb{R}^d \setminus \{f \geq 2t\}$ . Such a path necessarily crosses  $\partial S \subset \{f \geq 2t\}$ , which is absurd. We hence conclude that  $S \subset \{f \geq 2t\}$ , so that  $f(x) \geq 2t$ , and taking  $t$  to  $C_0(f, x)$ , we find that  $C_0(f, x) \leq f(x)/2$ , which concludes the proof.  $\square$

Hence, for densities  $f$  with simple enough landscapes, the continuum coreness is, as a measure of depth, equivalent to the likelihood depth. Otherwise, generically,  $C_0(\cdot, f)$  provides us with a new notion of depth that lies between  $f/2$  and  $f$  (see Figure 6.4.2). As is the case for  $C_r(\cdot, f)$ , the continuum coreness  $C_0(\cdot, f)$  depends on the values  $f$  on the entire space, at least in principle. This is apparent in the variational

formulation of Lemma 6.4.4 and is clearly illustrated by the plateau areas of Figure 6.4.2. And just like  $C_r(\cdot, f)$  for  $r > 0$ ,  $C_0(\cdot, f)$  satisfies Properties P1 for rigid transformations and P4 of [ZS00], and also P2 and P3 under the same conditions as in Remark 6.2.3.

We finally address the large-sample limit of  $C_r(x, \mathcal{X}_n)$  as  $r = r_n \rightarrow 0$ , which does coincide with the continuum coreness  $C_0(x, f)$ .

**Theorem 6.4.12.** *If  $r = r_n$  is such that  $r \rightarrow 0$  and  $nr^d \gg \log n$ , then almost surely,*

$$\frac{1}{N} C_r(x, \mathcal{X}_n) \xrightarrow[n \rightarrow \infty]{} C_0(x, f) \text{ uniformly in } x \in \mathbb{R}^d. \quad (6.12)$$

The remaining results are directed towards the proof of Theorem 6.4.12, which follows directly from Lemma 6.4.13 and Lemma 6.4.14. The usual decomposition in term of variance and bias that we used for instance in the proof of Theorem 6.3.7 does not work here, because the deviation term would be indexed by a class of subsets that is too rich (and which would not satisfy the assumptions of Lemma 6.2.1). Instead, we take advantage of the alternative definition of the coreness through  $C^\alpha$  introduced in the beginning of this Section 6.4.

**Lemma 6.4.13.** *If  $r = r_n$  is such that  $r \rightarrow 0$  and  $nr^d \gg \log n$ , then almost surely,*

$$\limsup_{n \rightarrow \infty} \frac{1}{N} C_r(x, \mathcal{X}_n) \leq C_0(x, f) \text{ uniformly in } x \in \mathbb{R}^d.$$

*Proof.* For short, write  $c_n = C_r(x, \mathcal{X}_n)$ , and  $S_n$  for the vertices of a subgraph of  $\mathcal{G}_r(x, \mathcal{X}_n)$  containing  $x$  with minimal degree  $c_n$ . Let  $\alpha > 0$  and consider  $S_n^\alpha \in \mathcal{B}_\alpha$ . For any  $y \in S_n^\alpha$ , there exists  $s \in S_n$  such that  $\|s - y\| \leq \alpha$ . We deduce that

$$f(y) \geq f(s) - \omega_f(\alpha) \geq \frac{1}{\omega r^d} P(\mathcal{B}(s, r)) - \omega_f(r) - \omega_f(\alpha) \geq \frac{c_n}{N} - \eta - \omega_f(r) - \omega_f(\alpha).$$

where we denoted by

$$\eta = \sup_{S \in \mathcal{S}_r} \frac{1}{\omega r^d} |P_n(S) - P(S)|, \quad (6.13)$$

$$\text{with } \mathcal{S}_r = \{\mathcal{B}(y, r) \cap \mathcal{B}(z, \alpha) \mid y, z \in \mathbb{R}^d\} \cup \{\mathcal{B}(y, r) \setminus \mathcal{B}(z, \alpha) \mid y, z \in \mathbb{R}^d\}.$$

The sets  $\mathcal{S}_r$  satisfy the assumptions of Lemma 6.2.1, so that  $\eta$  goes to 0 almost surely as  $n \rightarrow \infty$ . Now, let  $y \in \partial S_n^\alpha$ , and take  $s \in S_n$  among its nearest neighbors in  $S_n$ . This neighbor  $s$  is at distance exactly  $\alpha$  from  $y$ , so that  $|S_n \cap \{\mathcal{B}(s, r) \setminus \mathcal{B}(y, \alpha)\}| = c_n$ . But on the other hand, we have

$$\begin{aligned} \frac{1}{\omega r^d} \int \mathbb{1}_{\mathcal{B}(s, r) \setminus \mathcal{B}(y, \alpha)}(z) f(z) dz &\leq \frac{f(s) + \omega_f(r)}{\omega r^d} \int \mathbb{1}_{\mathcal{B}(s, r) \setminus \mathcal{B}(y, \alpha)}(z) dz \\ &\leq (f(s) + \omega_f(r)) (1/2 + O(r/\alpha)), \end{aligned}$$



so that

$$\begin{aligned}
 f(y) &\geq f(s) - \omega_f(\alpha) \geq (2 - O(r/\alpha)) \frac{1}{\omega r^d} P(\mathbf{B}(s, r) \setminus \mathbf{B}(y, \alpha)) - \omega_f(r) - \omega_f(\alpha) \\
 &\geq (2 - O(r/\alpha)) \left( \frac{c_n}{N} - \eta \right) - \omega_f(r) - \omega_f(\alpha) \\
 &\geq 2 \frac{c_n}{N} - O(r/\alpha) - 2\eta - \omega_f(r) - \omega_f(\alpha).
 \end{aligned}$$

Putting the two estimates of  $f$  over  $\partial S^\alpha$  and  $S^\alpha$  together, we have shown that

$$C^\alpha(x, f) \geq \frac{c_n}{N} - O(r/\alpha) - \eta - \omega_f(r) - \omega_f(\alpha),$$

so that, using Lemma 6.2.1, we have almost surely,

$$\limsup_{n \rightarrow \infty} \frac{c_n}{N} \leq C^\alpha(x, f) + \omega_f(\alpha) \text{ uniformly in } x \in \mathbb{R}^d.$$

Letting  $\alpha \rightarrow 0$  then concludes the proof.  $\square$

**Lemma 6.4.14.** *If  $r = r_n$  is such that  $r \rightarrow 0$  and  $nr^d \gg \log n$ , then almost surely,*

$$\liminf_{n \rightarrow \infty} \frac{1}{N} C_r(x, \mathcal{X}_n) \geq C_0(x, f) \text{ uniformly in } x \in \mathbb{R}^d.$$

*Proof.* Let  $\alpha > 0$  and  $\varepsilon > 0$ . Denoting  $t = C^\alpha(x, f)$ , there is  $S \in \mathcal{B}_\alpha$  with  $x \in S$  such that

$$\forall y \in S, f(y) \geq t - \varepsilon \quad \text{and} \quad \forall y \in \partial S, f(y) \geq 2t - 2\varepsilon.$$

Let  $H$  be the subgraph of  $\mathcal{G}_r(x, \mathcal{X}_n)$  with vertices in  $S$ , and let  $\deg_H(s)$  be the degree of a vertex  $s \in S$  in  $H$ . If  $s$  is at distance more than  $r$  from  $\partial S$ , then, using again  $\eta$  introduced in the proof of Lemma 6.4.13 at (6.13),

$$\deg_H(s) = n \times P_n(\mathbf{B}(s, r)) - 1 \geq N(f(s) - \omega_f(r) - \eta) - 1 \geq N(t - \varepsilon - \omega_f(r) - \eta) - 1.$$

Now if  $s$  is at distance less than  $r$  than  $\partial S$ , we can take  $y \in S$  such that  $s \in \mathbf{B}(y, \alpha) \subset S$ . The volume of  $\mathbf{B}(s, r) \cap \mathbf{B}(y, \alpha)$  is then at least  $\omega r^d(1/2 - O(r/\alpha))$  according to Lemma 6.4.8. We thus have,

$$\begin{aligned}
 \deg_H(s) &= n \times P_n(S \cap \mathbf{B}(s, r)) - 1 \\
 &\geq n \times P_n(\mathbf{B}(y, \alpha) \cap \mathbf{B}(s, r)) - 1 \\
 &\geq N \left( \frac{1}{\omega r^d} P(\mathbf{B}(y, \alpha) \cap \mathbf{B}(s, r)) - \eta \right) - 1 \\
 &\geq N \left( \left( \frac{1}{2} - O(r/\alpha) \right) (f(s) - \omega_f(r)) - \eta \right) - 1 \\
 &\geq N(t - \varepsilon - O(r/\alpha) - \omega_f(r) - \eta) - 1
 \end{aligned}$$

where we used the fact that  $f(s) \geq 2t - 2\varepsilon - \omega_f(r)$  because  $s$  is  $r$ -close to  $\partial S$ . We thus have shown here that

$$\frac{C_r(x, \mathcal{X}_n)}{N} \geq t - \varepsilon - O(r/\alpha) - \omega_f(r) - \eta - 1/N.$$

Now letting  $n \rightarrow \infty$  yields, almost surely,

$$\liminf_{n \rightarrow \infty} \frac{C_r(x, \mathcal{X}_n)}{N} \geq t - \varepsilon = C^\alpha(x, f) - \varepsilon \quad \text{uniformly in } x \in \mathbb{R}^d.$$

and letting  $\alpha, \varepsilon \rightarrow 0$  yields the result.  $\square$

## 6.5 Numerical Simulations

We performed some small-scale proof-of-concept computer experiments to probe into the convergences established earlier in the paper, as well as other questions of potential interest not addressed in this paper.

### 6.5.1 Illustrative Examples

In the regime where  $r = r_n \rightarrow 0$  and  $nr^d \gg \log(n)$ , Theorems 6.2.4, 6.3.7 and 6.4.12 show that only  $f(x)$  and  $C_0(x, f)$  can be obtained as limits of H-index iterates  $H_r^k(x, \mathcal{X}_n)$ , when  $k \in \{0, 1, \dots, \infty\}$  is fixed. Figures 6.5.1a and 6.5.1b both illustrate, for  $d = 1$  and  $d = 2$  respectively, the following convergence behavior:

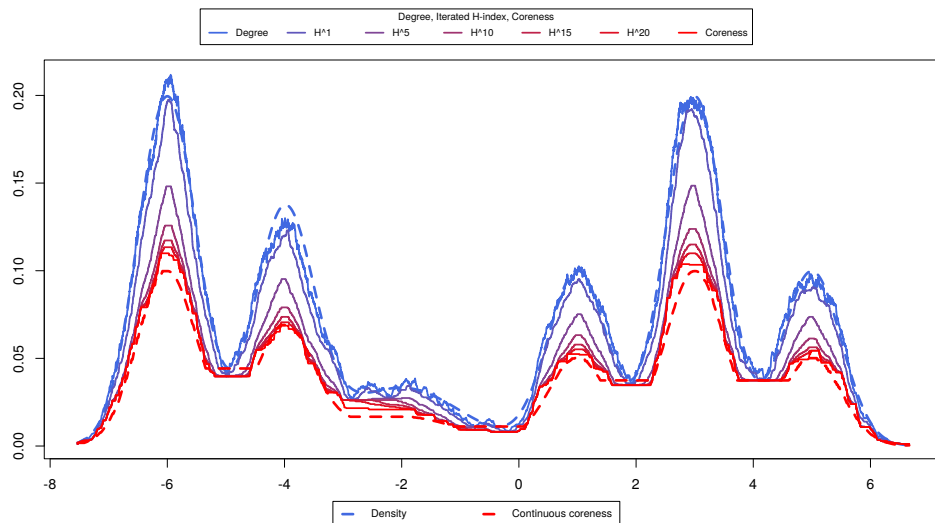
- $\frac{1}{N} \deg_r(x, \mathcal{X}_n) \xrightarrow{n \rightarrow \infty} f(x)$  (see Theorem 6.2.4);
- $H_r^k(x, \mathcal{X}_n) \xrightarrow{k \rightarrow \infty} C_r(x, \mathcal{X}_n)$  (see (6.8));
- $\frac{1}{N} C_r(x, \mathcal{X}_n) \xrightarrow{n \rightarrow \infty} C_0(x, f)$  (see Theorem 6.4.12).

The density functions have been chosen to exhibit non-trivial super-level sets, so that  $C_0(\cdot, f) \neq f/2$  (see Proposition 6.4.10).

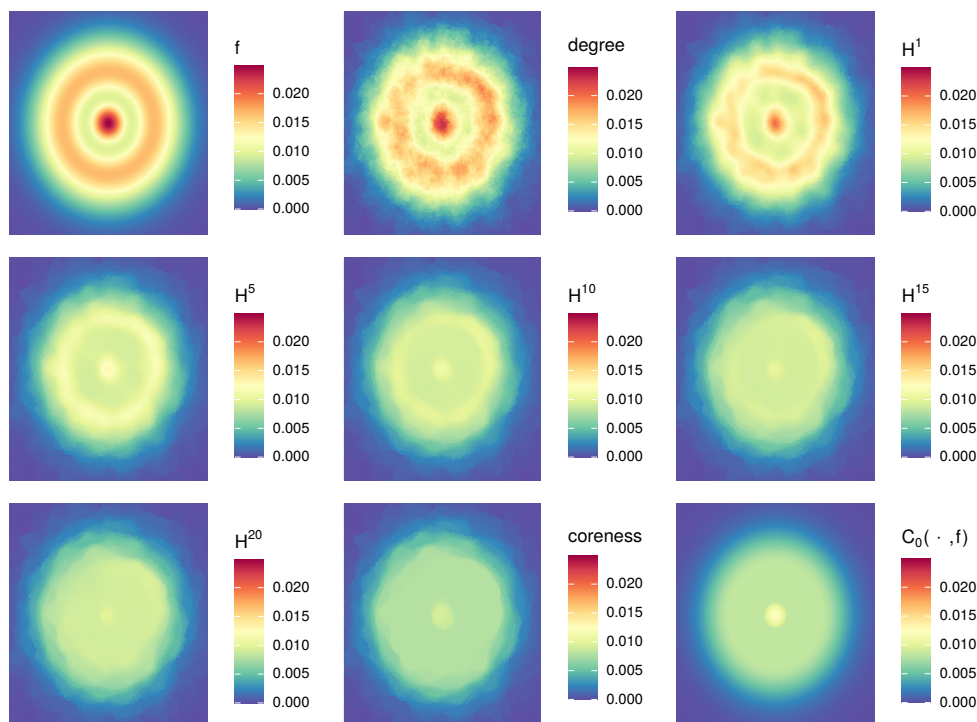
### 6.5.2 Convergence Rates

Intending to survey limiting properties of the degree, the H-index and the coreness, the above work does not provide convergence rates. We now discuss them numerically in the regime where  $r \rightarrow 0$ .

A close look at the proofs indicates that only bias terms of order  $O(r \vee \omega_f(r))$  appear in the centrality-to-depth convergences of Theorems 6.2.4, 6.3.7 and 6.4.12. For the degree, the stochastic term is known to be of order  $O(1/\sqrt{nr^d})$ . If  $f$  is Lipschitz (i.e.,  $\omega_f(r) = O(r)$ ), the bandwidth  $r_{\text{opt}}$  that achieves the best minimax possible convergence rate in Theorem 6.2.4 is  $r_{\text{opt}} = O(n^{-1/(d+2)})$ , yielding a pointwise error  $|N^{-1} \deg_r(x, \mathcal{X}_n) - f(x)| = O(r_{\text{opt}}) = O(n^{-1/(d+2)})$ . Naturally, larger values  $r \geq r_{\text{opt}}$  make the bias term lead, and smaller values  $r \leq r_{\text{opt}}$  make the stochastic term lead. Although it remains unclear how bias terms behave for H-indices and the coreness, simulations indicate a similar bias-variance tradeoff depending on  $n$  and  $r$ . Indeed, the sup-norms  $\|N^{-1} \deg_r(\cdot, \mathcal{X}_n) -$

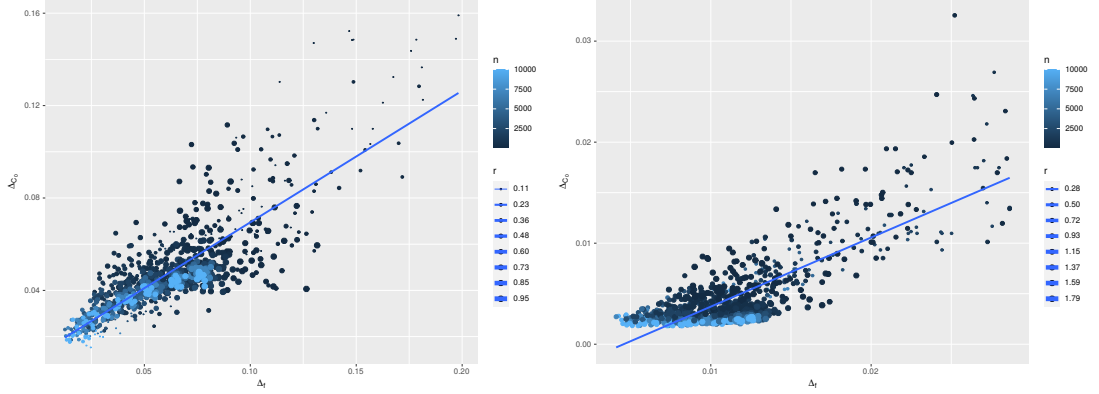


(a) The mixture  $f$  of six Gaussians in dimension  $d = 1$  from Figure 6.4.2 sampled  $n = 10000$  times. On the discrete side (solid), are displayed the degree ( $k = 0$ ), iterated H-indices for  $k \in \{1, 5, 10, 15, 20\}$ , and coreness ( $k = \infty$ ). On the continuous side (dashed), the density  $f$  and the continuum coreness  $C_0(\cdot, f)$  are plotted. Here,  $r \approx 0.13$  was picked proportional to the optimal kernel bandwidth  $r_{\text{opt}} \propto n^{-1/(d+2)} = n^{-1/3}$ .



(b) A plot similar to Figure 6.5.1a for  $d = 2$ . The generating density function  $f$  exhibits a crater-like shape enclosing a peak, yielding a continuum coreness  $C_0(\cdot, f)$  that plateaus, and in particular differs from  $f/2$  within the crater area. Here,  $n = 20000$ ,  $k \in \{0, 1, 5, 10, 15, 20, \infty\}$  and  $r \propto r_{\text{opt}} \propto n^{-1/(d+2)} = n^{-1/4}$ .

Figure 6.5.1 – Illustrative examples in dimension  $d = 1$  and in dimension  $d = 2$ .


 (a) Distribution of Figure 6.5.1a ( $d = 1$ ).

 (b) Distribution of Figure 6.5.1b ( $d = 2$ ).

Figure 6.5.2 – Scatterplot of values  $(\|N^{-1}\deg_r(\cdot, \mathcal{X}_n) - f\|_\infty, \|N^{-1}C_r(\cdot, \mathcal{X}_n) - C_0(\cdot, f)\|_\infty)$  with data generated according to (a) the Gaussian mixture distribution depicted in Figure 6.5.1a, and (b) the crater-like density of Figure 6.5.1b. Sample size values  $n$  take 9 different values in  $[100, 10000]$ , while connection radii  $r$  take 8 different values within the interval  $[0.1, 0.97]$  for (a) and  $[0.27, 1.80]$  for (b). For each such pair  $(n, r)$ , simulations are repeated 10 to 20 times, depending on the value of  $n$ .

$f\|_\infty$  and  $\|N^{-1}C_r(\cdot, \mathcal{X}_n) - C_0(\cdot, f)\|_\infty$  appear to be linearly correlated (see Figure 6.5.2). As a result, with a choice  $r \asymp r_{\text{opt}} = O(n^{-1/(d+2)})$ , we anticipate

$$\begin{aligned} |N^{-1}C_r(x, \mathcal{X}_n) - C_0(x)| &= O(|N^{-1}\deg_r(x, \mathcal{X}_n) - f(x)|) && \text{(Rate Conjecture)} \\ &= O(n^{-1/(d+2)}), && (6.14) \end{aligned}$$

with high probability. Furthermore, Figure 6.5.2 suggests that the slope relating  $\|N^{-1}\deg_r(\cdot, \mathcal{X}_n) - f\|_\infty$  and  $\|N^{-1}C_r(\cdot, \mathcal{X}_n) - C_0(\cdot, f)\|_\infty$  is of constant order, in fact between  $1/2$  and  $1$ , which suggests very moderate constants hidden in the term  $O(|N^{-1}\deg_r(x, \mathcal{X}_n) - f(x)|)$ .

### 6.5.3 Iterations of the H-Index

Seen as the limit (6.8) of H-index iterations, the coreness  $C_r(x, \mathcal{X}_n) = H_r^\infty(x, \mathcal{X}_n)$  raises computational questions. One of them resides in determining whether it is reasonable to compute it naively, by iterating the H-index over the graph until stationarity at all the vertices.

More generally, given a graph  $G = (V, E)$  and a vertex  $v \in V$  of  $G$ , and similarly as what we did in Section 6.3 for random geometric graphs, we can study the H-index  $H_G(v)$ , its iterations  $H_G^k(v)$  for  $k \in \mathbb{N}$ , and the coreness  $C_G(v)$ . The *max-iteration*  $k^\infty(G)$  of the H-index of  $G$  is then defined as the minimal number of iterations for which the iterated H-index  $H_G^k$  coincides with the coreness  $C_G$ . That is,

$$k^\infty(G) := \min \{k \in \mathbb{N} \mid \forall v \in V, C_G(v) = H_G^k(v)\}.$$

Known bounds for  $k^\infty(G)$  are of the form

$$k^\infty(G) \leq 1 + \sum_{v \in V} |\deg_G(v) - C_G(v)| \quad \text{and} \quad k^\infty(G) \leq |V|,$$

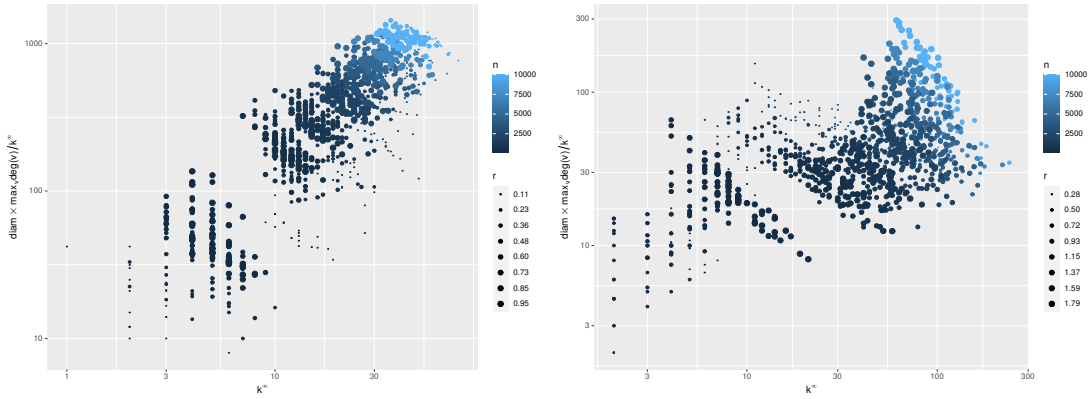
and can be found in [MDPM13, Thm 4 & Thm 5]. For random geometric graphs, this yields probabilistic bounds of order  $O(n^2 r^d)$  and  $O(n)$  respectively, with one or the other prevailing depending on whether we are in a sub-critical or super-critical regime.

However, for the random geometric graphs  $\mathcal{G}(x, \mathcal{X}_n)$ , numerical simulations suggest that an even stronger bound of order  $k^\infty(\mathcal{G}_r(x, \mathcal{X}_n)) = O(nr^{d-1})$  may hold with high probability (see Figure 6.5.3). Indeed, in the regime where  $r = r_n$  is large enough that  $\mathcal{G}_r(x, \mathcal{X}_n)$  is connected, this latter quantity appears to coincide with its diameter — which is of order  $O(1/r)$  — multiplied by its maximal degree — which is of order  $O(nr^d)$ .

Coming back to the general deterministic case, this observation leads us to conjecture that

$$k^\infty(G) \leq \max_{\substack{H \subset G \\ \text{connected}}} \text{diam}(H) \times \max_{v \in V} \deg_G(v), \quad (\text{Max-Iter. Conjecture})$$

where  $\text{diam}(H)$  is the diameter of  $H$  seen as a combinatorial graph (with edge weight 1). This conjecture, clearly satisfied in simulations (see Figure 6.5.3), would shed some light — if correct — on the dependency of the H-index iteration process with respect to the graph's geometry.



(a) Distribution of Figure 6.5.1a ( $d = 1$ ).

(b) Distribution of Figure 6.5.1b ( $d = 2$ ).

Figure 6.5.3 – Scatterplot of values  $(k^\infty(\mathcal{G}_r), \text{diam}(\mathcal{G}_r) \times \max_v \deg_{\mathcal{G}_r}(v) / k^\infty(\mathcal{G}_r))$  in log-log scale, with data generated according to (a) the Gaussian mixture distribution of Figure 6.5.1a, and (b) the crater-like density of Figure 6.5.1b. Values all appear to satisfy  $k^\infty(\mathcal{G}_r) \leq \text{diam}(\mathcal{G}_r) \times \max_v \deg_{\mathcal{G}_r}(v)$  widely (i.e., points with ordinate at least 1 in these plots), even for small values of  $r$  and  $n$ .

## 6.6 Concluding Remarks and Open Questions

**New Depths from Other Centralities** We obtained limits for the centrality measures given by the degree, H-index, and coreness. We noticed that they shared the desirable properties of *depths* from [ZS00], but only in the specific cases where the likelihood is itself a relevant notion of depth. This is not surprising, in the sense that we observed here that degree-based centralities yield density-based ‘depths’. Seeking for new notions of data depths, it would be interesting to study the limiting objects associated to other notions of centrality, such as the closeness centrality of [Fre78], the betweenness of [Fre77], and other ‘spectral’ notions [Kat53, Bon72, PBMW99, Kle99]. Similarly, we only focused on a  $r$ -ball neighborhood graph construction, but there are other graphs that could play that role, such as nearest-neighbor (possibly oriented) graphs or Delaunay triangulations, possibly yielding new limiting objects.

**Conjectures** Beyond the first-order almost-sure convergence results that we obtained, we could consider deriving convergence rates. In this regard, we left the conjecture displayed in (Rate Conjecture), but all the convergence rates associated with the results displayed in Figure 6.1.1 remain to be established. Another conjecture that we leave open is the bound on  $k^\infty(G)$  displayed in (Max-Iter. Conjecture).

# Bibliography

- [AACB21] Eddie Aamari, Ery Arias-Castro, and Clément Berenfeld. From graph centrality to data depth. *arXiv preprint arXiv:2105.03122*, 2021.
- [AAL21] Eddie Aamari, Catherine Aaron, and Clément Levrard. Minimax boundary estimation and estimation with boundary. *arXiv preprint arXiv:2108.03135*, 2021.
- [Aam17] Eddie Aamari. *Vitesses de convergence en inférence géométrique*. PhD thesis, Paris Saclay, 2017.
- [AB06] Stephanie B Alexander and Richard L Bishop. Gauss equation and injectivity radii for subspaces in spaces of curvature bounded above. *Geometriae Dedicata*, 117(1):65–84, 2006.
- [AB17] Martin Arjovsky and Léon Bottou. Towards principled methods for training generative adversarial networks. *arXiv preprint arXiv:1701.04862*, 2017.
- [ABL22] Eddie Aamari, Clément Berenfeld, and Clément Levrard. Optimal reach estimation and metric learning. *arXiv preprint arXiv:2207.06074*, 2022.
- [AC11] Ery Arias-Castro. Clustering based on pairwise distances when the data is of mixed dimensions. *IEEE Transactions on Information Theory*, 57(3):1692–1706, 2011.
- [AC20] Catherine Aaron and Alejandro Cholaquidis. On boundary detection. In *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, volume 56, pages 2028–2050. Institut Henri Poincaré, 2020.
- [ACB17] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017.
- [ACC20] Ery Arias-Castro and Phong Alain Chau. Minimax estimation of distances on a surface and minimax manifold learning in the isometric-to-convex setting. *arXiv preprint arXiv:2011.12478*, 2020.
- [ACLG19] Ery Arias-Castro and Thibaut Le Gouic. Unconstrained and curvature-constrained shortest-path distances and their approximation. *Discrete & Computational Geometry*, 62(1):1–28, 2019.
- [ACP13] Ery Arias-Castro and Bruno Pelletier. On the convergence of Maximum Variance Unfolding. *Journal of Machine Learning Research*, 14(7), 2013.

- 
- [ACPP12] Ery Arias-Castro, Bruno Pelletier, and Pierre Pudlo. The normalized graph cut and Cheeger constant: from discrete to continuous. *Advances in Applied Probability*, 44(4):907–937, 2012.
  - [AKC<sup>+</sup>19] Eddie Aamari, Jisu Kim, Frédéric Chazal, Bertrand Michel, Alessandro Rinaldo, and Larry Wasserman. Estimating the reach of a manifold. *Electronic journal of statistics*, 13(1):1359–1399, 2019.
  - [AL15] Dominique Attali and André Lieutier. Geometry-driven collapses for converting a Čech complex into a triangulation of a nicely triangulable shape. *Discrete & Computational Geometry*, 54(4):798–825, 2015.
  - [AL18] Eddie Aamari and Clément Levrard. Stability and minimax optimality of tangential Delaunay complexes for manifold reconstruction. *Discrete & Computational Geometry*, 59(4):923–971, 2018.
  - [AL19] Eddie Aamari and Clément Levrard. Nonasymptotic rates for manifold, tangent space and curvature estimation. *The Annals of Statistics*, 47(1):177–204, 2019.
  - [Alm86] F. Almgren. Optimal isoperimetric inequalities. *Indiana Univ*, 35:451–547, 1986.
  - [Alo06] Greg Aloupis. Geometric measures of data depth. In Regina Y Liu, Robert Joseph Serfling, and Diane L Souvaine, editors, *Data Depth: Robust Multivariate Analysis, Computational Geometry, and Applications*, volume 72, pages 147–158. American Mathematical Society, 2006.
  - [ALS13] Dominique Attali, André Lieutier, and David Salinas. Vietoris–Rips complexes also provide topologically correct reconstructions of sampled shapes. *Computational Geometry*, 46(4):448–465, 2013.
  - [And55] Theodore W Anderson. The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proceedings of the American Mathematical Society*, 6(2):170–176, 1955.
  - [And09] Tessa K Anderson. Kernel density estimation and k-means clustering to profile road accident hotspots. *Accident Analysis & Prevention*, 41(3):359–364, 2009.
  - [AS21] Yariv Aizenbud and Barak Sober. Non-Parametric Estimation of Manifolds from Noisy Data. *arXiv e-prints*, page arXiv:2105.04754, May 2021.
  - [AST93] Martin Anthony and John Shawe-Taylor. A result of Vapnik with applications. *Discrete Applied Mathematics*, 47(3):207–217, 1993.
  - [Bar76] Vic Barnett. The ordering of multivariate data. *Journal of the Royal Statistical Society: Series A (General)*, 139(3):318–344, 1976.
  - [BBM99] Andrew Barron, Lucien Birgé, and Pascal Massart. Risk bounds for model selection via penalization. *Probability theory and related fields*, 113(3):301–413, 1999.
  - [BCJ<sup>+</sup>19] Eli Bingham, Jonathan P Chen, Martin Jankowiak, Fritz Obermeyer, Neeraj Pradhan, Theofanis Karaletsos, Rohit Singh, Paul Szerlip, Paul Horsfall, and Noah D Goodman. Pyro: Deep universal probabilistic programming. *The Journal of Machine Learning Research*, 20(1):973–978, 2019.



- [BCQY97] MR Brito, EL Chavez, AJ Quiroz, and JE Yukich. Connectivity of the mutual  $k$ -nearest-neighbor graph in clustering and outlier detection. *Statistics & Probability Letters*, 35(1):33–42, 1997.
- [BDSLT00] Mira Bernstein, Vin De Silva, John C Langford, and Joshua B Tenenbaum. Graph approximations to geodesics on embedded manifolds. Technical report, Citeseer, 2000.
- [BDSS21] Andy Brock, Soham De, Samuel L Smith, and Karen Simonyan. High-performance large-scale image recognition without normalization. In *International Conference on Machine Learning*, pages 1059–1071. PMLR, 2021.
- [BE06] Stephen P Borgatti and Martin G Everett. A graph-theoretic perspective on centrality. *Social Networks*, 28(4):466–484, 2006.
- [Bes78] AL Besse. *Manifolds all of whose Geodesics are Closed*, volume 93. Springer Science & Business Media, 1978.
- [BG14] Jean-Daniel Boissonnat and Arijit Ghosh. Manifold reconstruction using tangential Delaunay complexes. *Discrete & Computational Geometry*, 51(1):221–267, 2014.
- [BH21] Clément Berenfeld and Marc Hoffmann. Density estimation on an unknown submanifold. *Electronic Journal of Statistics*, 15(1):2179–2223, 2021.
- [BHHS22] Clément Berenfeld, John Harvey, Marc Hoffmann, and Krishnan Shankar. Estimating the reach of a manifold via its convexity defect function. *Discrete & Computational Geometry*, 67(2):403–438, 2022.
- [Bit90] John F Bithell. An application of density estimation to geographical epidemiology. *Statistics in medicine*, 9(6):691–701, 1990.
- [BKMP09] Paolo Baldi, Gérard Kerkycharian, Domenico Marinucci, and Dominique Picard. Adaptive density estimation for directional data using needlets. *The Annals of Statistics*, 37(6A):3362–3395, 2009.
- [BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.
- [BLW19] Jean-Daniel Boissonnat, André Lieutier, and Mathijs Wintraecken. The reach, metric distortion, geodesic convexity and the variation of tangent spaces. *Journal of applied and computational topology*, 3(1):29–58, 2019.
- [BM21] Dalya Baron and Brice Ménard. Extracting the main trend in a data set: The sequencer algorithm. *The Astrophysical Journal*, 916(2):91, 2021.
- [BN01] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. *Advances in neural information processing systems*, 14, 2001.
- [BN03] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.

- [BN08] Mikhail Belkin and Partha Niyogi. Towards a theoretical foundation for Laplacian-based manifold methods. *Journal of Computer and System Sciences*, 74(8):1289–1308, 2008.
- [Bon72] Phillip Bonacich. Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, 2(1):113–120, 1972.
- [Bou85] Jean Bourgain. On Lipschitz embedding of finite metric spaces in hilbert space. *Israel Journal of Mathematics*, 52(1):46–52, 1985.
- [Bow84] Adrian W. Bowman. An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71(2):353–360, 1984.
- [BRR22] Clément Berenfeld, Paul Rosa, and Judith Rousseau. Estimating a density near an unknown manifold: a bayesian nonparametric approach. *arXiv preprint arXiv:2205.15717*, 2022.
- [BRS<sup>+</sup>12] Sivaraman Balakrishnan, Alesandro Rinaldo, Don Sheehy, Aarti Singh, and Larry Wasserman. Minimax rates for homology inference. In *Artificial Intelligence and Statistics*, pages 64–72. PMLR, 2012.
- [BS17] Tyrus Berry and Timothy Sauer. Density estimation on manifolds with boundary. *Computational Statistics & Data Analysis*, 107:1–17, 2017.
- [CC08] Michael AA Cox and Trevor F Cox. Multidimensional scaling. In *Handbook of data visualization*, pages 315–347. Springer, 2008.
- [CCSL09] Frédéric Chazal, David Cohen-Steiner, and André Lieutier. A sampling theory for compact sets in euclidean space. *Discrete & Computational Geometry*, 41(3):461–479, 2009.
- [CCSM11] Frédéric Chazal, David Cohen-Steiner, and Quentin Mérigot. Geometric inference for probability measures. *Foundations of Computational Mathematics*, 11(6):733–751, 2011.
- [CDB17] Antonio Canale and Pierpaolo De Blasi. Posterior asymptotics of nonparametric location-scale mixtures for multivariate density estimation. *Bernoulli*, 23(1):379–404, 2017.
- [CDFM22] Federico Camerlenghi, Emanuele Dolera, Stefano Favaro, and Edoardo Mainini. Wasserstein posterior contraction rates in non-dominated bayesian nonparametric models. *arXiv preprint arXiv:2201.12225*, 2022.
- [CFGM20] Alejandro Cholaquidis, Ricardo Fraiman, Fabrice Gamboa, and Leonardo Moreno. Weighted lens depth: Some applications to supervised classification. *arXiv preprint arXiv:2011.11140*, 2020.
- [CFM21] Alejandro Cholaquidis, Ricardo Fraiman, and Leonardo Moreno. Universally consistent estimation of the reach. *arXiv preprint arXiv:2110.12208*, 2021.
- [CFM22] Alejandro Cholaquidis, Ricardo Fraiman, and Leonardo Moreno. Level set and density estimation on manifolds. *Journal of Multivariate Analysis*, 189:104925, 2022.

- [CGK<sup>+</sup>20] Galatia Cleanthous, Athanasios G Georgiadis, Gerard Kerkyacharian, Pencho Petrushev, and Dominique Picard. Kernel and wavelet density estimators on manifolds and more general metric spaces. *Bernoulli*, 26(3):1832–1862, 2020.
- [CGLM14] Frédéric Chazal, Marc Glisse, Catherine Labruère, and Bertrand Michel. Convergence rates for persistence diagram estimation in topological data analysis. In *International Conference on Machine Learning*, pages 163–171. PMLR, 2014.
- [CGW15] Yen-Chi Chen, Christopher R Genovese, and Larry Wasserman. Asymptotic theory for density ridges. *The Annals of Statistics*, 43(5):1896–1928, 2015.
- [Chi91] Shean-Tsong Chiu. Bandwidth selection for kernel density estimation. *Ann. Statist.*, 19(4):1883–1905, 1991.
- [CHM11] Dong Chen, Peter Hall, and Hans-Georg Müller. Single and multiple index functional regression models with nonparametric link. *The Annals of Statistics*, 39(3):1720–1747, 2011.
- [Chu97] Fan Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.
- [CKKL21] Minwoo Chae, Dongha Kim, Yongdai Kim, and Lizhen Lin. A likelihood approach to nonparametric estimation of a singular distribution using deep generative models. *arXiv preprint arXiv:2105.04046*, 2021.
- [CKP14] Ismaël Castillo, Gérard Kerkyacharian, and Dominique Picard. Thomas Bayes’ walk on manifolds. *Probability Theory and Related Fields*, 158(3):665–710, 2014.
- [CKTK10] Ratthachat Chatpatanasiri, Teesid Korsrilabutr, Pasakorn Tangchanachaianan, and Boonserm Kijirikul. A new kernelization framework for Mahalanobis distance learning algorithms. *Neurocomputing*, 73(10-12):1570–1579, 2010.
- [CL04] Frédéric Chazal and André Lieutier. Stability and homotopy of a subset of the medial axis. In *Symposium on Solid Modeling and Applications*, pages 243–248. Citeseer, 2004.
- [CL05] Frédéric Chazal and André Lieutier. The “ $\lambda$ -medial axis”. *Graphical models*, 67(4):304–331, 2005.
- [CL13] Fabienne Comte and Claire Lacour. Anisotropic adaptive kernel deconvolution. In *Annales de l’IHP Probabilités et statistiques*, volume 49, pages 569–609, 2013.
- [CM21] Frédéric Chazal and Bertrand Michel. An introduction to topological data analysis: fundamental and practical aspects for data scientists. *Frontiers in Artificial Intelligence*, 4, 2021.
- [CMG22] Jeremie Capitao-Miniconi and Élisabeth Gassiat. Deconvolution of spherical data corrupted with unknown noise. *arXiv preprint arXiv:2203.00654*, 2022.
- [Cot12] Frederick Steven Cottle. *Statistical human body form classification: Methodology development and application*. Auburn University, 2012.
- [CS96] G Constantine and T Savits. A multivariate Faa di Bruno formula with applications. *Transactions of the American Mathematical Society*, 348(2):503–520, 1996.

- [CS20] Jeff Calder and Charles K Smart. The limit shape of convex hull peeling. *Duke Mathematical Journal*, 169(11):2079–2124, 2020.
- [CSQ<sup>+</sup>19] Junyue Cao, Malte Spielmann, Xiaojie Qiu, Xingfan Huang, Daniel M Ibrahim, Andrew J Hill, Fan Zhang, Stefan Mundlos, Lena Christiansen, Frank J Steemers, et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature*, 566(7745):496–502, 2019.
- [CSW95] F H. Clarke, R. J. Stern, and P R. Wolenski. Proximal smoothness and the lower- $C^2$  property. *J. Convex Anal.*, 2(1-2):117–144, 1995.
- [CVvL18] Leena Chennuru Vankadara and Ulrike von Luxburg. Measures of distortion for machine learning. *Advances in Neural Information Processing Systems*, 31, 2018.
- [CZ05] Olivier Chapelle and Alexander Zien. Semi-supervised classification by low density separation. In *International workshop on artificial intelligence and statistics*, pages 57–64. PMLR, 2005.
- [DCFF92] Manfredo Perdigao Do Carmo and J Flaherty Francis. *Riemannian geometry*, volume 6. Springer, 1992.
- [DCS<sup>+</sup>10] David Dunson, Minhua Chen, Jorge Silva, John Paisley, Chunping Wang, and Lawrence Carin. Compressive sensing on manifolds using a nonparametric mixture of factor analyzers: Algorithm and performance bounds. *IEEE Transactions on Signal Processing*, 58(12):6140–6155, 2010.
- [DG85] Luc Devroye and László Györfi. *Nonparametric density estimation*. Wiley Series in Probability and Mathematical Statistics: Tracts on Probability and Statistics. John Wiley & Sons, Inc., New York, 1985. The  $L_1$  view.
- [DG92] David L Donoho and Miriam Gasko. Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *The Annals of Statistics*, 20(4):1803–1827, 1992.
- [DG05] David L Donoho and Carrie Grimes. Image manifolds which are isometric to euclidean space. *Journal of mathematical imaging and vision*, 23(1):5–24, 2005.
- [DGK02] Inderjit S Dhillon, Yuqiang Guan, and Jacob Kogan. Iterative clustering of high dimensional text data augmented by local search. In *ICDM*, volume 2, page 131. Citeseer, 2002.
- [Div21a] Vincent Divol. Minimax adaptive estimation in manifold inference. *Electronic Journal of Statistics*, 15(2):5888–5932, 2021.
- [Div21b] Vincent Divol. Reconstructing measures on manifolds: an optimal transport approach. *arXiv preprint arXiv:2102.07595*, 2021.
- [DMR18] Luc Devroye, Abbas Mehrabian, and Tommy Reddad. The total variation distance between high-dimensional gaussians. *arXiv preprint arXiv:1810.08693*, 2018.
- [DNHTY14] Sophie Dabo-Niang, Leila Hamdad, Camille Ternynck, and Anne-Françoise Yao. A kernel spatial density estimation allowing for the analysis of spatial clustering. Application to Monsoon Asia Drought Atlas data. *Stochastic environmental research and risk assessment*, 28(8):2075–2099, 2014.

## BIBLIOGRAPHY

---

- [DW21] David B Dunson and Nan Wu. Inferring manifolds from noisy data using Gaussian processes. *arXiv preprint arXiv:2110.07478*, 2021.
- [Edd82] William F Eddy. Convex hull peeling. In *COMPSTAT Symposium*, pages 42–47, Toulouse, 1982. Springer.
- [EG92] Lawrence Craig Evans and Ronald F Gariepy. *Measure theory and fine properties of functions*. Chapman and Hall/CRC, 1992.
- [ES00] Ömer Eugeneçioğlu and Ashok Srinivasan. Efficient nonparametric density estimation on the sphere with applications in fluid mechanics. *SIAM Journal on Scientific Computing*, 22(1):152–176, 2000.
- [ET94] B. Efron and R. J. Tibshirani. *An introduction to the Bootstrap*. Chapman & Hall, Monograph on Statistics and Probability 57. London, 1994.
- [EW95] Michael D Escobar and Mike West. Bayesian density estimation and inference using mixtures. *Journal of the american statistical association*, 90(430):577–588, 1995.
- [Fed59] Herbert Federer. Curvature measures. *Transactions of the American Mathematical Society*, 93(3):418–491, 1959.
- [Fed69] Herbert Federer. *Geometric measure theory*. Springer, 1969.
- [Fer83] Thomas S Ferguson. Bayesian density estimation by mixtures of normal distributions. In *Recent advances in statistics*, pages 287–302. Elsevier, 1983.
- [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.
- [FKRD09] Alan Frieze, Jon Kleinberg, R Ravi, and Warren DeBany. Line-of-sight networks. *Combinatorics, Probability and Computing*, 18(1-2):145–163, 2009.
- [FLM97] Ricardo Fraiman, Regina Y Liu, and Jean Meloche. Multivariate density estimation by probing depth. In Yadolah Dodge, editor,  *$L^1$ -Statistical Procedures and Related Topics*, pages 415–430. Institute of Mathematical Statistics, 1997.
- [FM99] Ricardo Fraiman and Jean Meloche. Multivariate L-estimation. *Test*, 8(2):255–317, 1999.
- [FMN16a] Charles Fefferman, Sanjoy Mitter, and Hariharan Narayanan. Testing the manifold hypothesis. *Journal of the American Mathematical Society*, 29(4):983–1049, 2016.
- [FMN16b] Charles Fefferman, Sanjoy Mitter, and Hariharan Narayanan. Testing the manifold hypothesis. *Journal of the American Mathematical Society*, 29(4):983–1049, 2016.
- [Fre77] Linton C Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41, 1977.
- [Fre78] Linton C Freeman. Centrality in social networks: Conceptual clarification. *Social Networks*, 1(3):215–239, 1978.

- [FSA07] Amir Massoud Farahmand, Csaba Szepesvári, and Jean-Yves Audibert. Manifold-adaptive dimension estimation. In *Proceedings of the 24th international conference on Machine learning*, pages 265–272, 2007.
- [GBC16] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.
- [GGVDV00] Subhashis Ghosal, Jayanta K Ghosh, and Aad W Van Der Vaart. Convergence rates of posterior distributions. *Annals of Statistics*, pages 500–531, 2000.
- [GH96] Zoubin Ghahramani and Geoffrey E Hinton. The EM algorithm for mixtures of factor analyzers. Technical report, Technical Report CRG-TR-96-1, University of Toronto, 1996.
- [GHL90] Sylvestre Gallot, Dominique Hulin, and Jacques Lafontaine. *Riemannian geometry*, volume 2. Springer, 1990.
- [Gir21] Christophe Giraud. *Introduction to high-dimensional statistics*. Chapman and Hall/CRC, 2021.
- [GK06] Evarist Giné and Vladimir Koltchinskii. Empirical graph Laplacian approximation of Laplace–Beltrami operators: Large sample results. In Evarist Giné, Vladimir Koltchinskii, Wenbo Li, and Joel Zinn, editors, *High-Dimensional Probability*, pages 238–259. Institute of Mathematical Statistics, 2006.
- [GL07] Stéphane Gaïffas and Guillaume Lecué. Optimal rates and adaptation in the single-index model using aggregation. *Electron. J. Stat.*, 1:538–573, 2007.
- [GL08] Alexander Goldenshluger and Oleg Lepski. Universal pointwise selection rule in multivariate function estimation. *Bernoulli*, 14(4):1150–1190, 2008.
- [GL11] Alexander Goldenshluger and Oleg Lepski. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.*, 39(3):1608–1632, 2011.
- [GL14] Alexander Goldenshluger and Oleg Lepski. On adaptive minimax density estimation on  $R^d$ . *Probab. Theory Related Fields*, 159(3-4):479–543, 2014.
- [GN10] Evarist Giné and Richard Nickl. Confidence bands in density estimation. *The Annals of Statistics*, 38(2):1122–1170, 2010.
- [GN16] Evarist Giné and Richard Nickl. *Mathematical foundations of infinite-dimensional statistical models*, volume 40. Cambridge University Press, 2016.
- [GPAM<sup>+</sup>14] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.
- [GPPVW12a] Christopher R Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman. Manifold estimation and singular deconvolution under Hausdorff loss. *The Annals of Statistics*, 40(2):941–963, 2012.

## BIBLIOGRAPHY

---

- [GPPVW12b] Christopher R. Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman. Minimax manifold estimation. *J. Mach. Learn. Res.*, 13:1263–1291, 2012.
- [GPPVW14] Christopher R Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman. Nonparametric ridge estimation. *The Annals of Statistics*, 42(4):1511–1545, 2014.
- [GR70] Mikhail Leonidovich Gromov and Vladimir Abramovich Rokhlin. Embeddings and immersions in Riemannian geometry. *Russian Mathematical Surveys*, 25(5):1, 1970.
- [GR05] Amir Globerson and Sam Roweis. Metric learning by collapsing classes. *Advances in neural information processing systems*, 18, 2005.
- [Gro94] K. Grove. Critical point theory for distance functions. in *Proceedings of Symposia in Pure Mathematics*, 54:357–386, 1994.
- [GVDV01] Subhashis Ghosal and Aad W Van Der Vaart. Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *The Annals of Statistics*, 29(5):1233–1263, 2001.
- [GVDV07] Subhashis Ghosal and Aad Van Der Vaart. Posterior convergence rates of dirichlet mixtures at smooth densities. *The Annals of Statistics*, pages 697–723, 2007.
- [GVS09] Matthieu Guillaumin, Jakob Verbeek, and Cordelia Schmid. Is that you? Metric learning approaches for face identification. In *2009 IEEE 12th international conference on computer vision*, pages 498–505. IEEE, 2009.
- [Hat02] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2002.
- [Hen90] Harrie Hendriks. Nonparametric estimation of a probability density on a Riemannian manifold using Fourier expansions. *The Annals of Statistics*, pages 832–849, 1990.
- [HG07] Alexander Hinneburg and Hans-Henning Gabriel. Denclue 2.0: Fast clustering based on kernel density estimation. In *International symposium on intelligent data analysis*, pages 70–80. Springer, 2007.
- [Hir05] Jorge E Hirsch. An index to quantify an individual’s scientific research output. *Proceedings of the National Academy of Sciences*, 102(46):16569–16572, 2005.
- [HJS01] Marian Hristache, Anatoli Juditsky, and Vladimir Spokoiny. Direct estimation of the index coefficient in a single-index model. *Annals of Statistics*, pages 595–623, 2001.
- [HL02] M Hoffman and Oleg Lepski. Random rates in anisotropic regression (with a discussion and a rejoinder by the authors). *The Annals of Statistics*, 30(2):325–396, 2002.
- [HP21] Christian Horvat and Jean-Pascal Pfister. Density estimation on low-dimensional manifolds: an inflation-deflation approach. *arXiv preprint arXiv:2105.12152*, 2021.

- [HR02] Geoffrey E Hinton and Sam Roweis. Stochastic neighbor embedding. *Advances in neural information processing systems*, 15, 2002.
- [HWC87] Peter Hall, GS Watson, and Javier Cabrera. Kernel density estimation with spherical data. *Biometrika*, 74(4):751–762, 1987.
- [JHD21] Michael Jauch, Peter D Hoff, and David B Dunson. Monte Carlo simulation on the Stiefel manifold via polar expansion. *Journal of Computational and Graphical Statistics*, 30(3):622–631, 2021.
- [JL07] Svante Janson and Malwina J Luczak. A simple solution to the k-core problem. *Random Structures & Algorithms*, 30(1-2):50–62, 2007.
- [JL08] Svante Janson and Malwina J Luczak. Asymptotic normality of the k-core in random graphs. *The Annals of Applied Probability*, 18(3):1085–1137, 2008.
- [Kat53] Leo Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 1953.
- [KK02] Peter T. Kim and Ja-Yong Koo. Optimal spherical deconvolution. *J. Multivariate Anal.*, 80(1):21–42, 2002.
- [KKL09] Peter T. Kim, Ja-Yong Koo, and Zhi-Ming Luo. Weyl eigenvalue asymptotics and sharp adaptation on vector bundles. *J. Multivariate Anal.*, 100(9):1962–1978, 2009.
- [Kle99] Jon M Kleinberg. Hubs, authorities, and communities. *ACM Computing Surveys (CSUR)*, 31(4es), 1999.
- [KLM<sup>+</sup>20] Doyeon Kim, V Lekić, Brice Ménard, Dalya Baron, and Manuchehr Taghizadeh-Popp. Sequencing seismograms: A panoptic view of scattering in the core-mantle boundary region. *Science*, 368(6496):1223–1228, 2020.
- [KLP01] Gérard Kerkycharian, Oleg Lepski, and Dominique Picard. Nonlinear estimation in anisotropic multi-index denoising. *Probability theory and related fields*, 121(2):137–170, 2001.
- [KNP12] Gerard Kerkycharian, Richard Nickl, and Dominique Picard. Concentration inequalities and confidence bands for needlet density estimators on compact homogeneous manifolds. *Probability Theory and Related Fields*, 153(1-2):363–404, 2012.
- [Kol09] Eric D Kolaczyk. *Statistical Analysis of Network Data: Methods and Models*. Springer Science & Business Media, 2009.
- [KPNP11] Gérard Kerkycharian, Thanh Mai Pham Ngoc, and Dominique Picard. Localized spherical deconvolution. *Ann. Statist.*, 39(2):1042–1068, 2011.
- [KPP70] Wolfgang Klein, Reinier Plomp, and Louis CW Pols. Vowel spectra, vowel spaces, and vowel identification. *The Journal of the Acoustical Society of America*, 48(4B):999–1009, 1970.
- [KPS14] Teresa Klein, Sven Proppert, and Markus Sauer. Eight years of single-molecule localization microscopy. *Histochemistry and cell biology*, 141(6):561–575, 2014.



- [KRVDV10] Willem Kruijer, Judith Rousseau, and Aad Van Der Vaart. Adaptive Bayesian density estimation with location-scale mixtures. *Electronic Journal of Statistics*, 4:1225–1257, 2010.
- [KRW16] Jisu Kim, Alessandro Rinaldo, and Larry Wasserman. Minimax rates for estimating the dimension of a manifold. *arXiv preprint arXiv:1605.01011*, 2016.
- [KT03] James T Kwok and Ivor W Tsang. Learning with idealized kernels. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pages 400–407, 2003.
- [KVL17] Matthäus Kleindessner and Ulrike Von Luxburg. Lens depth function and k-relative neighborhood graph: versatile tools for ordinal data analysis. *The Journal of Machine Learning Research*, 18(1):1889–1940, 2017.
- [KW13] Diederik P Kingma and Max Welling. Auto-encoding variational Bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- [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.
- [LB04] Elizaveta Levina and Peter Bickel. Maximum likelihood estimation of intrinsic dimension. *Advances in neural information processing systems*, 17, 2004.
- [LBBH98] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [Lee06] John M Lee. *Riemannian manifolds: an introduction to curvature*, volume 176. Springer Science & Business Media, 2006.
- [Lep90] O. V. Lepskii. A problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatnost. i Primenen.*, 35(3):459–470, 1990.
- [Lep92] OV Lepskii. Asymptotically minimax adaptive estimation. i: Upper bounds. optimally adaptive estimates. *Theory of Probability & Its Applications*, 36(4):682–697, 1992.
- [Liu90] Regina Y Liu. On a notion of data depth based on random simplices. *The Annals of Statistics*, 18(1):405–414, 1990.
- [LMR17] Claire Lacour, Pascal Massart, and Vincent Rivoirard. Estimator selection: a new method with applications to kernel density estimation. *Sankhya A*, 79(2):298–335, 2017.
- [LMS97] Oleg V Lepski, Enno Mammen, and Vladimir G Spokoiny. Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *The Annals of Statistics*, 25(3):929–947, 1997.
- [Low92] Mark G. Low. Nonexistence of an adaptive estimator for the value of an unknown probability density. *Ann. Statist.*, 20(1):598–602, 1992.
- [Low97] Mark G Low. On nonparametric confidence intervals. *The Annals of Statistics*, 25(6):2547–2554, 1997.

- [LPS99] Regina Y Liu, Jesse M Parelius, and Kesar Singh. Multivariate analysis by data depth: descriptive statistics, graphics and inference. *The Annals of Statistics*, 27(3):783–858, 1999.
- [LS14] Oleg Lepski and Nora Serdyukova. Adaptive estimation under single-index constraint in a regression model. *Ann. Statist.*, 42(1):1–28, 2014.
- [LSS06] Regina Y Liu, Robert Joseph Serfling, and Diane L Souvaine, editors. *Data Depth: Robust Multivariate Analysis, Computational Geometry, and Applications*. American Mathematical Society, 2006.
- [Łuc91] Tomasz Łuczak. Size and connectivity of the k-core of a random graph. *Discrete Mathematics*, 91(1):61–68, 1991.
- [LV07] John A Lee and Michel Verleysen. *Nonlinear dimensionality reduction*, volume 1. Springer, 2007.
- [Lyt04] A. Lytchak. On the geometry of subsets of positive reach. *Manuscripta Math*, 115:199–205, 2004.
- [LZZS16] Linyuan Lü, Tao Zhou, Qian-Ming Zhang, and H Eugene Stanley. The H-index of a network node and its relation to degree and coreness. *Nature Communications*, 7:10168, 2016.
- [MDPM13] Alberto Montresor, Francesco De Pellegrini, and Daniele Miorandi. Distributed k-core decomposition. *IEEE Transactions on Parallel and Distributed Systems*, 24(2):288–300, 2013.
- [MF12] Yunqian Ma and Yun Fu. *Manifold learning theory and applications*, volume 434. CRC press Boca Raton, 2012.
- [MGPV20] Fragkiskos D Malliaros, Christos Giatsidis, Apostolos N Papadopoulos, and Michalis Vazirgiannis. The core decomposition of networks: Theory, algorithms and applications. *The VLDB Journal*, 29(1):61–92, 2020.
- [MHM18] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.
- [MHvL09] Markus Maier, Matthias Hein, and Ulrike von Luxburg. Optimal construction of k-nearest-neighbor graphs for identifying noisy clusters. *Theoretical Computer Science*, 410(19):1749–1764, 2009.
- [MLD20] Minerva Mukhopadhyay, Didong Li, and David B Dunson. Estimating densities with non-linear support by using Fisher–Gaussian kernels. *Journal of the Royal Statistical Society Series B*, 82(5):1249–1271, 2020.
- [Mos13] Karl Mosler. Depth statistics. In Becker C., Fried R., and Kuhnt S., editors, *Robustness and Complex Data Structures*, pages 17–34. Springer, 2013.
- [MP20] Tobias Müller and Mathew D Penrose. Optimal Cheeger cuts and bisections of random geometric graphs. *The Annals of Applied Probability*, 30(3):1458–1483, 2020.

## BIBLIOGRAPHY

---

- [MPRR04] Peter Müller, Giovanni Parmigiani, Christian Robert, and Judith Rousseau. Optimal sample size for multiple testing: the case of gene expression microarrays. *Journal of the American Statistical Association*, 99(468):990–1001, 2004.
- [MRM13] Cathy Maugis-Rabusseau and Bertrand Michel. Adaptive density estimation for clustering with Gaussian mixtures. *ESAIM: Probability and Statistics*, 17:698–724, 2013.
- [Nik12] Sergei Mihailovic Nikol'skii. *Approximation of functions of several variables and imbedding theorems*, volume 205. Springer Science & Business Media, 2012.
- [NJW02] Andrew Y Ng, Michael I Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, pages 849–856, 2002.
- [NMDB17] Bac Nguyen, Carlos Morell, and Bernard De Baets. Supervised distance metric learning through maximization of the Jeffrey divergence. *Pattern Recognition*, 64:215–225, 2017.
- [NNM96] SA Nene, SK Nayar, and H Murase. Columbia object image library (coil-20), 1996. *Data available at <http://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php>*, 1996.
- [NR17] Zacharie Naulet and Judith Rousseau. Posterior concentration rates for mixtures of normals in random design regression. *Electronic Journal of Statistics*, 11(2):4065–4102, 2017.
- [NSW08] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete & Computational Geometry*, 39(1):419–441, 2008.
- [OG09] Arkadas Ozakin and Alexander Gray. Submanifold density estimation. *Advances in Neural Information Processing Systems*, 22, 2009.
- [Oja83] Hannu Oja. Descriptive statistics for multivariate distributions. *Statistics & Probability Letters*, 1(6):327–332, 1983.
- [Par62] Emanuel Parzen. On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33:1065–1076, 1962.
- [PBMW99] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The PageRank citation ranking: Bringing order to the Web. Technical report, Stanford InfoLab, 1999.
- [Pel05] Bruno Pelletier. Kernel density estimation on Riemannian manifolds. *Statistics & probability letters*, 73(3):297–304, 2005.
- [Pen03] Mathew Penrose. *Random Geometric Graphs*, volume 5. Oxford university press, 2003.
- [PRT00] R. A. Poliquin, R. T. Rockafellar, and L. Thibault. Local differentiability of distance functions. *Trans. Amer. Math. Soc.*, 352(11):5231–5249, 2000.

- [PS22] Nikita Puchkin and Vladimir Spokoiny. Structure-adaptive manifold estimation. *Journal of Machine Learning Research*, 23(40):1–62, 2022.
- [PSW96] Boris Pittel, Joel Spencer, and Nicholas Wormald. Sudden emergence of a giant  $k$ -core in a random graph. *Journal of Combinatorial Theory, Series B*, 67(1):111–151, 1996.
- [PVI<sup>+</sup>16] Charlotte Pelletier, Silvia Valero, Jordi Inglada, Nicolas Champion, and Gérard Dedieu. Assessing the robustness of random forests to map land cover with high resolution satellite image time series over large areas. *Remote Sensing of Environment*, 187:156–168, 2016.
- [PZA<sup>+</sup>21] Phillip Pope, Chen Zhu, Ahmed Abdelkader, Micah Goldblum, and Tom Goldstein. The intrinsic dimension of images and its impact on learning. *arXiv preprint arXiv:2104.08894*, 2021.
- [Ric11] Lewis Fry Richardson. IX. the approximate arithmetical solution by finite differences of physical problems involving differential equations, with an application to the stresses in a masonry dam. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 210(459-470):307–357, 1911.
- [Rio08] Oliver Riordan. The  $k$ -core and branching processes. *Combinatorics, Probability and Computing*, 17(1):111–136, 2008.
- [RLSMG08] Branko Ristic, Barbara La Scala, Mark Morelande, and Neil Gordon. Statistical analysis of motion patterns in ais data: Anomaly detection and motion prediction. In *2008 11th International Conference on Information Fusion*, pages 1–7. IEEE, 2008.
- [Rou10] Judith Rousseau. Rates of convergence for the posterior distributions of mixtures of betas and adaptive nonparametric estimation of the density. *The Annals of Statistics*, 38(1):146–180, 2010.
- [RR12] Vincent Rivoirard and Judith Rousseau. Posterior concentration rates for infinite dimensional exponential families. *Bayesian Analysis*, 7(2):311–334, 2012.
- [RS00] Sam T Roweis and Lawrence K Saul. Nonlinear dimensionality reduction by locally linear embedding. *science*, 290(5500):2323–2326, 2000.
- [Sam12] Richard J Samworth. Optimal weighted nearest neighbour classifiers. *The Annals of Statistics*, 40(5):2733–2763, 2012.
- [Sei83] Stephen B Seidman. Network structure and minimum degree. *Social Networks*, 5(3):269–287, 1983.
- [SGH21] Juan Luis Suárez, Salvador García, and Francisco Herrera. A tutorial on distance metric learning: Mathematical foundations, algorithms, experimental analysis, prospects and challenges. *Neurocomputing*, 425:300–322, 2021.
- [Sil86] B. W. Silverman. *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London, 1986.

- [Sin06] Amit Singer. From graph to manifold Laplacian: The convergence rate. *Applied and Computational Harmonic Analysis*, 21(1):128–134, 2006.
- [SJF14] Jun Shi, Zhiguo Jiang, and Hao Feng. Adaptive graph embedding discriminant projections. *Neural processing letters*, 40(3):211–226, 2014.
- [STG13] Weining Shen, Surya T Tokdar, and Subhashis Ghosal. Adaptive Bayesian multi-variate density estimation with Dirichlet mixtures. *Biometrika*, 100(3):623–640, 2013.
- [Sto82] Charles J. Stone. Optimal global rates of convergence for nonparametric regression. *Ann. Statist.*, 10(4):1040–1053, 1982.
- [TF79] Godfried T Toussaint and Ron S Foutsen. Some new algorithms and software implementation methods for pattern recognition research. In *Computer Software and Applications Conference*, pages 55–58. IEEE, 1979.
- [Tri87] Hans Triebel. Characterizations of function spaces on a complete Riemannian manifold with bounded geometry. *Mathematische Nachrichten*, 130(1):321–346, 1987.
- [TSAY19] Nicolás García Trillos, Daniel Sanz-Alonso, and Ruiyi Yang. Local regularization of noisy point clouds: Improved global geometric estimates and data analysis. *J. Mach. Learn. Res.*, 20(136):1–37, 2019.
- [TSL00] Joshua B Tenenbaum, Vin de Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.
- [TSVB<sup>+</sup>16] Nicolás García Trillos, Dejan Slepčev, James Von Brecht, Thomas Laurent, and Xavier Bresson. Consistency of Cheeger and ratio graph cuts. *The Journal of Machine Learning Research*, 17(1):6268–6313, 2016.
- [Tsy08] A.B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer New York, 2008.
- [Tuk75] John Tukey. Mathematics and picturing data. In *International Congress of Mathematics*, volume 2, pages 523–531, 1975.
- [TY22] Rong Tang and Yun Yang. Minimax rate of distribution estimation on unknown submanifold under adversarial losses. *arXiv preprint arXiv:2202.09030*, 2022.
- [VB02] Pascal Vincent and Yoshua Bengio. Manifold Parzen windows. *Advances in neural information processing systems*, 15, 2002.
- [VDL<sup>+</sup>21] Hippolyte Verdier, Maxime Duval, François Laurent, Alhassan Cassé, Christian L Vestergaard, and Jean-Baptiste Masson. Learning physical properties of anomalous random walks using graph neural networks. *Journal of Physics A: Mathematical and Theoretical*, 54(23):234001, 2021.
- [vdVvZ08] Aad W van der Vaart and J Harry van Zanten. Rates of contraction of posterior distributions based on Gaussian process priors. *The Annals of Statistics*, 36(3):1435–1463, 2008.

- [VF05] Michel Verleysen and Damien François. The curse of dimensionality in data mining and time series prediction. In *International work-conference on artificial neural networks*, pages 758–770. Springer, 2005.
- [VSFD<sup>+</sup>22] Abhishek Varghese, Edgar Santos-Fernandez, Francesco Denti, Antonietta Mira, and Kerrie Mengersen. On the intrinsic dimensionality of Covid-19 data: a global perspective. *arXiv preprint arXiv:2203.04165*, 2022.
- [Was18] Larry Wasserman. Topological data analysis. *Annual Review of Statistics and Its Application*, 5:501–532, 2018.
- [WEG87] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.
- [WPS05] K.Q. Weinberger, B. D. Packer, and L. K. Saul. Nonlinear dimensionality reduction by semidefinite programming and kernel matrix factorization. In *Artificial Intelligence and Statistics*, 2005.
- [WS06a] Kilian Q Weinberger and Lawrence K Saul. An introduction to nonlinear dimensionality reduction by maximum variance unfolding. In *AAAI*, volume 6, pages 1683–1686, 2006.
- [WS06b] Kilian Q Weinberger and Lawrence K Saul. Unsupervised learning of image manifolds by semidefinite programming. *International journal of computer vision*, 70(1):77–90, 2006.
- [WW20] Hau-Tieng Wu and Nan Wu. Strong uniform consistency with rates for kernel density estimators with general kernels on manifolds. *arXiv preprint arXiv:2007.06408*, 2020.
- [XJRN02] Eric Xing, Michael Jordan, Stuart J Russell, and Andrew Ng. Distance metric learning with application to clustering with side-information. *Advances in neural information processing systems*, 15, 2002.
- [XY15] Fuyong Xing and Lin Yang. Fast cell segmentation using scalable sparse manifold learning and affine transform-approximated active contour. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 332–339. Springer, 2015.
- [YJ06] Liu Yang and Rong Jin. Distance metric learning: A comprehensive survey. *Michigan State University*, 2(2):4, 2006.
- [Yu97] Bin Yu. Assouad, Fano, and Le Cam. In *Festschrift for Lucien Le Cam*, pages 423–435. Springer, 1997.
- [ZS00] Yijun Zuo and Robert Serfling. General notions of statistical depth function. *Annals of statistics*, pages 461–482, 2000.



## RÉSUMÉ

---

En statistique, l'*hypothèse des variétés* suppose que les données observées se répartissent autour de structures de faible dimension, appelées variétés. Ce postulat permet d'expliquer pourquoi les algorithmes d'apprentissage fonctionnent bien même sur des données en grande dimension, et est naturellement satisfait pour de nombreux jeux de données issus de la vie réelle.

Nous présentons dans cette thèse quelques contributions aux problèmes d'estimation de deux quantités sous cette hypothèse : la densité de la distribution sous-jacente, et le reach de son support. Pour l'estimation du reach, nous élaborons des stratégies basées sur des invariants géométriques, avec d'une part la fonction de défaut de convexité, et d'autre part, des mesures de distortion métrique, desquels nous obtenons des vitesses de convergence optimales au sens minimax. Concernant l'estimation de la densité, nous proposons deux approches : l'une s'appuyant sur l'étude fréquentiste d'un estimateur à noyaux, et une approche bayésienne non-paramétrique se reposant sur des mélanges de gaussiennes. Nous montrons que ces deux méthodes sont optimales et adaptatives en la régularité de la densité.

Enfin, nous examinons le comportement de certaines mesures de centralité dans des graphes aléatoires géométriques, l'étude duquel, bien que sans lien avec l'hypothèse des variétés, a des implications méthodologiques et théoriques qui peuvent être intéressantes dans tout cadre statistique.

## MOTS CLÉS

---

Statistique non-paramétrique, statistique en grande dimension, apprentissage sur variété, inférence géométrique, estimation de densité, théorie minimax, estimation adaptative, estimation du reach, inférence bayésienne, mesure de centralité, profondeur statistique

## ABSTRACT

---

In high-dimensional statistics, the *manifold hypothesis* presumes that the data lie near low-dimensional structures, called manifolds. This assumption helps explain why machine learning algorithms work so well on high-dimensional data, and is satisfied for many real-life data sets.

We present in this thesis some contributions regarding the estimation of two quantities in this framework: the density of the underlying distribution, and the reach of its support. For the problem of reach estimation, we suggest different strategies based on important geometric invariants — namely the convexity defect functions, and measures of metric distortions — from which we derive minimax-optimal rates of convergence. Regarding the problem of density estimation, we propose two approaches: one relying on the frequentist study of a kernel density estimator, and a Bayesian nonparametric approach based on location-scale mixtures of Gaussians. Both methods are shown to be optimal in most settings, and adaptive to the smoothness of the density.

Lastly, we examine the behavior of some centrality measures in random geometric graph, the study of which, although unrelated to the manifold hypothesis, bears methodological and theoretical implications that can be of interest in any statistical framework.

## KEYWORDS

---

Nonparametric statistics, high-dimensional statistics, manifold learning, manifold hypothesis, geometric inference, density estimation, minimax theory, adaptive estimation, reach estimation, Bayesian inference, centrality measure, data depth