Reconstructing measures on manifolds: an optimal transport approach

Vincent Divol

RECONSTRUCTING MEASURES ON MANIFOLDS: AN OPTIMAL TRANSPORT APPROACH

Vincent Divol *

Abstract. Assume that we observe i.i.d. points lying close to some unknown d-dimensional $\mathcal{C}^k$ submanifold M in a possibly high-dimensional space. We study the problem of reconstructing the probability distribution generating the sample. After remarking that this problem is degenerate for a large class of standard losses (L_p, Hellinger, total variation, etc.), we focus on the Wasserstein loss, for which we build an estimator, based on kernel density estimation, whose rate of convergence depends on d and the regularity $s \leq k-1$ of the underlying density, but not on the ambient dimension. In particular, we show that the estimator is minimax and matches previous rates in the literature in the case where the manifold M is a d-dimensional cube. The related problem of the estimation of the volume measure of M for the Wasserstein loss is also considered, for which a minimax estimator is exhibited.

1 Introduction

Density estimation is one of the most fundamental tasks in non-parametric statistics. If efficient methods (from both a theoretical and a practical point of view) exist when the ambient space is of low dimension, minimax rates of estimation become increasingly slow as the dimension increases. To overcome this so-called *curse of dimensionality*, some structural assumptions on the underlying probability are to be made in moderate to high dimensions, which may take different forms, including e.g. the existence of a parametric component [LLW07], the single-index model [LZZL13], sparsity assumptions [Tib96], or constraints on the shape of the support. We focus in this work on the latter, namely on the case where the probability distribution μ generating the observations is assumed to be concentrated around a submanifold M of $\mathbb{R}^D$, of dimension d smaller than D. This assumption, known as the manifold assumption, has been fruitfully studied, with an emphasis put on reconstructing different geometric quantities related to the manifold, such as M itself [GPPVW12, AL18, AL19, Div20], its homology groups [NSW08, BRS⁺12], its dimension [HA05, LJM09, KRW16] or its reach [AKC⁺19, BHHS20]. The topic of density estimation in the manifold setting has itself been studied for over thirty years, with the emphasis initially being put on reconstructing the density in the case where the manifold M is given—think for instance of datasets lying on the space of orthogonal matrices—notable works including [Hen90, Pel05, CGK⁺20]. Less attention has been dedicated to the more general

* *Inria Saclay & Université Paris-Saclay*, `firstname.lastname@inria.fr`

setting where the manifold M is unknown and acts as a nuisance parameter. Kernel density estimators on manifolds are designed in [BS17, WW20], where rates are exhibited, respectively in the case where the manifold has a boundary and in the case where the density is Hölder continuous. In [BH19], kernel density estimators are shown to be minimax, and an adaptive procedure is designed, based on Lepski's method, to estimate the unknown density in a point $x \in \mathbb{R}^D$ which is known to belong to the unknown (and possibly nonsmooth) manifold M.

To go beyond the pointwise estimation of μ, even the choice of a relevant loss is nontrivial. Indeed, most standard losses between probability measures (e.g. the L_p distance, the Hellinger distance or the Kullback-Leibler divergence) are degenerate when comparing mutually singular measures, which will typically be the case for measures on two distinct manifolds, even if they are very close to each other with respect to the Hausdorff distance. This implies that the estimation problem is degenerate from a minimax perspective when choosing such losses (see Theorem 2.13). On the contrary, the Wasserstein distances W_p, $1 \le p \le \infty$ are particularly adapted to this problem, as they are by design robust to small metric perturbations of the support of a measure.

Apart from this first motivation, the use of Wasserstein distances, and more generally of the theory of optimal transport, has shown to be an efficient tool in widely different recent problems of machine learning, with fast implementations and sound theoretical results (see e.g. [PC19] for a survey). From a statistical perspective, most of the attention has been dedicated to studying rates of convergence between a probability distribution μ and its empirical counterpart μ_n [Dud69, DSS13, FG15, SP18, WB19a, L$^+$20]. Unsurprisingly, if more regularity is assumed on μ, then it is possible to build estimators with smaller risks than the empirical measure μ_n. Assume for instance that μ is a probability distribution on the cube $[-1, 1]^D$, with density f of regularity s (measured through the Besov scale $B^s_{p,q}$). In this setting, it has been shown in [WB19b] that, given n i.i.d. points of law μ, the minimax rate (up to logarithmic factors) for the estimation of μ with respect to the Wasserstein distance W_p is of order

$$\begin{cases} n^{-\frac{s+1}{2s+D}} & \text{if } D \ge 3 \\ n^{-\frac{1}{2}} \log n & \text{if } D = 2 \\ n^{-\frac{1}{2}} & \text{if } D = 1, \end{cases} \tag{1.1}$$

and that this rate is attained by a modified linear wavelet density estimator. Our main contribution consists in extending the results of [WB19b] by allowing the support of the probability to be any d-dimensional compact $\mathcal{C}^k$ submanifold $M \subset \mathbb{R}^D$ for $k \ge 2$. More precisely, assume that some probability μ on M has a lower and upper bounded density f which belongs to the Besov space $B^s_{p,q}(M)$ for some $0 < s \le k-1$, $1 \le p < \infty$, $1 \le q \le \infty$ (see Section 2 for details). We first show (Theorem 3.1) that some weighted kernel density estimator that we integrate against the volume measure vol_M on M attains, for the W_p distance, the rate of estimation

$$\begin{cases} n^{-\frac{s+1}{2s+d}} & \text{if } d \geq 3 \\ n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}} & \text{if } d = 2 \\ n^{-\frac{1}{2}} & \text{if } d = 1. \end{cases} \tag{1.2}$$

In the case where the manifold M is unknown, we do not have access to the volume measure vol_M, so that the latter estimator is not computable. We therefore propose to estimate the volume measure vol_M in a preliminary step. Such an estimator $\widehat{\mathrm{vol}}_M$ is defined by using local polynomial estimation techniques from [AL19]. We show that this estimator is a minimax estimator of the volume measure up to logarithmic factors (Theorem 3.6), with a risk of order $(\log n/n)^{k/d}$. We then show (Theorem 3.7) that a weighted kernel density estimator integrated against $\widehat{\mathrm{vol}}_M$ attains the rate (1.2). Those rates are significantly faster than the rates of (1.1) if $d \ll D$ and are shown to be minimax up to logarithmic factors.

In Section 2, we define our statistical model and give some preliminary results on Wasserstein distances. In Section 3, we define kernel density estimators on a manifold M, and state our main results. Proofs of the main theorems are then given in Section 4 while additional proofs are found in the Appendix.

2 Preliminaries

2.1 Regularity of manifolds

For any $d > 0$, we write $\cdot$ for the dot product and $|v|$ for the norm of a vector $v \in \mathbb{R}^d$. The ball centered at $x \in \mathbb{R}^d$ of radius $h > 0$ is denoted by $\mathcal{B}(x, h)$. For $\Omega \subset \mathbb{R}^d$ a set and $x \in \mathbb{R}^d$, we let $d(x, \Omega) := \inf\{|x - y|,\ y \in \Omega\}$ be the distance from x to Ω and we write $\mathcal{B}_\Omega(x, h)$ for $\mathcal{B}(x, h) \cap \Omega$. Also, we let $\Omega^h := \{x \in \mathbb{R}^d,\ d(x, \Omega) < h\}$ be the h-tubular neighborhood of Ω. Given a tensor $A : (\mathbb{R}^{d_1})^i \to \mathbb{R}^{d_2}$ of order $i \geq 0$, the operator norm $\|A\|_{\mathrm{op}}$ is defined as $\|A\|_{\mathrm{op}} := \max\{A[v_1, \dots, v_i],\ |v_1|, \dots, |v_i| \leq 1\}$. Also, we let $A^* : \mathbb{R}^{d_2} \to \mathbb{R}^{d_1}$ denote the adjoint of the operator $A : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$.

Let $D > 0$ and let $\mathcal{M}_d$ be the set of all smooth d-dimensional connected submanifolds in $\mathbb{R}^D$ without boundary, endowed with the metric induced by the standard metric on $\mathbb{R}^D$. We denote by d_g the geodesic distance on M. The tangent space at a point $x \in M$ is denoted by $T_x M$. It is identified with a d-dimensional subspace of $\mathbb{R}^D$, and the orthogonal projection on $T_x M$ is denoted by π_x. We also let $\tilde{\pi}_x : \mathbb{R}^D \to T_x M$ be defined by $\tilde{\pi}_x(y) = \pi_x(y - x)$. We denote by $T_x M^\perp$ the normal space at $x \in M$. If $M_1 \in \mathcal{M}_{d_1}$, $M_2 \in \mathcal{M}_{d_2}$, $x \in M_1$ and $f : M_1 \to M_2$ is a $\mathcal{C}^l$ function, then we let $d^l f(x) : (T_x M_1)^l \to T_{f(x)} M_2$ be the lth differential of f at x and $\|f\|_{\mathcal{C}^l(M_1)} := \max_{0 \leq i \leq l} \sup_{x \in M_1} \|d^i f(x)\|_{\mathrm{op}}$. For $i = 1$, we write df for $d^1 f$, and if $d_1 \leq d_2$, then we define the Jacobian of f at $x \in M_1$ as $Jf(x) = \sqrt{\det(df(x)^* df(x))}$. We let $\mathcal{C}^l(M)$ be the space of all $\mathcal{C}^l$ functions $f : M \to \mathbb{R}$ (with possibly $l = \infty$) and for $f \in \mathcal{C}^1(M)$, we let ∇f denote the gradient of f. We also denote by $\nabla \cdot$ the divergence operator on M.

Let vol_M be the volume measure associated with the Riemannian metric on M (or equivalently vol_M is the d-dimensional Hausdorff measure restricted to M). We will denote the integration with respect to $\mathrm{dvol}_M(x)$ by $\mathrm{d}x$ when the context is clear. For $1 \le p \le \infty$, we let $L_p(M)$ be the set of measurable functions $f : M \to \mathbb{R}$ with finite p-norm $\|f\|_{L_p(M)} := (\int f \mathrm{dvol}_M)^{1/p}$ (and usual modification if $p = \infty$). We say that a locally integrable function is weakly differentiable if there exists a measurable section ∇f of the tangent bundle TM (uniquely defined almost everywhere) such that for all smooth vector fields w on M with compact support, we have

$$\int f(\nabla \cdot w)\, \mathrm{dvol}_M = -\int (\nabla f) \cdot w \, \mathrm{dvol}_M.$$

Furthermore, we will denote by $p^* \in [1, \infty]$ the number satisfying $\frac{1}{p} + \frac{1}{p^*} = 1$.

The key quantity used to describe the regularity of a manifold M is its reach $\tau(M)$. It is defined as the distance between M and its medial axis, that is the set of points $x \in \mathbb{R}^D$ for which there are at least two points of M which attain the distance from x to M. In particular, the projection π_M on the manifold M is defined on $M^{\tau(M)}$. Originally introduced in [Fed59], the reach $\tau(M)$ measures both the local regularity of M (namely its curvature) and its global regularity, see e.g. [AKC+19, BHHS20] or [DZ01, Section 6.6] for precise results on the relationships between the reach of a manifold and its geometry. We then measure the regularity of M through the regularity of local parametrizations of M (see [AL19]).

Definition 2.1. *Let $M \in \mathcal{M}_d$, and $\tau_{\min}, L > 0$, $k \ge 2$. Let $r_0 = (\tau_{\min} \wedge L)/4$. We say that M is in $\mathcal{M}^k_{d,\tau_{\min},L}$ if M is closed, of reach larger than $\tau_{\min}$ and if, for all $x \in M$, the projection $\tilde{\pi}_x : M \to T_x M$ is a local diffeomorphism in x, with inverse Ψ_x defined on $\mathcal{B}_{T_x M}(0, r_0)$, satisfying $\|\Psi_x\|_{\mathcal{C}^k(\mathcal{B}_{T_x M}(0,r_0))} \le L$.*

Remark 2.2. (i) For the sake of convenience, we use a definition slightly different from the definition of [AL19], where authors assume the existence of local parametrizations $\tilde{\Psi}_x$ having controlled $\mathcal{C}^k$ norms, with $\tilde{\Psi}_x$ not necessarily equal to the inverse Ψ_x of the orthogonal projection. However, our definition is not restrictive. Indeed, on can write $\Psi_x = \tilde{\Psi}_x \circ (\tilde{\pi}_x \circ \tilde{\Psi}_x)^{-1}$, where the $\mathcal{C}^k$ norm of $(\tilde{\pi}_x \circ \tilde{\Psi}_x)^{-1}$ is controlled by the inverse function theorem. Therefore, the $\mathcal{C}^k$ norm of Ψ_x can always be controlled by the $\mathcal{C}^k$ norms of other parametrizations $\tilde{\Psi}_x$. Both definitions can also be proven to be equivalent to assuming that the function $d^2(\cdot, M)$ has a controlled $\mathcal{C}^k$ norm on $M^{\tau(M)}$, see e.g. [PR84].

(ii) The value of the scale parameter r_0 is used for convenience. Other small scales could be used, or the radius r_0 could also be added as another parameter of the model, without any substantial gain in doing so.

2.2 Besov spaces on manifolds

Let $M \in \mathcal{M}^k_{d,\tau_{\min},L}$ for some $k \geq 2$, $\tau_{\min}, L > 0$. As stated in the introduction, minimax rates for the estimation of a given probability will depend crucially on the regularity of its density f, which is assumed to belong to some Besov space $B^s_{p,q}(M)$. We first introduce Sobolev spaces $H^l_p(M)$ on M for $l \leq k$ an integer, and Besov spaces on M are then defined by real interpolation.

Definition 2.3 (Sobolev space on a manifold)**.** *Let $0 \leq l \leq k$, $1 \leq p < \infty$ and let $f \in \mathcal{C}^\infty(M)$ function. We let*

$$\|f\|_{H^l_p(M)} := \max_{0\leq i\leq l} \left(\int \left\| d^i f(x) \right\|^p_{\mathrm{op}} \mathrm{dvol}_M(x) \right)^{1/p}. \tag{2.1}$$

The space $H^l_p(M)$ is the completion of $\mathcal{C}^\infty(M)$ for the norm $\|\cdot\|_{H^l_p(M)}$.

Remark 2.4 (On the case $p = \infty$)*.* The previous definition cannot be extended to the case $p = \infty$. Indeed, the completion of $\mathcal{C}^\infty(M)$ for the norm $\|\cdot\|_{H^l_\infty(M)}$ is equal to $\mathcal{C}^l(M)$, whereas for instance $H^0_\infty(M)$ should be equal to $L_\infty(M)$. For $l = 1$, the space $H^1_p(M)$ can equivalently be defined as the space of weakly differentiable functions f with $\|f\|_{H^1_p(M)} < \infty$, while this definition can be easily extended to the case $p = \infty$. In particular, if $f \in H^1_\infty(M)$, then one can verify that $f \circ \Psi_x \in H^1_\infty(\mathcal{B}_{T_xM}(0, r_0))$ for any $x \in M$. It follows from standard results on Sobolev spaces on domains that $f \circ \Psi_x$ is Lipschitz continuous (see e.g. [Bre10, Proposition 9.3]). Hence, f is also locally Lipschitz continuous. By Rademacher theorem, f is therefore almost everywhere differentiable, and its differential coincides with the weak differential. As a consequence, a function $f \in H^1_\infty(M)$ is Lipschitz continuous, with Lipschitz constant for the distance d_g equal to $\|f\|_{H^1_\infty(M)}$.

For $1 \leq p < \infty$, we introduce the negative homogeneous Sobolev norm $\|\cdot\|_{\dot{H}^{-1}_p(M)}$, defined, for $f \in L_p(M)$ with $\int f \mathrm{dvol}_M = 0$, by

$$\|f\|_{\dot{H}^{-1}_p(M)} := \sup\left\{ \int fg\mathrm{dvol}_M,\ \|\nabla g\|_{L_{p^*}(M)} \leq 1 \right\}, \tag{2.2}$$

where the supremum is taken over all functions $g \in H^1_{p^*}(M)$. For $f \in L_p(M)$, the negative Sobolev norm is defined by

$$\|f\|_{H^{-1}_p(M)} := \sup\left\{ \int fg\mathrm{dvol}_M,\ \|g\|_{H^1_{p^*}(M)} \leq 1 \right\}, \tag{2.3}$$

and the corresponding Banach space is denoted by $H^{-1}_p(M)$.

Proposition 2.5. *Let $1 \leq p < \infty$ and $f \in H^{-1}_p(M)$ with $\int f \mathrm{dvol}_M = 0$.*

(i) We have $C_{d,\tau_{\min}} |\mathrm{vol}_M|^{\frac{d-1}{p}-d} \|f\|_{\dot{H}^{-1}_p(M)} \leq \|f\|_{H^{-1}_p(M)} \leq \|f\|_{\dot{H}^{-1}_p(M)}$ for some positive constant $C_{d,\tau_{\min}}$ depending on d and $\tau_{\min}$.

(ii) *We have* $\|f\|_{\dot{H}_p^{-1}(M)} = \inf\{\|w\|_{L_p(M)},\ \nabla\cdot w = f\}$*, where the infimum is taken over all measurable vector fields* w *on* M *with finite* p*-norm, and where* $\nabla \cdot w = f$ *means that* $\int f g \mathrm{dvol}_M = -\int w \cdot \nabla g \mathrm{dvol}_M$ *for all* $g \in \mathcal{C}^\infty(M)$*.*

Following [Tri92], Besov spaces on a manifold M are defined as real interpolation of Sobolev spaces.

Definition 2.6 (Real interpolation of spaces)**.** *Let* A_0*,* A_1 *be two Banach spaces, which continuously embed into some Banach space* A*. We endow the space* $A_0 \cap A_1$ *with the norm* $\|x\|_{A_0\cap A_1} = \max\{\|x\|_{A_0}, \|x\|_{A_1}\}$ *for* $x \in A_0 \cap A_1$ *and the space* $A_0 + A_1$ *with the norm* $K(x,1)$ *for* $x \in A_0 + A_1$*, where*

$$K(x,\lambda) := \inf\{\|x_0\|_{A_0} + \lambda\|x_1\|_{A_1},\ x = x_0 + x_1,\ x_0 \in A_0,\ x_1 \in A_1\},\quad \lambda \geq 0. \tag{2.4}$$

For $\theta \in [0,1]$ *and* $1 \leq q \leq \infty$*, we let*

$$\|x\|_{(A_0,A_1)_{\theta,q}} := \left(\int_0^\infty \lambda^{-\theta q} K(x,\lambda)^q \frac{\mathrm{d}\lambda}{\lambda}\right)^{1/q},\quad x \in A_0 + A_1, \tag{2.5}$$

and $(A_0,A_1)_{\theta,q} := \{x \in A_0 + A_1,\ \|x\|_{(A_0,A_1)_{\theta,q}} < \infty\}$ *(with usual modification if* $q = \infty$*). The pair* (A_0, A_1) *is called a compatible pair, and* $(A_0,A_1)_{\theta,q}$ *is the real interpolation between* A_0 *and* A_1 *of exponents* θ *and* q*.*

For A, B two Banach spaces and $F : A \to B$ a bounded operator, we let $\|F\|_{A,B}$ be the operator norm of F. Let (A_0, A_1) and (B_0, B_1) be two compatible pairs. Let $F : A_0 + A_1 \to B_0 + B_1$ be a linear map such that the restriction of F to A_j is a bounded linear map into B_j $(j = 0, 1)$. Then, the following interpolation inequality holds [Lun18, Theorem 1.1.6]

$$\|F\|_{(A_0,A_1)_{\theta,q},(B_0,B_1)_{\theta,q}} \leq \|F\|_{A_0,B_0}^{1-\theta}\|F\|_{A_1,B_1}^{\theta}. \tag{2.6}$$

Definition 2.7 (Besov space on a manifold)**.** *Let* $1 \leq p < \infty$ *and* $0 < s < k$*. The Besov space* $B^s_{p,q}(M)$ *is defined as* $B^s_{p,q}(M) := (L_p(M), H^k_p(M))_{s/k,q}$*.*

Basic results from interpolation theory then imply that $\|\cdot\|_{B^s_{p,q}(M)} \leq \|\cdot\|_{B^{s'}_{p,q}(M)}$ if $0 < s \leq s' < k$ (see e.g. [Lun18]).

2.3 Wasserstein distances and negative Sobolev distances

Let $\mathcal{P}$ be the set of finite Borel measures μ on $\mathbb{R}^D$, with $|\mu|$ the total mass of μ. Let $\mathcal{P}_1$ be the set of measures in $\mathcal{P}$ with $|\mu| = 1$. For $1 \leq p \leq \infty$, let $\mathcal{P}^p$ be the set of measures $\mu \in \mathcal{P}$ satisfying $(\int |x|^p \mathrm{d}\mu(x))^{1/p} < \infty$ (usual modification if $p = \infty$) and let $\mathcal{P}^p_1 = \mathcal{P}^p \cap \mathcal{P}_1$. The pushforward of a measure μ by a measurable application $\phi : \mathbb{R}^D \to \mathbb{R}^D$ is defined by

$$\phi_\# \mu(A) := \mu(\phi^{-1}(A)) \tag{2.7}$$

for any Borel set $A \subset \mathbb{R}^D$. For $\rho : \mathbb{R}^D \to [0, \infty)$ a measurable function, we denote by $\rho \cdot \mu$ the measure having density ρ with respect to μ.

Definition 2.8 (Wasserstein distance)**.** *Let $1 \le p \le \infty$ and let $\mu, \nu \in \mathcal{P}^p$ with the same total mass. Let $\Pi(\mu, \nu)$ be the set of transport plans between μ and ν, i.e. probability measures on $\mathbb{R}^D \times \mathbb{R}^D$ with first marginal π^1 (resp. second marginal π^2) equal to μ (resp. ν). The cost $C_p(\pi)$ of $\pi \in \Pi(\mu, \nu)$ is defined as $\int |x - y|^p \mathrm{d}\pi(x, y)$. The p-Wasserstein distance between μ and ν is defined as*

$$W_p(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} C_p(\pi)^{1/p}, \tag{2.8}$$

with usual modification if $p = \infty$.

A crucial point in the study conducted in the next sections is the relation between Wasserstein distances and negative Sobolev norms.

Proposition 2.9 (Wasserstein distances and negative Sobolev norms)**.** *Let $1 \le p < \infty$. Let $M \in \mathcal{M}_d$ be a manifold with reach $\tau(M) \ge \tau_{\min}$, and let $\mu, \nu \in \mathcal{P}_1^p$ be two probability measures supported on M, absolutely continuous with respect to vol_M. Assume that $\mu, \nu \ge f_{\min} \cdot \mathrm{vol}_M$ for some $f_{\min} > 0$. Then, identifying measures with their densities, we have*

$$\begin{aligned} W_p(\mu, \nu) &\le p^{-1/p} f_{\min}^{1/p-1} \|\mu - \nu\|_{\dot{H}_p^{-1}(M)} \\ &\le p^{-1/p} C_{d,\tau_{\min},f_{\min}} \|\mu - \nu\|_{H_p^{-1}(M)}, \end{aligned} \tag{2.9}$$

for some constant $C_{d,\tau_{\min},f_{\min}}$ depending on d, $\tau_{\min}$ and $f_{\min}$.

In particular, if $p = 1$, then the first inequality in (2.9) is actually an equality by the Kantorovitch-Rubinstein duality formula [Vil08, Particular Case 5.16]. This inequality appears in [Pey18] for $p = 2$ and in [San15, Section 5.5.1] for measures having density with respect to the Lebesgue measure. We carefully adapt their proofs in Appendix B.

2.4 Statistical models

Let $(\mathcal{Y}, \mathcal{H})$, $(\mathcal{X}, \mathcal{G})$ be measurable spaces and let $\mathcal{Q}$ be a subset of the space of probability measures on $(\mathcal{Y}, \mathcal{H})$. Assume that there is a measurable function $\iota : (\mathcal{Y}, \mathcal{H}) \to (\mathcal{X}, \mathcal{G})$ such that we observe i.i.d. variables $X_1, \dots, X_n \sim \iota_\# \xi$ for some $\xi \in \mathcal{Q}$. Those random variables are all defined on some probabilistic space $(\Omega, \mathcal{F}, \mathbb{P})$, and the integration with respect to $\mathbb{P}$ is denoted by $\mathbb{E}$. Let ϑ be a functional of interest defined on $\mathcal{P}$, taking its value in some measurable space $(E, \mathcal{E})$. The tuple $(\mathcal{Y}, \mathcal{H}, \mathcal{Q}, \iota, \vartheta)$ is a *statistical model*. Given i.i.d. observations with law $\iota_\# \xi$ where $\xi \in \mathcal{Q}$, the goal is to produce an estimator $\hat{\vartheta}$ (depending on the observations and the parameters defining $\mathcal{Q}$) such that its risk $\mathbb{E}\mathcal{L}(\hat{\vartheta}, \vartheta)$ is as small as possible, where $\mathcal{L}$ is some measurable loss function $\mathcal{L} : E \times E \to [0, \infty]$. The infimum of the risk over the estimators $\hat{\vartheta}$ is called the *minimax risk* for the estimation of ϑ on $\mathcal{Q}$ with respect to the loss $\mathcal{L}$:

$$\mathcal{R}_n(\vartheta, \mathcal{Q}, \mathcal{L}) := \inf_{\hat{\vartheta}} \sup_{\mu \in \mathcal{Q}} \mathbb{E}\mathcal{L}(\hat{\vartheta}, \vartheta), \tag{2.10}$$

where $\hat{\vartheta} = \hat{\vartheta}(\iota(X_1), \dots, \iota(X_n))$ and $X_1, \dots, X_n$ is an i.i.d. sample with law ξ. We consider the following models, where points are sampled on a manifold, with possibly tubular noise. We fix in the following some parameters $\tau_{\min}, L_s, L_k > 0$, $1 \le q \le \infty$ and $0 < f_{\min} < f_{\max} < \infty$. We also write $\mathcal{M}_d^k$ instead of $\mathcal{M}_{d,\tau_{\min},L_k}^k$.

Definition 2.10 (Noise free model)**.** *Let $d \le D$ be integers, $k \ge 2$, $0 \le s < k$ and $1 \le p < \infty$. Let $M \in \mathcal{M}_d^k$. For $s = 0$, the set $\mathcal{Q}^0(M)$ is the set of probability distributions μ on $\mathbb{R}^D$ absolutely continuous with respect to the volume measure vol_M, with a density f satisfying $f_{\min} \le f \le f_{\max}$ almost everywhere. For $s > 0$, the set $\mathcal{Q}^s(M)$ is the set of distributions $\mu \in \mathcal{Q}^0(M)$, with density $f \in B^s_{p,q}(M)$ satisfying $\|f\|_{B^s_{p,q}(M)} \le L_s$. The model $\mathcal{Q}_d^{s,k}$ is equal to the union of the sets $\mathcal{Q}^s(M)$ for $M \in \mathcal{M}_d^k$. The statistical model is completed by letting $(\mathcal{Y}, \mathcal{H})$ be $\mathbb{R}^D$ endowed with its Borel σ-algebra and ι and ϑ be the identity.*

Remark 2.11. If $\mu \in \mathcal{Q}^s(M)$, then, as $\mu \ge f_{\min}\mathrm{vol}_M$, one has $\mathrm{diam}(M) \le C_d/(f_{\min}\tau_{\min}^{d-1})$ for some constant C_d depending only on d [AL18, Lemma 2.2]. In particular, the manifold M is automatically compact.

Definition 2.12 (Tubular noise model)**.** *Let $d \le D$ be integers, $k \ge 2$, $0 \le s < k$, $1 \le p < \infty$ and $\gamma \ge 0$. The set $\mathcal{Q}_d^{s,k}(\gamma)$ is the set of probability distributions ξ of random variables (Y, Z) where $Y \sim \mu \in \mathcal{Q}_d^{s,k}$ and $Z \in \mathcal{B}(0, \gamma)$ is such that $Z \in T_Y M^\perp$. The statistical model is completed by letting $(\mathcal{Y}, \mathcal{H})$ be $\mathbb{R}^D \times \mathbb{R}^D$ endowed with its Borel σ-algebra, ι be the addition $\mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}^D$ and $\vartheta(\xi)$ be the first marginal μ of ξ.*

Concretely, a n-sample in the tubular noise model is given by $X_1, \dots, X_n$, where X_i is equal to $Y_i + Z_i$ with Y_i supported on some manifold M and $Z_i \in T_{Y_i} M^\perp$ is of norm smaller than γ. The goal is then to reconstruct the law μ of Y_i. We first show that such a task is impossible if the loss function $\mathcal{L}$ is larger than the total variation distance TV, which is defined by $\mathrm{TV}(\mu, \nu) := \sup_A |\mu(A) - \nu(A)|$ for $\mu, \nu \in \mathcal{P}_1$, where the supremum is taken over all measurable sets $A \subset \mathbb{R}^D$.

Theorem 2.13. *Let $d \le D$ be integers, $k \ge 2$, $0 \le s < k$, $1 \le p < \infty$. Let $\mathcal{L} : \mathcal{P} \times \mathcal{P} \to [0, \infty]$ be a measurable map with respect to the Borel σ-algebra associated to the total variation distance on $\mathcal{P} \times \mathcal{P}$. Assume that $\mathcal{L}(\mu, \nu) \ge g(\mathrm{TV}(\mu, \nu))$ for a convex nondecreasing function $g : \mathbb{R} \to [0, \infty]$ with $g(0) = 0$. Then, for any $\tau_{\min} > 0$, if $f_{\min}$ is small enough and $L_k, L_s, f_{\max}$ are large enough, we have*

$$\mathcal{R}_n(\mu, \mathcal{Q}_d^{s,k}, \mathcal{L}) \ge g(c_d), \tag{2.11}$$

for some constant $c_d > 0$.

Examples of such losses include the total variation distance, the Hellinger distance (with $g(x) = x$), the Kullback-Leibler divergence (with $g(x) = x^2/2$), and the L_p distance with respect to some dominating measure (with $g(x) = x^p$). We give a proof of Theorem 2.13, based on Assouad's lemma, in Appendix G. A simple example of loss $\mathcal{L}$ which is not degenerate for mutually singular measures is given by the W_p distance. As stated in the introduction, we will therefore choose this loss, and study $\mathcal{R}_n(\mu, \mathcal{Q}_d^{s,k}(\gamma), W_p)$, the minimax rate of estimation for μ with respect to W_p, where $\vartheta(\xi) = \mu$ is the first marginal of $\xi \in \mathcal{Q}_d^{s,k}(\gamma)$.

Remark 2.14. For $\gamma > 0$, the statistical model $\mathcal{Q}_d^{s,k}(\gamma)$ is not identifiable, in the sense that there exist ξ, ξ' in the model for which $\iota_\# \xi = \iota_\# \xi'$. Having such an equality implies that $W_p(\vartheta(\xi), \vartheta(\xi')) \leq W_p(\vartheta(\xi), \iota_\# \xi) + W_p(\iota_\# \xi', \vartheta(\xi')) \leq 2\gamma$. This inequality is tight up to a constant. Indeed, take Y an uniform random variable on the unit sphere, let ξ be the law of $(Y, 0)$ and ξ' be the law of $((1+\gamma)Y, -\gamma Y)$. Then, ξ and ξ' are in $\mathcal{Q}_d^{s,k}(\gamma)$ and $\iota_\# \xi = \iota_\# \xi'$, whereas, by the Kantorovitch-Rubinstein duality formula,

$$W_p(\vartheta(\xi), \vartheta(\xi')) \geq W_1(\vartheta(\xi), \vartheta(\xi')) \geq \mathbb{E}[\phi((1+\gamma)Y) - \phi(Y)]$$

for any 1-Lipschitz function ϕ. Letting ϕ be the distance to the unit sphere, we obtain that this distance is larger than γ. In that sense, γ represents the maximal precision for the estimation of $\vartheta(\xi)$.

Remark 2.15. For ease of notation, we will write in the following $a \lesssim b$ to indicate that there exists a constant C depending on the parameters $p, k, \tau_{\min}, L_s, L_k, f_{\min}, f_{\max}$, **but not on s and D**, such that $a \leq Cb$, and write $a \asymp b$ to indicate that $a \lesssim b$ and $b \lesssim a$. Also, we will write c_α to indicate that a constant c depends on some parameter α.

3 Kernel density estimation on an unknown manifold

Before building an estimator in the model $\mathcal{Q}_d^{s,k}(\gamma)$, let us consider the easier problem of the estimation of μ in the case where $\gamma = 0$ (noise free model) and the support M is known. Let $\mu \in \mathcal{Q}^s(M)$ and $Y_1, \dots, Y_n$ be a n-sample of law μ. Let $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{Y_i}$ be the empirical measure of the sample. Identify $\mathbb{R}^d$ with $\mathbb{R}^d \times \{0\}^{D-d}$ and consider a kernel $K : \mathbb{R}^D \to \mathbb{R}$ satisfying the following conditions:

- **Condition A:** The kernel K is a smooth radial function with support $\mathcal{B}(0,1)$ such that $\int_{\mathbb{R}^d} K = 1$.
- **Condition $B(m)$:** The kernel K is of order $m \geq 0$ in the following sense. Let $|\alpha| := \sum_{j=1}^d \alpha_j$ be the length of a multiindex $\alpha = (\alpha_1, \dots, \alpha_d)$. Then, for all multiindexes α^0, α^1 with $0 \leq |\alpha^0| < m$, $0 \leq |\alpha^1| < m + |\alpha_0|$, and with $|\alpha^1| > 0$ if $\alpha^0 = 0$, we have

$$\int_{\mathbb{R}^d} \partial^{\alpha^0} K(v) v^{\alpha^1} \mathrm{d}v = 0, \tag{3.1}$$

where $v^\alpha = \prod_{j=1}^d v_j^{\alpha_j}$ and $\partial^\alpha K$ is the partial derivative of K in the direction α.

- **Condition** $C(\beta)$**:** The negative part K_- of K satisfies $\int_{\mathbb{R}^d} K_- \leq \beta$.

We show in Appendix H that for every integer $m \geq 0$ and real number $\beta > 0$, there exists a kernel K satisfying conditions A, $B(m)$ and $C(\beta)$. Define the convolution of K with a measure $\nu \in \mathcal{P}$ as

$$K * \nu(x) := \int K(x-y)\mathrm{d}\nu(y), \quad x \in \mathbb{R}^D, \tag{3.2}$$

and, for $h > 0$, let $K_h := h^{-d}K(\cdot/h)$. Let $\rho_h := K_h * \mathrm{vol}_M$ and let $\mu_{n,h}$ be the measure with density $K_h * (\mu_n/\rho_h)$ with respect to vol_M. Dividing by ρ_h ensures that $\mu_{n,h}$ is a measure of mass 1. Remark that the computation of $\mu_{n,h}$ requires to have access to M, that is $\mu_{n,h}$ is an estimator on $\mathcal{Q}^s(M)$ but not on $\mathcal{Q}_d^{s,k}$. By linearity, the expectation of $\mu_{n,h}$ is given by μ_h, the measure having for density $K_h * (\mu/\rho_h)$ on M.

Theorem 3.1. *Let $d \leq D$ be integers, $0 < s \leq k-1$ with $k \geq 2$ and $1 \leq p < \infty$. Let $M \in \mathcal{M}_d^k$ and $\mu \in \mathcal{Q}^s(M)$ with $Y_1, \dots, Y_n$ a n-sample of law μ. There exists a constant β depending on the parameters of the model such that, if K is a kernel satisfying conditions A, $B(k)$ and $C(\beta)$, then the measure $\mu_{n,h}$ satisfies the following:*

(i) If $(\log n/n)^{1/d} \lesssim h \lesssim 1$, then, with probability larger than $1 - cn^{-k/d}$, the density of $\mu_{n,h}$ is larger than $f_{\min}/2$ and smaller than $2f_{\max}$ everywhere on M.

(ii) If $n^{-1/d} \lesssim h \lesssim 1$, then we have

$$\mathbb{E}\|\mu - \mu_{n,h}\|_{H_p^{-1}(M)} \leq \|\mu - \mu_h\|_{H_p^{-1}(M)} + \mathbb{E}\|\mu_{n,h} - \mu_h\|_{H_p^{-1}(M)} \tag{3.3}$$

$$\lesssim h^{s+1} + \frac{h^{1-d/2} I_d(h)}{\sqrt{n}}, \tag{3.4}$$

where $I_d(h) = 1$ if $d \geq 3$, $(-\log(h))^{1/2}$ if $d = 2$ and $h^{-1/2}$ if $d = 1$.

(iii) Let $h \asymp n^{-1/(2s+d)}$ if $d \geq 3$, $h \asymp (\log n/n)^{1/d}$ if $d \leq 2$. Define $\mu_{n,h}^0 = \mu_{n,h}$ if $\mu_{n,h}$ is a probability measure and $\mu_{n,h}^0 = \delta_{X_1}$ otherwise. Then,

$$\mathbb{E}W_p(\mu_{n,h}^0, \mu) \lesssim \begin{cases} n^{-\frac{s+1}{2s+d}} & \text{if } d \geq 3, \\ n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}} & \text{if } d = 2, \\ n^{-\frac{1}{2}} & \text{if } d = 1. \end{cases} \tag{3.5}$$

(iv) Furthermore, for any $0 \leq s < k$ and $\tau_{\min} > 0$, if $f_{\min}$ is small enough and if $f_{\max}$ and L_s are large enough, then there exists a manifold $M \in \mathcal{M}_d^k$ such that

$$\mathcal{R}_n(\mu, W_p, \mathcal{Q}^s(M)) \gtrsim \begin{cases} n^{-\frac{s+1}{2s+d}} & \text{if } d \geq 3, \\ n^{-\frac{1}{2}} & \text{if } d \leq 2. \end{cases} \tag{3.6}$$

Remark 3.2. The condition $C(\beta)$ on the kernel is only used to ensure that the measure $\mu_{n,h}$ has a lower and upper bounded density on M. An alternative possibility to ensure this property is to assume that the density of μ is Hölder continuous of exponent δ for some $\delta > 0$. Techniques from [BH19] then imply that $\|\mu_{n,h} - \mu\|_{L^\infty(M)} \lesssim h^\delta + n^{-1/2}h^{-d/2} \ll 1$ with high probability, ensuring in particular that the density is lowerbounded. If $sp > d$, then every element of $B^s_{p,q}(M)$ is Hölder continuous (by [Tri92, Theorem 7.4.2]), and condition $C(\beta)$ is no longer required. However, Theorem 3.1 also holds for non-continuous densities.

Remark 3.3. Let K be a nonnegative kernel satisfying conditions A, $B(0)$ and $C(\beta)$. It is straightforward to check that $W_p(\mu_n, \mu_{n,h}) \lesssim h$. Therefore, Theorem 3.1(ii) and Proposition 2.9 imply in particular that $W_p(\mu_n, \mu) \lesssim h + \frac{h^{1-d/2} I_d(h)}{\sqrt{n}}$. By choosing h of the order $n^{-1/d}$, we obtain that

$$W_p(\mu_n, \mu) \lesssim \begin{cases} n^{-\frac{1}{d}} \text{ if } d \geq 3 \\ n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}} \text{ if } d = 2 \\ n^{-\frac{1}{2}} \text{ if } d = 1. \end{cases} \tag{3.7}$$

Such a result was already shown for $p = \infty$ [TGHS20] with additional logarithmic factors, with a proof very different than ours. See also [Div21] for a short proof of this result when M is the flat torus.

In (3.4), a classical bias-variance trade-off appears. Namely, the bias of the estimator is of order h^{s+1}, whereas its fluctuations are of order $h^{1-d/2}/\sqrt{n}$ (at least for $d \geq 3$). This decomposition can be compared to the classical bias-variance decomposition for a kernel density estimator of bandwidth h, say for the pointwise estimation of a function of class $\mathcal{C}^s$ on the cube $[0,1]^d$. It is then well-known (see e.g. [Tsy08, Chapter 1]) that the bias of the estimator is of order h^s whereas its variance is of order $h^{-d/2}/\sqrt{n}$. The supplementary factor h appearing both in the bias and fluctuation terms can be explained by the fact that we are using a norm $H_p^{-1}(M)$ instead of a pointwise norm to quantify the risk of the estimator: in some sense, we are estimating the antiderivative of the density rather than the density itself. This is particularly true if $d = 1$ and $p = 1$, where the Wasserstein distance between two measures is given by the L_1 distance between the cumulative distribution functions of the two measures [San15, Proposition 2.17].

Before giving a proof of Theorem 3.1, let us explain how to extend it to the case where the manifold M is unknown and in the presence of tubular noise. The measure $\mu_{n,h}$ is the measure having density $K_h * (\mu_n/\rho_h)$ with respect to vol_M. Of course, if M is unknown, then so is vol_M, and we therefore propose the following estimation procedure of vol_M, using local polynomial estimation techniques from [AL19]. Let $X_1, \dots, X_n$ be a n-sample in the model with tubular noise $\mathcal{Q}_d^{s,k}(\gamma)$, with $X_i = Y_i + Z_i$, Y_i of law μ and $Z_i \in T_{Y_i}M^\perp$ with $|Z_i| \leq \gamma$. Let $\nu_n^{(i)}$ be the empirical measure $\frac{1}{n-1}\sum_{j\neq i} \delta_{X_j - X_i}$. For two positive parameters ℓ, ε, the local polynomial estimator $(\hat{\pi}_i, \hat{V}_{2,i}, \dots, \hat{V}_{m-1,i})$ of order m at X_j is defined as an element of

$$\operatorname*{arg\,min}_{\pi, \sup_{2\le j\le m-1}\|V_j\|_{\mathrm{op}}\le \ell} \nu_n^{(i)}\left(\left|x-\pi(x)-\sum_{j=2}^{m-1} V_j[\pi(x)^{\otimes j}]\right|^2 \mathbf{1}\{x\in\mathcal{B}(0,\varepsilon)\}\right), \tag{3.8}$$

where the argmin is taken over all orthogonal projectors π of rank d and symmetric tensors $V_j : (\mathbb{R}^D)^j \to \mathbb{R}^D$ of order j. Let $\hat{T}_i$ be the image of $\hat{\pi}_i$ and $\hat{\Psi}_i : v \in \mathbb{R}^D \mapsto X_i + v + \sum_{j=2}^{m-1} \hat{V}_{j,i}[v^{\otimes j}]$. Let $\angle(T_1, T_2)$ denote the angle between two d-dimensional subspaces, defined by $\|\pi_{T_1} - \pi_{T_2}\|_{\mathrm{op}}$, where π_{T_i} is the orthogonal projection on T_i for $i = 1, 2$. We summarize the results of [AL19] in the following proposition (see Appendix A for details).

Proposition 3.4. *With probability at least $1 - cn^{-k/d}$, if $m \le k$, $(\log n/n)^{1/d} \lesssim \varepsilon \lesssim 1$, $\gamma \lesssim \varepsilon$ and $1 \lesssim \ell \lesssim \varepsilon^{-1}$, then,*

$$\max_{1\le i\le n} \angle(T_{Y_i}M, \hat{T}_i) \lesssim \varepsilon^{m-1} + \gamma\varepsilon^{-1} \tag{3.9}$$

and, for all $1 \le i \le n$, if $v \in \hat{T}_i$ with $|v| \le 3\varepsilon$, we have

$$|\hat{\Psi}_i(v) - \Psi_{Y_i} \circ \pi_{Y_i}(v)| \lesssim \varepsilon^m + \gamma \tag{3.10}$$

$$\left\| d\hat{\Psi}_i(v) - d(\Psi_{Y_i} \circ \pi_{Y_i})(v) \right\|_{\mathrm{op}} \lesssim \varepsilon^{m-1} + \gamma\varepsilon^{-1}. \tag{3.11}$$

Hence, if γ is of order at most ε^k, then it is possible to approximate the tangent space at Y_i with precision ε^{k-1} and the local parametrization with precision ε^k. In particular, authors in [AL19] show that, with high probability, $\bigcup_{i=1}^n \mathcal{B}_{\hat{\Psi}_i(\hat{T}_i)}(X_i, \varepsilon)$ is at Hausdorff distance less than $\varepsilon^k + \gamma$ from M (up to a constant). We now define an estimator $\widehat{\mathrm{vol}}_M$ of vol_M by using an appropriate partition of unity $(\chi_j)_j$, which is built thanks to the next lemma. For $A, B \subset \mathbb{R}^D$, introduce the asymmetric Hausdorff distance $d_H(A|B) := \sup_{x\in A} d(x, B)$ and the Hausdorff distance $d_H(A, B) := d_H(A|B) \vee d_H(B|A)$. We say that a set S is δ-sparse if $|x - y| \ge \delta$ for all distinct points $x, y \in S$.

Lemma 3.5 (Construction of partitions of unity)**.** *Let $\delta \lesssim 1$. Let $S \subset M^\delta$ be a set which is $\frac{7}{3}\delta$-sparse, with $d_H(M^\delta|S) \le 4\delta$. Let $\theta : \mathbb{R}^D \to [0, 1]$ be a smooth radial function supported on $\mathcal{B}(0, 1)$, which is equal to 1 on $\mathcal{B}(0, 1/2)$. Define, for $y \in M^\delta$ and $x \in S$,*

$$\chi_x(y) = \frac{\theta\left(\frac{y-x}{8\delta}\right)}{\sum_{x'\in S} \theta\left(\frac{y-x'}{8\delta}\right)}. \tag{3.12}$$

Then, the sequence of functions $\chi_x : M^\delta \to [0, 1]$ for $x \in S$, satisfies (i) $\sum_{x\in S} \chi_x \equiv 1$, with at most c_d non zero terms in the sum at any given point of M^δ, (ii) $\|\chi_x\|_{\mathcal{C}^l(M^\delta)} \le C_{l,d}\delta^{-l}$ for any $l \ge 0$ and, (iii) χ_x is supported on $\mathcal{B}_{M^\delta}(x, 8\delta)$.

A proof of Lemma 3.5 is given in Appendix A. Given a set $S_0 \subset M^\delta$ with $d_H(M^\delta | S_0) \leq 5\delta/3$, the farthest sampling algorithm with parameter $7\delta/3$ (see e.g. [AL18, Section 3.3]) ouputs a set $S \subset S_0$ which is $7\delta/3$-sparse and $7\delta/3$-close from S: the set S then satisfies the hypothesis of Lemma 3.5. The next proposition describes how we may define a minimax estimator $\widehat{\mathrm{vol}}_M$ of the volume measure on M (up to logarithmic factors) using such a partition of unity.

Theorem 3.6 (Minimax estimation of the volume measure on M)**.** *Let $d \leq D$ be integers and $k \geq 2$. Let $\xi \in \mathcal{Q}_d^{0,k}(\gamma)$ and let $X_1, \dots, X_n$ be a n-sample of law $\iota_\# \xi$. Let $(\log n/n)^{1/d} \lesssim \varepsilon \lesssim 1$, $\gamma \lesssim \varepsilon$, $1 \lesssim \ell \lesssim \varepsilon^{-1}$.*

(i) Let $\{X_{i_1}, \dots, X_{i_J}\}$ be the output of the farthest point sampling algorithm with parameter $7\varepsilon/24$ and input $\{X_1, \dots, X_n\}$. With probability larger than $1 - cn^{-k/d}$, there exists a sequence of smooth nonnegative functions $\chi_j : M^{\varepsilon/8} \to [0,1]$ for $1 \leq j \leq J$, such that χ_j is supported on $\mathcal{B}_{M^{\varepsilon/8}}(X_{i_j}, \varepsilon)$, $\|\chi_j\|_{\mathcal{C}^1(M^{\varepsilon/8})} \lesssim \varepsilon^{-1}$ and $\sum_{j=1}^J \chi_j(z) = 1$ for $z \in M^{\varepsilon/8}$, with at most c_d non-zero terms in the sum.

(ii) Let $\hat{\Psi}_i$ be the local polynomial estimator of order $m \leq k$ with parameter ε and ℓ, and $\hat{T}_i$ the associated tangent space. Let $\widehat{\mathrm{vol}}_M$ be the measure defined by, for all continuous bounded functions $f : \mathbb{R}^D \to \mathbb{R}$,

$$\int f(x) \mathrm{d}\widehat{\mathrm{vol}}_M(x) = \sum_{j=1}^J \int_{\hat{\Psi}_{i_j}(\hat{T}_{i_j})} f(x) \chi_j(x) \mathrm{d}x, \tag{3.13}$$

where the integration is taken against the d-dimensional Hausdorff measure on $\hat{\Psi}_{i_j}(\hat{T}_{i_j})$. Then, for $1 \leq r \leq \infty$, with probability larger than $1 - cn^{-k/d}$, we have, for $\gamma \lesssim \varepsilon^2$,

$$W_r\left(\frac{\widehat{\mathrm{vol}}_M}{|\widehat{\mathrm{vol}}_M|}, \frac{\mathrm{vol}_M}{|\mathrm{vol}_M|}\right) \lesssim \gamma + \varepsilon^m. \tag{3.14}$$

(iii) In particular, if $m = k$, $\varepsilon \asymp (\log n/n)^{1/d}$ and $\gamma \lesssim \varepsilon^2$, we obtain that

$$\mathbb{E} W_r\left(\frac{\widehat{\mathrm{vol}}_M}{|\widehat{\mathrm{vol}}_M|}, \frac{\mathrm{vol}_M}{|\mathrm{vol}_M|}\right) \lesssim \gamma + \left(\frac{\log n}{n}\right)^{\frac{k}{d}}. \tag{3.15}$$

Also, for any $\tau_{\min} > 0$ and $0 \leq s < k$, if $f_{\min}$ is small enough, and if $f_{\max}, L_k, L_s$ are large enough, then

$$\mathcal{R}_n\left(\frac{\mathrm{vol}_M}{|\mathrm{vol}_M|}, \mathcal{Q}_d^{s,k}(\gamma), W_r\right) \gtrsim \gamma + \left(\frac{1}{n}\right)^{\frac{k}{d}}. \tag{3.16}$$

Let $\hat{\rho}_h := K_h * \widehat{\mathrm{vol}}_M$. We define $\hat{\nu}_{n,h}$ as the measure having density $K_h * (\nu_n / \hat{\rho}_h)$ with respect to the measure $\widehat{\mathrm{vol}}_M$, where $\nu_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ is the empirical measure of the sample $(X_1, \dots, X_n)$.

Theorem 3.7. *Let $d \leq D$ be integers, $0 < s \leq k-1$ with $k \geq 2$ and $1 \leq p < \infty$. Let $\xi \in \mathcal{Q}_d^{s,k}(\gamma)$, with μ the first marginal of ξ and let $X_1, \dots, X_n$ be a n-sample of law $\iota_\# \xi$. There exists a constant β depending on the parameters of the model such that the following holds. Assume that K is a kernel satisfying conditions A, B(k) and C(β), that $(\log n/n)^{1/d} \lesssim \varepsilon \lesssim h \lesssim 1$, $\gamma \lesssim \varepsilon^2$, $1 \lesssim \ell \lesssim \varepsilon^{-1}$ and consider the estimator $\widehat{\mathrm{vol}}_M$ defined in (3.13) with parameters m, ε and ℓ. Then,*

(i) The measure $\hat{\nu}_{n,h}$ is a nonnegative measure with probability larger than $1 - cn^{-k/d}$.

(ii) Define $\hat{\nu}^0_{n,h} = \hat{\nu}_{n,h}$ if $\hat{\nu}_{n,h}$ is a nonnegative measure and $\hat{\nu}^0_{n,h} = \delta_{X_1}$ otherwise. Then, with probability larger than $1 - cn^{-k/d}$,

$$W_p(\hat{\nu}^0_{n,h}, \mu^0_{n,h}) \lesssim \gamma + \varepsilon^m. \tag{3.17}$$

(iii) In particular, let $m = \lceil s+1 \rceil$, $\varepsilon \asymp (\ln n/n)^{1/d}$, $\ell \asymp \varepsilon^{-1}$ and $h \asymp n^{-1/(2s+d)}$ if $d \geq 3$, $h \asymp (\log n/n)^{1/d}$ if $d \leq 2$. Then,

$$\mathbb{E} W_p(\hat{\nu}^0_{n,h}, \mu) \lesssim \gamma + \begin{cases} n^{-\frac{s+1}{2s+d}} & \text{if } d \geq 3 \\ n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}} & \text{if } d = 2 \\ n^{-\frac{1}{2}} & \text{if } d = 1. \end{cases} \tag{3.18}$$

(iv) Furthermore, if $0 \leq s < k$ and $\tau_{\min} > 0$, for any $f_{\min}$ small enough and $f_{\max}$, L_s, L_k large enough, we have

$$\mathcal{R}_n(\mu, \mathcal{Q}_d^{s,k}(\gamma), W_p) \gtrsim \gamma + \begin{cases} n^{-\frac{s+1}{2s+d}} + n^{-\frac{k}{d}} & \text{if } d \geq 3, \\ n^{-\frac{1}{2}} & \text{if } d \leq 2. \end{cases} \tag{3.19}$$

Remark 3.8 (Numerical considerations). There are several considerations worth of interest concerning the numerical implementations of the estimators $\widehat{\mathrm{vol}}_M$ and $\hat{\nu}_{n,h}$. In a preprocessing step, one must first solve the optimization problem (3.8) for each element X_{i_j} of the output of the farthest point sampling algorithm. Let N_j be the number of points of the sample at distance less than ε from X_{i_j} (which is with high probability of order $n\varepsilon^d \asymp \log n$). For $k = 2$, minimizing (3.8) is equivalent to performing a PCA on the N_j neighbors of X_{i_j}, with a corresponding time complexity of order $\mathcal{O}(N_j^3)$ with high probability. For $k \geq 3$, as the space of orthogonal projectors of rank d is a non-convex manifold, the minimization of the objective function is more delicate. In [ZJRS16], a Riemannian SVRG procedure is proposed to minimize a functional defined on some Riemannian manifold. Their procedure outputs values whose costs are provably close to the minimal value of the objective function, even for non-convex smooth functions. The implementation of such an algorithm is a promising way to minimize (3.8) in practice.

The uniform measure on M can be approximated by considering the empirical measure $(\hat{U}_M)_N$ of a N-sample of law $\hat{U}_M := \widehat{\mathrm{vol}}_M / |\widehat{\mathrm{vol}}_M|$. To create such a sample, we may use importance sampling techniques to sample according to the measure with density χ_j on $\hat{\Psi}_{i_j}(\hat{T}_{i_j})$. Eventually, the measure $\hat{\nu}_{n,h}^{(N)}$ with density $K_h * (\nu_n/\hat{\rho}_h)$ with respect to $(\hat{U}_M)_N$ may be used as a proxy for $\hat{\nu}_{n,h}$.

4 Proofs of the main theorems

4.1 Bias of the kernel density estimator

The first step to prove Theorem 3.1 is to study the bias of the estimator, given by the distance $\|\cdot\|_{H_p^{-1}(M)}$ between μ_h and μ. Write $\tilde{\phi}$ for ϕ/ρ_h. Introduce the operator $A_h : B_{p,q}^s(M) \to H_p^{-1}(M)$ defined for $\phi \in L_1(M)$ and $x \in M$ by

$$A_h\phi(x) := K_h * \left(\frac{\phi(x)}{\rho_h(x)}\right) - \phi(x) = \int_M K_h(x-y)\left(\tilde{\phi}(y) - \tilde{\phi}(x)\right) \mathrm{dvol}_M(x). \tag{4.1}$$

Then,

$$\begin{aligned} \|\mu_h - \mu\|_{H_p^{-1}(M)} = \|A_h f\|_{H_p^{-1}(M)} &\leq \|A_h\|_{B_{p,q}^s(M), H_p^{-1}(M)} \|f\|_{B_{p,q}^s(M)} \\ &\leq \|A_h\|_{B_{p,q}^s(M), H_p^{-1}(M)} L_s. \end{aligned} \tag{4.2}$$

Proposition 4.1. *Let $0 < s \leq k-1$, $1 \leq p < \infty$, and assume that the kernel K is of order k. Then, if $h \lesssim 1$,*

$$\|A_h\|_{B_{p,q}^s(M), H_p^{-1}(M)} \lesssim h^{s+1}. \tag{4.3}$$

The proof of Proposition 4.1 consists in using the Taylor expansion of a function $\phi \in B_{p,q}^s(M)$, and by using that all polynomial terms of low order in the Taylor expansion disappear when integrated against K, as the kernel K is of sufficiently large order. Namely, we have the following property, whose proof is given in Appendix C.

Lemma 4.2. *Assume that the kernel K is of order k and let $B : (\mathbb{R}^D)^j \to \mathbb{R}$ be a tensor of order $1 \leq j < k$. Then, for all $x \in M$,*

$$\left|\int_M K_h(x-y) B[(x-y)^{\otimes j}] \mathrm{d}y\right| \lesssim \|B\|_{\mathrm{op}}\, h^k \tag{4.4}$$

$$|\rho_h(x) - 1| \lesssim h^{k-1} \quad \textit{and} \quad \|\rho_h\|_{\mathcal{C}^j(M)} \lesssim h^{k-1-j} \tag{4.5}$$

Let us now give a sketch of proof of Proposition 4.1 in the case $0 < s \leq 1$. The $H_p^{-1}(M)$-norm of $A_h\phi$ is by definition equal to

$$\|A_h\phi\|_{H_p^{-1}(M)} = \sup\left\{\int (A_h\phi) g \mathrm{dvol}_M,\ \|g\|_{H_{p^*}^1(M)} \leq 1\right\}.$$

Let $g \in H^1_{p^*}(M)$ with $\|g\|_{H^1_{p^*}(M)} \leq 1$. We use the following symmetrization trick:

$$\begin{aligned}
\int A_h\phi(x)g(x)\mathrm{d}x &= \iint K_h(x-y)(\tilde{\phi}(y)-\tilde{\phi}(x))g(x)\mathrm{d}y\mathrm{d}x \\
&= \iint K_h(y-x)(\tilde{\phi}(x)-\tilde{\phi}(y))g(y)\mathrm{d}y\mathrm{d}x \qquad \text{(by swapping the indexes } x \text{ and } y\text{)} \\
&= \frac{1}{2}\iint K_h(x-y)(\tilde{\phi}(y)-\tilde{\phi}(x))(g(x)-g(y))\mathrm{d}y\mathrm{d}x \qquad (4.6)
\end{aligned}$$

where, at the last line, we averaged the two previous lines and used that K is an even function. Informally, as $K_h(x-y)=0$ if $|x-y|\geq h$, and as ρ_h is roughly constant, we expect $|\tilde{\phi}(y)-\tilde{\phi}(x)|$ to be of order h^s and $|g(x)-g(y)|$ to be of order h, leading to a bound of $\int A_h\phi(x)g(x)\mathrm{d}x$ of order h^{s+1}. For $l\geq 1$, the following analogue of the symmetrization trick holds.

Lemma 4.3 (Symmetrization trick)**.** *There exists $h_0 \lesssim 1$ such that the following holds. Let $0\leq l\leq k-1$ be even and let $K^{(l)}(x)=\int_0^1 K_\lambda(x)\frac{(1-\lambda)^{l-1}\lambda^{-l}}{(l-1)!}\mathrm{d}\lambda$ for $x\in\mathbb{R}^D$. Fix $x_0\in M$ and let $\phi\in\mathcal{C}^\infty(M)$ be a function supported in $\mathcal{B}_M(x_0,h_0)$. Define $\tilde{\phi}_l := d^l(\tilde{\phi}\circ\Psi_{x_0})\circ\tilde{\pi}_{x_0}$. Let $g\in L_{p^*}(M)$ with $\|g\|_{L_{p^*}(M)}\leq 1$. Then, for $h\lesssim 1$, $\int A_h\phi(x)g(x)\mathrm{d}x$ is equal to*

$$\frac{1}{2}\iint_{\mathcal{B}_M(x_0,h_0)^2} K_h^{(l)}(x-y)(\tilde{\phi}_l(y)-\tilde{\phi}_l(x))\left[\pi_{x_0}(x-y)\right]^{\otimes l}(g(x)-g(y))\,\mathrm{d}y\mathrm{d}x + R, \qquad (4.7)$$

where R is a remainder term satisfying $|R|\lesssim \|\tilde{\phi}\|_{H^l_p(M)}h^{l+1}$. Furthermore, if $l\leq k-2$ is even, we have $|R|\lesssim\|\tilde{\phi}\|_{H^{l+1}_p(M)}h^{l+2}$.

Lemma 4.4. *Let $\eta\in\mathcal{C}^\infty(M)$ and let $0\leq l\leq k-2$. Assume that either $l=0$ or that η is supported on $\mathcal{B}_M(x_0,h_0)$. Let $\eta_l=d^l(\eta\circ\Psi_{x_0})\circ\tilde{\pi}_{x_0}$. Then, for any $h\lesssim 1$,*

$$\begin{aligned}
&\left(h^{-d}\iint_{\mathcal{B}_M(x_0,h_0)^2}\mathbf{1}\{|x-y|\leq h\}\frac{\|\eta_l(x)-\eta_l(y)\|^p_{\mathrm{op}}}{|x-y|^p}\mathrm{d}x\mathrm{d}y\right)^{1/p} \\
&\qquad\lesssim\left(\int_{\mathcal{B}_M(x_0,h_0)}\|\eta_{l+1}(x)\|^p_{\mathrm{op}}\,\mathrm{d}x\right)^{1/p}\lesssim\|\eta\|_{H^{l+1}_p(M)}.
\end{aligned} \qquad (4.8)$$

Proofs of Lemma 4.3 and Lemma 4.4 are found in Appendix C. Let $\phi\in\mathcal{C}^\infty(M)$ be a function supported in $\mathcal{B}_M(x_0,h_0)$ and $g\in H^1_{p^*}(M)$ with $\|g\|_{H^1_{p^*}(M)}\leq 1$.

Case 1: s is even Let $l=s$. Assume first that $p>1$ and that g is smooth. We have

$$\iint_{\mathcal{B}_M(x_0,h_0)^2}|K_{\lambda h}(x-y)|\left\|\tilde{\phi}_l(y)-\tilde{\phi}_l(x)\right\|_{\mathrm{op}}|g(x)-g(y)||x-y|^l\mathrm{d}x\mathrm{d}y \qquad (4.9)$$

$$
\begin{aligned}
&\leq \|K\|_{\mathcal{C}^0(\mathbb{R}^D)}(\lambda h)^{l+1-d}\iint_{\mathcal{B}_M(x_0,h_0)^2}\mathbf{1}\{|x-y|\leq\lambda h\}\left\|\tilde{\phi}_l(y)-\tilde{\phi}_l(x)\right\|_{\mathrm{op}}\frac{|g(x)-g(y)|}{|x-y|}\mathrm{d}x\mathrm{d}y\\
&\leq \|K\|_{\mathcal{C}^0(\mathbb{R}^D)}(\lambda h)^{l+1}\left((\lambda h)^{-d}\iint_{\mathcal{B}_M(x_0,h_0)^2}\mathbf{1}\{|x-y|\leq\lambda h\}\left\|\tilde{\phi}_l(y)-\tilde{\phi}_l(x)\right\|_{\mathrm{op}}^p\mathrm{d}x\mathrm{d}y\right)^{1/p}\\
&\qquad\times\left((\lambda h)^{-d}\iint_{\mathcal{B}_M(x_0,h_0)^2}\mathbf{1}\{|x-y|\leq\lambda h\}\frac{|g(x)-g(y)|^{p^*}}{|x-y|^{p^*}}\mathrm{d}x\mathrm{d}y\right)^{1/p^*}\\
&\lesssim(\lambda h)^{l+1}\left(2^p(\lambda h)^{-d}\int_{x\in\mathcal{B}_M(x_0,h_0)}\left\|\tilde{\phi}^l(x)\right\|_{\mathrm{op}}^p\mathrm{vol}_M(\mathcal{B}_M(x,\lambda h))\mathrm{d}x\right)^{1/p}\|g\|_{H^1_{p^*}(M)}\\
&\lesssim\|\tilde{\phi}\|_{H^l_p(M)}(\lambda h)^{l+1}\lesssim\|\phi\|_{H^l_p(M)}(\lambda h)^{l+1},
\end{aligned}
\tag{4.10}
$$

where at the last line, we used Lemma A.1(iii) to control the volume of $\mathcal{B}_M(x,\lambda h)$ and, at the second to last line, we used Lemma 4.4. Furthermore, it follows from Leibniz formula for the derivative of a product and Lemma 4.2 that $\|\tilde{\phi}\|_{H^l_p(M)}\lesssim\|\phi\|_{H^l_p(M)}$.

As $\mathcal{C}^\infty(M)$ is dense in $H^1_{p^*}(M)$, inequality (4.10) actually holds for every $g\in H^1_{p^*}(M)$. If $p=1$, then every function $g\in H^1_{p^*}(M)$ with $\|g\|_{H^1_{p^*}(M)}\leq 1$ is Lipschitz continuous for the distance d_g (see Remark 2.4). Using that $d_g(x,y)\leq 2|x-y|$ if $|x-y|\leq\tau_{\min}/4$ (see [AL18, Proposition 30]), a similar computation than in the case $p<\infty$ shows that inequality (4.10) also holds if $p=\infty$.

By integrating inequality (4.10) against $\lambda\in(0,1)$ and by using Lemma 4.3, we obtain the inequality $\|A_h\phi\|_{H^{-1}_p(M)}\lesssim h^{s+1}\|\phi\|_{H^s_p(M)}$.

Case 2: s is odd Similarly, we treat the case where $s\leq k-1$ is odd. Let $l=s-1$. Once again, assume first that $p>1$ and that g is smooth. Then,

$$
\begin{aligned}
&\iint_{\mathcal{B}_M(x_0,h_0)^2}|K_{\lambda h}(x-y)|\left\|\tilde{\phi}_l(y)-\tilde{\phi}^l(x)\right\|_{\mathrm{op}}|g(x)-g(y)||x-y|^l\mathrm{d}x\mathrm{d}y\\
&\leq\iint_{\mathcal{B}_M(x_0,h_0)^2}|K_{\lambda h}(x-y)|\frac{\left\|\tilde{\phi}_l(y)-\tilde{\phi}_l(x)\right\|_{\mathrm{op}}}{|x-y|}\frac{|g(x)-g(y)|}{|x-y|}|x-y|^{l+2}\mathrm{d}x\mathrm{d}y\\
&\leq\|K\|_{\mathcal{C}^0(\mathbb{R}^D)}(\lambda h)^{l+2-d}\iint_{\mathcal{B}_M(x_0,h_0)^2}\mathbf{1}\{|x-y|\leq\lambda h\}\frac{\left\|\tilde{\phi}_l(y)-\tilde{\phi}_l(x)\right\|_{\mathrm{op}}}{|x-y|}\frac{|g(x)-g(y)|}{|x-y|}\mathrm{d}x\mathrm{d}y\\
&\leq\|K\|_{\mathcal{C}^0(\mathbb{R}^D)}(\lambda h)^{l+2}\left((\lambda h)^{-d}\iint_{\mathcal{B}_M(x_0,h_0)^2}\mathbf{1}\{|x-y|\leq\lambda h\}\frac{\left\|\tilde{\phi}_l(y)-\tilde{\phi}_l(x)\right\|_{\mathrm{op}}^p}{|x-y|^p}\mathrm{d}x\mathrm{d}y\right)^{1/p}\\
&\qquad\times\left((\lambda h)^{-d}\iint_{\mathcal{B}_M(x_0,h_0)^2}\mathbf{1}\{|x-y|\leq\lambda h\}\frac{|g(x)-g(y)|^{p^*}}{|x-y|^{p^*}}\mathrm{d}x\mathrm{d}y\right)^{1/p^*}
\end{aligned}
$$

$$\lesssim (\lambda h)^{l+2} \|\phi\|_{H_p^s(M)}, \tag{4.11}$$

where at last line we used Lemma 4.4 and the inequality $\|\tilde{\phi}\|_{H_p^l(M)} \lesssim \|\phi\|_{H_p^l(M)}$. As in the previous case, the same inequality holds for $g \in H_{p^*}^1(M)$ non necessarily smooth and if $p = 1$. By using Lemma 4.3 and by integrating (4.11) against $\lambda \in (0,1)$, we obtain that $\|A_h\phi\|_{H_p^{-1}(M)} \lesssim h^{s+1}\|\phi\|_{H_p^s(M)}$.

So far, we have proven that

$$\|A_h\phi\|_{H_p^{-1}(M)} \lesssim h^{s+1}\|\phi\|_{H_p^s(M)} \tag{4.12}$$

for all integers $0 \le s \le k-1$ and ϕ a smooth function supported on $\mathcal{B}_M(x_0, h_0)$. To obtain the result when ϕ is not supported on some ball $\mathcal{B}_M(x_0, h_0)$, we use an appropriate partition of unity. Indeed, for $\delta = h_0/8$, standard packing arguments show the existence of a set S_0 of cardinality $N \le c_d|\mathrm{vol}_M|\delta^{-d}$ with $d_H(M^\delta|S_0) \le 5\delta/3$. By the remark following Lemma 3.5, the output S of the farthest point sampling algorithm with parameter $7\delta/3$ satisfies the assumption of Lemma 3.5, and is of cardinality smaller than $N \lesssim 1$. We consider such a covering $(\mathcal{B}_M(x, h_0))_{x\in S}$, with associated partition of unity $(\chi_x)_{x\in S}$ given by Lemma 3.5. Then, $\|A_h\phi\|_{H_p^{-1}(M)}$ is bounded by

$$\sum_{x\in S} \|A_h(\chi_x\phi)\|_{H_p^{-1}(M)} \lesssim h^{s+1}\sum_{x\in S}\|\chi_x\phi\|_{H_p^s(M)} \lesssim h^{s+1}\sum_{x\in S}\|\chi_x\|_{\mathcal{C}^s(M)}\|\phi\|_{H_p^s(M)} \lesssim h^{s+1}\|\phi\|_{H_p^s(M)},$$

where the second to last inequality follows from Leibniz rule for the derivative of a product. Also, the last inequality follows from the fact that $(\chi_x)_{|M} = \chi_x \circ i_M$, where $i_M : M \to M^\delta$ is the inclusion, which is a $\mathcal{C}^k$ function with controlled $\mathcal{C}^k$-norm. Hence, $\|\chi_x\|_{\mathcal{C}^s(M)} \lesssim \|\chi_x\|_{\mathcal{C}^s(M^\delta)} \lesssim 1$ by the chain rule.

As $\mathcal{C}^\infty(M)$ is dense in $H_p^s(M)$, this gives the desired bound on the operator norm of $A_h : H_p^s(M) \to H_p^{-1}(M)$ for $0 \le s \le k-1$ an integer. To obtain the conclusion for Besov spaces $B_{p,q}^s(M)$, we use the interpolation inequality (2.6). By the reiteration theorem [Lun18, Theorem 1.3.5], for $0 < s < k-1$, $B_{p,q}^s(M) = (L_p(M), H_p^{k-1}(M))_{s/(k-1),q}$, with an equivalent norm. Hence, we have, for $0 < s < k-1$,

$$\begin{aligned}\|A_h\|_{B_{p,q}^s(M), H_p^{-1}(M)} &\lesssim \|A_h\|_{L_p(M), H_p^{-1}(M)}^{1-\theta}\|A_h\|_{H_p^{k-1}(M), H_p^{-1}(M)}^{\theta} \\ &\lesssim h^{1-\frac{s}{k-1}}h^{k\frac{s}{k-1}} \lesssim h^{s+1},\end{aligned}$$

so that Proposition 4.1 is proven for $s < k-1$. It remains to prove the inequality in the case $s = k-1$. By Fatou's lemma and the definition of interpolation spaces (2.5), we have, for some constant C not depending on s,

$$\|A_h f\|_{B_{p,q}^{k-1}(M)} \le \liminf_{\substack{s\to k-1 \\ s<k-1}} \|A_h f\|_{B_{p,q}^s(M)} \le \liminf_{\substack{s\to k-1 \\ s<k-1}} \left(Ch^{s+1}\|f\|_{B_{p,q}^s(M)}\right) \le Ch^k\|f\|_{B_{p,q}^{k-1}(M)},$$

where we used that $\|f\|_{B_{p,q}^s(M)} \le \|f\|_{B_{p,q}^{k-1}(M)}$. This concludes the proof of Proposition 4.1.

4.2 Fluctuations of the kernel density estimator

The purpose of this section is to prove the following bound on the fluctuations of the kernel density estimator.

Proposition 4.5. *Let $\mu \in \mathcal{Q}^s(M)$ with $Y_1, \dots, Y_n$ a n-sample of law μ. Assume that $h \lesssim 1$ and that $nh^d \gtrsim 1$. Then,*

$$\mathbb{E}\|\mu_{n,h} - \mu_h\|_{H_p^{-1}(M)} \lesssim n^{-1/2} h^{1-d/2} I_d(h), \tag{4.13}$$

where $I_d(h)$ is defined in Theorem 3.1.

Let Δ be the Laplace-Beltrami operator on M and $G : U_M \to \mathbb{R}$ be a Green's function, defined on $\{(x,y) \in M \times M,\ x \neq y\}$ (see [Aub82, Chapter 4]). By definition, if $f \in \mathcal{C}^\infty(M)$, then the function $Gf : x \in M \mapsto \int G(x,y)f(y)\mathrm{d}y$ is a smooth function satisfying $\Delta Gf = f$, with $\nabla Gf(x) = \int \nabla_x G(x,y) f(y)\mathrm{d}y$ for $x \in M$. Hence, if $w = \nabla Gf$, then $\nabla \cdot w = f$, so that, Proposition 2.5 yields

$$\|f\|_{H_p^{-1}(M)} \leq \|f\|_{\dot{H}_p^{-1}(M)} \leq \|\nabla G f\|_{L_p(M)}.$$

By linearity, we have

$$\begin{aligned} \|\mu_{n,h} - \mu_h\|_{H_p^{-1}(M)} &= \|K_h(\mu_n - \mu)\|_{H_p^{-1}(M)} \\ &\leq \left\| \frac{1}{n}\sum_{i=1}^n \nabla G\left(K_h * \left(\frac{\delta_{Y_i}}{\rho_h(Y_i)}\right)\right) - \mathbb{E}\left[\nabla G\left(K_h * \left(\frac{\delta_{Y_i}}{\rho_h(Y_i)}\right)\right)\right]\right\|_{L_p(M)}. \end{aligned} \tag{4.14}$$

The expectation of the L_p-norm of the sum of i.i.d. centered functions is controlled thanks to the next lemma.

Lemma 4.6. *Let $U_1, \dots, U_n$ be i.i.d. functions on $L_p(M)$. Then, $\mathbb{E}\left\|\frac{1}{n}\sum_{i=1}^n (U_i - \mathbb{E}U_i)\right\|_{L_p(M)}^p$ is smaller than*

$$\begin{cases} n^{-p/2}\int \left(\mathbb{E}\left[|U_1(z)|^2\right]\right)^{p/2}\mathrm{d}z & \text{if } p \leq 2, \\ C_p n^{-p/2}\int \left(\mathbb{E}|U_1(z)|^2\right)^{p/2}\mathrm{d}z + C_p n^{1-p}\int_M \mathbb{E}\left[|U_1(z)|^p\right]\mathrm{d}z & \text{if } p > 2. \end{cases} \tag{4.15}$$

Proof. If $p \leq 2$, one has by Jensen's inequality

$$\mathbb{E}\left|\sum_{i=1}^n (U_i(z) - \mathbb{E}U_i(z))\right|^p \leq \left(\mathbb{E}\left|\sum_{i=1}^n (U_i(z) - \mathbb{E}U_i(z))\right|^2\right)^{p/2} \leq n^{p/2}\left(\mathbb{E}|U_1(z)|^2\right)^{p/2}$$

and (4.15) follows by integrating this inequality against $z \in M$. For $p > 2$, we use Rosenthal inequality [Ros70, Theorem 3] for a fixed $z \in M$, and then integrate the inequality against $z \in M$. □

It remains to bound $\mathbb{E}\left[\left|\nabla G\left(K_h * \left(\frac{\delta_Y}{\rho_h(Y)}\right)\right)(z)\right|^p\right]$ where $Y \sim \mu$, $z \in M$ and $p \geq 2$.

Lemma 4.7. *Let $p \geq 2$. Then, for all $z \in M$ and $h \lesssim 1$,*

$$\mathbb{E}\left[\left|\nabla G\left(K_h * \left(\frac{\delta_Y}{\rho_h(Y)}\right)\right)(z)\right|^p\right] \lesssim \begin{cases} 1 & \text{if } d = 1 \\ -\log h & \text{if } p = d = 2 \\ h^{p+d-dp} & \text{else.} \end{cases} \tag{4.16}$$

A proof of Lemma 4.7 is found in Appendix D. From (4.14), Lemma 4.6 and Lemma 4.7, we obtain, in the case $p \geq 2$ and $d \geq 3$

$$\begin{aligned} \mathbb{E}\|\mu_{n,h} - \mu_h\|_{H_p^{-1}(M)} &\leq \left(\mathbb{E}\|\mu_{n,h} - \mu_h\|^p_{H_p^{-1}(M)}\right)^{1/p} \\ &\leq C_p n^{-1/2}\left(\int\left(\mathbb{E}\left|\nabla G\left(K_h * \left(\frac{\delta_Y}{\rho_h(Y)}\right)\right)(z)\right|^2\right)^{p/2} \mathrm{d}z\right)^{1/p} \\ &\quad + C_p n^{1/p-1}\left(\int \mathbb{E}\left[\left|\nabla G\left(K_h * \left(\frac{\delta_Y}{\omega_h(Y)}\right)\right)(z)\right|^p\right] \mathrm{d}z\right)^{1/p} \\ &\lesssim n^{-1/2}|\mathrm{vol}_M|^{1/p} h^{1-d/2} + n^{1/p-1}|\mathrm{vol}_M|^{1/p} h^{1+d/p-d}. \end{aligned}$$

Recalling that $|\mathrm{vol}_M| \leq f_{\min}^{-1} \lesssim 1$ and that $nh^d \gtrsim 1$, one can check that this quantity is smaller up to a constant than $n^{-1/2}h^{1-d/2}$, proving Proposition 4.5 in the case $p \geq 2$ and $d \geq 3$. A similar computation shows that Proposition 4.5 also holds if $p \leq 2$ or $d \leq 2$.

4.3 Proof of Theorem 3.1

The proof of (i) is found in Appendix E. Let us now prove (ii). If $0 < s \leq k-1$, by Proposition 4.1 and (4.2), we have

$$\|\mu - \mu_h\|_{H_p^{-1}(M)} \leq L_s \|A_h\|_{B^s_{p,q}(M), H_p^{-1}(M)} \lesssim h^{s+1}.$$

Combining this inequality with Proposition 4.5 yields (3.4).

Let us prove (iii). Let E be the event described in (i). If E is realized, then $\mu^0_{n,h}$ is equal to $\mu_{n,h}$, and it satisfies $\mu^0_{n,h} \geq \frac{f_{\min}}{2}\mathrm{vol}_M$. Thus, Proposition 2.9 yields $W_p(\mu^0_{n,h}, \mu) \lesssim \|\mu_{n,h} - \mu\|_{H_p^{-1}(M)}$. If E is not realized, we bound $W_p(\mu^0_{n,h}, \mu)$ by $\mathrm{diam}(M)$, which is itself bounded by a constant depending only on the parameters of the model (see [AL18, Lemma 2.2]). Hence,

$$\mathbb{E}W_p(\mu^0_{n,h}, \mu) \leq \mathbb{E}\left[W_p(\mu^0_{n,h}, \mu)\mathbf{1}\{E\}\right] + \mathrm{diam}(M)\mathbb{P}(E^c) \lesssim \mathbb{E}\|\mu_{n,h} - \mu\|_{H_p^{-1}(M)} + n^{-k/d},$$

and we conclude thanks to (3.4).

Eventually, a proof of (iv) is found in Appendix G.

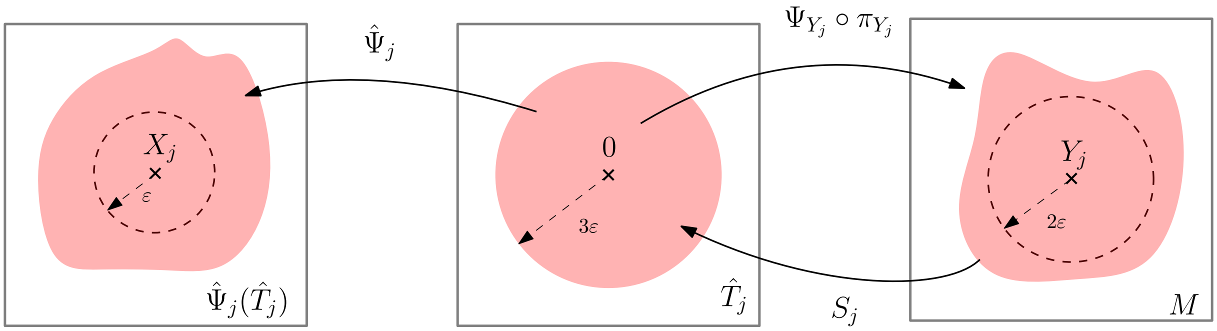

Figure 1 – Illustration of Lemma 4.8(a)

4.4 Proofs of Theorem 3.6 and Theorem 3.7

Proof of Theorem 3.6(i).

Assume that $\gamma \le \varepsilon/24$. Let $\mathbb{X} = \{X_1, \dots, X_n\}$ and $\mathbb{Y} = \{Y_1, \dots, Y_n\}$. By the remark following Lemma 3.5, the existence of a partition of unity satisfying the requirements of Theorem 3.6(i) is ensured as long as $d_H(M^{\varepsilon/8}|\mathbb{X}) \le 5\varepsilon/24$. We have $d_H(M^{\varepsilon/8}|\mathbb{X}) \le d_H(M^{\varepsilon/8}|\mathbb{Y}) + \varepsilon/24 \le d_H(M|\mathbb{Y}) + 4\varepsilon/24$. Hence, the partition of unity exists if $d_H(M|\mathbb{Y}) \le \varepsilon/24$. This is satisfied with probability larger than $1 - cn^{-k/d}$ if $\varepsilon \gtrsim (\log n/n)^{1/d}$ by [Aam17, Lemma III.23].

Proof of Theorem 3.6(ii).

For ease of notation, we will assume that the output $\{X_{i_1}, \dots, X_{i_J}\}$ of the farthest point sampling algorithm is equal to $\{X_1, \dots, X_J\}$. Write ν_j for the measure having density χ_j with respect to the d-dimensional Hausdorff measure on $\hat\Psi_j(\hat T_j)$.

Lemma 4.8. *If $(\log n/n)^{1/d} \lesssim \varepsilon \lesssim 1$ and $\gamma \lesssim \varepsilon^2$, with probability larger than $1 - cn^{-k/d}$, for all $j = 1, \dots, J$:*

(a) The map $\Psi_{Y_j} \circ \pi_{Y_j} : \mathcal{B}_{\hat T_j}(0, 3\varepsilon) \to M$ is a diffeomorphism on its image, which contains $\mathcal{B}_M(Y_j, 2\varepsilon)$. Let $S_j : \mathcal{B}_M(Y_j, 2\varepsilon) \to \mathcal{B}_{\hat T_j}(0, 3\varepsilon)$ be the inverse of $\Psi_{Y_j} \circ \pi_{Y_j}$. Then, $\hat\Psi_j \circ S_j : \mathcal{B}_M(Y_j, 2\varepsilon) \to \hat\Psi_j(\hat T_j)$ is also a diffeomorphism on its image, which contains $\mathcal{B}_{\hat\Psi_j(\hat T_j)}(X_j, \varepsilon)$. Furthermore, for all $z \in \mathcal{B}_M(Y_j, 2\varepsilon)$, we have $|\hat\Psi_j \circ S_j(z) - X_j| \ge \frac{7}{8}|z - Y_j|$.

(b) The measure $(\hat\Psi_j \circ S_j)^{-1}_{\#}\nu_j$ has a density $\tilde\chi_j$ on M equal to

$$\tilde\chi_j(z) = \chi_j(\hat\Psi_j \circ S_j(z)) J(\hat\Psi_j \circ S_j)(z) \text{ for } z \in M, \tag{4.17}$$

where the function is extended by 0 for $z \in M \backslash \mathcal{B}_M(Y_j, 2\varepsilon)$.

(c) For $z \in \mathcal{B}_M(Y_j, 2\varepsilon)$, we have

$$|\hat{\Psi}_j \circ S_j(z) - z| \lesssim \varepsilon^m + \gamma, \tag{4.18}$$

$$|\tilde{\chi}_j(z) - \chi_j(z)| \lesssim \varepsilon^m + \gamma. \tag{4.19}$$

A proof of Lemma 4.8 is found in Appendix F. Let $\hat{M}_\varepsilon = \bigcup_{j=1}^J \mathcal{B}_{\hat{\Psi}_j(\hat{T}_j)}(X_j, \varepsilon)$ be the support of $\widehat{\mathrm{vol}}_M$.

Lemma 4.9. *Let $(\log n/n)^{1/d} \lesssim \varepsilon \lesssim 1$ and $\gamma \lesssim \varepsilon^2$. Fix $1 \le r \le \infty$ and let $\phi : M \to \mathbb{R}$, $\tilde{\phi} : \hat{M}_\varepsilon \to \mathbb{R}$ be functions satisfying $\phi_{\min} \le \phi, \tilde{\phi} \le \phi_{\max}$ for some positive constants $\phi_{\min}, \phi_{\max} > 0$. Assume further that for all $j = 1, \dots, J$ and for all $z \in M$ we have, $|\tilde{\phi}(\hat{\Psi}_j \circ S_j(z)) - \phi(z)| \le T \lesssim 1$. Then, with probability larger than $1 - cn^{-k/d}$, we have*

$$W_r\left(\frac{\tilde{\phi} \cdot \widehat{\mathrm{vol}}_M}{|\tilde{\phi} \cdot \widehat{\mathrm{vol}}_M|}, \frac{\phi \cdot \mathrm{vol}_M}{|\phi \cdot \mathrm{vol}_M|}\right) \lesssim C_0(T + \varepsilon^m + \gamma), \tag{4.20}$$

where C_0 depends on $\phi_{\min}$ and $\phi_{\max}$.

In particular, inequality (3.14) is a consequence of Lemma 4.9 with $\phi \equiv \tilde{\phi} \equiv 1$.

Proof. Assume first that $r < \infty$. We have the bound

$$\begin{aligned} W_r\left(\frac{\tilde{\phi} \cdot \widehat{\mathrm{vol}}_M}{|\tilde{\phi} \cdot \widehat{\mathrm{vol}}_M|}, \frac{\phi \cdot \mathrm{vol}_M}{|\phi \cdot \mathrm{vol}_M|}\right) &= \frac{1}{|\tilde{\phi} \cdot \widehat{\mathrm{vol}}_M|^{1/r}} W_r\left(\tilde{\phi} \cdot \widehat{\mathrm{vol}}_M, \phi \cdot \mathrm{vol}_M \frac{|\tilde{\phi} \cdot \widehat{\mathrm{vol}}_M|}{|\phi \cdot \mathrm{vol}_M|}\right) \\ &\le \frac{1}{|\tilde{\phi} \cdot \widehat{\mathrm{vol}}_M|^{1/r}} \left(W_r\left(\sum_{j=1}^J \tilde{\phi} \cdot \nu_j, \sum_{j=1}^J (\hat{\Psi}_j \circ S_j)_{\#}^{-1}(\tilde{\phi} \cdot \nu_j)\right) \right. \\ &\left. + W_r\left(\sum_{j=1}^J (\hat{\Psi}_j \circ S_j)_{\#}^{-1}(\tilde{\phi} \cdot \nu_j), \phi \cdot \mathrm{vol}_M \frac{|\tilde{\phi} \cdot \widehat{\mathrm{vol}}_M|}{|\phi \cdot \mathrm{vol}_M|}\right) \right) \end{aligned} \tag{4.21}$$

We use Proposition 2.9 to bound the second term in (4.21). By a change of variables, the density of $(\hat{\Psi}_j \circ S_j)_{\#}^{-1}(\tilde{\phi} \cdot \nu_j)$ is given by $\tilde{\phi}_j : z \mapsto \tilde{\phi}(\hat{\Psi}_j \circ S_j(z))\tilde{\chi}_j(z)$. With probability larger than $1 - cn^{-k/d}$, we have for $z \in M$, should $\varepsilon^m + \gamma$ be small enough,

$$\sum_{j=1}^J \tilde{\chi}_j(z) \ge \sum_{j=1}^J \chi_j(z) - Cc_d(\varepsilon^m + \gamma) \ge 1 - \frac{1}{2} = \frac{1}{2},$$

where c_d is the constant of Lemma 3.5. Therefore, the density of $\sum_{j=1}^J (\hat{\Psi}_j \circ S_j)_{\#}^{-1}(\tilde{\phi} \cdot \nu_j)$ is larger than $\phi_{\min}/2$. Remark also that $\tilde{\chi}_j(z) \le 2$ for any $z \in M$. Hence, we have according to Lemma

4.8, $|\tilde{\phi}_j(z) - \phi(z)\chi_j(z)| \leq T + 2\phi_{\max}|\chi_j(z) - \tilde{\chi}_j(z)| \lesssim T + \phi_{\max}(\varepsilon^m + \gamma)$ for some constant C_0. This gives the bound,

$$\begin{aligned} ||\tilde{\phi} \cdot \widehat{\mathrm{vol}}_M| - |\phi \cdot \mathrm{vol}_M|| &\leq \left\| \sum_{j=1}^J \tilde{\phi}_j - \phi \right\|_{L_1(M)} \leq \left\| \sum_{j=1}^J \tilde{\phi}_j - \phi \right\|_{L_r(M)} |\mathrm{vol}_M|^{1-1/r} \\ &\leq \left\| \sum_{j=1}^J \tilde{\phi}_j - \phi \right\|_{L_\infty(M)} |\mathrm{vol}_M| \leq C_0 |\mathrm{vol}_M| (T + \phi_{\max}(\varepsilon^m + \gamma)). \end{aligned} \tag{4.22}$$

Therefore, $\phi \frac{|\tilde{\phi} \cdot \widehat{\mathrm{vol}}_M|}{|\phi \cdot \mathrm{vol}_M|}$ is larger than

$$\phi_{\min} \left(1 - C_0 |\mathrm{vol}_M| \frac{T + \phi_{\max}(\varepsilon^m + \gamma)}{\phi_{\min}|\mathrm{vol}_M|} \right) \geq \phi_{\min} - C_0(T + \phi_{\max}(\varepsilon^m + \gamma)) \geq \frac{\phi_{\min}}{2}$$

if T, ε and γ are small enough. Hence, by Proposition 2.9 and using (4.22),

$$\begin{aligned} &W_r \left(\sum_{j=1}^J (\hat{\Psi}_j \circ S_j)^{-1}_{\#} (\tilde{\phi} \cdot \nu_j), \phi \cdot \mathrm{vol}_M \frac{|\tilde{\phi} \cdot \widehat{\mathrm{vol}}_M|}{|\phi \cdot \mathrm{vol}_M|} \right) \\ &\leq r^{-1/r} \left(\frac{2}{\phi_{\min}} \right)^{1-1/r} \left\| \sum_{j=1}^J \tilde{\phi}_j - \phi \frac{|\tilde{\phi} \cdot \widehat{\mathrm{vol}}_M|}{|\phi \cdot \mathrm{vol}_M|} \right\|_{H_r^{-1}(M)} \leq \left(\frac{2}{\phi_{\min}} \vee 1 \right) \left\| \sum_{j=1}^n \tilde{\phi}_j - \phi \frac{|\tilde{\phi} \cdot \widehat{\mathrm{vol}}_M|}{|\phi \cdot \mathrm{vol}_M|} \right\|_{L_r(M)} \\ &\leq \left(\frac{2}{\phi_{\min}} \vee 1 \right) \left(\left\| \sum_{j=1}^J \tilde{\phi}_j - \phi \right\|_{L_r(M)} + \frac{||\phi \cdot \mathrm{vol}_M| - |\tilde{\phi} \cdot \widehat{\mathrm{vol}}_M||}{|\phi \cdot \mathrm{vol}_M|} \|\phi\|_{L_r(M)} \right) \\ &\leq \left(\frac{2}{\phi_{\min}} \vee 1 \right) C_0(T + \phi_{\max}(\varepsilon^m + \gamma)) \left(|\mathrm{vol}_M|^{1/r} + \frac{|\mathrm{vol}_M|}{\phi_{\min}|\mathrm{vol}_M|} |\mathrm{vol}_M|^{1/r} \phi_{\max} \right) \\ &\leq C_{\phi_{\min},\phi_{\max}} \left(T + \varepsilon^m + \gamma \right), \end{aligned}$$

where we used that $|\mathrm{vol}_M| \leq f_{\min}^{-1} \lesssim 1$, and the constant $C_{\phi_{\min},\phi_{\max}}$ in the upper bound depending on $\phi_{\min}$ and $\phi_{\max}$, but not on r.

To bound the first term in (4.21), consider the transport plan $\sum_{j=1}^J (\mathrm{id}, (\hat{\Psi}_j \circ S_j)^{-1})_{\#}(\tilde{\phi} \cdot \nu_j)$, which has, according to Lemma 4.8, a cost bounded by

$$\sum_{j=1}^J \int |y - (\hat{\Psi}_j \circ S_j)^{-1}(y)|^r \mathrm{d}(\tilde{\phi} \cdot \nu_j)(y) \lesssim \phi_{\max} \left(\varepsilon^m + \gamma \right)^r |\widehat{\mathrm{vol}}_M|.$$

As $|\widehat{\mathrm{vol}}_M| \lesssim |\mathrm{vol}_M| + T + \phi_{\max}(\varepsilon^m + \gamma) \lesssim 1$, we obtain the desired bound. By letting $r \to \infty$, and remarking that the different constants involved are independent of r, we observe that the same bound holds for $r = \infty$. □

Remark 4.10. Inequality (4.22) with $\phi \equiv \phi' \equiv 1$ gives a bound on the distance between the total mass of $\widehat{\mathrm{vol}}_M$ and the volume $|\mathrm{vol}_M|$ of M: choosing $k = m$, it is of order $\varepsilon^k + \gamma$ with probability larger than $1 - cn^{-k/d}$.

Proof of Theorem 3.6(iii).

Inequality (3.15) is a consequence of Theorem 3.6(ii), whereas the lower bound on the minimax risk (3.16) is proven in Appendix G.

Proof of Theorem 3.7.

Note first that $\hat{\nu}_{n,h}$ is indeed a measure of mass 1. We show in Lemma F.2 that

$$T := \max_{j=1\dots J} \sup_{z\in\mathcal{B}(Y_j,\varepsilon)} \left| K_h * \left(\frac{\nu_n}{\hat{\rho}_h}\right)(\hat{\Psi}_j \circ S_j(z)) - K_h * \left(\frac{\mu_n}{\rho_h}\right)(z)\right|$$

satisfies $T \lesssim \varepsilon^m + \gamma$ with probability larger than $1 - cn^{-k/d}$. As $f_{\min}/2 \leq K_h * \mu_n \leq 2f_{\max}$ on M by Theorem 3.1(i), and as every $y \in \hat{M}_\varepsilon$ is in the image of $\hat{\Psi}_j \circ S_j$ for some $j = 1\dots J$, we have $f_{\min}/3 \leq K_h * \nu_n \leq 3f_{\max}$ on $\hat{M}_\varepsilon$ should $\varepsilon^k + \gamma$ be small enough. This proves Theorem 3.1(i) and, together with Lemma 4.9, this also proves Theorem 3.7(ii). Theorem 3.7(iii) is a consequence of Theorem 3.7(ii) and Theorem 3.7(iv) is proven in Appendix G.

Acknowledgements

I am grateful to Fréderic Chazal, Pascal Massart, Eddie Aamari, Clément Berenfeld and Clément Levrard for helpful discussions and valuable comments on different mathematical aspects of this work.

APPENDIX

A Geometric properties of $\mathcal{C}^k$ manifolds with positive reach and their estimators

Let $M \in \mathcal{M}^k_{d,\tau_{\min},L}$ for some $k \geq 2$ and $\tau_{\min}, L > 0$. Recall that the angle between two d-dimensional subspaces T_1 and T_2 is given by $\angle(T_1, T_2) := \|\pi_{T_1} - \pi_{T_2}\|_{\mathrm{op}} = \|\pi^{\perp}_{T_1} \circ \pi_{T_2}\|_{\mathrm{op}}$, where π_{T_1} (resp. π_{T_2}) is the orthogonal projection on T_1 (resp. T_2) and $\pi^{\perp}_{T_1} := \mathrm{id} - \pi_{T_1}$.

Lemma A.1. *Let $x, y \in M$. The following properties hold:*

(i) One has $|\pi^{\perp}_y(x - y)| \leq \frac{|x-y|^2}{2\tau_{\min}}$ and $\angle(T_xM, T_yM) \leq 2\frac{|y-x|}{\tau_{\min}}$.

(ii) *If $\pi_M(z) = x$ for some $z \in M^{\tau_{\min}}$, then $z - x \in T_xM^\perp$.*

(iii) *If $h \le \tau_{\min}/4$, then $c_d h^d \le \mathrm{vol}_M(\mathcal{B}_M(x,h)) \le C_d h^d$.*

(iv) *If $h \le r_0$, then $\mathcal{B}_M(x,h) \subset \Psi_x(\mathcal{B}_{T_xM}(0,h)) \subset \mathcal{B}_M(x, 8h/7)$. Also, if $u \in \mathcal{B}_{T_xM}(0,r_0)$, then $|u| \le |\Psi_x(u) - x| \le 8|u|/7$.*

(v) *There exists a map $N_x : \mathcal{B}_{T_xM}(0,r_0) \to T_xM^\perp$ satisfying $dN_x(0) = 0$, and such that, for $u \in \mathcal{B}_{T_xM}(0,r_0)$, we have $\Psi_x(u) = x + u + N_x(u)$ with $|N_x(u)| \le L|u|^2$.*

(vi) *There exist tensors $B_x^1, \dots, B_x^{k-1}$ of operator norm controlled by a constant depending on L, d, k and $\tau_{\min}$, such that, if $u \in T_xM$ satisfies $|u| \le C_{k,d,L}$, then $J\Psi_x(u) = 1 + \sum_{i=2}^{k-1} B_x^i[u^{\otimes i}] + R_x(u)$, with $|R_x(u)| \le C'_{k,d,L}|u|^k$.*

Proof. See Theorem 4.18 in [Fed59], Lemma 6 in [GW03] for (i), Theorem 4.8 in [Fed59] for (ii), and Proposition 8.7 in [AL18] for (iii). See Lemma A.2 in [AL19] for the second inclusion of balls in (iv), which also implies the second inequality in (iv). The first inclusion as well as the first inequality in (iv) follow from the fact that Ψ_x is the inverse of $\tilde{\pi}_x$, which is 1-Lipschitz.

By a Taylor expansion of Ψ_x at $u = 0$, we have $\Psi_x(u) = x + u + N_x(u)$, with $N_x(u) = \int_0^1 d^2\Psi_x(tu)[u^{\otimes 2}]\mathrm{d}t$. Hence, $|N_x(u)| \le L|u|^2$. Furthermore, as $\tilde{\pi}_x \circ \Psi_x(u) = u$, we have $\pi_x(N_x(u)) = 0$, i.e. N_x takes its values in $T_xM^\perp$. This proves (v).

Eventually, we prove (vi). We have $d\Psi_x(u) = \mathrm{id}_{T_xM} + dN_x(u)$, and $d\Psi_x(u)^* d\Psi_x(u) = \mathrm{id}_{T_xM} + (dN_x(u))^* dN_x(u)$. Therefore,

$$J\Psi_x(u) = \sqrt{\det(d\Psi_x(u)^* d\Psi_x(u))} = \sqrt{\det(\mathrm{id}_{T_xM} + (dN_x(u))^* dN_x(u))}.$$

One has $dN_x(u) = dN_x(0) + \sum_{j=2}^{k-1} \frac{d^j N_x(0)}{(j-1)!}[u^{\otimes(j-1)}] + R_x(u)$, with $|R_x(u)| \le C_{k,L}|u|^{k-1}$ and $dN_x(0) = 0$. Hence, $(dN_x(u))^* dN_x(u)$ is written as $\sum_{j=2}^{k-1} B_j[u^{\otimes j}] + R'_x(u)$, with $|R'_x(u)| \le C'_{k,l}|u|^k$. The operator norm of this operator is smaller than, say, 1/2 for $|u|$ sufficiently small, and we conclude the proof by writing a Taylor expansion at 0 of the function $F \mapsto \sqrt{\det(\mathrm{id} + F)}$. ☐

We now prove Lemma 3.5, on the construction of smooth partitions of unity based on some set S which is sufficiently sparse and dense over a tubular neighborhood of M.

Proof of Lemma 3.5. Consider the functions θ and $(\chi_x)_{x\in S}$ as in the statement of the lemma, and, for $y \in M^\delta$, let $Z(y) = \sum_{x'\in S} \theta\left(\frac{y-x'}{8\delta}\right)$. As $d_H(M^\delta|S) \le 4\delta$, we have $Z(y) \ge 1$ and the quantity $\chi_x(y)$ is well defined. The function χ_x is smooth and we have $\sum_{x\in S} \chi_x \equiv 1$ on M^δ. One has $d^l\chi_x(y)$ which is written as a sum of terms of the form $d^{l-j}\theta\left(\frac{y-x}{8\delta}\right) d^j(Z^{-1})(y)$, and $d^j(Z^{-1})(y)$ is equal to a sum of terms of the form $Z^{j'-j-2}(y) d^{j'} Z(y)$ for $1 \le j' \le j$. Also,

$\left\| d^j\theta\left(\frac{y-x'}{8\delta}\right)\right\|_{\mathrm{op}} \le C_j\delta^{-j}$ and $\|d^jZ(y)\|_{\mathrm{op}} \le C_j\delta^{-j}\sum_{x\in S}\mathbf{1}\{|x-y|\le 8\delta\}$. Hence, as $Z\ge 1$, we have for any $l\ge 0$

$$\left\| d^l\chi_x(y)\right\|_{\mathrm{op}} \le C_l'\delta^{-l}\sum_{x\in S}\mathbf{1}\{|x-y|\le 8\delta\}.$$

It remains to bound this sum. If $x\in\mathcal{B}(y,8\delta)$, then $\pi_M(x)\in\mathcal{B}(\pi_M(y),10\delta)$. Also, for $x\neq x'\in S$, we have $|\pi_M(x)-\pi_M(x')|\ge |x-x'|-2\delta\ge 2\delta$. In particular, the balls $\mathcal{B}_M(\pi_M(x),\delta)$ for $x\in S$ are pairwise disjoint, and are all included in $\mathcal{B}_M(\pi_M(y),11\delta)$. Therefore, if $11\delta\le\tau(M)/4$, using Lemma A.1(iii) twice, we obtain that $\mathrm{vol}_M(\mathcal{B}_M(\pi_M(x),\delta))\ge c_d\delta^d$, and that

$$\sum_{x\in S}\mathbf{1}\{|x-y|\le 8\delta\}\le\sum_{x\in S}\mathbf{1}\{|x-y|\le 8\delta\}\frac{\mathrm{vol}_M(\mathcal{B}_M(\pi_M(x),\delta))}{c_d\delta^d}\le\frac{\mathrm{vol}_M(\mathcal{B}_M(\pi_M(y),11\delta))}{c_d\delta^d}\le c_d'.$$

This concludes the proof. □

We end this section by detailing the properties of the local polynomial estimators $\hat{\Psi}_i$ and $\hat{T}_i$ defined in [AL19]. In particular, we prove Proposition 3.4. Recall that $X_i=Y_i+Z_i$ with $Y_i\in M$ and $|Z_i|\le\gamma$. Aamari and Levrard introduce tensors $V_{j,i}^*$ which are defined as $d^j\Psi_{X_i}(0)/j!$, where $d^j\Psi_{X_i}(0)$ is the jth differential of Ψ_{X_i} at 0 (see the proof of Lemma 2 in [AL19] for details). In particular, we have $V_{1,i}^*=\pi_{Y_i}$. Furthermore, as $\tilde{\pi}_{Y_j}\circ\Psi_{Y_j}=\mathrm{id}$, we have $\pi_{Y_j}\circ V_{j,i}^*=0$ for $j\ge 2$.

Lemma A.2. *With probability larger than $1-cn^{-k/d}$, for any $1\le i\le n$,*

(i) We have $\angle(T_{Y_i}M,\hat{T}_i)\lesssim\varepsilon^{m-1}+\gamma\varepsilon^{-1}$.

(ii) For $v\in\hat{T}_i$, we have $\hat{\Psi}_i(v)=X_i+v+\hat{N}_i(v)$, where $\hat{N}_i:\hat{T}_i\to\hat{T}_i^{\perp}$ is defined by $\hat{N}_i(v)=\sum_{j=2}^{m-1}\hat{V}_{j,i}[v^{\otimes j}]$.

(iii) For any $2\le j<m$, $\left\|\hat{V}_{j,i}\circ\hat{\pi}_i-V_{j,i}^*\circ\pi_{Y_i}\right\|_{\mathrm{op}}\lesssim\varepsilon^{m-j}+\gamma\varepsilon^{-j}$.

(iv) For $v\in\mathcal{B}_{\hat{T}_i}(0,3\varepsilon)$, we have

$$|\hat{\Psi}_i(v)-\Psi_{Y_i}(\pi_{Y_i}(v))|\lesssim\varepsilon^m+\gamma, \tag{A.1}$$
$$|\hat{N}_i(v)-N_{Y_i}(\pi_{Y_i}(v))|\lesssim\varepsilon^m+\gamma, \tag{A.2}$$
$$\left\|d\hat{\Psi}_i(v)-d(\Psi_{Y_i}\circ\pi_{Y_i})(v)\right\|_{\mathrm{op}}\lesssim\varepsilon^{m-1}+\gamma\varepsilon^{-1} \tag{A.3}$$
$$\left\|d\hat{N}_i(v)-d(N_{Y_i}\circ\pi_{Y_i})(v)\right\|_{\mathrm{op}}\lesssim\varepsilon^{m-1}+\gamma\varepsilon^{-1}. \tag{A.4}$$

Proof of Proposition 3.4. Lemma A.2(i) is stated in Theorem 2 in [AL19]. Remark that for $x\in\mathcal{B}(X_i,\varepsilon)$, with $\tilde{x}=x-X_i$,

$$\left|\tilde{x}-\pi(\tilde{x})-\sum_{j=2}^{m-1}V_j[\pi(\tilde{x})^{\otimes j}]\right|^2=\left|\tilde{x}-\pi(\tilde{x})-\sum_{j=2}^{m-1}\pi^{\perp}\circ V_j[\pi(\tilde{x})^{\otimes j}]\right|^2+\left|\sum_{j=2}^{m-1}\pi\circ V_j[\pi(\tilde{x})^{\otimes j}]\right|^2$$

so that we may always assume that the tensors $\hat{V}_{j,i}$ minimizing the criterion (3.8) satisfy $\hat{\pi}_i \circ \hat{V}_{j,i} = 0$ for $j \geq 2$. This proves Lemma A.2(ii).

We prove Lemma A.2(iii) by induction on $2 \leq j < m$. The result for $j = 2$ is stated in [AL19, Theorem 2]. It is shown in [AL19] (see Equation (3)) that there exist tensors $V'_{j,i}$ for $1 \leq j < m$ satisfying with probability larger than $1 - cn^{-k/d}$,

$$\left\| V'_{j,i} \circ \pi_{Y_i} \right\|_{\text{op}} \lesssim \varepsilon^{m-j} + \gamma \varepsilon^{-j}. \tag{A.5}$$

The tensors $V'_{j,i}$ are defined by the relations, for $y \in M$ close enough to Y_i,

$$\begin{cases} y - Y_i = \pi_{Y_i}(y - Y_i) + \sum_{j=2}^{m-1} V^*_{j,i}[\pi_{Y_i}(y - Y_i)^{\otimes j}] + R(y - Y_i) \\ y - Y_i - \hat{\pi}_i(y - Y_i) - \sum_{j=2}^{m-1} \hat{V}_{j,i}[\hat{\pi}_i(y - Y_i)^{\otimes j}] = \sum_{j=1}^{m-1} V'_{j,i}[\pi_{Y_i}(y - Y_i)^{\otimes j}] + R'(y - Y_i), \end{cases}$$

with $|R(y-Y_i)|, |R'(y-Y_i)| \lesssim \varepsilon^m$, see the proof of Lemma 3 in [AL19]. In particular, for $j \geq 2$, noting that $\pi_{Y_i} \circ V^*_{j,i} = 0$, we see that $V'_{j,i} \circ \pi_{Y_i}$ is written as the sum of $(\pi_{Y_i} - \hat{\pi}_i) \circ V^*_{j,i} + (V^*_{j,i} \circ \pi_{Y_i} - \hat{V}_{j,i} \circ \hat{\pi}_i)$ and of a sum of terms proportional to

$$\hat{V}_{j',i}[\hat{\pi}_i \circ V^*_{a_1,i} \circ \pi_{Y_i}, \ldots, \hat{\pi}_i \circ V^*_{a_{j'},i} \circ \pi_{Y_i}], \tag{A.6}$$

where $2 \leq j' < j$ and $a_1 + \cdots + a_{j'} = j$, $1 \leq a_1, \ldots, a_{j'} < j$. There exists in particular an index in the sum which is larger than 2. Assume without loss of generality that $a_1, \ldots, a_l > 1$ and $a_{l+1}, \ldots, a_{j'} = 1$, so that $\hat{\pi}_i \circ \hat{V}_{a_u,i} = 0$ for $1 \leq u \leq l$. Then,

$$\begin{aligned}
&\left\| \hat{V}_{j',i}[\hat{\pi}_i \circ V^*_{a_1,i} \circ \pi_{Y_i}, \ldots, \hat{\pi}_i \circ V^*_{a_l,i} \circ \pi_{Y_i}, \ldots, \hat{\pi}_i \circ V^*_{a_{j'},i} \circ \pi_{Y_i}] \right\|_{\text{op}} \\
&= \left\| \hat{V}_{j',i}[\hat{\pi}_i \circ (V^*_{a_1,i} - \hat{V}_{a_1,i}) \circ \pi_{Y_i}, \ldots, \hat{\pi}_i \circ (V^*_{a_l,i} - \hat{V}_{a_l,i}) \circ \pi_{Y_i}, \ldots, \hat{\pi}_i \circ V^*_{a_{j'},i} \circ \pi_{Y_i}] \right\|_{\text{op}} \\
&\lesssim \ell \prod_{u=1}^{l} \left\| V^*_{a_u,i} \circ \pi_{Y_i} - \hat{V}_{a_u,i} \circ \pi_{Y_i} \right\|_{\text{op}} \\
&\lesssim \ell \prod_{u=1}^{l} \left(\left\| V^*_{a_u,i} \circ \pi_{Y_i} - \hat{V}_{a_u,i} \circ \hat{\pi}_i \right\|_{\text{op}} + \ell \left\| \pi_{Y_i} - \hat{\pi}_i \right\|_{\text{op}} \right) \\
&\lesssim \varepsilon^{-1} \prod_{u=1}^{l} \left(\varepsilon^{m-a_u} + \gamma \varepsilon^{-a_u} + \varepsilon^{m-2} + \gamma \varepsilon^{-2} \right) \lesssim \varepsilon^{-1}(\varepsilon^{lm-(j-l)} + \gamma^l \varepsilon^{-(j-l)}) \lesssim \varepsilon^{m-j} + \gamma \varepsilon^{-j},
\end{aligned}$$

where at the last line we use the induction hypothesis as well as Lemma A.2(i), the fact that $\sum_{u=1}^{l} a_u = j - l$ and that $\ell \lesssim \varepsilon^{-1}$. As $\left\| (\pi_{Y_i} - \hat{\pi}_i) \circ V^*_{j,i} \right\|_{\text{op}} \lesssim \varepsilon^{m-1} + \gamma \varepsilon^{-1}$, we obtain that

$$\left\| (V^*_{j,i} \circ \pi_{Y_i} - \hat{V}_{j,i} \circ \hat{\pi}_i) - V'_{j,i} \circ \pi_{Y_i} \right\|_{\text{op}} \lesssim \varepsilon^{m-j} + \gamma \varepsilon^{-j}.$$

Hence, using (A.5),

$$\left\|V^*_{j,i}\circ\pi_{Y_i}-\hat{V}_{j,i}\circ\hat{\pi}_i\right\|_{\mathrm{op}}\leq\left\|(V^*_{j,i}\circ\pi_{Y_i}-\hat{V}_{j,i}\circ\hat{\pi}_i)-V'_{j,i}\circ\pi_{Y_i}\right\|_{\mathrm{op}}+\left\|V'_{j,i}\circ\pi_{Y_i}\right\|_{\mathrm{op}}$$
$$\lesssim\varepsilon^{m-j}+\gamma\varepsilon^{-j}.$$

We now may prove (A.1). Indeed, for $v\in\mathcal{B}_{\hat{T}_i}(0,3\varepsilon)$, $\hat{\Psi}_i(v)=X_i+v+\sum_{j=2}^{m-1}\hat{V}_{j,i}[v^{\otimes j}]$, whereas by a Taylor expansion, $\Psi_{Y_i}\circ\pi_{Y_i}(v)=Y_i+\pi_{Y_i}(v)+\sum_{j=2}^{m-1}V_{j,i}[\pi_{Y_i}(v)^{\otimes j}]+R(v)$, with $|R(v)|\lesssim\varepsilon^m$. By Lemma A.2(iii), the difference between the two quantities is bounded with high probability by a sum of terms of order $(\varepsilon^{m-j}+\gamma\varepsilon^{-j})|v|^j\lesssim\varepsilon^m+\gamma$. Inequality (A.2) is directly implied by (A.1) and Lemma A.2(i). Inequality (A.3) is proven as (A.1), by noting that, for $h\in\hat{T}_i$,

$$\begin{cases} d(\Psi_{Y_j}\circ\pi_{Y_j})(v)[h]=\pi_{Y_j}(h)+\sum_{j=2}^{m-1}jV^*_{j,i}[\pi_{Y_j}(v),\pi_{Y_j}(h)^{\otimes(j-1)}]+R'(v)h \\ d\hat{\Psi}_j(v)[h]=h+\sum_{j=2}^{m-1}j\hat{V}_{j,i}[v,h^{\otimes(j-1)}], \end{cases}$$

with $\|R'(v)\|_{\mathrm{op}}\lesssim\varepsilon^{m-1}$. Equation (A.4) is shown in a similar way. □

B Properties of negative Sobolev norms

Proof of Proposition 2.5. The second inequality in (i) is trivial. The assertion (ii) is stated in [BCS10, Theorem 2.1] for an open set $\Omega\subset\mathbb{R}^d$, and their proof can be straightforwardly adapted to the manifold setting. It remains to prove the first inequality in (i). Note that for any g with $\|\nabla g\|_{L_{p^*}(M)}\leq 1$, one has $\int fg\mathrm{dvol}_M=\int f(g-\int g\mathrm{dvol}_M)\mathrm{dvol}_M$ as $\int f\mathrm{dvol}_M=0$. Also, by Poincaré inequality (see [BCH18, Theorem 0.6]),

$$\left\|g-\int_M g\right\|_{L_{p^*}(M)}\leq C^{\frac{1}{p}}R^{\frac{d}{p^*}+\frac{1}{p}}\|\nabla g\|_{L_{p^*}(M)}\leq C^{\frac{1}{p}}R^{\frac{d}{p^*}+\frac{1}{p}},$$

where $R=\max\{d_g(x,y),\ x,y\in M\}$ and C depends on d and on a lower bound κ on the Ricci curvature of M. Therefore, $\|g-\int_M g\|_{H^1_{p^*}(M)}\leq C^{\frac{1}{p}}R^{\frac{d}{p^*}+\frac{1}{p}}$. The quantity κ can be further lower bounded by a constant depending on $\tau_{\min}$ and d. Indeed, a bound on the second fundamental form of M entails a bound on the Ricci curvature according to Gauss equation (see e.g. [dC92, Chapter 6]), and the second fundamental form is controlled by the reach of M, see [NSW08, Proposition 6.1]. As $C^{\frac{1}{p}}\leq C\vee 1$, to conclude, it suffices to bound the geodesic diameter of M. This is done in the following lemma. □

Lemma B.1. *The geodesic diameter of M satisfies* $\sup_{x,y\in M}d_g(x,y)\leq c_d|\mathrm{vol}_M|\tau_{\min}^{1-d}$.

Proof. Consider a covering of M by N open balls of radius $r_1=\tau_{\min}/4$ (for the Euclidean distance) and let $x,y\in M$. Such a covering exists with $N\leq c_d|\mathrm{vol}_M|r_1^{-d}$ by standard packing

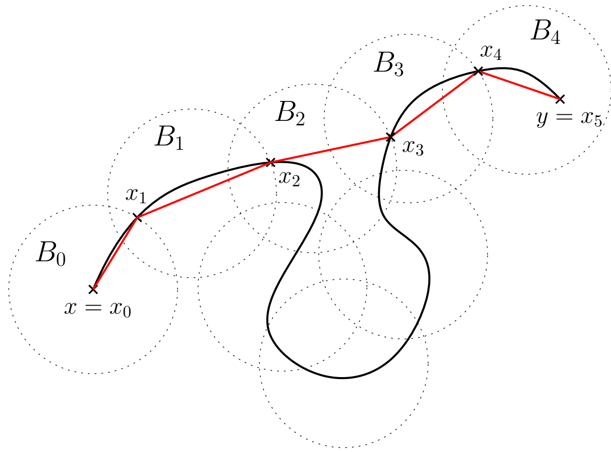

Figure 2 – Illustration of the construction in the proof of Lemma B.1

arguments. Let $\gamma : [0, \ell] \to M$ be a unit speed curve between x and y. Let B_0 be the ball of the covering such that $x \in B_0$. If $y \in B_0$, then $|x - y| \leq 2r_1$, and by [NSW08, Proposition 6.3], we have $d_g(x, y) \leq 4r_1$. Otherwise, let $t_0 = \inf\{t \in [0, \ell],\ \forall t' \geq t,\ \gamma(t') \notin B_0\}$. Then $x_1 := \gamma(t_0)$ belong to the boundary of B_0, and is also in some other ball B_1. By the previous argument, we have $d_g(x, x_1) \leq 4r_1$. If $y \in B_1$, then $d_g(x_1, y) \leq 4r_1$ and $d_g(x, y) \leq 8r_1$. Otherwise, we define $t_1 = \inf\{t \in [t_0, \ell],\ \forall t' \geq t,\ \gamma(t') \notin B_1\}$ and we iterate the same argument. At the end, we obtain a sequence $x = x_0, x_1, \ldots, x_I$ of points in M with associated balls B_i which contain x_i, such that $y \in B_I$ and $d_g(x_i, x_{i+1}) \leq 4r_1$. Furthermore, all the balls B_i are pairwise distinct. As $d_g(x_I, y) \leq 4r_1$, we have $\ell \leq (I + 1)4r_1 \leq (N + 1)4r_1 \leq 8Nr_1$. By letting γ be a geodesic, we obtain in particular $\ell = d_g(x, y) \leq 8Nr_1 \leq 8c_d|\mathrm{vol}_M|r_1^{1-d}$. □

Proof of Proposition 2.9. Given a measurable map $\rho : [0, 1] \to \mathcal{P}^p$, E_t a vectorial measure absolutely continuous with respect to ρ_t (see [San15, Box 4.2]) and $v(x, t)$ a time-depending vector field, defined as the density of E_t with respect to ρ_t, we define the Benamou-Brenier functional

$$\mathcal{B}_p(\rho, E) := \int |v(x, t)|^p \mathrm{d}\rho_t(x)\mathrm{d}t. \tag{B.1}$$

The Benamou-Brenier formula [BB00, Bre03] asserts that for $\mu, \nu \in \mathcal{P}_1^p$ supported on some ball of radius R,

$$W_p^p(\mu, \nu) = \min\left\{\mathcal{B}_p(\rho, E),\ \partial_t\rho_t + \nabla \cdot E_t = 0, \rho_0 = \mu, \rho_1 = \nu\right\}, \tag{B.2}$$

where ρ_t is supported on the ball of radius R, and the continuity equation $\partial_t\rho + \nabla \cdot E = \mu - \nu$ has to be understood in the distributional sense, i.e.

$$\int_{[0,1]\times\mathbb{R}^D} \partial_t\phi(t, x)\mathrm{d}\rho(t, x) + \int_{[0,1]\times\mathbb{R}^D} \nabla\phi(t, x) \cdot \mathrm{d}E(t, x) = 0, \tag{B.3}$$

for all $\phi \in \mathcal{C}^1((0,1)\times\mathcal{B}(0,R))$ with compact support.

Assume that μ has a density f_0 and ν has a density f_1 on M. As $\tau(M) > 0$, the existence of a probability measure of mass 1, supported on M, with density larger than $f_{\min}$ implies that M is compact, see Remark 2.11. It is in particular included in a ball $\mathcal{B}(0,R)$ for some R large enough. Let w be a vector field on M with $\nabla\cdot w = \mu-\nu$ in a distributional sense, i.e. $\int \nabla g\cdot w = -\int g(\mu-\nu)$ for all $g\in\mathcal{C}^1(M)$. Let $\rho_t = (1-t)\mu + t\nu$ and define E the vector measure having density w with respect to $\mathrm{Leb}_1\times \mathrm{vol}_M$, where Leb_1 is the Lebesgue measure on $[0,1]$. Then (ρ,E) satisfies the continuity equation and $E = v\cdot\rho$ where $v(t,x) = \frac{w(x)}{(1-t)f_0(x)+tf_1(x)}$ for $t\in[0,1]$, $x\in M$. Hence,

$$\begin{aligned} W_p^p(\mu,\nu) &\le \int_0^1\int \frac{1}{p}|v|^p\mathrm{d}\rho = \frac{1}{p}\int_0^1\int \frac{|w(x)|^p}{|(1-t)f_0(x)+tf_1(x)|^p}((1-t)f_0(x)+tf_1(x))\mathrm{d}x\mathrm{d}t \\ &\le \frac{1}{p}\int |w(x)|^p\mathrm{d}x\frac{1}{f_{\min}^{p-1}}. \end{aligned}$$

By taking the infimum on vector fields w on M satisfying $\nabla\cdot w = \mu-\nu$ and using Proposition 2.5, we obtain the conclusion. The second inequality in (2.9) follows from Proposition 2.5. □

C Proofs of Section 4.1

Proof of Lemma 4.2. We first prove (4.4). Note that if $|x-y|\ge h$ for $x,y\in M$, then $K_h(x-y) = 0$. Hence, by a change of variable, using that $\mathcal{B}_M(x,h)\subset\Psi_x(\mathcal{B}_{T_xM}(0,h))$ according to Lemma A.1(iv),

$$\begin{aligned} \int_M K_h(x-y)B[(x-y)^{\otimes j}]\mathrm{d}y &= \int_{\mathcal{B}_{T_xM}(0,h)} K_h(x-\Psi_x(v))B[(x-\Psi_x(v))^{\otimes j}]J\Psi_x(v)\mathrm{d}v \\ &= \int_{\mathcal{B}_{T_xM}(0,1)} K\left(\frac{x-\Psi_x(hv)}{h}\right)B[(x-\Psi_x(hv))^{\otimes j}]J\Psi_x(hv)\mathrm{d}v. \end{aligned}$$

As the functions Ψ_x and K are $\mathcal{C}^k$, according to Lemma A.1(v) and Lemma A.1(vi), we can write by a Taylor expansion, for $v,u\in\mathcal{B}_{T_xM}(0,r_0)$,

$$\begin{cases} \Psi_x(v) = x + v + \sum_{i=2}^{k-1}\frac{d^i\Psi_x(0)}{i!}[v^{\otimes i}] + R_1(x,v) \\ J\Psi_x(v) = 1 + \sum_{i=2}^{k-1}B_x^i[v^{\otimes i}] + R_2(x,v) \\ K(v+u) = K(v) + \sum_{i=1}^{k-1}\frac{d^iK(v)}{i!}[u^{\otimes i}] + R_3(v,u) \\ B[(v+u)^{\otimes j}] = B[v^{\otimes j}] + \sum_{\emptyset\neq\sigma\subset\{1,\dots,j\}}B[v^\sigma,u^{\sigma^c}], \end{cases} \tag{C.1}$$

where $|R_j(x,v)|\le C_j|v|^k$ for $j=1,2$, $|R_3(v,u)|\le C_3|u|^k$ and (v^σ,u^{σ^c}) is the j-tuple whose lth entry is equal to v if $l\in\sigma$, u otherwise. We obtain that

$$\frac{x-\Psi_x(hv)}{h} = -v - \sum_{i=2}^{k-1}\frac{d^i\Psi_x(0)}{i!}[(hv)^{\otimes i}]h^{-1} - R_1(x,hv)h^{-1},$$

and that the expression $K\left(\frac{x-\Psi_x(hv)}{h}\right)B[(x-\Psi_x(hv))^{\otimes j}]J\Psi_x(hv)$ is written as a sum of terms of the form

$$C_{i_0,i_1,i_2}h^{-i_0}d^{i_0}K(v)[(d^{i_1}\Psi_x(0)[(hv)^{\otimes i_1}])^{\otimes i_0}]F_{i_2}[(hv)^{\otimes i_2}] \tag{C.2}$$

for $0 \leq i_0 \leq k-1$, $2 \leq i_1 \leq k-1$ and $j \leq i_2 \leq k'$, where F_{i_2} is some tensor of order i_2 and k' is some integer depending on k and j, plus a remainder term smaller than $\|B\|_{\mathrm{op}}|hv|^{k-1+j}$ up to a constant depending on k, j, L_k and K. The terms for which $i_0i_1+i_2-i_0 \geq k$ are smaller than $\|B\|_{\mathrm{op}}h^k$ up to a constant, whereas the integrals of the other the terms are null as the kernel is of order k. The first inequality in (4.5) is proven in a similar manner. Let us now bound $\|\rho_h\|_{\mathcal{C}^j(M)}$. Given $x \in M$, it suffices to bound $\left\|d^j(\rho_h\circ\Psi_x)(0)\right\|_{\mathrm{op}}$. We have

$$d^j(\rho_h\circ\Psi_x)(0) = h^{-j}\int_{\mathcal{B}_{T_xM}(0,h)}(d^jK)_h(x-\Psi_x(v))J\Psi_x(v)\mathrm{d}v.$$

Therefore, using the same argument as before, we obtain that $\left\|d^j(\rho_h\circ\Psi_x)(0)\right\|_{\mathrm{op}} \lesssim h^{k-1-j}$. □

Proof of Lemma 4.3. Let $0 \leq l \leq k-1$ be even, $\phi \in \mathcal{C}^\infty(M)$ be supported in $\mathcal{B}_M(x_0,h_0)$ for some h_0 small enough and $g \in L_{p^*}(M)$ with $\|g\|_{L_{p^*}(M)} \leq 1$. Let $x = \Psi_{x_0}(u) \in \mathcal{B}_M(x_0,h_0)$ and let $\tilde{\phi}_{x_0} = \tilde{\phi}\circ\Psi_{x_0}$. Recall that $\tilde{\phi}_l = d^l\tilde{\phi}_{x_0}\circ\tilde{\pi}_{x_0}$. We have $K_h(x-\Psi_{x_0}(v)) \neq 0$ only if $|x-\Psi_{x_0}(v)| \leq h$. Hence, as $|x-\Psi_{x_0}(v)| \geq |u-v|$ (recall that Ψ_{x_0} is the inverse of the projection $\tilde{\pi}_{x_0}$), the function $K_h(x-\Psi_{x_0}(\cdot))$ is supported on $\mathcal{B}_{T_{x_0}M}(u,h) \subset \mathcal{B}_{T_{x_0}M}(0,r_0) =: B_0$ for h, h_0 small enough. Thus,

$$\begin{aligned} A_h\phi(x) &= \int_{\mathcal{B}_M(x,h)} K_h(x-y)(\tilde{\phi}(y)-\tilde{\phi}(x))\mathrm{d}y \\ &= \int_{B_0} K_h(x-\Psi_{x_0}(v))(\tilde{\phi}_{x_0}(v)-\tilde{\phi}_{x_0}(u))J\Psi_{x_0}(v)\mathrm{d}v. \end{aligned}$$

We may write

$$\tilde{\phi}_{x_0}(v)-\tilde{\phi}_{x_0}(u) = \sum_{i=1}^{l-1}\frac{d^i\tilde{\phi}_{x_0}(u)}{i!}[(v-u)^{\otimes i}] + \int_0^1 d^l\tilde{\phi}_{x_0}(u+\lambda(v-u))[(v-u)^{\otimes l}]\frac{(1-\lambda)^{l-1}}{(l-1)!}\mathrm{d}\lambda.$$

Each term $\int_{B_0}K_h(x-\Psi_{x_0}(v))\frac{d^i\tilde{\phi}_{x_0}(u)}{i!}[(v-u)^{\otimes i}]J\Psi_{x_0}(v)\mathrm{d}v$ is equal to

$$\int_M K_h(x-y)\frac{d^i\tilde{\phi}_{x_0}(\tilde{\pi}_{x_0}(x))}{i!}[(\pi_{x_0}(y-x))^{\otimes i}]\mathrm{d}y,$$

and is therefore of order smaller than $h^k\max_{1\leq i\leq l}\left\|\tilde{\phi}_i(x)\right\|_{\mathrm{op}}$ by Lemma 4.2. Hence, $A_h\phi(x)$ is equal to the sum of a remainder term of order $h^k\max_{1\leq i\leq l}\left\|\tilde{\phi}_i(x)\right\|_{\mathrm{op}}$ and of

$$\int_0^1\int_{B_0}K_h(x-\Psi_{x_0}(v))d^l\tilde{\phi}_{x_0}(u+\lambda(v-u))[(v-u)^{\otimes l}]\frac{(1-\lambda)^{l-1}}{(l-1)!}J\Psi_{x_0}(v)\mathrm{d}v\mathrm{d}\lambda$$

$$= \int_0^1 \int_{B_0} K_h(x - \Psi_{x_0}(v)) \left(d^l \tilde{\phi}_{x_0}(u + \lambda(v-u)) - d^l \tilde{\phi}_{x_0}(u) \right) [(v-u)^{\otimes l}] \frac{(1-\lambda)^{l-1}}{(l-1)!} J\Psi_{x_0}(v) \mathrm{d}v \mathrm{d}\lambda$$
$$+ R_1(x),$$

where $|R_1(x)| \lesssim h^k \max_{1 \le i \le l} \left\| \tilde{\phi}_i(x) \right\|_{\mathrm{op}}$ by Lemma 4.2. We now fix $\lambda \in (0,1)$ and write, by a change of variables, and as $\mathcal{B}_{T_{x_0}M}(u,h) \subset B_0$ for h_0, h small enough,

$$U(x) := \int_{B_0} K_h(x - \Psi_{x_0}(v)) \left(d^l \tilde{\phi}_{x_0}(u + \lambda(v-u)) - d^l \tilde{\phi}_{x_0}(u) \right) [(v-u)^{\otimes l}] J\Psi_{x_0}(v) \mathrm{d}v$$
$$= \int_{B_0} K_h \left(x - \Psi_{x_0} \left(u + \frac{w-u}{\lambda} \right) \right) \left(d^l \tilde{\phi}_{x_0}(w) - d^l \tilde{\phi}_{x_0}(u) \right) \left[\frac{(w-u)^{\otimes l}}{\lambda^l} \right] J\Psi_{x_0} \left(u + \frac{w-u}{\lambda} \right) \frac{\mathrm{d}w}{\lambda^d}$$

Note that $|K_h(u) - K_h(v)| \lesssim h^{-d-1}|u-v| \mathbf{1}\{|u| \le h \text{ or } |v| \le h\}$, and that, as Ψ_{x_0} is $\mathcal{C}^2$,

$$\left| x - \Psi_{x_0} \left(u + \frac{w-u}{\lambda} \right) - \frac{x - \Psi_{x_0}(w)}{\lambda} \right| \le \left| \frac{d\Psi_{x_0}(u)[w-u] - (x - \Psi_{x_0}(w))}{\lambda} \right| + \frac{L_k |w-u|^2}{2\lambda^2}$$
$$\le \frac{L_k |w-u|^2}{\lambda} \lesssim \frac{|w-u|^2}{\lambda},$$

whereas, as $J\Psi_{x_0}$ is Lipschitz continuous,

$$\left| J\Psi_{x_0} \left(u + \frac{w-u}{\lambda} \right) - J\Psi_{x_0}(w) \right| \lesssim \left| u + \frac{w-u}{\lambda} - w \right| \lesssim \frac{|w-u|}{\lambda}.$$

Hence, $U(x)$ is equal to the sum of

$$\lambda^{-l} \int_{B_0} K_{h\lambda} (x - \Psi_{x_0}(w)) \left(d^l \tilde{\phi}_{x_0}(w) - d^l \tilde{\phi}_{x_0}(u) \right) [(w-u)^{\otimes l}] J\Psi_{x_0}(w) \, \mathrm{d}w$$
$$= \lambda^{-l} \int_M K_{h\lambda} (x - y) \left(\tilde{\phi}_l(y) - \tilde{\phi}_l(x) \right) [(\pi_{x_0}(y-x))^{\otimes l}] \mathrm{d}y,$$

and of a remainder term smaller than

$$\lambda^{-l} \int_{B_0} \left| \lambda^{-d} K_h \left(x - \Psi_{x_0} \left(u + \frac{w-u}{\lambda} \right) \right) J\Psi_{x_0} \left(u + \frac{w-u}{\lambda} \right) - K_{h\lambda} (x - \Psi_{x_0}(w)) J\Psi_{x_0}(w) \right|$$
$$\times \left\| d^l \tilde{\phi}_{x_0}(w) - d^l \tilde{\phi}_{x_0}(u) \right\|_{\mathrm{op}} |w-u|^l \mathrm{d}w$$
$$\lesssim \lambda^{-l} \int_{|w-u| \lesssim \lambda h} \left(\frac{|w-u|^2}{(\lambda h)^{d+1}} J\Psi_{x_0} \left(u + \frac{w-u}{\lambda} \right) + |K_{h\lambda}(x - \Psi_{x_0}(w))| \frac{|w-u|}{\lambda} \right)$$
$$\times \left\| d^l \tilde{\phi}_{x_0}(w) - d^l \tilde{\phi}_{x_0}(u) \right\|_{\mathrm{op}} |w-u|^l \mathrm{d}w$$
$$\lesssim h^{l+1} (\lambda h)^{-d} \int_{|w-u| \lesssim \lambda h} \left\| d^l \tilde{\phi}_{x_0}(w) - d^l \tilde{\phi}_{x_0}(u) \right\|_{\mathrm{op}} \mathrm{d}w.$$

Putting all the estimates together, we may now write $\int_M A_h\phi(x)g(x)\mathrm{d}x$ as $S + R_2$, where, by the symmetrization trick (using that l is even)

$$\begin{aligned}
S &= \iint_{M\times M} K_h^{(l)}(x-y)\left(\tilde\phi_l(y)-\tilde\phi_l(x)\right)[(\pi_{x_0}(y-x))^{\otimes l}]g(x)\mathrm{d}y\mathrm{d}x \\
&= \iint_{M\times M} K_h^{(l)}(x-y)\left(\tilde\phi_l(x)-\tilde\phi_l(y)\right)[(\pi_{x_0}(x-y))^{\otimes l}]g(y)\mathrm{d}y\mathrm{d}x \\
&= \frac{1}{2}\iint_{M\times M} K_h^{(l)}(x-y)\left(\tilde\phi_l(y)-\tilde\phi_l(x)\right)[(\pi_{x_0}(x-y))^{\otimes l}](g(x)-g(y))\mathrm{d}y\mathrm{d}x,
\end{aligned}$$

and, as $A_h\phi$ is supported on $\mathcal{B}_M(x_0, h_0+h) \subset \mathcal{B}_M(x, 2h_0)$ if h is small enough, R_2 is smaller up to a constant than,

$$h^{l+1}(\lambda h)^{-d}\int_{x\in\mathcal{B}_M(x,2h_0)}\int_{|w-\tilde\pi_{x_0}(x)|\lesssim\lambda h}\left\|d^l\tilde\phi_{x_0}(w)-d^l\tilde\phi_{x_0}(\tilde\pi_{x_0}(x))\right\|_{\mathrm{op}}|g(x)|\mathrm{d}w\mathrm{d}x \tag{C.3}$$

$$+\int_M h^k \max_{1\le i\le l}\left\|\tilde\phi_i(x)\right\|_{\mathrm{op}}|g(x)|\mathrm{d}x$$

$$\lesssim h^{l+1}(\lambda h)^{-d}\int_{w\in\mathcal{B}_M(x,3h_0)}\left\|d^l\tilde\phi_{x_0}(w)\right\|_{\mathrm{op}}\int_{|w-\tilde\pi_{x_0}(x)|\lesssim\lambda h}|g(x)|\mathrm{d}x\mathrm{d}w \tag{C.4}$$

$$+h^{l+1}\int_{x\in\mathcal{B}_M(x,2h_0)}\left\|\tilde\phi_l(x)\right\|_{\mathrm{op}}|g(x)|\mathrm{d}x+\int_M h^k\max_{1\le i\le l}\left\|\tilde\phi_i(x)\right\|_{\mathrm{op}}|g(x)|\mathrm{d}x,$$

where we also used Lemma A.1(iii). By the chain rule,

$$\max_{1\le i\le l}\left\|\tilde\phi_i(x)\right\|_{\mathrm{op}} \lesssim \max_{1\le i\le l}\left\|d^i\tilde\phi(x)\right\|_{\mathrm{op}} \lesssim \sum_{i=1}^{l}\left\|d^i\tilde\phi(x)\right\|_{\mathrm{op}}.$$

Hence, applying Hölder's inequality and using that $\|g\|_{L_{p^*}(M)} \le 1$ show that the two last terms in (C.4) are of order $h^{l+1}\|\tilde\phi\|_{H_p^l(M)}$. To bound the first term in (C.4), remark that by Young's inequality for integral operators [Sog17, Theorem 0.3.1], if $\mathcal{T}_{\lambda h}(g)(y) = (\lambda h)^{-d}\int_{|x-y|\lesssim\lambda h}|g(x)|\mathrm{d}x$, then $\|\mathcal{T}_{\lambda h}g\|_{L_{p^*}(M)} \lesssim \|g\|_{L_{p^*}(M)}$. This yields, by Hölder's inequality,

$$h^{l+1}\int_{w\in\mathcal{B}_M(x,3h_0)}\left\|d^l\tilde\phi_{x_0}(w)\right\|_{\mathrm{op}}\mathcal{T}_{h\lambda}(g)(\Psi_{x_0}(w))\mathrm{d}w \lesssim h^{l+1}\|\tilde\phi\|_{H_p^l(M)},$$

which concludes the proof of the first statement of Lemma 4.3. To bound the remainder term in terms of $\|\tilde\phi\|_{H_p^{l+1}(M)}$, we bound the second term in (C.3) in the same fashion, while, to bound the first term, we write, by a change of variables,

$$\begin{aligned}
&\int_{\mathcal{B}_M(x_0,2h_0)}\int_{|w-\tilde\pi_{x_0}(x)|\lesssim\lambda h}\left\|d^l\tilde\phi_{x_0}(w)-d^l\tilde\phi_{x_0}(\tilde\pi_{x_0}(x))\right\|_{\mathrm{op}}|g(x)|\mathrm{d}x\mathrm{d}w \\
&\le \int_0^1\int_{\mathcal{B}_M(x_0,2h_0)}\int_{|w-\tilde\pi_{x_0}(x)|\lesssim\lambda h}\left\|d^{l+1}\tilde\phi_{x_0}(\tilde\pi_{x_0}(x)+\lambda'(w-\tilde\pi_{x_0}(x)))\right\|_{\mathrm{op}}|\tilde\pi_{x_0}(x)-w||g(x)|\mathrm{d}x\mathrm{d}w\mathrm{d}\lambda'
\end{aligned}$$

$$\lesssim h\int_0^1 \int_{\mathcal{B}_M(x_0,2h_0)} \int_{|u-\tilde{\pi}_{x_0}(x)|\lesssim \lambda'\lambda h} \left\|d^{l+1}\tilde{\phi}_{x_0}(u)\right\|_{\mathrm{op}} |g(x)|\mathrm{d}x \frac{\mathrm{d}u}{\lambda'^d}\mathrm{d}\lambda',$$

and this term is bounded as the first term in (C.4) by $h(h\lambda)^d\|\tilde{\phi}\|_{H_p^{l+1}(M)}$, concluding the proof of Lemma 4.3. □

Proof of Lemma 4.4. By the chain rule, $\left\|d^{l+1}(\eta\circ\Psi_{x_0})(u)\right\|_{\mathrm{op}} \lesssim \max_{1\le i\le l+1}\left\|d^i\eta(\Psi_{x_0}(u))\right\|_{\mathrm{op}}$ for any $u\in\mathcal{B}_{T_{x_0}M}(0,h_0)$. Hence, by a change of variables,

$$\begin{aligned}\int_{\mathcal{B}_M(x_0,h_0)} \|\eta_{l+1}(x)\|_{\mathrm{op}}^p \mathrm{d}x &\lesssim \int_{\mathcal{B}_{T_{x_0}M}(0,h_0)} \max_{1\le i\le l+1}\left\|d^i\eta(\Psi_{x_0}(u))\right\|_{\mathrm{op}}^p \mathrm{d}u\\ &\lesssim \sum_{i=1}^{l+1}\int_{\mathcal{B}_{T_{x_0}M}(0,h_0)} \left\|d^i\eta(\Psi_{x_0}(u))\right\|_{\mathrm{op}}^p \mathrm{d}u\\ &\lesssim \sum_{i=1}^{l+1}\int_{\mathcal{B}_{T_{x_0}M}(0,h_0)} \left\|d^i\eta(\Psi_{x_0}(u))\right\|_{\mathrm{op}}^p J\Psi_{x_0}(u)\mathrm{d}u \lesssim \|\eta\|_{H_p^{l+1}(M)}^p,\end{aligned}$$

where we used at last line that, by Lemma A.1(vi), $J\Psi_{x_0}(u)\ge 1/2$ for $|u|\le h_0$ if h_0 is small enough. To prove the first inequality, write

$$\begin{aligned}&h^{-d}\iint_{\mathcal{B}_M(x_0,h_0)^2} \mathbf{1}\{|x-y|\le h\}\frac{\|\eta_l(x)-\eta_l(y)\|_{\mathrm{op}}^p}{|x-y|^p}\mathrm{d}x\mathrm{d}y\\ &\lesssim h^{-d}\iint_{\mathcal{B}_{T_{x_0}M}(0,h_0)^2} \mathbf{1}\{|\Psi_{x_0}(u)-\Psi_{x_0}(v)|\le h\}\frac{\left\|d^l(\eta\circ\Psi_{x_0})(u)-d^l(\eta\circ\Psi_{x_0})(v)\right\|_{\mathrm{op}}^p}{|\Psi_{x_0}(u)-\Psi_{x_0}(v)|^p}\mathrm{d}u\mathrm{d}v\\ &\lesssim h^{-d}\int_0^1\iint_{\mathcal{B}_{T_{x_0}M}(0,h_0)^2} \mathbf{1}\{|u-v|\le h\}\left\|d^{l+1}(\eta\circ\Psi_{x_0})(u+\lambda(v-u))\right\|_{\mathrm{op}}^p \mathrm{d}u\mathrm{d}v\mathrm{d}\lambda\\ &\lesssim h^{-d}\int_0^1\iint_{\mathcal{B}_{T_{x_0}M}(0,2h_0)^2} \mathbf{1}\{|w-u|\le\lambda h\}\left\|d^{l+1}(\eta\circ\Psi_{x_0})(w)\right\|_{\mathrm{op}}^p \mathrm{d}u\mathrm{d}w\lambda^{-d}\mathrm{d}\lambda\\ &\lesssim \int_0^1\int_{\mathcal{B}_{T_{x_0}M}(0,2h_0)} \left\|d^{l+1}(\eta\circ\Psi_{x_0})(w)\right\|_{\mathrm{op}}^p \mathrm{d}w \lesssim \int_{\mathcal{B}_M(x_0,h_0)}\|\eta_{l+1}(x)\|_{\mathrm{op}}^p\mathrm{d}x,\end{aligned}$$

where at the second to last line, we used that $w=u+\lambda(v-u)$ is of norm smaller than $2h_0$ if $|u|\le h_0$ and $|v-u|\le h\le h_0$, and, at the last line, we used that $J\Psi_{x_0}(w)\ge 1/2$ for $|w|$ small enough. □

D Proof of Lemma 4.7

Lemma 4.7 is heavily based on the following classical control on the gradient of the Green function.

Lemma D.1. *Let $x, y \in M$, then*

$$|\nabla_x G(x,y)| \lesssim \frac{1}{d_g(x,y)^{d-1}} \leq \frac{1}{|x-y|^{d-1}}. \tag{D.1}$$

Proof. For $d \geq 2$, a proof of Lemma D.1 is found in [Aub82, Theorem 4.13]. See also [H+96, Theorem 5.2] for a proof with more explicit constants in the case $d \geq 3$. Constants in their proofs depend on d, bounds on the curvature of M, $|\mathrm{vol}_M|$ and the geodesic diameter of M. As, those three last quantities can be further bounded by constants depending on $\tau_{\min}$, $f_{\min}$ and d, see Lemma B.1 and [NSW08, Proposition 6.1], this concludes the proof. For $d = 1$, M is isometric to a circle, for which a closed formula for G exists [Bur94], and satisfies $|\nabla_x G(x,y)| \leq 1$. □

Recall that, by Lemma 4.2, $|\rho_h(x)| \geq 1/2$ for all $x \in M$. Therefore, Lemma D.1 yields

$$\left|\nabla G\left(K_h * \left(\frac{\delta_x}{\rho_h}\right)\right)(z)\right| = \left|\int_M \nabla_z G(z,y)\frac{K_h(x-y)}{\rho_h(x)}\mathrm{d}y\right| \lesssim \int_{\mathcal{B}_M(x,h)} \frac{\|K\|_\infty h^{-d}}{|z-y|^{d-1}}\mathrm{d}y.$$

If $d = 1$, this quantity is smaller than a constant as $\mathrm{vol}_M(\mathcal{B}_M(x,h)) \lesssim h^d$ by Lemma A.1(iii). We then obtain directly the result in this case by integrating this inequality against $f(x)\mathrm{d}x$. If $d \geq 2$, we use the following argument.

- If $|x-z| \geq 2h$ and $y \in \mathcal{B}_M(x,h)$, then $|z-y| \geq |x-z| - h \geq |x-z|/2$. Therefore, by Lemma A.1(iii),

$$\int_{\mathcal{B}_M(x,h)} \frac{\|K\|_\infty h^{-d}}{|z-y|^{d-1}}\mathrm{d}y \leq \frac{2^{1-d}\|K\|_\infty h^{-d}}{|x-z|^{d-1}}\mathrm{vol}_M(\mathcal{B}_M(x,h)) \lesssim \frac{1}{|x-z|^{d-1}}.$$

- If $|x-z| \leq 2h$, then

$$\begin{aligned}\int_{\mathcal{B}_M(x,h)} \frac{\|K\|_\infty h^{-d}}{|z-y|^{d-1}}\mathrm{d}y &\leq \int_{\mathcal{B}_M(z,3h)} \frac{\|K\|_\infty h^{-d}}{|z-y|^{d-1}}\mathrm{d}y \leq \int_{\mathcal{B}_{T_zM}(0,3h)} \frac{\|K\|_\infty h^{-d} J\Psi_z(u)}{|z-\Psi_z(u)|^{d-1}}\mathrm{d}u \\ &\lesssim h^{-d}\int_{\mathcal{B}_{T_zM}(0,3h)} \frac{\mathrm{d}u}{|u|^{d-1}} \lesssim h^{1-d},\end{aligned}$$

  where at the last line we used that $|z-\Psi_z(u)| \geq |u|$ and that $J\Psi_z(u) \lesssim 1$ by Lemma A.1.

Hence,

$$\begin{aligned}&\mathbb{E}\left[|\nabla(G(K_h * \delta_X))(z)|^p\right] = \int_M f(x)|\nabla(G(K_h*\delta_x))(z)|^p\mathrm{d}x \\ &\leq f_{\max}\left(\int_{\mathcal{B}_M(z,2h)} |\nabla(G(K_h*\delta_x))(z)|^p\mathrm{d}x + \int_{M\setminus\mathcal{B}_M(z,2h)} |\nabla(G(K_h*\delta_x))(z)|^p\mathrm{d}x\right)\end{aligned}$$

$$\lesssim \int_{\mathcal{B}_M(z,2h)} h^{(1-d)p}\mathrm{d}x + \int_{M\setminus\mathcal{B}_M(z,2h)} |z-x|^{(1-d)p}\mathrm{d}x \lesssim h^{(1-d)p+d} + \int_{M\setminus\mathcal{B}_M(z,2h)} |z-x|^{(1-d)p}\mathrm{d}x.$$

The latter integral is bounded by

$$\begin{aligned}
&\int_{2h\le|x-z|\le r_0} |z-x|^{(1-d)p}\mathrm{d}x + \int_{|x-z|\ge r_0} |z-x|^{(1-d)p}\mathrm{d}x \\
&\quad\le \int_{2h\le|\Psi_z(u)-z|\le r_0} |z-\Psi_z(u)|^{(1-d)p} J\Psi_z(u)\mathrm{d}u + |\mathrm{vol}_M| r_0^{(1-d)p} \\
&\quad\lesssim \int_{14h/8\le|u|\le r_0} |u|^{(1-d)p}\mathrm{d}u + 1 \lesssim h^{(1-d)p+d} \text{ if } (1-d)p+d<0,
\end{aligned}$$

where at the last line we use that $|u| \le |z-\Psi_z(u)| \le 8|u|/7$ by Lemma A.1. If $d>2$ or if $d=2$ and $p>2$, the condition $(1-d)p+d<0$ is always satisfied. If $d=2$ and $p=2$, then $\int_{14h/8\le|u|\le h_0} |u|^{(1-d)p}\mathrm{d}u$ is of order $-\log h$, concluding the proof.

E Proof of Theorem 3.1(i)

Let f be the density of μ and $\tilde{f} = f/\rho_h$. By Lemma 4.2, $f_{\min}(1-c_0h^{k-1}) \le \tilde{f} \le f_{\max}(1+c_0h^{k-1})$ for h small enough. We have

$$\begin{aligned}
&K_h * f(x) = \int_M K_h(x-y)\tilde{f}(y)\mathrm{d}y = \int_{\mathcal{B}_{T_xM}(0,h)} K_h(x-\Psi_x(v))\tilde{f}\circ\Psi_x(v)J\Psi_x(v)\mathrm{d}v \\
&\ge \int_{\mathcal{B}_{T_xM}(0,h)} K_h(v)\tilde{f}\circ\Psi_x(v)J\Psi_x(v)\mathrm{d}v - \int_{\mathcal{B}_{T_xM}(0,h)} |K_h(x-\Psi_x(v)) - K_h(v)|\tilde{f}\circ\Psi_x(v)J\Psi_x(v)\mathrm{d}v.
\end{aligned} \tag{E.1}$$

By Lemma A.1(v), the quantity $|K_h(x-\Psi_x(v))-K_h(v)|$ is bounded by $\frac{\|K\|_{\mathcal{C}^1(\mathbb{R}^d)}}{h^{d+1}}|x-v-\Psi_x(v)| \lesssim \frac{|v|^2}{h^{d+1}}$, so that the second term in (E.1) is bounded by $Cf_{\max}\int_{\mathcal{B}_{T_xM}(0,h)} \frac{|v|^2}{h^{d+1}}\mathrm{d}v \lesssim h$. Also, using that $|J\Psi_x(v)-1| \le c_1|v|$ by Lemma A.1, the first term is larger than

$$\begin{aligned}
&f_{\min}(1-c_0h^{k-1})(1-c_1h)\int_{\mathbb{R}^d} K_+(v)\mathrm{d}v - f_{\max}(1+c_1h)(1+c_0h^{k-1})\int_{\mathbb{R}^d} K_-(v)\mathrm{d}v \\
&= f_{\min}(1-c_2h)\left(1+\int_{\mathbb{R}^d} K_-(v)\mathrm{d}v\right) - f_{\max}(1+c_2h)\int_{\mathbb{R}^d} K_-(v)\mathrm{d}v \\
&= f_{\min}(1-c_2h) - (f_{\max}(1+c_2h) - f_{\min}(1-c_2h))\int_{\mathbb{R}^d} K_-(v)\mathrm{d}v \\
&\ge f_{\min}(1-c_2h) - (f_{\max}(1+c_2h) - f_{\min}(1-c_2h))\beta \\
&\ge 3f_{\min}/4,
\end{aligned}$$

if $\beta < f_{\min}/(4(f_{\max}-f_{\min}))$ and h is small enough. Likewise, we show that $K_h*\tilde{f}(x) \le 3f_{\max}/2$. It remains to show that $|K_h*\tilde{f}(x) - K_h*(\mu_n/\rho_h)(x)|$ is small enough for all $x\in M$ with high probability. Note that $K_h * \tilde{f} - K_h * (\mu_n/\rho_h)$ is L-Lipschitz with $L \lesssim h^{-d-1}$. Let $t = f_{\min}/4$ and consider a covering of M by N balls $\mathcal{B}_M(x_j, t/(2L))$. By standard packing arguments, such a covering exists with $N \lesssim (L/t)^d$. If $|K_h * \tilde{f}(x_j) - K_h * \mu_n(x_j)| \leq t/2$ for all $j = 1, \ldots, N$, then $\|K_h * \tilde{f} - K_h * \mu_n\|_{L_\infty(M)} \leq t/2 + Lt/(2L) \leq t$. Hence, using Bernstein inequality [GN15, Theorem 3.1.7], as $|K_h(x_j - Y_i)| \leq \|K\|_{\mathcal{C}^0(\mathbb{R}^D)} h^{-d}$ and $\mathrm{Var}(K_h(x_j - Y_i)) \leq \|K^2\|_{\mathcal{C}^0(\mathbb{R}^D)} h^{-d}$, we obtain

$$\begin{aligned}\mathbb{P}(\|K_h * \tilde{f} - K_h * \mu_n\|_{L_\infty(M)} \geq t) &\leq \mathbb{P}(\exists j,\ |K_h * \tilde{f}(x_j) - K_h * \mu_n(x_j)| \geq t/2) \\ &\lesssim (L/t)^d \mathbb{P}(|K_h * \tilde{f}(x_j) - K_h * \mu_n(x_j)| \geq t/2) \lesssim h^{-d(d+1)} \exp(-Cnh^d).\end{aligned}$$

Choosing $nh^d = C' \log n$ for C' large enough yields the conclusion.

F Proofs of Section 4.4

We first prove Lemma 4.8.

Proof of (a). The application $\Psi_{Y_j} \circ \pi_{Y_j} : \mathcal{B}_{\hat{T}_j}(0, 3\varepsilon) \to M$ is a diffeomorphism on $\mathcal{B}_{\hat{T}_j}(0, 3\varepsilon)$, as the composition of the diffeomorphisms Ψ_{Y_j} and $(\pi_{Y_j})_{|\hat{T}_j}$ (recall that $\angle(\hat{T}_j, T_{Y_j}M) \lesssim \varepsilon^{m-1} + \gamma\varepsilon^{-1} \lesssim 1$ by Proposition 3.4). Furthermore, by Lemma A.1(iv) and the bound on the angle,

$$\mathcal{B}_M(Y_j, 2\varepsilon) \subset \Psi_{Y_j}(\mathcal{B}_{T_{Y_j}M}(0, 2\varepsilon)) \subset (\Psi_{Y_j} \circ \pi_{Y_j})(\mathcal{B}_{\hat{T}_j}(0, 3\varepsilon)).$$

This proves the first part of Lemma 4.8(a). Let $S_j : \mathcal{B}_M(Y_j, 2\varepsilon) \to \mathcal{B}_{\hat{T}_j}(0, 3\varepsilon)$ be the inverse of $\Psi_{Y_j} \circ \pi_{Y_j}$. By Lemma A.2(ii), $\hat{\Psi}_j$ is injective on $\hat{T}_j$, while, for $v \in \hat{T}_j$ with $|v| \leq 3\varepsilon$,

$$\left\|\mathrm{id} - d\hat{\Psi}_j(v)\right\|_{\mathrm{op}} \leq \left\|\sum_{a=2}^{m-1} a\hat{V}_{a,j}[\cdot, v^{\otimes(a-1)}]\right\| \lesssim \ell\varepsilon \leq 1/2 \tag{F.1}$$

if $\ell \lesssim \varepsilon^{-1}$ is small enough. Hence, $\hat{\Psi}_j : \mathcal{B}_{\hat{T}_j}(0, 3\varepsilon) \to \hat{\Psi}_j(\hat{T}_j)$ is a diffeomorphism on its image, and $\hat{\Psi}_j \circ S_j$ is a diffeomorphism as a composition of diffeomorphisms. Note that the inverse of $\hat{\Psi}_j$ is given by $\hat{\pi}_j(\cdot - X_j)$, so that $\mathcal{B}_{\hat{\Psi}_j(\hat{T}_j)}(X_j, \varepsilon) \subset \hat{\Psi}_j(\mathcal{B}_{\hat{T}_j}(0, \varepsilon))$. Furthermore, by Lemma A.1,

$$(\Psi_{Y_j} \circ \pi_{Y_j})(\mathcal{B}_{\hat{T}_j}(0, \varepsilon)) \subset \Psi_{Y_j}(\mathcal{B}_{T_{Y_j}}(0, \varepsilon)) \subset \mathcal{B}_M(Y_j, 8\varepsilon/7),$$

so that $(\hat{\Psi}_j \circ S_j)(\mathcal{B}_M(Y_j, 2\varepsilon))$ contains $\mathcal{B}_{\hat{\Psi}_j(\hat{T}_j)}(X_j, \varepsilon)$. Furthermore, these inclusions of balls also hold for any $\varepsilon' \leq \varepsilon$, proving that $|\hat{\Psi}_j \circ S_j(z) - X_j| \geq (7/8)|z - Y_j|$ for any $z \in \mathcal{B}_M(Y_j, 2\varepsilon)$.

Proof of (b). The formula for the density $\tilde{\chi}_j$ follows from a change of variables.

Proof of (c). The inequality (4.18) follows from Proposition 3.4. We now prove that, for $z \in \mathcal{B}_M(Y_j, 2\varepsilon)$,

$$|\pi_{Y_i}(z - \hat{\Psi}_j \circ S_j(z))| \lesssim \varepsilon(\varepsilon^m + \gamma). \tag{F.2}$$

Let $u \in \hat{T}_j$ be such that $z = \Psi_{Y_j} \circ \pi_{Y_j}(u)$ and $y = \hat{\Psi}_j(u)$. Recall that $X_j \in T_{Y_j}M^\perp$ by assumption, so that $\pi_{Y_j}(X_j - Y_j) = 0$. Also, by Lemma A.1(v), we have $\Psi_{Y_j}(\pi_{Y_j}(u)) = Y_j + \pi_{Y_j}(u) + N_{Y_j}(\pi_{Y_j}(u))$ with $N_{Y_j}(\pi_{Y_j}(u)) \in T_{Y_j}M^\perp$, while by Lemma A.2(ii), we have $\hat{\Psi}_j(u) = X_j + u + \hat{N}_j(u)$ with $\hat{N}_j(u) \in \hat{T}_j^\perp$. Hence,

$$\begin{aligned}
|\pi_{Y_j}(z-y)| &= |\pi_{Y_j}(Y_j + \pi_{Y_j}(u) + N_{Y_j}(\pi_{Y_j}(u)) - (X_j + u + \hat{N}_j(u)))| \\
&= |\pi_{Y_j}(N_{Y_j}(\pi_{Y_j}(u)) - \hat{N}_j(u))| \\
&\leq \angle(T_{Y_j}M, \hat{T}_j)|N_{Y_j}(\pi_{Y_j}(u)) - \hat{N}_j(u)| + |\hat{\pi}_j(N_{Y_j}(\pi_{Y_j}(u)) - \hat{N}_j(u))| \\
&\lesssim (\varepsilon^{m-1} + \gamma\varepsilon^{-1})(\varepsilon^m + \gamma) + |\hat{\pi}_j(\pi_{Y_j}^\perp(N_{Y_j}(\pi_{Y_j}(u))))| \\
&\lesssim (\varepsilon^{m-1} + \gamma\varepsilon^{-1})(\varepsilon^m + \gamma) + \angle(T_{Y_j}M, \hat{T}_j)|N_{Y_j}(\pi_{Y_j}(u))| \\
&\lesssim (\varepsilon^{m-1} + \gamma\varepsilon^{-1})(\varepsilon^m + \gamma + \varepsilon^2) \lesssim (\varepsilon^{m-1} + \gamma\varepsilon^{-1})(\varepsilon^2 + \gamma),
\end{aligned}$$

where we used Proposition 3.4 to bound $\angle(T_{Y_j}M, \hat{T}_j)$, Lemma A.2 to bound $|N_{Y_j}(\pi_{Y_j}(u)) - \hat{N}_j(u)|$ and Lemma A.1 to bound $|N_{Y_j}(\pi_{Y_j}(u))|$. Recalling that $\gamma \lesssim \varepsilon^2$ by assumption, we obtain (F.2).

To prove inequality (4.19), we first bound $|\chi_j(\hat{\Psi}_j \circ S_j(z)) - \chi_j(z)|$ and then bound $|J(\hat{\Psi}_j \circ S_j)(z) - 1|$. The first bound is based on the following elementary lemma.

Lemma F.1. *Let $\theta : \mathbb{R}^D \to \mathbb{R}$ be a smooth radial function. Then, $|\theta(x) - \theta(y)| \leq \frac{\|\theta\|_{\mathcal{C}_2(\mathbb{R}^D)}}{2}||x|^2 - |y|^2|$.*

Proof. As $d\theta(0) = 0$, one can write $\theta(x) = \tilde{\theta}(|x|^2)$ for some function $\tilde{\theta}$ which is Lipschitz continuous with Lipschitz constant $\frac{\|d^2\theta\|_{\mathcal{C}^0(\mathbb{R}^D)}}{2}$. This implies the conclusion. □

Recall from the proof of Lemma 3.5 that we have $\chi_j(z) = \zeta_j(z)/\sum_{i=1}^J \zeta_i(z)$ where $\zeta_i = \theta\left(\frac{z - X_i}{\varepsilon}\right)$ for some smooth radial function θ, and that furthermore, there is at most c_d non-zero terms in the sum in the denominator, which is always larger than 1. Hence, if we control for every $i = 1, \ldots, J$ the difference $||z - X_i|^2 - |\hat{\Psi}_j \circ S_j(z) - X_i|^2|$, then we obtain a control on $|\chi_j(z) - \chi_j(\hat{\Psi}_j \circ S_j(z))|$. We have by (4.18) and (F.2),

$$\begin{aligned}
&||\hat{\Psi}_j \circ S_j(z) - X_i|^2 - |z - X_i|^2| = ||\hat{\Psi}_j \circ S_j(z) - z|^2 + 2(\hat{\Psi}_j \circ S_j(z) - z) \cdot (z - X_i)| \\
&\lesssim (\varepsilon^m + \gamma)^2 + |(\hat{\Psi}_j \circ S_j(z) - z) \cdot (z - Y_i)| + |(\hat{\Psi}_j \circ S_j(z) - z) \cdot (X_i - Y_i)| \\
&\lesssim (\varepsilon^m + \gamma)^2 + |\pi_{Y_j}(\hat{\Psi}_j \circ S_j(z) - z) \cdot \pi_{Y_j}(z - Y_i)| + |\pi_{Y_j}^\perp(\hat{\Psi}_j \circ S_j(z) - z) \cdot \pi_{Y_j}^\perp(z - Y_i)| + (\varepsilon^m + \gamma)\gamma \\
&\lesssim (\varepsilon^m + \gamma)^2 + \varepsilon(\varepsilon^m + \gamma)|z - Y_i| + (\varepsilon^k + \gamma)|\pi_{Y_j}^\perp(z - Y_i)| + (\varepsilon^m + \gamma)\gamma.
\end{aligned}$$

By Lemma A.1(i), $|\pi^\perp_{Y_j}(z-Y_i)| \leq |\tilde{\pi}^\perp_{Y_j}(z)| + |\tilde{\pi}^\perp_{Y_j}(Y_i)| \lesssim \varepsilon^2 + |Y_i - Y_j|^2$ and $\gamma, \varepsilon^m \lesssim \varepsilon^2$. Hence, we obtain that

$$||\hat{\Psi}_j \circ S_j(z) - X_i|^2 - |z - X_i|^2| \lesssim (\varepsilon^m + \gamma)(\varepsilon^2 + |Y_i - Y_j|^2). \tag{F.3}$$

Therefore,

$$\left|\theta\left(\frac{z - X_i}{\varepsilon}\right) - \theta\left(\frac{\hat{\Psi}_j \circ S_j(z) - X_i}{\varepsilon}\right)\right| \lesssim \frac{(\varepsilon^m + \gamma)(\varepsilon^2 + |Y_i - Y_j|^2)}{\varepsilon^2} = (\varepsilon^m + \gamma)\left(1 + \frac{|Y_i - Y_j|^2}{\varepsilon^2}\right). \tag{F.4}$$

Note also that if $|Y_i - Y_j| \geq 3\varepsilon$, then $|z - X_i| \geq |X_i - X_j| - |z - X_j| \geq 3\varepsilon - \varepsilon - 3\gamma \geq \varepsilon$, while by the same argument $|\hat{\Psi}_j \circ S_j(z) - X_i| \geq \varepsilon$. Hence, both terms in the left-hand side of (F.4) are null in that case. Thus, we may assume that $|Y_i - Y_j| \leq 3\varepsilon$, so that $\left|\theta\left(\frac{z-X_i}{\varepsilon}\right) - \theta\left(\frac{\hat{\Psi}_j \circ S_j(z) - X_i}{\varepsilon}\right)\right| \lesssim \varepsilon^m + \gamma$. From the definition of $\chi_j(z)$, and as the function $t \mapsto 1/t$ is Lipschitz on $[1, \infty[$, we obtain that $|\chi_j(z) - \chi_j(\hat{\Psi}_j \circ S_j(z))| \lesssim \varepsilon^m + \gamma$.

We now prove a bound on $|J(\hat{\Psi}_j \circ S_j)(z) - 1|$. One has, for $u = S_j(z) \in \hat{T}_j$,

$$|J(\hat{\Psi}_j \circ S_j)(z) - 1| = \frac{|J\hat{\Psi}_j(u) - J(\Psi_{Y_j} \circ \pi_{Y_j})(u)|}{J(\Psi_{Y_j} \circ \pi_{Y_j})(u)}.$$

By Lemma A.1(v) and Lemma A.2(ii), $\left\|\mathrm{id}_{\hat{T}_j} - d(\Psi_{Y_j} \circ \pi_{Y_j})(u)\right\|_{\mathrm{op}} \lesssim |u|$ and $\left\|\mathrm{id}_{\hat{T}_j} - d\hat{\Psi}_j(u)\right\|_{\mathrm{op}} \lesssim |u|$. As a consequence, both Jacobians are larger than, say $1/2$ for u small enough, and, as the function $A \in \mathbb{R}^{d\times d} \mapsto \sqrt{\det(A)}$ is c_d-Lipschitz continuous on the set of matrices with $\det(A) \geq 1/2$ and $\|A\|_{\mathrm{op}} \leq 2$, we have

$$|J(\hat{\Psi}_j \circ S_j)(z) - 1| \leq 2c_d \left\|d\hat{\Psi}_j(u)^* d\hat{\Psi}_j(u) - d(\Psi_{Y_j} \circ \pi_{Y_j})(u)^* d(\Psi_{Y_j} \circ \pi_{Y_j})(u)\right\|_{\mathrm{op}}. \tag{F.5}$$

Recall that $\hat{\Psi}_j(u) = X_j + u + \hat{N}_j(u)$ and $\Psi_{Y_j} \circ \pi_{Y_j}(u) = Y_j + \pi_{Y_j}(u) + N_{Y_j} \circ \pi_{Y_j}(u)$. We may write

$$\begin{aligned} &d\hat{\Psi}_j(u)^* d\hat{\Psi}_j(u) = \mathrm{id}_{\hat{T}_j} + (d\hat{N}_j(u))^* d\hat{N}_j(u) \quad \text{and} \\ &d(\Psi_{Y_j} \circ \pi_{Y_j})(u)^* d(\Psi_{Y_j} \circ \pi_{Y_j})(u) = \hat{\pi}_j \pi_{Y_j} \hat{\pi}_j + (d(N_{Y_j} \circ \pi_{Y_j})(u))^* d(N_{Y_j} \circ \pi_{Y_j})(u). \end{aligned}$$

One has $\left\|\mathrm{id}_{\hat{T}_j} - \hat{\pi}_j \pi_{Y_j} \hat{\pi}_j\right\|_{\mathrm{op}} = \left\|\hat{\pi}_j \pi^\perp_{Y_j} \pi^\perp_{Y_j} \hat{\pi}_j\right\|_{\mathrm{op}} \leq \angle(T_{Y_j}M, \hat{T}_j)^2 \lesssim (\varepsilon^{m-1} + \gamma\varepsilon^{-1})^2 \lesssim \varepsilon^m + \gamma$ (recall that $\gamma \lesssim \varepsilon^2$). Furthermore, by Lemma A.2(iv),

$$\begin{aligned} &\left\|(d\hat{N}_j(u))^* d\hat{N}_j(u) - (d(N_{Y_j} \circ \pi_{Y_j})(u))^* d(N_{Y_j} \circ \pi_{Y_j})(u)\right\|_{\mathrm{op}} \\ &\quad \leq \left(\left\|d\hat{N}_j(u)\right\|_{\mathrm{op}} + \left\|d(N_{Y_j} \circ \pi_{Y_j})(u)\right\|_{\mathrm{op}}\right) \left\|d\hat{N}_j(u) - d(N_{Y_j} \circ \pi_{Y_j})(u)\right\|_{\mathrm{op}} \end{aligned}$$

$$\lesssim \varepsilon(\varepsilon^{m-1} + \gamma\varepsilon^{-1}) \lesssim \varepsilon^m + \gamma.$$

Putting together (F.5) with those two inequalities, we obtain that $|J(\hat{\Psi}_j \circ S_j)(z) - 1| \lesssim \varepsilon^m + \gamma$, concluding the proof of Lemma 4.8.

To conclude the section, we state and prove Lemma F.2, which gives an upper bound on the quantity T appearing in Lemma 4.9 for $\phi = K_h * (\nu_n/\hat{\rho}_h)$ and $\phi' = K_h * (\mu_n/\rho_h)$.

Lemma F.2. *The quantity $T = \max_{j=1\dots J} \sup_{z \in \mathcal{B}(Y_j,\varepsilon)} |\phi(\hat{\Psi}_j \circ S_j(z)) - \phi'(z)|$ satisfies $T \lesssim \varepsilon^m + \gamma$ with probability larger than $1 - cn^{-k/d}$.*

Proof. For $z \in \mathcal{B}(Y_j, \varepsilon)$, we have

$$|\phi(\hat{\Psi}_j \circ S_j(z)) - \phi'(z)| \leq \frac{1}{n} \sum_{i=1}^{n} \left| \frac{K_h * \delta_{X_i}(\hat{\Psi}_j \circ S_j(z))}{\hat{\rho}_h(X_i)} - \frac{K_h * \delta_{Y_i}(z)}{\rho_h(Y_i)} \right|.$$

Fix an index $i \in \{1, \dots, n\}$. By Lemma A.1(i), as $X_i - Y_i \in T_{Y_i}M^{\perp}$, we have for $z \in M$,

$$||z - Y_i|^2 - |z - X_i|^2| = ||X_i - Y_i|^2 - 2(z - Y_i) \cdot (X_i - Y_i)| \leq \gamma^2 + |z - Y_i|^2 \frac{\gamma}{\tau_{\min}}.$$

This inequality together with (F.3) and Lemma F.1 yield

$$\begin{aligned}
&|K_h(X_i - \hat{\Psi}_j \circ S_j(z)) - K_h(Y_i - z)| \\
&\quad \leq |K_h(X_i - \hat{\Psi}_j \circ S_j(z)) - K_h(X_i - z)| + |K_h(X_i - z) - K_h(Y_i - z)| \\
&\quad \lesssim h^{-d-2} \left((\varepsilon^m + \gamma)(\varepsilon^2 + |Y_i - Y_j|^2) + \gamma^2 + \gamma |z - Y_i|^2 \right).
\end{aligned}$$

We may assume that $|Y_i - Y_j| \leq 3h$ and $|z - Y_i| \leq 2h$, for otherwise both quantities in the left-hand site of the above equation are equal to zero. Hence, as $\gamma \lesssim \varepsilon \lesssim h$ by assumption, we have

$$|K_h(X_i - \hat{\Psi}_j \circ S_j(z)) - K_h(Y_i - z)| \lesssim h^{-d}(\varepsilon^m + \gamma)\mathbf{1}\{Y_i \in \mathcal{B}_M(z, 2h)\}. \tag{F.6}$$

Let us now bound $|\hat{\rho}_h(\hat{\Psi}_j \circ S_j(X_i)) - \rho_h(Y_i)|$. By the triangle inequality, and using (4.19) and (F.6), we obtain that this quantity is smaller than

$$\begin{aligned}
&\sum_{j=1}^{J} \int_M |\tilde{\chi}_j(z) K_h(X_i - \hat{\Psi}_j \circ S_j(z)) - \chi_j(z) K_h(Y_i - z)| \mathrm{d}z \\
&\lesssim \sum_{j=1}^{J} \int_M \left(\mathbf{1}\{z \in \mathcal{B}_M(Y_j, 2\varepsilon)\}(\varepsilon^m + \gamma)|K_h(Y_i - z)| + \tilde{\chi}_j(z) h^{-d} (\varepsilon^m + \gamma) \mathbf{1}\{z \in \mathcal{B}_M(Y_i, 2h)\} \right) \mathrm{d}z \\
&\lesssim h^{-d}(\varepsilon^m + \gamma) \sum_{j=1}^{J} \int_M \mathbf{1}\{z \in \mathcal{B}_M(Y_j, 2\varepsilon)\} \mathbf{1}\{z \in \mathcal{B}_M(Y_i, 2h)\} \mathrm{d}z
\end{aligned}$$

$$\lesssim \varepsilon^d h^{-d}(\varepsilon^m+\gamma)\sum_{j=1}^J \mathbf{1}\{|Y_j-Y_i|\leq 4h\} \lesssim h^{-d}(\varepsilon^m+\gamma)\sum_{j=1}^J \mathbf{1}\{|Y_j-Y_i|\leq 4h\}\mathrm{vol}_M(\mathcal{B}_M(Y_j,\varepsilon/8))$$
$$\lesssim h^{-d}(\varepsilon^m+\gamma)\mathrm{vol}_M(\mathcal{B}_M(Y_i,5h)) \lesssim \varepsilon^m+\gamma,$$

where we use that $\{X_1,\dots,X_J\}$ is $7\varepsilon/24$-sparse, so that $\{Y_1,\dots,Y_J\}$ is $\varepsilon/4$-sparse. Therefore, the balls $\mathcal{B}_M(Y_j,\varepsilon/8)$ for $|Y_j-Y_i|\leq 4h$ are pairwise distincts, and are all included in $\mathcal{B}_M(Y_i,4h+\varepsilon/8)\subset\mathcal{B}_M(Y_i,5h)$. We conclude by Lemma A.1(iii). Letting $N(z,2h)$ be the number of points Y_i belonging to $\mathcal{B}_M(z,2h)$, we obtain

$$|\phi(\hat{\Psi}_j\circ S_j(z))-\phi'(z)| \lesssim \frac{1}{n}\sum_{i=1}^n \Big(|K_h(Y_i-z)|(\varepsilon^m+\gamma)+h^{-d}(\varepsilon^m+\gamma)\mathbf{1}\{Y_i\in\mathcal{B}_M(z,2h)\}\Big)$$
$$\lesssim \frac{N(z,2h)}{nh^d}(\varepsilon^m+\gamma).$$

If, for every $z\in M$ and some $\lambda>0$, $N(z,2h)\leq\lambda nh^d$, then we have the conclusion. Let us bound

$$P_0=\mathbb{P}(\exists z\in M,\ N(z,2h)>\lambda nh^d).$$

If $N(z,2h)>\lambda nh^d$, then there exists a point Y_i with $N(Y_i,4h)\geq N(z,2h)>\lambda nh^d$. Hence, $P_0\leq n\mathbb{P}(N(Y_1,4h)>\lambda nh^d)$. Conditionally on Y_1, $N(Y_1,4h)=1+U$ with U a binomial random variable of parameters $n-1$ and $\mu(\mathcal{B}_M(Y_1,4h))\leq f_{\max}\mathrm{vol}_M(\mathcal{B}_M(Y_1,4h))\lesssim h^d$ (see Lemma A.1(iii)). In particular, for λ large enough, the probability P_0 is smaller than $n^{-k/d}$ by Hoeffding's inequality. □

G Lower bounds on minimax risks

In this section, we prove the different lower bounds on minimax risks stated in the article. The main tool used will be Assouad's lemma. Fix as in Section 2.4 a statistical model $(\mathcal{Y},\mathcal{H},\mathcal{Q},\iota,\vartheta)$, where $\vartheta:\mathcal{Y}\to(E,\mathcal{L})$ is a measurable function taking its values in some semi-metric space $(E,\mathcal{L})$.

Lemma G.1 (Assouad's lemma [Yu97])**.** *Let $m\geq 1$ be an integer and $\mathcal{Q}_m=\{\xi_\sigma,\ \sigma\in\{-1,1\}^m\}\subset\mathcal{Q}$ be a set of probability measures. Assume that for all $\sigma,\sigma'\in\{-1,1\}^m$,*

$$\mathcal{L}(\vartheta(\xi_\sigma),\vartheta(\xi_{\sigma'}))\geq|\sigma-\sigma'|\delta, \tag{G.1}$$

where $|\sigma-\sigma'|=\sum_{i=1}^m\mathbf{1}\{\sigma(i)\neq\sigma'(i)\}$ is the Hamming distance between σ and σ'. Then,

$$\mathcal{R}_n(\vartheta,\mathcal{Q},\mathcal{L})\geq m\frac{\delta}{16}\left(1-\max\left\{TV(\iota_\#\xi_\sigma,\iota_\#\xi_{\sigma'}),\ |\sigma-\sigma'|=1\right\}\right)^{2n}. \tag{G.2}$$

The lower bound on the minimax rates we prove are actually going to hold on the smaller model of uniform distributions on manifolds.

Definition G.2. *Let $k \geq 2$ and $\gamma \geq 0$. The set $\mathcal{Q}_d^k(\gamma)$ is the set of probability distributions ξ of random variables (Y, Z), where Y follows the uniform distribution on some manifold $M \in \mathcal{M}_d^k$ with $f_{\max}^{-1} \leq |\mathrm{vol}_M| \leq f_{\min}^{-1}$, and $Z \in \mathcal{B}(0,\gamma)$ is such that $Z \in T_Y M^\perp$. The statistical model is completed by letting $(\mathcal{Y}, \mathcal{H})$ be $\mathbb{R}^D \times \mathbb{R}^D$ endowed with its Borel σ-algebra, ι be the addition $\mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}^D$ and $\vartheta(\xi)$ be the first marginal μ of ξ.*

We write $\mathcal{Q}_d^k$ for $\mathcal{Q}_d^k(0)$. One can check that $\mathcal{Q}_d^k(\gamma) \subset \mathcal{Q}_d^{k,s}(\gamma)$, with parameter $L_s = f_{\min}^{-1/p} \vee f_{\max}^{1-1/p}$. Therefore, a lowerbound on the minimax risk on the model $\mathcal{Q}_d^k(\gamma)$ yields a lowerbound on the minimax risk on the model $\mathcal{Q}_d^{k,s}(\gamma)$ should the parameter L_s be large enough.

We build a subfamily of manifolds indexed by $\sigma \in \{-1,1\}^m$ following [AL19]. By [AL19, Section C.2], there exists a manifold $M \subset \mathbb{R}^{d+1}$ of reach $2\tau_{\min}$, of volume $C_d \tau_{\min}^d$ which contains $\mathcal{B}_{\mathbb{R}^d}(0, \tau_{\min})$. Let $\delta > 0$ and consider a family of m points $x_1, \dots, x_m \in \mathcal{B}_{\mathbb{R}^d}(0, \tau_{\min}/2)$, with $|x_i - x_{i'}| \geq 4\delta$ for $i \neq i'$ and $c_d(\tau_{\min}/\delta)^d \leq m \leq C_d(\tau_{\min}/\delta)^d$. Let $0 < \Lambda < \delta$ and let $\phi : \mathbb{R}^{d+1} \to [0,1]$ be a smooth radial function supported on $\mathcal{B}(0,1)$, with $\phi \equiv 1$ on $\mathcal{B}(0,1/2)$. Let e be the unit vector in the $(d+1)$th direction. We then let, for $\sigma \in \{-1,1\}^m$,

$$\Phi_\sigma^\Lambda(x) = x + \sum_{i=1}^m \frac{\sigma_i + 1}{2} \Lambda \phi\left(\frac{x - x_i}{\delta}\right) e. \tag{G.3}$$

Let $M_\sigma^\Lambda = \Phi_\sigma^\Lambda(M)$ and μ_σ^Λ be the the uniform measure on M_σ^Λ. If $\Lambda \leq c_{k,d,\tau_{\min}} \delta^k$, then $\mu_\sigma^\Lambda \in \mathcal{Q}_d^k$, provided that L_k is large enough [AL19, Lemma C.13]. If $\sigma_i = 1$, the volume of $\Phi_\sigma^\Lambda(\mathcal{B}_{\mathbb{R}^d}(x_i, \delta))$ satisfies, with ω_d the volume of the d-dimensional unit ball,

$$\begin{aligned} \left|\mathrm{vol}_{M_\sigma^\Lambda}(\Phi_\sigma^\Lambda(\mathcal{B}_{\mathbb{R}^d}(x_i,\delta)) - \omega_d \delta^d\right| &\leq \int_{\mathcal{B}_{\mathbb{R}^d}(x_i,\delta)} |J\Phi_\sigma^\Lambda(x) - 1| \mathrm{d}x \\ &\leq \int_{\mathcal{B}_{\mathbb{R}^d}(x_i,\delta)} \left| \sqrt{1 + \Lambda^2 \delta^{-2} \left|\nabla \phi\left(\frac{x - x_i}{\delta}\right)\right|^2} - 1 \right| \mathrm{d}x \leq C_d \delta^d \Lambda^2 \delta^{-2}. \end{aligned}$$

Hence, for δ small enough, we have $||\mathrm{vol}_{M_\sigma^\Lambda}| - C_d \tau_{\min}^d| \leq m C_d \delta^d \Lambda^2 \delta^{-2} \leq C_d \tau_{\min}^d / 3$, as $m \leq C_d(\tau_{\min}/\delta)^d$ and $\Lambda \leq c_{k,d,\tau_{\min}} \delta^k$. As a consequence, if $|\sigma - \sigma'| = 1$, with for instance $\sigma_i = 1$ and $\sigma_i' = -1$, then

$$\mathrm{TV}(\mu_\sigma^\Lambda, \mu_\sigma'^\Lambda) \leq \max(\mu_\sigma^\Lambda(\Phi_\sigma^\Lambda(\mathcal{B}_{\mathbb{R}^d}(x_i,\delta))), \mu_{\sigma'}^\Lambda(\mathcal{B}_{\mathbb{R}^d}(x_i,\delta)) \leq C_{d,\tau_{\min}} \delta^d. \tag{G.4}$$

We may now prove the different minimax lower bounds using Assouad's Lemma on the family $\{\mu_\sigma^\Lambda,\ \sigma \in \{-1,1\}^m\}$.

Proof of Theorem 2.13. As g is nondecreasing and convex, by Jensen's inequality, we may assume without loss of generality that $\mathcal{L} = \mathrm{TV}$. Let $\Gamma = |(\mu_\sigma^\Lambda - \mu_{\sigma'}^\Lambda)(B_i)|$, where $B_i = \mathcal{B}_{\mathbb{R}^d}(x_i, \delta)$

and $\sigma(i) \neq \sigma'(i)$. Then, $\mathrm{TV}(\mu_\sigma^\Lambda, \mu_{\sigma'}^\Lambda) \geq |\sigma - \sigma'|\Gamma$. Furthermore, if for instance $\sigma'(i) = 1$, $\Gamma \geq \mu_{\sigma'}^\Lambda(B_i) = (\omega_d \delta^d)/|\mathrm{vol}_{M_{\sigma'}^\Lambda}| \geq c_d \delta^d/\tau_{\min}^d$. By Assouad's Lemma,

$$\mathcal{R}_n(\mu; \mathcal{Q}_d^{s,k}; \mathrm{TV}) \geq \mathcal{R}_n(\mu; \mathcal{Q}_d^{k}; \mathrm{TV}) \geq \frac{m}{16} c_d \frac{\delta^d}{\tau_{\min}^d} \left(1 - C_{d,\tau_{\min}}\delta^d\right)^{2n} \geq C_d \left(1 - C_{d,\tau_{\min}}\delta^d\right)^{2n}.$$

We obtain the conclusion by letting δ go to 0. □

Lemma G.3. *For any $\tau_{\min} > 0$ and $1 \leq r \leq \infty$, for $f_{\min}$ small enough and $f_{\max}$, L_k large enough, one has*

$$\mathcal{R}_n\left(\frac{\mathrm{vol}_M}{|\mathrm{vol}_M|}, \mathcal{Q}_d^k(\gamma), W_r\right) \gtrsim \gamma + n^{-k/d}. \tag{G.5}$$

Proof. As, $W_r \geq W_1$, we may assume that $r = 1$. Let $\sigma, \sigma' \in \{-1,1\}^m$ with $\sigma(i) \neq \sigma'(i)$. Let $p_{\sigma(i)} = \mathrm{vol}_{M_\sigma^\Lambda}(\mathcal{B}(x_i, \delta))$ and $U_{\sigma,i}^\Lambda = p_{\sigma(i)}^{-1}(\mathrm{vol}_{M_\sigma^\Lambda})_{|\mathcal{B}(x_i,\delta)}$. By the Kantorovitch-Rubinstein duality formula, $W_1(\mu, \nu) = \max \int f \mathrm{d}(\mu - \nu)$, where the maximum is taken over all 1-Lipschitz continuous functions $f : \mathbb{R}^D \to \mathbb{R}$. Let $f : x \mapsto x \cdot e$. Assume for instance that $\sigma(i) = -1$ and $\sigma'(i) = 1$. We have $f(x) = 0$ for $x \in \mathcal{B}_{M_\sigma^\Lambda}(x_i, \delta)$ and $f(x) = \Lambda$ for $x \in \mathcal{B}_{M_{\sigma'}^\Lambda}(x_i, \delta/2)$. Therefore, we have, as $p_{\sigma'(i)} \leq c\delta^{-d}$,

$$W_1(U_{\sigma,i}^\Lambda, U_{\sigma',i}^\Lambda) \geq p_{\sigma'(i)}^{-1} \Lambda \omega_d (\delta/2)^d \geq c_1 \Lambda.$$

Note also that $|p_{\sigma(i)} - p_{\sigma'(i)}| \leq \left|\mathrm{vol}_{M_\sigma^\Lambda}(\Phi_\sigma^\Lambda(\mathcal{B}_{\mathbb{R}^d}(x_i, \delta)) - \omega_d \delta^d\right| \leq C_d \delta^d \Lambda^2 \delta^{-2}$. Furthermore, $||\mathrm{vol}_{M_\sigma^\Lambda}| - |\mathrm{vol}_{M_{\sigma'}^\Lambda}|| \leq \sum_{i=1}^m |p_{\sigma(i)} - p_{\sigma'(i)}| \leq |\sigma - \sigma'| C_d \delta^d \Lambda^2 \delta^{-2}$. Let f_i be a 1-Lipschitz continuous function such that $W_1(U_{\sigma,i}^\Lambda, U_{\sigma',i}^\Lambda) = \int f_i d(U_{\sigma,i}^\Lambda - U_{\sigma',i}^\Lambda)$. One can choose f_i such that $f_i(x_i) = 0$, so that the maximum of $|f_i|$ on $\mathcal{B}(x_i, \delta)$ is at most δ. One can then change the value of f_i outside the ball without changing the value of the integral, so that f_i is supported on $\mathcal{B}(x_i, 2\delta)$ and is 1-Lipschitz continuous. Consider the function f obtained by gluing together the different functions f_i. The function f is 1-Lipschitz continuous, so that

$$\begin{aligned}
W_1\left(\mu_\sigma^\Lambda, \mu_{\sigma'}^\Lambda\right) &\geq \sum_{i=1}^m \left(\frac{p_{\sigma(i)}}{|\mathrm{vol}_{M_\sigma^\Lambda}|} U_{\sigma,i}^\Lambda - \frac{p_{\sigma'(i)}}{|\mathrm{vol}_{M_{\sigma'}^\Lambda}|} U_{\sigma',i}^\Lambda\right)(f) \\
&\geq \sum_{i=1}^m \frac{p_{\sigma(i)}}{|\mathrm{vol}_{M_\sigma^\Lambda}|}(U_{\sigma,i}^\Lambda - U_{\sigma',i}^\Lambda)(f) - |p_{\sigma(i)} - p_{\sigma'(i)}| \frac{|U_{\sigma',i}^\Lambda(f)|}{|\mathrm{vol}_{M_\sigma^\Lambda}|} - p_{\sigma'(i)} |U_{\sigma',i}^\Lambda(f)| \left|\frac{1}{|\mathrm{vol}_{M_\sigma^\Lambda}|} - \frac{1}{|\mathrm{vol}_{M_{\sigma'}^\Lambda}|}\right| \\
&\geq \sum_{i=1}^m \frac{p_{\sigma(i)}}{|\mathrm{vol}_{M_\sigma^\Lambda}|} W_1(U_{\sigma,i}^\Lambda, U_{\sigma',i}^\Lambda) - \sum_{i=1}^m c_4 |p_{\sigma(i)} - p_{\sigma'(i)}| \delta \mathbf{1}\{\sigma(i) \neq \sigma'(i)\} - c_5 \delta |\sigma - \sigma'| \delta^d \Lambda^2 \delta^{-2} \\
&\geq \sum_{i=1}^m \mathbf{1}\{\sigma(i) \neq \sigma'(i)\}(c_6 \delta^d \Lambda - c_4 \delta^d \Lambda^2 \delta^{-1}) - c_5 \delta |\sigma - \sigma'| \delta^d \Lambda^2 \delta^{-2} \geq c_7 \delta^d \Lambda |\sigma - \sigma'|.
\end{aligned}$$

Hence, letting $\Lambda = c_{k,d,\tau_{\min},L_k}\delta^k$ and $\delta = n^{-1}$, we have, by Assouad's Lemma,

$$\mathcal{R}_n\left(\frac{\mathrm{vol}_M}{|\mathrm{vol}_M|}, \mathcal{Q}_d^k(\gamma), W_r\right) \gtrsim n^{-k/d}.$$

Consider now the case $\gamma > 0$. Let M_0 be the d-dimensional sphere of radius $\tau_{\min}$ and M_1 be the d-dimensional sphere of radius $\tau_{\min} + \delta$. Let Y be uniform on M_1, and let ξ be the law of $(Y, 0)$. Also, let ξ' be the law of $((1+\gamma/\tau_{\min})\,Y, -\gamma/\tau_{\min}Y)$. Then, $\iota_\#\xi = \iota_\#\xi'$, whereas $W_1\left(\frac{\mathrm{vol}_{M_0}}{|\mathrm{vol}_{M_0}|}, \frac{\mathrm{vol}_{M_1}}{|\mathrm{vol}_{M_1}|}\right) \geq \gamma$. We conclude by Le Cam lemma [Yu97] that $\mathcal{R}_n\left(\frac{\mathrm{vol}_M}{|\mathrm{vol}_M|}, \mathcal{Q}_d^k(\gamma), W_r\right) \gtrsim \gamma$. □

Proof of Theorem 3.1(iv). Let $a_n = n^{-\frac{s+1}{2s+d}}$ if $d \geq 3$ and $a_n = n^{-1/2}$ if $d \leq 2$. As $W_p \geq W_1$, we may assume without loss of generality that $p = 1$, and up to rescaling, we assume that $\tau_{\min} = \sqrt{d}$. Consider the manifold $M \subset \mathbb{R}^{d+1}$ containing $\mathcal{B}_{\mathbb{R}^d}(0, \sqrt{d})$ of the previous proof. In particular, M contains the cube $[-1,1]^d$. We adapt the proof of Theorem 3 in [WB19b], where authors consider a family of functions $f_\sigma : [-1,1]^d \to M$ indexed by $\sigma \in \{-1,1\}^m$, with $f_\sigma = 1 + n^{-1/2}\sum_{j=1}^m \sigma_j\psi_j$, where $(\psi_j)_{j=1,\dots,m}$ are elements of a wavelet basis of $[-1,1]^d$ (see [WB19b, Appendix E] for details on the construction of the wavelet basis). If $m \lesssim n^{d/(2s+d)}$, then $t_0 \leq f_\sigma \leq t_1$ for some positive constants $t_0 < 1 < t_1$, and $\|f_\sigma\|_{B^s_{p,q}([-1,1]^d)} \lesssim 1$. Define a function g_σ by $g_\sigma(x) = f_\sigma(x)$ if $x \in [-1,1]^d$ and $g_\sigma(x) = 1$ otherwise. The function g_σ satisfies $t_0 \leq g_\sigma \leq t_1$, as well as $\|g_\sigma\|_{B^s_{p,q}(M)} \lesssim \|f_\sigma\|_{B^s_{p,q}} + |\mathrm{vol}_M|^{1/p} \lesssim 1$. Such an inequality is clear for the $\|\cdot\|_{H^l_p(M)}$ norm for l an integer, as $\|g_\sigma\|^p_{H^l_p(M)} = \|g_\sigma\|^p_{H^l_p([-1,1]^d)} + \|g_\sigma\|^p_{H^l_p(M\setminus[-1,1]^d)}$, while the result follows from interpolation for Besov spaces [Lun18, Corollary 1.1.7]. Also, as $\int_{[-1,1]^d} f_\sigma = 1$, we have $\int g_\sigma = |\mathrm{vol}_M|$, and $g_\sigma/|\mathrm{vol}_M|$ is larger than $f_{\min} = t_0/|\mathrm{vol}_M|$ and smaller than $f_{\max} = t_1/|\mathrm{vol}_M|$. Hence, identifying measures with their densities, the set

$$\mathcal{Q}_m = \{\mu_\sigma = g_\sigma/|\mathrm{vol}_M|,\ \sigma \in \{-1,1\}^m\}$$

is a subset of $\mathcal{Q}_d^{s,k}$ for $f_{\min}$ small enough and L_k, L_s, $f_{\max}$ large enough. Furthermore, for $\sigma, \sigma' \in \{-1,1\}^m$, $\mathrm{TV}(\mu_\sigma, \mu_{\sigma'}) = \mathrm{TV}(f_\sigma, f'_\sigma)$, while $W_1(\mu_\sigma, \mu_{\sigma'}) = W_1(f_\sigma, f_{\sigma'})$ by the Kantorovitch-Rubinstein duality formula. Hence, applying Assouad's inequality in the same fashion than in [WB19b, Theorem 3] yields that $\mathcal{R}_n(\mu, \mathcal{Q}_d^{s,k}, W_1) \gtrsim a_n$. □

Proof of Theorem 3.7(iv). According to Lemma G.3, $\mathcal{R}_n(\mu, \mathcal{Q}_d^{s,k}, W_p) \geq \mathcal{R}_n(\mu, \mathcal{Q}_d^k, W_p) \gtrsim \gamma + n^{-k/d}$, and according to Theorem 3.1(iv), $\mathcal{R}_n(\mu, \mathcal{Q}_d^{s,k}, W_p) \gtrsim a_n$. □

H Existence of kernels satisfying conditions A, $B(m)$ and $C(\beta)$

The goal of the section is to prove the existence of a kernel K satisfying the conditions A, $B(m)$ and $C(\beta)$ stated at the beginning of Section 3.

If K is a radial kernel, we have by integration by parts, as K is smooth with compact support,

$$\int_{\mathbb{R}^d} \partial^{\alpha_0} K(v) v^{\alpha_1} \mathrm{d}v = C_{\alpha_0,\alpha_1} \int_{\mathbb{R}^d} K(v) v^{\alpha_1+\alpha_0} \mathrm{d}v = C'_{\alpha_0,\alpha_1} \int_{\mathbb{R}} K(r) r^{d+|\alpha_0|+|\alpha_1|-1} \mathrm{d}r.$$

Hence, to show the existence of such a kernel, it suffices to find, for every $m \geq 0$ and every positive constant κ, a smooth even function $K : \mathbb{R} \to \mathbb{R}$ supported on $[-1, 1]$ satisfying

- **Condition** A'**:** $\int_{\mathbb{R}} K(r) r^{d-1} \mathrm{d}r = \kappa$,
- **Condition** $B'(m)$**:** $\int_{\mathbb{R}} K(r) r^{d+i-1} \mathrm{d}r = 0$ for $i = 1, \dots, m$,
- **Condition** $C'(\beta)$**:** $\int_{\mathbb{R}} K(r)^{-} r^{d-1} \mathrm{d}r \leq \beta$.

We show by recursion on m that for any $\beta > 0$, there exists a such a kernel. For $m = 0$, let K_0 be any smooth even nonnegative function supported on $[-1, 1]$. Then, letting $K = \kappa K_0 / \int_{\mathbb{R}} K_0$, we obtain a kernel K satisfying the desired conditions for any $\beta > 0$. Consider now the case $m > 0$. Let $\beta > 0$.

- If $m + d$ is even, then any K satisfying conditions A', $B'(m-1)$ and $C'(\beta)$ will also satisfy $B'(m)$. Indeed, as K is even, we have $\int_{\mathbb{R}} K(r) r^{m+d-1} \mathrm{d}r = 0$, so that the induction step is proven.
- If $m + d$ is odd, let K be a kernel satisfying conditions A', $B'(m-1)$ and $C'(\beta/2)$. We use the following lemma.

Lemma H.1. *For $i \geq 0$, let $e_i : x \in \mathbb{R} \mapsto x^{i+d-1}$ and fix an integer $m > 0$. Then, for any $a \in \mathbb{R}$, let F_a be the set of smooth functions $f : (1, \infty) \to \mathbb{R}$ with compact support satisfying $\int f e_i = 0$ for $0 \leq i < m$ and $\int f e_m = a$. Then,*

$$\inf \left\{ \int |f(r)| r^{d-1} \mathrm{d}r, \ f \in F_a \right\} = 0. \tag{H.1}$$

Assume first the lemma. Let $a = -\frac{1}{2} \int_{\mathbb{R}} K(r) r^{m+d-1}$ and $f \in F_a$. Then,

$$\begin{cases} \int (K(r) + f(|r|)) r^{d-1} \mathrm{d}r = \kappa + \int f(|r|) r^{d-1} \mathrm{d}r = \kappa \\ \int (K(r) + f(|r|)) r^{i+d-1} \mathrm{d}r = \int f(|r|) r^{i+d-1} \mathrm{d}r = 0 \text{ for } 0 < i < m \\ \int (K(r) + f(|r|)) r^{m+d-1} \mathrm{d}r = \int K(r) r^{m+d-1} \mathrm{d}r + 2 \int_1^\infty f(r) r^{m+d-1} \mathrm{d}r = 0. \end{cases}$$

Hence, the kernel $K + f(|\cdot|)$ satisfies the conditions A and $B'(m)$. Also, we have, as $K(r) = 0$ if $|r| \geq 1$,

$$\int_{\mathbb{R}} (K(r) + f(|r|))_{-} r^{d-1} \mathrm{d}r = \int_{\mathbb{R}} K(r)_{-} \mathrm{d}r + 2 \int_1^\infty f(r)_{-} r^{d-1} \mathrm{d}r$$

$$\leq \beta/2 + \int_1^\infty |f(r)| r^{d-1} \mathrm{d}r,$$

where we used at the last line that $\int_1^\infty f(r)_- r^{d-1}\mathrm{d}r = \int_1^\infty f(r)_+ r^{d-1}\mathrm{d}r = \frac{1}{2}\int_1^\infty |f(r)| r^{d-1}\mathrm{d}r$. Lemma H.1 asserts the existence of $f \in F_a$ with $\int |f(r)| r^{d-1}\mathrm{d}r \leq \beta/2$. For such a choice of f, the kernel $\tilde{K} = K + f(|\cdot|)$ satisfies also $C'(\beta)$. Finally, f has a compact support, included in $[0, R]$ for some $R > 0$. The kernel $\tilde{K}_{1/R}$ is supported on $\mathcal{B}(0,1)$, and satisfies conditions A', $B'(m)$ and $C'(\beta)$. This concludes the induction step, and the proof of the existence of kernels satisfying conditions A, $B(m)$ and $C(\beta)$.

Proof of Lemma H.1. Consider functions f supported on $[r_0, r_1]$ for some $1 < r_0 \leq r_1$ to fix. Let G_{r_0,r_1} be the subspace of $L_2([r_0,r_1])$ spanned by the functions e_i for $0 \leq i \leq m-1$ and let g_m be the projection of e_m on $G^\perp_{r_0,r_1}$ the orthogonal space of G_{r_0,r_1}, with L_2 norm ℓ. The function $f = \frac{a g_m}{\ell^2}$ is a polynomial of degree m restricted to $[r_0,r_1]$ and satisfies $\int f e_i = 0$ for $0 \leq i \leq m-1$ by construction, with $\int f e_m = \frac{a}{\ell^2}\int e_m g_m = a$. Also, we have for any polynomial $P \in G_{r_0,r_1}$,

$$\begin{aligned}
\|e_m - P\|^2_{L_2([r_0,r_1])} &= \int_{r_0}^{r_1} |r^{m+d-1} - P(r)|^2 \mathrm{d}r = \int_1^{\frac{r_1}{r_0}} r_0 |(r_0 r)^{d+m-1} - P(r r_0)|^2 \mathrm{d}r \\
&= r_0^{2(d+m)-1} \int_1^{\frac{r_1}{r_0}} |r^{d+m-1} - r_0^{-(d+m-1)} P(r r_0)|^2 \mathrm{d}r.
\end{aligned}$$

As $r \mapsto r_0^{-(d+m-1)} P(r r_0)$ is an element of $G_{1,r_1/r_0}$, letting $r_1 = 2r_0$, we obtain

$$\begin{aligned}
\ell^2 = \|g_m\|^2_{L_2([r_0,r_1])} &= \min_{P \in G_{r_0,r_1}} \|e_m - P\|^2_{L_2([r_0,r_1])} \\
&= r_0^{2(d+m)-1} \min_{P \in G_{1,2}} \|e_m - P\|^2_{L_2([1,2])} = C r_0^{2(d+m)-1},
\end{aligned}$$

where $C = C_m > 0$ is the distance between e_m restricted to $[1,2]$ and $G_{1,2}$. The function f is not smooth so that it does not belong to F_a. To overcome this issue, we consider a smooth kernel ρ on $\mathbb{R}$ satisfying $\int \rho = 1$ and $\int \rho(r) r^i \mathrm{d}r = 0$ for $i = 1, \dots, m+d-1$, with support included in $\mathcal{B}_{\mathbb{R}}(0, r_0/2)$. See e.g. [BH19, Section 3.2] for the construction of such a kernel ρ. The map $\rho * f$ is supported on $(1,\infty)$ and it is straighforward to check that $\rho * f \in F_a$ for $r_0 > 2$. By Young's inequality, $\|\rho * f\|_{L_2(\mathbb{R})} \leq \|\rho\|_{L_\infty(\mathbb{R})} \|f\|_{L_2(\mathbb{R})}$, so that

$$\begin{aligned}
\int |\rho * f(r)| r^{d-1} \mathrm{d}r &\leq \left(\int_{r_0/2}^{5r_0/2} r^{2d-2} \mathrm{d}r \right)^{1/2} \|\rho * f\|_{L_2(\mathbb{R})} \leq \left(c_d r_0^{2d-1} \right)^{1/2} \|\rho\|_{L_\infty(\mathbb{R})} \|f\|_{L_2(\mathbb{R})} \\
&\leq C_{d,m} a r_0^{-m}
\end{aligned}$$

By letting r_0 goes to ∞, we see that $\inf \left\{ \int |f(r)| r^{d-1} \mathrm{d}r,\ f \in F_a \right\} = 0$. □

References

[Aam17] Eddie Aamari. *Vitesses de convergence en inférence géométrique*. PhD thesis, Paris Saclay, 2017.

[AKC+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.

[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.

[Aub82] Thierry Aubin. *Nonlinear Analysis on Manifolds. Monge-Ampère Equations*. Grundlehren der mathematischen Wissenschaften. Springer New York, 1982.

[BB00] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numerische Mathematik*, 84:375–393, 01 2000.

[BCH18] Gérard Besson, Gilles Courtois, and Sa'ar Hersonsky. Poincaré inequality on complete Riemannian manifolds with Ricci curvature bounded below. *Mathematical Research Letters*, 25(6):1741–1769, 2018.

[BCS10] Lorenzo Brasco, Guillaume Carlier, and Filippo Santambrogio. Congested traffic dynamics, weak flows and very degenerate elliptic equations. *Journal de mathématiques pures et appliquées*, 93(6):652–671, 2010.

[BH19] Clément Berenfeld and Marc Hoffmann. Density estimation on an unknown submanifold. *arXiv preprint arXiv:1910.08477*, 2019.

[BHHS20] Clément Berenfeld, John Harvey, Marc Hoffmann, and Krishnan Shankar. Estimating the reach of a manifold via its convexity defect function. *arXiv preprint arXiv:2001.08006*, 2020.

[Bre03] Yann Brenier. Extended Monge-Kantorovich theory. In *Optimal transportation and applications*, pages 91–121. Springer, 2003.

[Bre10] Haim Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Universitext. Springer New York, 2010.

[BRS+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, 2012.

[BS17] Tyrus Berry and Timothy Sauer. Density estimation on manifolds with boundary. *Computational Statistics & Data Analysis*, 107:1–17, 2017.

[Bur94] Heinrich Burkhardt. Sur les fonctions de Green relatives à un domaine d'une dimension. *Bulletin de la Société Mathématique de France*, 22:71–75, 1894.

[CGK+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.

[dC92] Manfredo Perdigao do Carmo. *Riemannian Geometry*. Mathematics (Boston, Mass.). Birkhäuser, 1992.

[Div20] Vincent Divol. Minimax adaptive estimation in manifold inference. *arXiv preprint arXiv:2001.04896*, 2020.

[Div21] Vincent Divol. A short proof on the rate of convergence of the empirical measure for the Wasserstein distance. *arXiv preprint arXiv:2101.08126*, 2021.

[DSS13] Steffen Dereich, Michael Scheutzow, and Reik Schottstedt. Constructive quantization: Approximation by empirical measures. In *Annales de l'IHP Probabilités et statistiques*, volume 49, pages 1183–1203, 2013.

[Dud69] Richard Mansfield Dudley. The speed of mean Glivenko-Cantelli convergence. *The Annals of Mathematical Statistics*, 40(1):40–50, 1969.

[DZ01] M.C. Delfour and J.P. Zolésio. *Shapes and Geometries: Analysis, Differential Calculus, and Optimization*. Advances in Design and Control. Society for Industrial and Applied Mathematics, 2001.

[Fed59] Herbert Federer. Curvature measures. *Transactions of the American Mathematical Society*, 93(3):418–491, 1959.

[FG15] Nicolas Fournier and Arnaud Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3-4):707–738, 2015.

[GN15] Evarist Giné and Richard Nickl. *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2015.

[GPPVW12] Christopher Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman. Minimax manifold estimation. *Journal of machine learning research*, 13(May):1263–1291, 2012.

[GW03] Joachim Giesen and Uli Wagner. Shape dimension and intrinsic metric from samples of manifolds with high co-dimension. In *Proceedings of the nineteenth annual symposium on Computational geometry*, pages 329–337, 2003.

[H$^+$96] Tsutomu Hiroshima et al. Construction of the Green function on Riemannian manifold using harmonic coordinates. *Journal of Mathematics of Kyoto University*, 36(1):1–30, 1996.

[HA05] Matthias Hein and Jean-Yves Audibert. Intrinsic dimensionality estimation of submanifolds in $\mathbb{R}^d$. In *Proceedings of the 22nd international conference on Machine learning*, pages 289–296. ACM, 2005.

[Hen90] Harrie Hendriks. Nonparametric estimation of a probability density on a Riemannian manifold using Fourier expansions. *The Annals of Statistics*, pages 832–849, 1990.

[KRW16] Jisu Kim, Alessandro Rinaldo, and Larry Wasserman. Minimax rates for estimating the dimension of a manifold. *arXiv preprint arXiv:1605.01011*, 2016.

[L$^+$20] Jing Lei et al. Convergence and concentration of empirical measures under Wasserstein distance in unbounded functional spaces. *Bernoulli*, 26(1):767–798, 2020.

[LJM09] Anna V Little, Yoon-Mo Jung, and Mauro Maggioni. Multiscale estimation of intrinsic dimensionality of data sets. In *2009 AAAI Fall Symposium Series*, 2009.

[LLW07] Han Liu, John Lafferty, and Larry Wasserman. Sparse nonparametric density estimation in high dimensions using the rodeo. In *Artificial Intelligence and Statistics*, pages 283–290, 2007.

[Lun18] A. Lunardi. *Interpolation Theory*. Publications of the Scuola Normale Superiore. Scuola Normale Superiore, 2018.

[LZZL13] Jicai Liu, Riquan Zhang, Weihua Zhao, and Yazhao Lv. A robust and efficient estimation method for single index models. *Journal of Multivariate Analysis*, 122:226–238, 2013.

[NSW08] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete & Computational Geometry*, 39(1-3):419–441, 2008.

[PC19] Gabriel Peyré and Marco Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

[Pel05] Bruno Pelletier. Kernel density estimation on Riemannian manifolds. *Statistics & probability letters*, 73(3):297–304, 2005.

[Pey18] Rémi Peyre. Comparison between W_2 distance and $\dot{H}^{-1}$ norm, and localization of Wasserstein distance. *ESAIM. Control, Optimisation and Calculus of Variations*, 24(4), 2018.

[PR84] Jean-Baptiste Poly and Gilles Raby. Fonction distance et singularités. *Bulletin de Sciences Mathématiques*, 108:187–195, 1984.

[Ros70] Haskell P Rosenthal. On the subspaces of L_p $(p > 2)$ spanned by sequences of independent random variables. *Israel Journal of Mathematics*, 8(3):273–303, 1970.

[San15] Filippo Santambrogio. Optimal transport for applied mathematicians. *Birkäuser, NY*, 55:58–63, 2015.

[Sog17] Christopher D. Sogge. *Fourier Integrals in Classical Analysis.* Cambridge Tracts in Mathematics. Cambridge University Press, 2 edition, 2017.

[SP18] Shashank Singh and Barnabás Póczos. Minimax distribution estimation in Wasserstein distance. *arXiv preprint arXiv:1802.08855*, 2018.

[TGHS20] Nicolás García Trillos, Moritz Gerlach, Matthias Hein, and Dejan Slepčev. Error estimates for spectral convergence of the graph Laplacian on random geometric graphs toward the Laplace–Beltrami operator. *Foundations of Computational Mathematics*, 20(4):827–887, 2020.

[Tib96] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

[Tri92] H. Triebel. *Theory of Function Spaces II.* Monographs in Mathematics. Springer Basel, 1992.

[Tsy08] Alexandre B Tsybakov. *Introduction to nonparametric estimation.* Springer Science & Business Media, 2008.

[Vil08] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

[WB19a] Jonathan Weed and Francis Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli*, 25(4A):2620–2648, 2019.

[WB19b] Jonathan Weed and Quentin Berthet. Estimation of smooth densities in Wasserstein distance. *arXiv preprint arXiv:1902.01778*, 2019.

[WW20] Hau-Tieng Wu and Nan Wu. Strong uniform consistency with rates for kernel density estimators with general kernels on manifolds, 2020.

[Yu97] Bin Yu. Assouad, Fano, and Le Cam. In *Festschrift for Lucien Le Cam*, pages 423–435. Springer, 1997.

[ZJRS16] Hongyi Zhang, Sashank J Reddi, and Suvrit Sra. Riemannian SVRG: Fast stochastic optimization on Riemannian manifolds. *Advances in Neural Information Processing Systems*, 29:4592–4600, 2016.