Journal of Machine Learning Research 27 (2026) 1-55 Submitted 5/24; Revised 2/26; Published 9/26

# Integral Probability Metrics on Submanifolds: Interpolation Inequalities and Optimal Inference

**Arthur Stéphanovitch** STEPHANOVITCH@DMA.ENS.FR
*Département de Mathématiques et Applications*
*Ecole Normale Supérieure, Université PSL, CNRS*
*F-75005 Paris, France*

**Editor:** Kenji Fukumizu

## Abstract

We study interpolation inequalities between Hölder Integral Probability Metrics (IPMs) in the case where the measures have densities on closed submanifolds. Precisely, it is shown that if two probability measures $\mu$ and $\mu^\star$ have $\beta$-smooth densities with respect to the volume measure of some submanifolds $\mathcal{M}$ and $\mathcal{M}^\star$ respectively, then the Hölder IPMs $d_{\mathcal{H}_1^\gamma}$ of smoothness $\gamma \geq 1$ and $d_{\mathcal{H}_1^\eta}$ of smoothness $\eta > \gamma$, satisfy $d_{\mathcal{H}_1^\gamma}(\mu, \mu^\star) \lesssim d_{\mathcal{H}_1^\eta}(\mu, \mu^\star)^{\frac{\beta+\gamma}{\beta+\eta}}$, up to logarithmic factors. We provide an application of this result to high-dimensional inference. These functional inequalities turn out to be a key tool for density estimation on unknown submanifold. In particular, it allows to build the first estimator attaining optimal rates of estimation for all the distances $d_{\mathcal{H}_1^\gamma}$, $\gamma \in [1, \infty)$ simultaneously.

**Keywords:** integral probability metrics, interpolation inequality, manifold data, distribution estimation, minimax rate

## 1. Introduction

The development of methods for estimating probability measures from data has been a major focus of contemporary statistics and machine learning (Bishop, 2006). This pursuit has given rise to thriving research fields such as generative models (Ruthotto and Haber, 2021), reinforcement learning (Szepesvári, 2022) and geometrical inference (Chazal et al., 2011). In order to evaluate the precision of the estimation, many methods use Integral Probability Metrics (IPMs) (Müller, 1997) as the distances to compare probability measures. It consists in choosing a class of functions $\mathcal{F}$ and looking at the distance between two probability measures $\mu, \mu^\star$ defined by

$$d_{\mathcal{F}}(\mu, \mu^\star) := \sup_{f \in \mathcal{F}} \left| \int f(x) d\mu(x) - \int f(x) d\mu^\star(x) \right|. \tag{1}$$

Notable examples of IPMs include the total variation distance (Verdú, 2014) ($\mathcal{F}$ is the class of functions taking value in $[-1, 1]$), the Wasserstein distance (Villani, 2009) ($\mathcal{F}$ is the class of 1-Lipschitz functions) and Maximum Mean Discrepancies (Smola et al., 2006) ($\mathcal{F}$ is the unit ball of a RKHS). In this paper we focus on the Hölder IPMs that have been used in generative models (Arjovsky et al. (2017), Chakraborty and Bartlett (2024)) and geometric learning (Tang and Yang, 2023). For $\eta > 0$, $\mathcal{X}, \mathcal{Y}$ two subsets of Euclidean spaces and $f = (f_1, ..., f_p) \in C^{\lfloor \eta \rfloor}(\mathcal{X}, \mathcal{Y})$ the set of $\lfloor \eta \rfloor := \max\{k \in \mathbb{N}_0 |\ k \leq \eta\}$ times differentiable

©2026 Arthur Stéphanovitch.

License: CC-BY 4.0, see https://creativecommons.org/licenses/by/4.0/. Attribution requirements are provided at http://jmlr.org/papers/v27/24-0804.html.

functions, denote $\partial^\nu f_i = \frac{\partial^{|\nu|} f_i}{\partial x_1^{\nu_1} ... \partial x_d^{\nu_d}}$ the partial differential operator for any multi-index $\nu = (\nu_1, ..., \nu_d) \in \mathbb{N}_0^d$ with $|\nu| := \nu_1 + ... + \nu_d \leq \lfloor \eta \rfloor$. Write $\|f_i\|_{\eta - \lfloor \eta \rfloor} = \sup_{x \neq y} \frac{f_i(x) - f_i(y)}{\min\{1, \|x-y\|^{\eta - \lfloor \eta \rfloor}\}}$ and let

$$\mathcal{H}_K^\eta(\mathcal{X}, \mathcal{Y}) = \Big\{ f \in C^{\lfloor \eta \rfloor}(\mathcal{X}, \mathcal{Y}) \; \Big| \max_i \sum_{|\nu| \leq \lfloor \eta \rfloor} \|\partial^\nu f_i\|_{L^\infty(\mathcal{X}, \mathcal{Y})} + \sum_{|\nu| = \lfloor \eta \rfloor} \|\partial^\nu f_i\|_{\eta - \lfloor \eta \rfloor} \leq K \Big\}$$

denote the ball of radius $K$ of the Hölder space $\mathcal{H}^\eta(\mathcal{X}, \mathcal{Y})$, the set of functions $f : \mathcal{X} \to \mathcal{Y}$ of regularity $\eta$. In this work we are interested in comparing the Hölder IPMs

$$d_{\mathcal{H}_1^\eta}(\mu, \mu^\star) := \sup_{f \in \mathcal{H}_1^\eta} \int f(x) d\mu(x) - \int f(x) d\mu^\star(x) \tag{2}$$

of different regularities $\eta$. The popularity of the Hölder IPM comes from its equivalence with the Wasserstein metric in the case $\eta = 1$. This metric enjoys several useful properties including the metrization of weak convergence of measures Villani (2021). Unfortunately, the Wasserstein distance has been shown to be computationally expensive (Peyré et al., 2019). Then, the need of finding smaller classes than the Lipschitz class in the IPM (1) arises. Ideally one wants to have a class $\mathcal{F}$ that is smaller than the Lipschitz class but that still controls the Wasserstein distance. Classical interpolation inequalities (Lunardi, 2018) allow to show the connection between the different IPMs: if $\mu$ and $\mu^\star$ have densities with respect to the Lebesgue measure and that belong to $\mathcal{H}_1^\beta$ for $\beta > 0$, then for all $0 < \gamma < \eta$, we have

$$d_{\mathcal{H}_1^\gamma}(\mu, \mu^\star) \lesssim d_{\mathcal{H}_1^\eta}(\mu, \mu^\star)^{\frac{\beta + \gamma}{\beta + \eta}}, \tag{3}$$

up to logarithmic terms (see Theorem 5). These inequalities are key in the estimation of smooth densities as they allow to show that if an estimator $\hat{\mu}$ of $\mu^\star$ attains optimal rates of estimation for the $d_{\mathcal{H}_1^\eta}$ distance, then it also attains optimal rates for the $d_{\mathcal{H}_1^\gamma}$ distance for all $\gamma \in (0, \eta)$ (Stéphanovitch et al., 2023).

Recent use of high dimensional data have shed light on the fact that the optimal rates for the estimation of probability measures, degenerate exponentially with respect to the dimension of the data (Köppen, 2000). This is the so called "curse of dimensionality" and has led researchers to forsake density estimation to rather focus on measures with low intrinsic dimensional structure. In this context, a natural setting is to consider measures having densities with respect to the volume measure of a submanifold (Divol, 2022). However, classical interpolation inequalities like (3) no longer apply in this setting. In order to obtain efficient methods taking into account the low dimensional structure of the measures, it is then crucial to understand how these inequalities can be generalized.

Stéphanovitch et al. (2023) have recently shown a generalization of these inequalities in the case of push forward measures. Supposing that $\mu$ and $\mu^\star$ have $\beta$-regular densities with respect to the volume measure of submanifolds $\mathcal{M}$ and $\mathcal{M}^\star$ respectively and that $\mu$ and $\mu^\star$ are push forward measures of $\beta + 1$-regular maps from the $d$-dimensional torus $\mathbb{T}^d$, it is shown that inequality (3) still stands in the case $\gamma \in [1, \beta + 1]$ and $\eta = \beta + 1$. In this paper, this result is generalized to measures having $\beta$-regular densities with respect to the volume measure of any closed submanifolds (without the push forward assumption) and to any

$0 < \gamma < \eta$. This generalization allows measure estimation methods like Stéphanovitch et al. (2023) to be generalized to the manifold setting. Precisely, we build the first estimator of density on unknown manifold that attains optimal rates for all the distances $d_{\mathcal{H}_1^\gamma}$, $\gamma \in [1, \infty)$ simultaneously. Its construction is more simpler than previous estimators like Tang and Yang (2023) as its use of high regularity IPMs implicitly regularizes the data.

To motivate our result from a functional point of view, it is detailed in Section 3.1 how the generalization of inequality (3) is the extension of some interpolation inequalities between Besov norms of $L^2$ functions, to Besov norms of signed measures having low dimensional smooth structure.

The remainder of this paper is organised as follows. In Section 2, we set the notation, properly define our setting, and present the main results of the paper. In Section 3, we give a walk-through of the proof by stating the key intermediate results that allow to obtain Theorem 14. In Section 4 we present a direct application of Theorem 14 to the estimation of density on unknown manifold. In Section 5, we detail a simple example to get some intuition on Theorem 14. Technical points of the proofs are gathered in the supplementary material.

## 2. Preliminaries and Main Results

### 2.1 Setting

Let us start by defining the precise setting of our work. The results focus on $\beta$-regular densities with $\beta \geq 1$ with respect to the volume measure of closed (i.e. compact without boundary) manifolds that are of dimension $d \in \mathbb{N}^\star$ and that are immersed in the ambient space $\mathbb{R}^p$ with $p > d$ an integer.

Throughout the paper, all manifolds (including $\mathcal{M}$ and $\mathcal{M}^\star$) are assumed to have the same fixed intrinsic dimension $d$.

For $\mathcal{M}$ a closed $d$-dimensional submanifold embedded in $\mathbb{R}^p$ and $x \in \mathcal{M}$, note

$$U_x = B^p(x, K^{-1}) \cap \mathcal{M} \quad \text{and} \quad \varphi_x : U_x \to \mathcal{T}_x(\mathcal{M}),\ \varphi_x = \pi_{\mathcal{T}_x(\mathcal{M})} - x,$$

for $\pi_{\mathcal{T}_x(\mathcal{M})}$ the orthogonal projection onto the tangent space $\mathcal{T}_x(\mathcal{M})$ of $\mathcal{M}$ at the point $x$ and $B^p(x, K^{-1})$ the $p$-dimensional ball of radius $K^{-1} > 0$ around $x$. Let us fix a certain $K > 1$ and define the notion of regularity for manifold used in this work.

**Definition 1** *A closed $d$-dimensional submanifold $\mathcal{M}$ is said to verify the $(\beta + 1, K)$-manifold condition if $\mathcal{M} \subset B^p(0, K)$ and for all $x \in \mathcal{M}$, $\varphi_x : U_x \to \mathcal{T}_x(\mathcal{M})$ is a diffeomorphism and verifies*

$$\varphi_x^{-1} \in \mathcal{H}_K^{\beta+1}(\varphi_x(U_x), U_x).$$

Definition 1 is equivalent to a reach condition (Federer, 1959) for $\beta + 1$-smooth submanifolds. The reach $r_{\mathcal{M}}$ of a manifold $\mathcal{M}$ corresponds to the largest $\epsilon \geq 0$ such that its orthogonal projection $\pi_{\mathcal{M}}$ is well defined from $\mathcal{M}^\epsilon$ to $\mathcal{M}$ with $\mathcal{M}^\epsilon = \{x \in \mathbb{R}^p |\ d(x, \mathcal{M}) < \epsilon\}$ the set of points having their distance to the manifold $d(x, \mathcal{M})$ smaller than $\epsilon$. Definition 1 is very general as any $\beta + 1$-regular closed submanifold with reach uniformly bounded from below, verifies the $(\beta + 1, K)$-manifold condition for a sufficiently large $K$ (Divol, 2022).

For $\alpha \in [0, \beta+1]$ and $A \subset \mathcal{M}$ an open subset of a $(\beta+1)$-smooth manifold $\mathcal{M}$, the ball of radius $K > 0$ of the $\alpha$-smooth Hölder class of functions $\mathcal{H}^\alpha(A)$ consists of all functions $f : A \to \mathbb{R}$ such that for every local charts $(U, \varphi)$ such that $U \cap A \neq \varnothing$, its localization $f \circ \varphi^{-1} : \varphi(U \cap A) \to \mathbb{R}$ satisfies $f \circ \varphi^{-1} \in \mathcal{H}^\alpha_K(\varphi(U \cap A))$.

Note that if a manifold has regularity $\lambda$, it is not meaningful to define densities of smoothness $\alpha > \lambda - 1$, since higher-order derivatives may fail to be well defined globally: the change-of-variables formula for densities involves the differentials of the charts. Consequently, for a density of smoothness $\alpha$, the supporting manifold must have smoothness at least $\alpha+1$. Moreover, if the manifold is smoother, say $\lambda > \alpha+1$, the interpolation inequality of Theorem 14 remains unchanged: the IPMs are governed by the density regularity, not by any excess smoothness of the manifold. This is analogous to the Euclidean case where the ambient space is $C^\infty$: the interpolation inequality still depends solely on the density smoothness. Let us now define the notion of regularity for densities used in the paper.

**Definition 2** *A probability measure $\mu$ supported on a submanifold $\mathcal{M}$ is said to verify the $(\beta, K)$-density condition if*

*i) $\mu$ admits a density $f_\mu \in \mathcal{H}^\beta_K(\mathcal{M}, \mathbb{R})$ with respect to the volume measure on $\mathcal{M}$,*

*ii) $f_\mu$ is bounded below by $K^{-1}$.*

Supposing that the densities are bounded below is a classical assumption in the setting of density estimation on manifolds (Divol (2022), Tang and Yang (2023)). Here it is needed in order to obtain smooth optimal transport maps between some push forward measures in Proposition 6 using Caffarelli's regularity (Villani, 2009). For a measure $\mu$ supported on a set $\mathcal{X}$ and a function $g$ defined on $\mathcal{X}$ and taking values in a set $\mathcal{Y}$, we write $g_{\#\mu}$ for the push forward measure of $\mu$ by $g$ i.e. $\forall f \in \mathcal{H}^0_1(\mathcal{X}, \mathcal{Y})$,

$$\int_{\mathcal{Y}} f(x) dg_{\#\mu}(x) = \int_{\mathcal{X}} f(g(u)) d\mu(u).$$

In contrast to the classical Euclidean setting, where interpolation inequalities between Hölder IPMs hold without any lower bound on the densities (as the proof relies solely on wavelet coefficients), the manifold setting requires assuming that the density is bounded below. This assumption ensures that the support of the measure does not exhibit regions of vanishing mass, which would effectively behave like "holes" in the manifold. Such regions would violate the closedness assumption and lead to situations akin to manifolds with boundary, for which the validity of the interpolation inequality is unclear.

### 2.2 Main Results

Throughout, the quantities $C, C_\delta, C_{\delta,\eta}...$ represent constants that can vary from line to line and that only depend on $p, d, \beta, K$ and there underscore. The following theorem provides the core interpolation inequality of our work. Intuitively, it shows that if two probability measures are very close according to a smoother IPM, then they are also close for rougher IPMs. The exponent encodes how the regularity of the densities translates into stability of distances: smoother densities make the measures "harder to distinguish" using rough functions, and this is exactly what the inequality quantifies.

**Theorem 14:** Let $\mathcal{M}, \mathcal{M}^\star$ be two submanifolds satisfying the $(\beta+1, K)$-manifold condition (Definition 1) and $\mu, \mu^\star$ two probability measures satisfying the $(\beta, K)$-density condition (Definition 2) on $\mathcal{M}$ and $\mathcal{M}^\star$ respectively. Then for all $0 < \gamma < \eta$, we have

$$d_{\mathcal{H}_1^\gamma}(\mu, \mu^\star) \leq C \log(1 + d_{\mathcal{H}_1^\eta}(\mu, \mu^\star)^{-1})^{C_2} d_{\mathcal{H}_1^\eta}(\mu, \mu^\star)^\delta,$$

with

$$\delta = \begin{cases} \frac{\beta+\gamma}{\beta+\eta} & \text{if } \gamma \geq 1 \\ \frac{\beta\gamma+\gamma}{\beta+\eta} & \text{if } \gamma \leq 1 \leq \eta \\ \frac{\gamma}{\eta} & \text{if } \eta \leq 1. \end{cases} \tag{4}$$

This result is sharp up to the logarithmic factor in the sense that the exponent $\delta$ is optimal. More precisely, Theorem 17 (stated and discussed in Section 5) constructs an explicit family of manifold-supported measures such that $d_{\mathcal{H}_1^\gamma}(\mu_k, \mu_k^\star) \to 0$ when $k$ tends to the infinity and $d_{\mathcal{H}_1^\gamma}(\mu_k, \mu_k^\star) \geq C^{-1} d_{\mathcal{H}_1^\eta}(\mu_k, \mu_k^\star)^\delta$.

The logarithmic factor in Theorem 14 arises from the use of wavelet theory to describe Hölder spaces. Throughout the paper, we highlight the efficiency of the wavelet tool to describe smoothness as it allows to greatly simplify some geometric proofs. However, Hölder spaces with integer regularity exponent, can not be exactly described by wavelets Haroske (2006). This leads to the appearance of the term $\log\left(d_{\mathcal{H}_1^\eta}(\mu, \mu^\star)^{-1}\right)^{C_2}$ in the inequality, as it allows to have injection of Besov into Hölder spaces. We do not keep track of the extra exponent $C_2$ to avoid unnecessary heaviness, as the obtained exponent is very likely suboptimal.

Theorem 14 is an improvement on Theorem 5.1 in Stéphanovitch et al. (2023) which requires the additional assumptions that the measures are push forward measures from the torus and that $\eta = \beta + 1$. Precisely, it is supposed that there exist $g, g^\star \in \mathcal{H}_K^{\beta+1}(\mathbb{R}^d/\mathbb{Z}^d, \mathbb{R}^p)$ embeddings, such that $\mu = g_{\#\lambda^d}$ and $\mu^\star = g^\star_{\#\lambda^d}$. In particular, this model can only describe manifolds $\mathcal{M}$ and $\mathcal{M}^\star$ that are smooth deformations of the torus. On the other hand, our extension of the result for any $\eta \geq \gamma$, shows that the inequality behaves like in the classical case and the fact that the measures are on submanifolds (with possibly a complex topology) does not impact the IPMs. Theorem 14 shows that although the measures do not have any regularity from the point of view of the ambient space $\mathbb{R}^p$, their low dimensional regularity still intervenes the same way in their discrimination through IPMs. We also extend the result to the simpler case $\gamma \in (0, 1)$ in Proposition 13, where in this case the inequality is naturally different.

The motivation behind Theorem 14 is to extend the results of Stéphanovitch et al. (2023) to closed submanifolds that are not necessarily smooth deformations of the torus. Specifically, it is built in Section 4 an estimator of density on unknown submanifold attaining optimal rates.

Let us write $\mathcal{F}$ for the set of probability measures $\mu$ such that there exists a submanifold $\mathcal{M}$ satisfying the $(\beta+1, K)$-manifold condition and that $\mu$ satisfies the $(\beta, K)$-density condition on $\mathcal{M}$. Additionally, we write $g(n) = \tilde{O}(f(n))$ for $g(n) \leq C \log(n)^{C_2} f(n)$.

**Theorem 16 :** Let $n \in \mathbb{N}_{>0}$, $\mu^\star \in \mathcal{F}$ and $(X_1, ..., X_n)$ an i.i.d. sample of law $\mu^\star$. Then for $\mu_n^\star := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$, the estimator

$$\hat{\mu} \in \arg\min_{\mu \in \mathcal{F}} d_{\mathcal{H}_1^{d/2}}(\mu, \mu_n^\star), \tag{5}$$

satisfies for all $\gamma \geq 1$,

$$\sup_{\mu^\star \in \mathcal{F}} \mathbb{E}_{X_i}[d_{\mathcal{H}_1^\gamma}(\hat{\mu}, \mu^\star)] = \tilde{O}\left(\inf_{\hat{\theta} \in \Theta} \sup_{\mu^\star \in \mathcal{F}} \mathbb{E}_{X_i}[d_{\mathcal{H}_1^\gamma}(\hat{\theta}, \mu^\star)]\right),$$

where $\Theta$ denotes the set of all possible estimators of $\mu^\star$ based on n sample.

To the best of our knowledge, the estimator (5) is the first one attaining optimal rates for all the distances $d_{\mathcal{H}_1^\gamma}$, $\gamma \in [1, \infty)$ simultaneously. This estimator uses high regularity IPMs in the minimized loss which implicitly regularizes the data. Its construction is much simpler than previous optimal estimators (like Tang and Yang (2023) and Liang (2021)) because it does not need to compute a complex regularization of the empirical measure as the regularization is made through the use of high regularity IPMs.

### 2.2.1 Implications and Potential Applications

Although Theorem 14 is presented here primarily as a tool for constructing a simple estimator of densities on unknown manifolds achieving minimax rates, the interpolation inequality it establishes has broader relevance for machine learning tasks involving manifold-supported distributions. Its key feature is that small discrepancies in a high-regularity Hölder IPM $d_{\mathcal{H}_\eta}$ automatically imply small discrepancies in all lower-regularity IPMs $d_{\mathcal{H}_\gamma}$, with the sharp exponent $\delta$ capturing the smooth structure of the measures. This principle can be leveraged in several settings:

1. **Generative modeling on manifolds.** Many generative models such as GANs, and normalizing flows optimize an IPM between a measure reconstructed from the data $\hat{\mu}_n$ and the model distribution $\mu_\theta$. In practice, training often benefits from using smooth discriminators (high $\eta$) to stabilize optimization, while evaluation or theoretical guarantees are typically given in coarser metrics such as Wasserstein-1 ($\gamma = 1$). Theorem 14 provides a theoretical link: controlling a high-regularity IPM suffices to control lower-regularity IPMs, which justifies training with smooth losses without loosing guarantees. This principle has already been exploited in Stéphanovitch (2025) to design a generative model that achieves state-of-the-art results on manifold-supported data using high-regularity IPMs.

2. **Two-sample testing and goodness-of-fit.**
In manifold-aware two-sample tests (Wang et al., 2023), the test statistic is often an IPM that measures a smooth discrepancy between the empirical and reference distributions. Theorem 14 implies that a test based on a high-regularity IPM can detect differences that are meaningful in coarser distances as well, a theoretical justification for the power of smooth distances in low-dimensional manifold settings. Moreover, it might allow one to derive finite-sample power guarantees by transferring bounds from smooth metrics to Wasserstein metric.

3. **Model-based reinforcement learning.** In many control and RL applications, the true and learned transition kernels are supported on low-dimensional manifolds (e.g., in robotics, where trajectories lie on a configuration manifold). Performance guarantees for planning and policy evaluation are typically expressed in low-regularity IPMs such as Wasserstein-1, but model learning often provides control in a smoother metric (e.g., via kernel embeddings or smoothed likelihoods). Theorem 9 provides a principled tool to propagate smooth model errors to the coarser distances that directly govern policy regret, potentially tightening sample-complexity bounds for model-based RL (Shen et al., 2020).

More broadly, since high-regularity function classes have smaller covering numbers, they can often be estimated more efficiently from finite samples; Theorem 14 then ensures that this control transfers to larger, lower-regularity classes such as Lipschitz functions. This makes the result a useful tool whenever data or distributions evolve on low-dimensional structures, with potential applications in generative modeling, two-sample testing, domain adaptation, and model-based reinforcement learning.

### 2.3 Besov Spaces and Additional Notations

Throughout the paper, we use the connection between Hölder and Besov spaces extensively (Triebel, 2010). Let us define the Besov spaces through their wavelet characterization. Let $\psi, \phi \in \mathcal{H}^{\beta+2}(\mathbb{R}, \mathbb{R})$ be a compactly supported *scaling* and *wavelet* function respectively (see Daubechies wavelets (Daubechies, 1988)). For ease of notation, the functions $\psi, \phi$ will be written $\psi_0, \psi_1$ respectively. Then for $j \in \mathbb{N}, l \in \{1, ..., 2^p - 1\}, w \in \mathbb{Z}^p$, the family of functions

$$\psi_{0w}(x) = \prod_{i=1}^{p} \psi_0(x_i - w_i) \quad , \quad \psi_{jlw}(x) = 2^{jp/2} \prod_{i=1}^{p} \psi_{l_i}(2^j x_i - w_i)$$

form an orthonormal basis of $L^2(\mathbb{R}^p, \mathbb{R})$ (with $l_i$ the $i$-th digit of the base-2-decomposition of $l$). Let $q_1, q_2 \geq 1, s > 0, b \geq 0$ be such that $\beta + 2 > s$. The Besov space $\mathcal{B}^{s,b}_{q_1,q_2}(\mathbb{R}^p, \mathbb{R})$ consists of functions $f$ that admit a wavelet expansion in $L^2$:

$$f(x) = \sum_{w \in \mathbb{Z}^p} \alpha_f(w) \psi_{0w}(x) + \sum_{j=0}^{\infty} \sum_{l=1}^{2^p - 1} \sum_{w \in \mathbb{Z}^p} \alpha_f(j, l, w) \psi_{jlw}(x)$$

equipped with the norm

$$\begin{aligned} \|f\|_{\mathcal{B}^{s,b}_{q_1,q_2}} = & \left( \left( \sum_{w \in \mathbb{Z}^p} |\alpha_f(w)|^{q_1} \right)^{q_2/q_1} \right. \\ & \left. + \sum_{j=0}^{\infty} 2^{jq_2(s+p/2-p/q_1)} (1+j)^{bq_2} \sum_{l=1}^{2^p-1} \Big( \sum_{w \in \mathbb{Z}^p} |\alpha_f(j,l,w)|^{q_1} \Big)^{q_2/q_1} \right)^{1/q_2}. \end{aligned}$$

with the usual modification for $q_1, q_2 = \infty$. Note that for $b = 0$, $\mathcal{B}^{s,0}_{q_1,q_2}$ coincides with the classical Besov space $\mathcal{B}^{s}_{q_1,q_2}$ (Giné and Nickl, 2015). For simplicity of notation, we write

$$f(x) = \sum_{j=0}^{\infty} \sum_{l=1}^{2^p} \sum_{w \in \mathbb{Z}^p} \alpha_f(j, l, w) \psi_{jlw}(x)$$

with the convention that $\psi_{02^p w} = \psi_{0w}$ and for all $j \geq 1$, $\psi_{j2^p w} = 0$. The Besov spaces can be generalized for any $s \in \mathbb{R}$ as a subspace of the space of tempered distribution $\mathcal{S}^{'}(\mathbb{R}^p)$. Indeed for $f \in \mathcal{S}^{'}(\mathbb{R}^p)$, writing $\alpha_f(j,l,w) = \langle f, \psi_{jlw} \rangle$, the Besov space for $s \in \mathbb{R}$ is defined as

$$\mathcal{B}^{s,b}_{q_1,q_2} = \{f \in \mathcal{S}^{'}(\mathbb{R}^p) | \|f\|_{\mathcal{B}^{s,b}_{q_1,q_2}} < \infty\}.$$

The same way, we write $\mathcal{B}^{s,b}_{q_1,q_2}(K) = \{f \in \mathcal{S}^{'}(\mathbb{R}^p) | \|f\|_{\mathcal{B}^{s,b}_{q_1,q_2}} \leq K\}$. In the following we will use intensively the connection between Hölder and Besov spaces.

**Lemma 3** *(Proposition 4.3.23 Giné and Nickl (2015), (4.63) Haroske (2006)) If $\alpha > 0$ is a non integer, then*

$$\mathcal{H}^{\alpha}(\mathbb{R}^p, \mathbb{R}) = \mathcal{B}^{\alpha}_{\infty,\infty}(\mathbb{R}^p, \mathbb{R})$$

*with equivalent norms. If $\alpha \geq 0$ is an integer, then $\forall \epsilon > 0$*

$$\mathcal{B}^{\alpha,1+\epsilon}_{\infty,\infty}(\mathbb{R}^p, \mathbb{R}) \hookrightarrow \mathcal{H}^{\alpha}(\mathbb{R}^p, \mathbb{R}) \hookrightarrow \mathcal{B}^{\alpha}_{\infty,\infty}(\mathbb{R}^p, \mathbb{R}),$$

*where we write $A \hookrightarrow B$ if the function space $A$ compactly injects in the function space $B$.*

This lemma states that Besov infinity and Hölder spaces coincides except when the regularity exponent is an integer. In particular, it implies that Theorem 12 extends to Besov infinity IPMs up to additional logarithmic terms for integer regularity exponents.

We write $\langle \cdot, \cdot \rangle$ the dot product on $\mathbb{R}^p$, $\|x\|$ the Euclidean norm of a vector $x$. For $a, b \in \mathbb{R}$, $a \wedge b$ and $a \vee b$ denote the minimum and maximum value between $a$ and $b$ respectively. We write $\text{Lip}_1$ for the set of 1-Lipschitz functions. The support of a function $f$ is denoted by $supp(f)$. We denote by Id the identity application from a Euclidean space to itself.

For any map $f : \mathbb{R}^k \to \mathbb{R}^l$ we denote by $\nabla f$ the differential of $f$ and by $\|\nabla f(x)\|$ its operator norm at the point $x$. We denote $(\nabla f(x))^\top$ its transpose matrix and if $k \leq l$, $\lambda_{\min}((\nabla f(x))^\top \nabla f(x))$ corresponds to the smallest eigenvalue of the matrix $(\nabla f(x))^\top \nabla f(x)$. We write $\int_{\mathcal{M}} f(x) d\lambda_{\mathcal{M}}(x)$ for the integration of a function $f : \mathcal{M} \to \mathbb{R}$ with respect to the volume measure on $\mathcal{M}$ i.e. the $d$-dimensional Hausdorff measure. The Hausdorff distance between two submanifold $\mathcal{M}, \mathcal{M}^\star$ is denoted by $\mathbb{H}(\mathcal{M}, \mathcal{M}^\star)$.

## 3. Walk-Through of the Construction of Theorem 14

In this section the main sub-results allowing to prove Theorem 14 are detailed. We first prove the classical inequality in the full dimensional setting. Then, it is explained how it can be generalized to the submanifold case.

### 3.1 The Classical Inequality

To build intuition, we first revisit the full-dimensional case where the measures have smooth densities in $\mathbb{R}^p$. In this simpler setting, the interpolation inequality can be understood as a comparison between different levels of 'magnifying glasses' on the difference of the measures: a high-regularity IPM looks only at very smooth features, whereas a low-regularity IPM is more sensitive. By decomposing the measures into wavelets, we see how smoothness dampens high-frequency components, leading naturally to the power-law relation between IPMs of different orders.

Let us give a sketch of the proof of the inequality. Take two densities $f, f^\star \in L^2(\mathbb{R}^p, \mathbb{R})$ and a potential $h \in \mathcal{B}^s_{\infty,\infty}(\mathbb{R}^p, \mathbb{R})$ to compare them. We call $h$ a potential in reference to Kantorovich potentials in optimal transport (Santambrogio, 2015). The family of wavelets functions $(\psi_{jlw})_{jlw}$ being an orthonormal basis of $L^2$, we can write for $\tau > 0$ and $q \in (0,1), q^\star = (1-1/q)^{-1}$,

$$\int_{\mathbb{R}^p} h(x)(f(x) - f^\star(x))d\lambda^p(x) = \sum_{j=0}^{\infty}\sum_{l=1}^{2^p}\sum_{w\in\mathbb{Z}^p} \alpha_h(j,l,w)(\alpha_f(j,l,w) - \alpha_{f^\star}(j,l,w))$$
$$\leq \sum_{j=0}^{\infty}\sum_{l=1}^{2^p}\sum_{w\in\mathbb{Z}^p} 2^{-j\frac{\tau}{q}+j\frac{\tau}{q}}|\alpha_h(j,l,w)(\alpha_f(j,l,w) - \alpha_{f^\star}(j,l,w))|^{1/q+1/q^\star}.$$

Then, applying Hölder's inequality we get

$$\int_{\mathbb{R}^p} h(x)(f(x) - f^\star(x))d\lambda^p(x)$$
$$\leq \left(\sum_{j=0}^{\infty}\sum_{l=1}^{2^p}\sum_{w\in\mathbb{Z}^p} 2^{-j\tau}|\alpha_h(j,l,w)(\alpha_f(j,l,w) - \alpha_{f^\star}(j,l,w))|\right)^{\frac{1}{q}}$$
$$\times \left(\sum_{j=0}^{\infty}\sum_{l=1}^{2^p}\sum_{w\in\mathbb{Z}^p} 2^{j\tau\frac{q^\star}{q}}|\alpha_h(j,l,w)(\alpha_f(j,l,w) - \alpha_{f^\star}(j,l,w))|\right)^{\frac{1}{q^\star}}. \tag{6}$$

This gives us (see the proof of Proposition 4 for details) a first interpolation inequality between Besov IPMs:

$$d_{\mathcal{B}^s_{\infty,\infty}}(f, f^\star) \leq d_{\mathcal{B}^{s+\tau}_{\infty,\infty}}(f, f^\star)^{\frac{1}{q}}\ d_{\mathcal{B}^{s-\frac{q^\star}{q}\tau}_{\infty,\infty}}(f, f^\star)^{\frac{1}{q^\star}}. \tag{7}$$

Finding the smallest $q \in (0,1)$ such that the quantity $d_{\mathcal{B}^{s-\frac{q^\star}{q}\tau}_{\infty,\infty}}(f, f^\star)$ is finite, we obtain the following result.

**Proposition 4** *Let $f, f^\star \in \mathcal{B}^{\beta,2}_{\infty,\infty}(\mathbb{R}^p, \mathbb{R}, K)$ compactly supported in $B^p(0,K)$. Then for all $\alpha \geq \gamma > 0$, we have*

$$\sup_{h\in\mathcal{B}^\gamma_{\infty,\infty}(1)} \int_{\mathbb{R}^p} h(x)(f(x) - f^\star(x))d\lambda^p(x) \leq C \sup_{h\in\mathcal{B}^\alpha_{\infty,\infty}(1)} \left(\int_{\mathbb{R}^p} h(x)(f(x) - f^\star(x))d\lambda^p(x)\right)^{\frac{\beta+\gamma}{\beta+\alpha}}.$$

The proof of Proposition 4 can be found in Section C.7.1. Note that this result can be seen as a particular case of classical interpolation inequalities between Besov spaces (Hajaiej et al., 2010). Indeed, noticing that for $g \in L^2$ we have

$$\sup_{h\in\mathcal{B}^\gamma_{\infty,\infty}(1)} \int_{\mathbb{R}^p} h(x)g(x)d\lambda^p(x) = \|g\|_{\mathcal{B}^{-\gamma}_{1,1}},$$

Proposition 4 is the application to the function $g = f - f^\star$ of the inequality

$$\|g\|_{\mathcal{B}^{-\gamma}_{1,1}} \leq \|g\|^{\frac{\alpha-\gamma}{\beta+\alpha}}_{\mathcal{B}^{\beta,2}_{1,1}}\ \|g\|^{\frac{\beta+\gamma}{\beta+\alpha}}_{\mathcal{B}^{-\alpha}_{1,1}}, \tag{8}$$

which can be shown just like Proposition 4.

Finally, by paying a logarithmic factor, we obtain that Proposition 4 translates to Hölder regularity.

**Theorem 5** *For $f, f^\star \in \mathcal{H}^\beta_K(\mathbb{R}^p, \mathbb{R})$ with compact support in $B^p(0, K)$, we have*

$$d_{\mathcal{H}^\gamma_1}(f, f^\star) \leq C \log(1 + d_{\mathcal{H}^\alpha_1}(f, f^\star)^{-1})^2 d_{\mathcal{H}^\alpha_1}(f, f^\star)^{\frac{\beta+\gamma}{\beta+\alpha}}.$$

The proof of Theorem 5 can be found in Section C.7.2. The goal of this paper is to generalize this inequality to the case where $f$ and $f^\star$ are densities with respect to the volume measures of two different closed submanifolds. The main difficulty of this generalization is to exploit the smoothness of the densities although, from the point of view of the ambient space $\mathbb{R}^p$, they do not have any. In particular, in the case where all the regularity exponents are non integers, Theorem 14 can be seen as a generalization of the interpolation inequality between Besov norms on functions (8) to Besov norms on the signed measures $\mu - \mu^\star$. Indeed, although the tempered distribution $\mu - \mu^\star$ does not belong to any Besov space with positive regularity exponent, Theorem 12 states that if $\mu$ and $\mu^\star$ have low dimensional $\beta$-regularity then the inequality

$$\|\mu - \mu^\star\|_{\mathcal{B}^{-\gamma}_{1,1}} \lesssim \|\mu - \mu^\star\|^{\frac{\beta+\gamma}{\beta+\alpha}}_{\mathcal{B}^{-\alpha}_{1,1}}$$

stands up to logarithmic factors.

### 3.2 Proof Outline of the Main Result (Theorem 14)

In this Section, the principal steps of the proof of Theorem 14 are presented. We start by stating some properties resulting of the manifold assumption and use them to show that the integration with respect to a smooth density on $\mathcal{M}$, can be parameterized through the integration on $\mathbb{R}^d$ (Proposition 6). Then, we show that if two measures supported on two manifolds respectively, are close enough for a distance $d_{\mathcal{H}^\alpha_1}$, there exists a smooth diffeomorphism between their support having some keys properties (Theorem 9). Finally, this diffeomorphism is used to split the IPMs into two distances: a distance between the support of the manifolds and a distance between the densities. This decomposition allows to conclude the proof of Theorem 14.

#### 3.2.1 Integration on Manifolds via Smooth Transport Maps

The first step is to handle measures supported on submanifolds. A challenge here is that the densities are smooth along the manifold but may look extremely irregular from the ambient space. Our approach is to 'unfold' the manifold into local coordinates using smooth charts. In these coordinates, integration against the measure behaves like integration in $\mathbb{R}^d$, allowing us to leverage the same wavelet-based arguments as in the classical case.

For a manifold $\mathcal{M}$ satisfying the $(\beta + 1, K)$-manifold condition, there exists an atlas $\Big(A_i, \varphi_i\Big)_{i \in \{1,\dots,m\}}$ of $\mathcal{M}$ such that:

- for all $i \in \{1, ..., m\}$, $A_i = \varphi_i^{-1}(B^d(0, \tau))$ with $\tau = \frac{1}{4K}$,

- for all $i \in \{1, ..., m\}$, $\varphi_i \in \mathcal{H}_K^{\beta+1}(A_i, B^d(0,\tau))$ and $\varphi_i^{-1} \in \mathcal{H}_K^{\beta+1}(B^d(0,\tau), A_i)$,
- for all $x \in \mathcal{M}$, there exists $i \in \{1, ..., m\}$ such that $B^p(x, \tau/2) \cap \mathcal{M} \subset A_i$.

The existence of this atlas results from basic properties of the $(\beta+1, K)$-manifold condition that are derived in Appendix Section A.1. Using this parametrization we can show that the integration with respect to a smooth density on $\mathcal{M}$ can be parameterized through the integration on $\mathbb{R}^d$.

**Proposition 6** *Let $\mathcal{M}$ satisfying the $(\beta+1, K)$-manifold condition and $\mu$ a probability measure satisfying the $(\beta, K)$-density condition on $\mathcal{M}$. Then there exists a collection of constants $C_i > 0$, maps $\phi_i \in \mathcal{H}_C^{\beta+1}(\mathbb{R}^d, \mathbb{R}^p)$ with $\inf_{u \in B^d(0,\tau)} \lambda_{\min}((\nabla\phi_i(u))^\top \nabla\phi_i(u)) \geq C^{-1}$ and weight functions $\zeta_i \in \mathcal{H}_C^{\beta+1}(\mathbb{R}^d, \mathbb{R})$ with $supp(\zeta_i) \subset B^d(0,\tau)$, such that for all bounded continuous function $h : \mathbb{R}^p \to \mathbb{R}$ we have*

$$\int_{\mathcal{M}} h(x) d\mu(x) = \sum_{i=1}^{m} \frac{1}{C_i} \int_{\mathbb{R}^d} h(\phi_i(u)) \zeta_i(u) d\lambda^d(u).$$

The proof of Proposition 6 can be found in Section A.2. This is a key result that will enable to exploit the $\beta$-regularity of the densities defined on manifolds, using the wavelet decomposition on $\mathbb{R}^d$ of the $\beta+1$-regular transport maps $\phi_i$.

### 3.2.2 Existence of a Smooth Diffeomorphism Between the Manifolds

When the measures are very close in a high-regularity IPM, not only their densities are close, but their supporting manifolds must also be geometrically near. This motivates constructing a smooth map that deforms one manifold into the other. Conceptually, this map lets us 'align' the supports so that the problem reduces to comparing densities on a single manifold, where classical interpolation tools apply.

We shall distinguish two cases:

- the case where the distance $d_{\mathcal{H}_1^\eta}(\mu, \mu^\star)$ is small enough so that there exists a smooth diffeomorphism between $\mathcal{M}$ and $\mathcal{M}^\star$,
- the case where the distance $d_{\mathcal{H}_1^\eta}(\mu, \mu^\star)$ is not small enough.

In the second case, one has $d_{\mathcal{H}_1^\eta}(\mu, \mu^\star) \geq C^{-1}$ so it is easy to get $d_{\mathcal{H}_1^\gamma}(\mu, \mu^\star) \leq C d_{\mathcal{H}_1^\eta}(\mu, \mu^\star)$. Therefore, let us focus on the first case and gather the properties that the diffeomorphism needs to respect.

**Definition 7** *For $\mathcal{M}, \mathcal{M}^\star$ satisfying the $(\beta+1, K)$-manifold condition and $t \in (0, r^\star/2)$ for $r^\star$ the reach of $\mathcal{M}^\star$, such that $\mathbb{H}(\mathcal{M}, \mathcal{M}^\star) < t$. We say that a map $T : \mathcal{M}^{\star t} \to \mathcal{M}^\star$ is $(\mathcal{M}, \mathcal{M}^\star)_{K_T}$-compatible with a radius $t$ if:*

*i) $T \in \mathcal{H}_{K_T}^{\beta+1}(\mathcal{M}, \mathcal{M}^\star)$ and the restriction $T_{|\mathcal{M}} : \mathcal{M} \to \mathcal{M}^\star$ is a diffeomorphism such that for $T_{|\mathcal{M}}^{-1} : \mathcal{M}^\star \to \mathcal{M}$ the inverse application such that $T_{|\mathcal{M}}^{-1} \circ T_{|\mathcal{M}} = Id_{|\mathcal{M}}$, we have $T_{|\mathcal{M}}^{-1} \in \mathcal{H}_{K_T}^{\beta+1}(\mathcal{M}^\star, \mathcal{M})$.*

*ii) For all* $x \in \mathcal{M}^\star$, $T^{-1}(\{x\}) - x \subset E_x$ *a* $p-d$ *dimensional subspace of* $\mathbb{R}^p$ *and*

$$(T^{-1}(\{x\}) - x) \cap B^p(x,t) = B_{E_x}(0,t).$$

*iii) For any* $f \in \mathcal{H}_1^\eta(\mathcal{M}^{\star t}, \mathbb{R})$ *with* $\eta \in [1, \beta+1]$, *the map* $x \mapsto \int_{T^{-1}(\{x\})} f(y) d\lambda_{E_x}^{p-d}(y)$ *belongs to* $\mathcal{H}_{K_T}^\eta(\mathcal{M}^\star, \mathbb{R})$.

*iv) The map* $x \mapsto ap_d(\nabla T(x))$ *(Definition 27) is bounded below by* $K_T^{-1}$.

We define $T : \mathcal{M}^{\star t} \to \mathcal{M}^\star$ as a smooth tubular retraction around $\mathcal{M}^\star$. Its restriction $T|_{\mathcal{M}} : \mathcal{M} \to \mathcal{M}^\star$ is a $C^{\beta+1}$ diffeomorphism with $(T|_{\mathcal{M}})^{-1} \in \mathcal{H}_{K_T}^{\beta+1}(\mathcal{M}^\star, \mathcal{M})$, which is exactly the regularity required in item (i) to smoothly push measure on $\mathcal{M}$ onto $\mathcal{M}^\star$. The larger domain $\mathcal{M}^{\star t}$ is used only for items (ii)–(iv): it provides normal fibers $T^{-1}(\{x\})$ around $\mathcal{M}^\star$, enabling averaging along the tube and the use of the coarea formula when transferring integrals to $\mathcal{M}^\star$.

Definition 7 plays a central role in our proof strategy: it allows us to construct the pushforward measure $T_\# \mu$ supported on $\mathcal{M}^\star$, which inherits smoothness and geometric properties from both $\mu$ and $T$. This measure serves as an intermediate object between $\mu$ and $\mu^\star$: it has the same support as $\mu^\star$ but a density related to that of $\mu$ through $T$. This intermediate step is essential in the decomposition (9), where the total discrepancy between $\mu$ and $\mu^\star$ is split into a *support term* (comparing $\mu$ and $T_\# \mu$) and a *density term* (comparing $T_\# \mu$ and $\mu^\star$). The $(\mathcal{M}, \mathcal{M}^\star)$-compatibility of $T$ ensures that both terms can be controlled in terms of high-regularity IPMs, ultimately leading to Theorem 14.

Fulfilling compatibility may seem to be restrictive, but it is shown in Theorem 9, that as long as the $d_{\mathcal{H}^\alpha}$ distance between the measures is relatively small for any $\alpha > 0$, there exists a $(\mathcal{M}, \mathcal{M}^\star)$-compatible map. Let us first state that the distance $d_{\mathcal{H}_1^\alpha}$ controls the Wasserstein and Hausdorff distances.

**Lemma 8** *Let* $\mathcal{M}, \mathcal{M}^\star$ *satisfying the* $(0,K)$*-manifold condition and* $\mu, \mu^\star$ *two probability measures satisfying the* $(0,K)$*-regularity density condition on* $\mathcal{M}$ *and* $\mathcal{M}^\star$ *respectively. Then for all* $\alpha \geq 1$, *there exists a constant* $C_\alpha > 0$ *such that* $\forall t > 0$, *if*

$$d_{\mathcal{H}_1^\alpha}(\mu, \mu^\star) \leq t^{(d+1)(2\alpha-1)},$$

*then*

$$W_1(\mu, \mu^\star) \leq C_\alpha t^{d+1} \quad \text{and} \quad \mathbb{H}(\mathcal{M}, \mathcal{M}^\star) < C_\alpha t.$$

The proof of Lemma 8 can be found in Section B.1. We are now in position to state the existence of a $(\mathcal{M}, \mathcal{M}^\star)$-compatible map.

**Theorem 9** *Let* $\mathcal{M}, \mathcal{M}^\star$ *satisfying the* $(\beta+1, K)$*-manifold condition and* $\mu, \mu^\star$ *probability measures on* $\mathcal{M}$ *and* $\mathcal{M}^\star$ *respectively satisfying the* $(0,K)$*-density condition. For all* $\alpha > 0$, *there exists a constant* $C_\alpha > 0$ *such that if*

$$d_{\mathcal{H}_1^\alpha}(\mu, \mu^\star) \leq C_\alpha^{-1},$$

*then there exists a* $(\mathcal{M}, \mathcal{M}^\star)_{K_T}$*-compatible map with radius* $t = C_\alpha^{-1}$ *and* $K_T = C_\alpha$.

The proof of Theorem 9 can be found in Section B.2. We see that imposing that the densities are close for any IPM, implies that the manifolds are close in the sense that there exists a smooth diffeomorphism between them. This diffeomorphism is a key tool that is going to allow to decompose the IPMs into two distinct distances.

### 3.2.3 Decomposition of the IPMs into a Manifold Distance and a Density Distance

Once the manifolds are aligned, the IPM can be split into two contributions: one from the geometric difference between the manifolds, and one from the difference in the densities themselves. This separation mirrors our intuition: part of the discrepancy comes from moving mass along the manifold (geometry), and part comes from adjusting the mass distribution (density). Each term is then handled with a different regularity argument.

For $h \in \mathcal{H}_1^\gamma(\mathbb{R}^p, \mathbb{R})$, $T$ being $(\mathcal{M}, \mathcal{M}^\star)$-compatible and $f_T$ the density of $T_{\#\mu}$ with respect to the volume measure on $\mathcal{M}^\star$, we write

$$\begin{aligned}&\int_{\mathcal{M}} h(x) f(x) d\lambda_{\mathcal{M}}(x) - \int_{\mathcal{M}^\star} h(x) f^\star(x) d\lambda_{\mathcal{M}^\star}(x) \\ &= \int_{\mathcal{M}} (h(x) - h(T(x))) f(x) d\lambda_{\mathcal{M}}(x) + \int_{\mathcal{M}^\star} h(x)(f_T(x) - f^\star(x)) d\lambda_{\mathcal{M}^\star}(x). \end{aligned} \tag{9}$$

The cost is split this way to exploit the regularity of $f$ and $f^\star$. Indeed, the second term of (9) discriminates measures on the same support $\mathcal{M}^\star$. We are going to show that this case is equivalent to the classical inequality (Theorem 5) in full dimension. For the first term of (9), we will use Proposition 6 to write the integration against the density $f$ on $\mathcal{M}$, as the integration on $\mathbb{R}^d$ via $\beta+1$-regular transport maps. This will enable to exploit the regularity of the density through the wavelet coefficients of the transport maps.

Let us take care of the first term of (9) in the case $h \in \mathcal{H}_1^1(\mathbb{R}^p, \mathbb{R})$.

**Lemma 10** *Let $\mathcal{M}, \mathcal{M}^\star$ satisfying the $(\beta+1, K)$-manifold condition and $\mu, \mu^\star$ two probability measures satisfying the $(\beta, K)$-density condition on $\mathcal{M}$ and $\mathcal{M}^\star$ respectively. For all $\alpha > 0$, there exists a constant $C_\alpha > 0$ such that if*

$$d_{\mathcal{H}_1^\alpha}(\mu, \mu^\star) \leq C_\alpha^{-1},$$

*then for $T$ the $(\mathcal{M}, \mathcal{M}^\star)$-compatible map given by Theorem 9, we have for all $\epsilon \in (0,1)$ and $\eta \in (1, \beta+1]$,*

$$d_{\mathcal{H}_1^1}(\mu, T_{\#\mu}) \leq C \log(\epsilon^{-1})^4 \sup_{\substack{h \in \mathcal{H}_1^\eta \\ h_{|\mathcal{M}^\star} = 0}} \left( \int_{\mathcal{M}} h(x) d\mu(x) \right)^{\frac{\beta+1}{\beta+\eta}} + \epsilon.$$

The proof of Lemma 10 can be found in Section C.2. This result highlights the fact that the first term of (9) focuses on the distance between the manifolds, as it bounds it with a supremum over functions being equal to 0 on $\mathcal{M}^\star$. Although Lemma 10 bounds the distance $d_{\mathcal{H}_1^1}(\mu, T_{\#\mu})$ only in the case $\gamma = 1$, it is easy to generalize it to the case $\gamma \geq 1$ (see the detailed proof of Theorem 12 in Section C.4 for explicit arguments) using the classical interpolation inequality (Corollary 26). Let us now bound the second term of (9).

**Lemma 11** *Let $\mathcal{M}, \mathcal{M}^\star$ satisfying the $(\beta+1, K)$-manifold condition and $\mu, \mu^\star$ two probability measures satisfying the $(\beta, K)$-density condition on $\mathcal{M}$ and $\mathcal{M}^\star$ respectively. For all $\alpha > 0$, there exists a constant $C_\alpha > 0$ such that if*

$$d_{\mathcal{H}_1^\alpha}(\mu, \mu^\star) \leq C_\alpha^{-1},$$

*then for $T$ the $(\mathcal{M}, \mathcal{M}^\star)$-compatible map given by Theorem 9, we have for all $\epsilon \in (0,1)$ and $0 < \gamma \leq \eta \leq \beta + 1$,*

$$d_{\mathcal{H}_1^\gamma}(T_{\#\mu}, \mu^\star) \leq C \log(\epsilon^{-1})^2 d_{\mathcal{H}_1^\eta}(T_{\#\mu}, \mu^\star)^{\frac{\beta+\gamma}{\beta+\eta}} + \epsilon.$$

The proof of Lemma 11 can be found in Section C.3. An immediate corollary of this result is that if the two measures $\mu, \mu^\star$ live on the same manifold $\mathcal{M}^\star$ (so that $T_{\#\mu} = \mu$), then

$$d_{\mathcal{H}_1^\gamma}(\mu, \mu^\star) \leq C \log(\epsilon^{-1})^2 d_{\mathcal{H}_1^\eta}(\mu, \mu^\star)^{\frac{\beta+\gamma}{\beta+\eta}} + \epsilon.$$

This is due to the reach condition on $\mathcal{M}^\star$, which allows to thicken the manifold and apply the classical interpolation inequality as if we were in full dimension (see the proof of Lemma 11 for details). Note that contrary to Lemma 10, Lemma 11 allows to have $\gamma \in (0,1)$. This supports the fact that the discrimination of densities with supports on the same manifolds is equivalent to the discrimination in full dimension. We will see in Proposition 13 that this is not true for densities having supports on different manifolds as the optimal potential behaves differently in this case.

Putting Lemmas 10 and 11 together, we finally obtain the interpolation inequality for densities having supports on different submanifolds.

**Theorem 12** *Let $\mathcal{M}, \mathcal{M}^\star$ satisfying the $(\beta+1, K)$-manifold condition and $\mu, \mu^\star$ two probability measures satisfying the $(\beta, K)$-density condition on $\mathcal{M}$ and $\mathcal{M}^\star$ respectively. Then for all $1 \leq \gamma \leq \eta$, we have*

$$d_{\mathcal{H}_1^\gamma}(\mu, \mu^\star) \leq C \log\left(1 + d_{\mathcal{H}_1^\eta}(\mu, \mu^\star)^{-1}\right)^{C_2} d_{\mathcal{H}_1^\eta}(\mu, \mu^\star)^{\frac{\beta+\gamma}{\beta+\eta}}.$$

The detailed proof of Theorem 12 can be found in Section C.4. This result is the extension of Theorem 5 to the submanifold case for $\gamma \geq 1$. Let us now state the inequality between low-regularity IPMs.

**Proposition 13** *Let $\mathcal{M}, \mathcal{M}^\star$ satisfying the $(\beta+1, K)$-manifold condition and $\mu, \mu^\star$ two probability measures satisfying the $(\beta, K)$-density condition on $\mathcal{M}$ and $\mathcal{M}^\star$ respectively. Then for all $0 < \gamma \leq \eta \leq 1$, we have*

$$d_{\mathcal{H}_1^\gamma}(\mu, \mu^\star) \leq C d_{\mathcal{H}_1^\eta}(\mu, \mu^\star)^{\frac{\gamma}{\eta}}.$$

The proof of Proposition 13 can be found in Section C.5. Distances of very low regularity ($\gamma < 1$) behave differently. In this regime, the optimal test functions resemble distance-to-manifold functions rather than smooth oscillatory potentials. This is why the power law changes, and the inequality becomes essentially geometric: the IPM is dominated by how far the supports are from each other. The inequality of Proposition 13 is sharp as the

equality is attained in numerous cases. For example, if $\mathcal{M}$ is the $p-1$ sphere of radius 1, $\mathcal{M}^\star$ is the $p-1$ sphere of radius $1+\epsilon$ and $\mu, \mu^\star$ are rescaled volume measure on $\mathcal{M}, \mathcal{M}^\star$ respectively. In this case we have that the $d_{\mathcal{H}_1^\gamma}$ and $d_{\mathcal{H}_1^\eta}$ distances behave like $\epsilon^\gamma$ and $\epsilon^\eta$ respectively. The intuition behind this result is that for $\gamma \in (0,1)$, the optimal potential $h$ within $\mathcal{H}_1^\gamma$ (i.e. $d_{\mathcal{H}_1^\gamma}(\mu,\mu^\star) = \int h d\mu - \int h d\mu^\star$) behaves like $d(\cdot, \mathcal{M}^\star)^\gamma$.

Putting Theorem 12 and Proposition 13 together, we obtain a general result for all $\gamma > 0$.

**Theorem 14** *Let $\mathcal{M}, \mathcal{M}^\star$ satisfying the $(\beta+1, K)$-manifold condition and $\mu, \mu^\star$ two probability measures satisfying the $(\beta, K)$-density condition on $\mathcal{M}$ and $\mathcal{M}^\star$ respectively. Then for all $0 < \gamma < \eta$, we have*

$$d_{\mathcal{H}_1^\gamma}(\mu,\mu^\star) \leq C \log(1 + d_{\mathcal{H}_1^\eta}(\mu,\mu^\star)^{-1})^{C_2} d_{\mathcal{H}_1^\eta}(\mu,\mu^\star)^\delta,$$

*with*

$$\delta = \begin{cases} \frac{\beta+\gamma}{\beta+\eta} & \text{if } \gamma \geq 1 \\ \frac{\beta\gamma+\gamma}{\beta+\eta} & \text{if } \gamma \leq 1 \leq \eta \\ \frac{\gamma}{\eta} & \text{if } \eta \leq 1. \end{cases} \tag{10}$$

When $0 < \gamma < 1 < \eta$, the exponent $\delta$ is obtained by applying first Proposition 13 on $d_{\mathcal{H}_1^\gamma}$ and $d_{\mathcal{H}_1^1}$, then applying Theorem 12 on $d_{\mathcal{H}_1^1}$ and $d_{\mathcal{H}_1^\eta}$. The exponent $\delta$ is still optimal as outlined in (41) using the example of Section 5.

In the assumptions of Theorem 14, we suppose that both $\mu$ and $\mu^\star$ verify the point ii) of the density condition which is $f_\mu,\ f_{\mu^\star} \geq K^{-1}$. However, it is only necessary that one of the measures verifies this condition as stated in the next result.

**Proposition 15** *Let $\mathcal{M}, \mathcal{M}^\star$ satisfying the $(2, K)$-manifold condition and $\mu, \mu^\star$ two probability measures with densities with respect to the volume measure on $\mathcal{M}, \mathcal{M}^\star$ respectively. Suppose that $\mu^\star$ verifies the $(0, K)$-density condition and that $f_\mu \in \mathcal{H}_K^1(\mathcal{M}, \mathbb{R})$. Then for all $\eta > 0$, there exist constants $C, C_\eta > 0$ such that if $\inf_{x \in \mathcal{M}} f_\mu(x) \leq C_\eta^{-1}$, then*

$$d_{\mathcal{H}_1^\eta}(\mu,\mu^\star) \geq C^{-1}.$$

The proof of Proposition 15 can be found in Section C.6. Using this result we have that if $\mu$ does not verify point ii) of the density condition for $K = C_\eta$, we have

$$d_{\mathcal{H}_1^\gamma}(\mu,\mu^\star) \leq C \leq C d_{\mathcal{H}_1^\eta}(\mu,\mu^\star).$$

Therefore, we could only suppose that one of the measures verifies point ii) of the density condition in Theorem 12. We did not write it this way for simplicity of read.

### 3.2.4 Comparison with the Torus Case

The interpolation inequality established in Stéphanovitch et al. (2023) is derived under the assumption that both measures are prescribed as push–forwards of Lebesgue measure on the torus, $g_\sharp \lambda_d$ and $g^\star_\sharp \lambda_d$. This *global* parametrisation produces two decisive simplifications:

(i) the $\beta$-regularity of the measures can be inferred directly from the wavelet expansions of the maps $g, g^\star$; and (ii) the "support term" in the proof is integrated against the uniform measure, so no Jacobian factors appear.

Neither convenience survives when the supports are arbitrary compact $C^{\beta+1}$ manifolds $\mathcal{M}, \mathcal{M}^\star$. To replace them we introduce three new ingredients.

1. **Transport–chart representation.** Proposition 6 constructs a finite $\mathcal{H}^{\beta+1}$ atlas $\{(\phi_i, \zeta_i)\}_{i=1}^m$ such that, for every bounded test function $h$,

$$\int_M h \, \mathrm{d}\mu \;=\; \sum_{i=1}^m C_i^{-1} \int_{\mathbb{R}^d} h\big(\phi_i(u)\big)\, \zeta_i(u) \, \mathrm{d}\lambda^d(u).$$

   This change of variables recovers a Euclidean setting in which wavelet techniques can exploit the full $\beta$-smoothness of the densities.

2. **Density-aware control of the support term.** Lemma 10 bounds the distance between supports while keeping track of both the Jacobian of the compatible map $T$ and the Hölder norms of the densities $\zeta_i$. Its proof is complemented by a Whitney–type extension theorem (Proposition 32), which extends any $\mathcal{H}^\eta$ function defined on a tubular neighbourhood of $\mathcal{M}^\star$ to the whole $\mathbb{R}^p$ without loss of regularity.

3. **Construction of a compatible diffeomorphism.** In the torus setting the map $T$ was produced by convolving the generator $g^\star$. For arbitrary $\mathcal{M}, \mathcal{M}^\star$ we instead use Fefferman et al. (2020) to obtain a $C^\infty$ manifold $\mathcal{M}_s$ at Hausdorff distance $s^2$ from $\mathcal{M}^\star$. The orthogonal projection $\pi_s$ then yields $T = \pi_s|_{\mathcal{M}}$, whose inverse and Jacobian enjoy quantitative $\mathcal{H}^{\beta+1}$ bounds (Theorem 9 and Lemma 29).

Taken together, these three components replace the single push–forward toolkit used in the torus proof and are indispensable for extending the interpolation inequality to *any* pair of closed $C^{\beta+1}$ manifolds.

## 4. A Direct Application: Optimal Estimator of Density on Unknown Submanifold

Suppose that we observe an i.i.d. sample $X_1, ...X_n \in \mathbb{R}^p$ from a probability measure $\mu^\star$ satisfying the $(\beta, K)$ density condition on an unknown $\mathcal{M}^\star$ satisfying the $(\beta+1, K)$ manifold condition. A thriving field of Machine Learning/Statistics focuses on building an estimator $\hat{\mu}(X_1, ..., X_n)$ of $\mu^\star$ that achieves optimal rates of convergence (see for example Tang and Yang (2023), Divol (2022), Schreuder et al. (2021), De Bortoli (2022)). In particular, fixing a $\gamma \in (0, \infty)$, Tang and Yang (2023) build a theoretical estimator that attains the optimal rate of estimation (up to a logarithmic factor) for the distance $d_{\mathcal{H}_1^\gamma}$. In this section we show that Theorem 12 allows to build a much simpler estimator that attains optimal rates for all the distances $d_{\mathcal{H}_1^\gamma}$, $\gamma \in [1, \infty)$ simultaneously.

### 4.1 The Estimator

Let us write $\mathcal{F}$ for the set of probability measures $\mu$ such that there exists a submanifold $\mathcal{M}$ satisfying the $(\beta+1, K)$-manifold condition and that $\mu$ satisfies the $(\beta, K)$-density condition

on $\mathcal{M}$. Define the estimator

$$\hat{\mu} \in \arg\min_{\mu \in \mathcal{F}} d_{\mathcal{H}_1^{d/2}}(\mu, \mu_n^\star), \tag{11}$$

with $\mu_n^\star := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ the empirical measure from the data (the existence of $\hat{\mu}$ is proven in Proposition 38 in Section C.9). We first bound the expected error of our estimator, namely

$$\mathbb{E}_{X_1,\dots,X_n \sim \mu^\star}[d_{\mathcal{H}_1^{d/2}}(\hat{\mu}, \mu^\star)].$$

To this end, define for all $\mu \in \mathcal{F}$

$$h_\mu \in \arg\max_{h \in \mathcal{H}_1^{d/2}} \int h(x) d\mu(x) - \int h(x) d\mu^\star(x),$$

This is an optimal potential between $\mu$ and $\mu^\star$, well defined by compactness of the class $\mathcal{H}_1^{d/2}(B^p(0,K))$. We also define

$$h_\mu^n \in \arg\max_{h \in \mathcal{H}_1^{d/2}} \int h(x) d\mu(x) - \frac{1}{n}\sum_{i=1}^n h(X_i),$$

an optimal potential between $\mu$ and $\mu_n^\star$. We have

$$\begin{aligned}
\mathbb{E}[d_{\mathcal{H}_1^{d/2}}(\hat{\mu}, \mu^\star)] =& \mathbb{E}[\int h_{\hat{\mu}}(x) d\hat{\mu}(x) - \int h_{\hat{\mu}}(x) d\mu^\star(x)] \\
=& \mathbb{E}[\int h_{\hat{\mu}}(x) d\hat{\mu}(x) - \frac{1}{n}\sum_{i=1}^n h_{\hat{\mu}}(X_i) + \frac{1}{n}\sum_{i=1}^n h_{\hat{\mu}}(X_i) - \int h_{\hat{\mu}}(x) d\mu^\star(x)] \\
\leq& \mathbb{E}[\int h_{\hat{\mu}}^n(x) d\hat{\mu}(x) - \frac{1}{n}\sum_{i=1}^n h_{\hat{\mu}}^n(X_i)] + \mathbb{E}[d_{\mathcal{H}_1^{d/2}}(\mu_n^\star, \mu^\star)].
\end{aligned}$$

Furthermore, by definition of $\hat{\mu}$ we have

$$\mathbb{E}[\int h_{\hat{\mu}}^n(x) d\hat{\mu}(x) - \frac{1}{n}\sum_{i=1}^n h_{\hat{\mu}}^n(X_i)] \leq \mathbb{E}[\int h_{\mu^\star}^n(x) d\mu^\star(x) - \frac{1}{n}\sum_{i=1}^n h_{\mu^\star}^n(X_i)] \leq \mathbb{E}[d_{\mathcal{H}_1^{d/2}}(\mu_n^\star, \mu^\star)].$$

Then, using Lemma 13 from Tang and Yang (2023) we have that

$$\mathbb{E}[d_{\mathcal{H}_1^{d/2}}(\mu_n^\star, \mu^\star)] \leq C \log(n) n^{-1/2},$$

which gives us using the previous derivations

$$\mathbb{E}[d_{\mathcal{H}_1^{d/2}}(\hat{\mu}, \mu^\star)] \leq C \log(n) n^{-1/2}.$$

Now, using Theorem 12 we have that for all $\gamma \in [1, d/2]$,

$$\mathbb{E}[d_{\mathcal{H}_1^\gamma}(\hat{\mu}, \mu^\star)] \leq C\mathbb{E}[\log(d_{\mathcal{H}_1^{d/2}}(\hat{\mu}, \mu^\star)^{-1})^{C_2} d_{\mathcal{H}_1^{d/2}}(\hat{\mu}, \mu^\star)^{\frac{\beta+\gamma}{\beta+d/2}}], \tag{12}$$

so using Jensen's inequality, we finally get for all $\gamma \in [1, \infty)$,

$$\mathbb{E}[d_{\mathcal{H}_1^\gamma}(\hat{\mu}, \mu^\star)] \leq C \log(n)^{C_2} (n^{-\frac{\beta+\gamma}{2\beta+d}} \vee n^{-1/2}) \tag{13}$$

and this rate has been proven to be optimal (up to the logarithmic factor) in Tang and Yang (2023).

In all, we have shown the following minimax optimality for $\hat{\mu}$, where we write $g(n) = \tilde{O}(f(n))$ if the exist $C, C_2 > 0$ such that $g(n) \leq C \log(n)^{C_2} f(n)$.

**Theorem 16** *Let $n \in \mathbb{N}_{>0}$, $\mu^\star \in \mathcal{F}$ and $(X_1, ..., X_n)$ an i.i.d. sample of law $\mu^\star$. Then the estimator $\hat{\mu}$ from (11) satisfies for all $\gamma \geq 1$*

$$\sup_{\mu^\star \in \mathcal{F}} \mathbb{E}_{X_i}[d_{\mathcal{H}_1^\gamma}(\hat{\mu}, \mu^\star)] = \tilde{O}\left( \inf_{\hat{\theta} \in \Theta} \sup_{\mu^\star \in \mathcal{F}} \mathbb{E}_{X_i}[d_{\mathcal{H}_1^\gamma}(\hat{\theta}, \mu^\star)] \right),$$

*where $\Theta$ denotes the set of all possible estimators of $\mu^\star$ based on $n$ sample.*

### 4.2 Discussion

If we were to replace the IPM $d_{\mathcal{H}_1^{d/2}}$ by the IPM $d_{\mathcal{H}_1^\gamma}$ with $1 \leq \gamma < d/2$ in the estimator (11), we would only attain the rate $O(n^{-\frac{\gamma}{d}} \vee n^{-\frac{1}{2}})$ which is not optimal for $\gamma \in [1, d/2)$. The intuition is that computing $\hat{\mu}$ with a high regularity IPM, implicitly regularizes the measure $\mu_n^\star$ which allows to not have to compute a complex regularization of $\mu_n^\star$ like in Liang (2021) or Tang and Yang (2023). This highlights the need of a sharp inequality like Theorem 12 allowing to compare the IPMs.

In contrast to Tang and Yang (2023), the estimator $\hat{\mu}$ (11) does not attain optimal rates for the distance $d_{\mathcal{H}_1^\gamma}$ with $\gamma \in (0, 1)$. The intuition is that optimal potentials within $\mathcal{H}_1^{d/2}$ and $\mathcal{H}_1^\gamma$ tend to be very different when $\gamma$ is small so the estimator $\hat{\mu}$ is not adapted to IPMs of low regularity. This raises the question of whether there could exist an estimator being optimal for every $\gamma \in (0, \infty)$ simultaneously.

Like in Tang and Yang (2023), the estimator (11) is theoretical meaning it is not computable in practice. Nevertheless, the minimum over the class $\mathcal{F}$ and the supremum over the class $\mathcal{H}_1^{d/2}$ could be approximated using stochastic gradient descent on neural network classes like in Stéphanovitch et al. (2023). The computational aspect of this estimator will be investigated in a future work.

## 5. Lower Bound and Example

### 5.1 Sharpness of Theorem 14

In this section, we demonstrate that the bound provided by Theorem 14 is sharp, by constructing a simple example and analyzing it in detail. We begin with the following theorem, which shows that the exponent appearing in Theorem 14 cannot be improved.

**Theorem 17** *For all $\epsilon > 0$ and $0 < \gamma \leq \eta$ there exists $\mathcal{M}, \mathcal{M}^\star$ satisfying the $(\beta + 1, K)$-manifold condition and $\mu, \mu^\star$ two probability measures satisfying the $(\beta, K)$-density condition*

*on $\mathcal{M}$ and $\mathcal{M}^\star$ respectively such that $d_{\mathcal{H}_1^\gamma}(\mu, \mu^\star) \leq \epsilon$ and*

$$d_{\mathcal{H}_1^\gamma}(\mu, \mu^\star) \geq C^{-1} d_{\mathcal{H}_1^\eta}(\mu, \mu^\star)^\delta,$$

*with $\delta$ defined in (4) and $C > 0$ a constant depending on $d, \beta, K, \gamma, \eta$ but not on $\epsilon$.*

The proof of Theorem 17 is presented in Section C.8, together with a detailed analysis of an example that saturates the inequality. Below, we provide a sketch that outlines the main strategy and conveys the underlying geometric and analytic intuition.

### 5.2 Proof Sketch and Intuition

#### 5.2.1 Objective

To prove Theorem 17 we construct a family of manifolds $\mathcal{M}_n, \mathcal{M}^\star$ with measures $\mu_n, \mu_n^\star$ respectively that saturate the exponent in Theorem 14. Concretely, the sequence satisfies

$$d_{\mathcal{H}_1^\gamma}(\mu_n, \mu_n^\star) \to 0 \quad \text{while} \quad d_{\mathcal{H}_1^\eta}(\mu_n, \mu_n^\star) \asymp d_{\mathcal{H}_1^\gamma}(\mu_n, \mu_n^\star)^\delta,$$

so that the power $\delta$ cannot be improved.

#### 5.2.2 Geometric Construction

- *Reference curve.* $\mathcal{M}^\star = \big\{(\cos 2\pi t, \sin 2\pi t) : t \in [0,1]\big\}$.
- *Perturbed curve.* For $n \geq 1$, take

  $$g_n(t) = \Big(1 + (2\pi n)^{-(\beta+1)} \sin 2\pi n t\Big)\, g^\star(t), \qquad g^\star(t) = (\cos 2\pi t, \sin 2\pi t).$$

  Let $\mu_n$ be the uniform measure on $\mathcal{M}_n = \{g_n(t) : t \in [0,1]\}$. Denote by $T$ the radial projection onto $\mathcal{M}^\star$ and $\mu_n^\star$ is chosen as $\mu_n^\star = T_\# \mu_n$.
- *Low-order distance.* Each point of $\mathcal{M}_n$ is displaced radially by $(2\pi n)^{-(\beta+1)} \sin 2\pi n t$; hence

  $$d_{\mathcal{H}_1^1}(\mu_n, T_\#\mu_n) \simeq W_1(\mu_n, T_\#\mu_n) \asymp n^{-(\beta+1)}.$$

#### 5.2.3 Building a $\mathcal{H}^\eta$ Potential Adapted to the Oscillations for $\eta \geq 1$

- *Parameterisation.* For $\Theta : \mathcal{M}^\star \to \mathbb{R}^2$ such that $\Theta(x)$ is the unitary normal vector pointing outward the manifold $\mathcal{M}^\star$ at the point $x$, set

  $$h(x) = H_1 \circ g^{\star -1} \circ T(x) \langle \Theta \circ T(x), x - T(x) \rangle$$

  a function of regularity $\eta$ with $H_1$ such that $h(x)$ is orthogonal to the manifold $\mathcal{M}^\star$ and points towards $\mathcal{M}_n$.
- *Shape.* Choose $H_1$ so that on each semi-period of length $\frac{1}{2n}$, it matches the sign of $\sin 2\pi n t$, vanishes at the nodes, and varies smoothly inside, e.g. let

  $$H_1(t) = (-1)^{\lfloor 2nt \rfloor}\, n^{-(\eta-1)}\, \lambda\Big(2n\big[t - \tfrac{\lfloor 2nt \rfloor}{2n}\big]\Big),$$

  where $\lambda \in \mathcal{H}^\eta([0,1])$ a positive function satisfying $\lambda(0) = \lambda(1) = 0$.

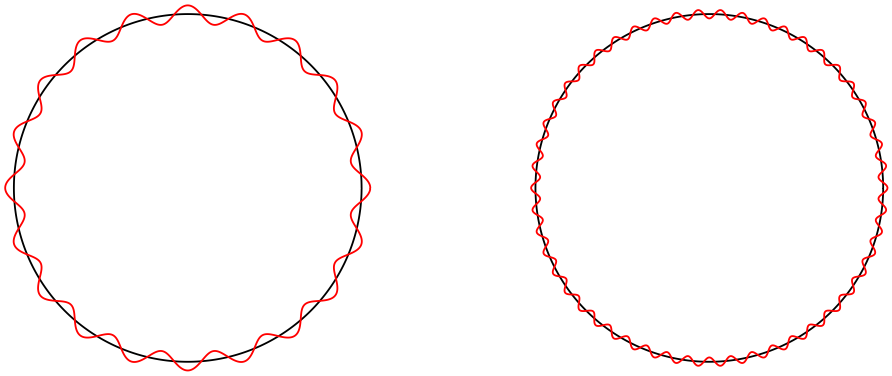


Figure 1: $\mathcal{M}_n$ and $\mathcal{M}^\star$ with $\beta = 0$, $n = 20$ (left) and $n = 50$ (right).

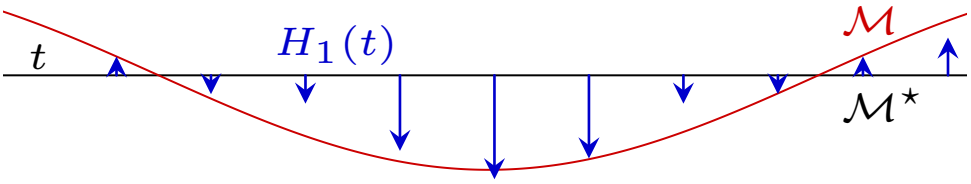



Figure 2: Zoom on a part of $\mathcal{M}_n, \mathcal{M}^\star$.

Then we obtain (in the case $\eta \geq 1$) that

$$\int_{\mathcal{M}_n} \big(h - h \circ T\big)\, d\mu_n \;\gtrsim\; n^{-(\beta+\eta)},$$

so proving that $h$ is optimal we get

$$d_{\mathcal{H}_1^\eta}(T_{\#}\mu_n, \mu_n) \asymp n^{-(\beta+\eta)}. \tag{14}$$

#### 5.2.4 Optimality of the Exponent

Using the estimate (14), we directly obtain for all $1 \leq \gamma < \eta$

$$d_{\mathcal{H}_1^\eta}(\mu_n, T_{\#}\mu_n) \;\asymp\; \big(d_{\mathcal{H}_1^\gamma}(\mu_n, T_{\#}\mu_n)\big)^{\frac{\beta+\eta}{\beta+\gamma}},$$

showing the exponent $\frac{\beta+\eta}{\beta+\gamma}$ of Theorem 14. The same sequence yields the optimal mixed-range exponent for all $0 < \gamma < \eta$.

## Appendix A. Proofs of the Properties of $\beta+1$-Smooth Submanifolds (Section 3.2.1)

In this Section, we give the detail of the proof of the integration on manifolds via smooth transport maps (Proposition 6) and list some properties of manifolds satisfying the $(2, K)$-manifold condition.

### A.1 Properties of $\beta+1$-Smooth Submanifolds

Recalling the notation of Definition 1, we obtain in the following result that the images of the charts contain the same ball around the origin.

**Proposition 18** *If $\mathcal{M}$ verifies the $(2,K)$-manifold condition, then for all $x \in \mathcal{M}$,*

$$B^d(0, \frac{1}{4K}) \in \varphi_x(\overline{B}^p(x, \sqrt{\frac{3}{4}}K^{-1}) \cap M).$$

**Proof** Let us first notice that for all $z \in \overline{B}^p(x, \sqrt{\frac{3}{4}}K^{-1}) \cap \mathcal{M}$ we have

$$\begin{aligned}
\|z - \pi_{\mathcal{T}_x(\mathcal{M})}(z)\| =& \|\varphi_x^{-1}\big(\pi_{\mathcal{T}_x(\mathcal{M})}(z) - x\big) - \pi_{\mathcal{T}_x(\mathcal{M})}(z)\| \\
=& \|\varphi_x^{-1}(0) + \nabla\varphi_x^{-1}(0)(\pi_{\mathcal{T}_x(\mathcal{M})}(z) - x) \\
&+ \int_0^1 \nabla^2\varphi_x^{-1}\big(t(\pi_{\mathcal{T}_x(\mathcal{M})}(z) - x)\big)(\pi_{\mathcal{T}_x(\mathcal{M})}(z) - x)dt - \pi_{\mathcal{T}_x(\mathcal{M})}(z)\| \\
=& \|\int_0^1 \nabla^2\varphi_x^{-1}\big(t(\pi_{\mathcal{T}_x(\mathcal{M})}(z) - x)\big)(\pi_{\mathcal{T}_x(\mathcal{M})}(z) - x)dt\| \\
\leq& K\|\pi_{\mathcal{T}_x(\mathcal{M})}(z) - x\|^2 \\
\leq& K\|z - x\|^2.
\end{aligned}$$

Then we obtain

$$\begin{aligned}
\|\varphi_x(z) - \varphi_x(x)\|^2 &= \|x - z\|^2 - \|z - \pi_{\mathcal{T}_x(\mathcal{M})}(z)\|^2 \\
&\geq \|x - z\|^2 - K^2\|x - z\|^4 \geq \frac{1}{4}\|x - z\|^2,
\end{aligned}$$

so if $\|z - x\| \geq \frac{1}{2K}$ we have

$$\|\varphi_x(z) - \varphi_x(x)\| \geq \frac{1}{4K}. \tag{15}$$

Suppose now the result is not true and take $y \in B^d(0, \frac{1}{4K}) \cap \varphi_x(\overline{B}^p(x, \sqrt{\frac{3}{4}}K^{-1}))^c$ with $A^c$ denoting the complementary of the set $A$. Define $\tilde{y} = t_y y$ for

$$t_y = \arg\max\{t \in [0,1) \mid ty \in \varphi_x(\overline{B}^p(x, \sqrt{\frac{3}{4}}K^{-1}))\}.$$

As $\|\tilde{y} - \varphi_x(x)\| = \|\tilde{y}\| \leq \|y\| < \frac{1}{4K}$, we have by (15) that $\varphi_x^{-1}(\tilde{y}) \in B^p(x, \frac{1}{2K})$. But as $\varphi_x$ is a local diffeomorphism, there exists $\epsilon > 0$ such that $B^d(\tilde{y}, \epsilon) \in \varphi_x(\overline{B}^p(x, \sqrt{\frac{3}{4}}K^{-1}))$ which is in contradiction with the definition of $\tilde{y}$. ■

This result enables to build a covering of $\mathcal{M}$ with images of the ball $B^d(0, \frac{1}{4K})$ by smooth diffeomorphisms. Write $\tau = \frac{1}{4K}$ and let $(x_i)_{i=1,\dots,m}$ be a finite sequence such that $\mathcal{M} \subset \bigcup_{i=1}^m B^p(x_i, \frac{\tau}{2})$.

**Corollary 19** *If $\mathcal{M}$ verifies the $(2,K)$-manifold condition then $\mathcal{M} \subset \bigcup_{i=1}^{m} \varphi_{x_i}^{-1}(B^d(0,\tau))$ and for all $x \in \mathcal{M}$, there exists $i \in \{1,...,m\}$ such that $B^p(x,\frac{\tau}{2}) \cap \mathcal{M} \subset \varphi_{x_i}^{-1}(B^d(0,\tau))$.*

**Proof** Let $x \in B^p(x_i,\tau/2) \cap \mathcal{M}$, as the projection $\pi_{\mathcal{T}_{x_i}(\mathcal{M})}$ is 1-Lipschitz, we have $\varphi_{x_i}(x) \in B^d(0,\tau/2)$. Furthermore

$$B^p(x,\frac{\tau}{2}) \cap \mathcal{M} \subset B^p(x_i,\tau) \cap \mathcal{M} \subset \varphi_{x_i}^{-1}(B^d(0,\tau)).$$

■

For ease of notation, we will write $\varphi_i$ instead of $\varphi_{x_i}$. Let us now bound below the norm of the differential of the $\varphi_i$'s.

**Proposition 20** *If $\mathcal{M}$ verifies the $(2,K)$-manifold condition of Definition 1, then the charts $\varphi_i^{-1} : B^d(0,\tau) \to \mathcal{M}$ verify that for all $u \in B^d(0,\tau)$ and $v \in \mathbb{R}^d$,*

$$\|\nabla\varphi_i^{-1}(u)\frac{v}{\|v\|}\| \geq 1/2.$$

**Proof** We have that $\nabla\varphi_i(x_i) = \mathrm{Id}$ so for all $u \in B^d(0,\tau)$ and $v \in \mathbb{R}^d$ we have

$$\|\nabla\varphi_i^{-1}(u)v\| = \|\nabla\varphi_i^{-1}(0)v + \int_0^1 \nabla^2\varphi_i^{-1}(tu)uvdt\| \geq \|v\| - K\|u\|\|v\| \geq (1-\frac{1}{4})\|v\|$$

recalling that $\tau = \frac{1}{4K}$. ■

In the following we will use the characterization of the reach for submanifolds.

**Lemma 21** *(Federer (1959), Theorem 4.18) The reach $r \in [0,\infty)$ of a submanifold $\mathcal{M}$ verifies*

$$r = \inf_{q \neq p \in \mathcal{M}} \frac{\|q-p\|^2}{2d(q-p,\mathcal{T}_p(\mathcal{M}))}.$$

Using this result, let us prove a lower bound on the reach resulting of the $(2,K)$-manifold condition.

**Proposition 22** *Let $\mathcal{M}$ satisfying the $(2,K)$-manifold condition, then $r$ the reach of $\mathcal{M}$ verifies $r \geq \frac{1}{2K}$.*

**Proof** Let $x, y \in \mathcal{M}$ such that $\|x-y\| \leq K^{-1}$, for $\varphi_y$ the orthogonal projection onto the tangent space $\mathcal{T}_y(\mathcal{M})$, we have

$$\begin{aligned}
&d(x-y,\mathcal{T}_y(\mathcal{M}))\\
&\quad= d(\varphi_y^{-1}(\varphi_y(x)) - \varphi_y^{-1}(0),\mathcal{T}_y(\mathcal{M}))\\
&\quad\leq d(\nabla\varphi_y^{-1}(0)\varphi_y(x),\mathcal{T}_y(\mathcal{M})) + d(\int_0^1 \nabla^2\varphi_y^{-1}(t\varphi_y(x))\varphi_y(x)^2(1-t)dt,\mathcal{T}_y(\mathcal{M}))\\
&\quad= d(\int_0^1 \nabla^2\varphi_y^{-1}(t\varphi_y(x))\varphi_y(x)^2(1-t)dt,\mathcal{T}_y(\mathcal{M}))\\
&\quad\leq \frac{K}{2}\|\varphi_y(x)\|^2 \leq \frac{K}{2}\|x-y\|^2.
\end{aligned}$$

Then $\|x-y\| \leq K^{-1}$ imply that

$$\frac{\|x-y\|^2}{2d(x-y, \mathcal{T}_y(\mathcal{M}))} \geq \frac{\|x-y\|^2}{2\frac{K}{2}\|x-y\|^2} = K^{-1}.$$

Now if $\|x-y\| \geq K^{-1}$ then

$$\frac{\|x-y\|^2}{2d(x-y, \mathcal{T}_y(\mathcal{M}))} \geq \frac{\|x-y\|^2}{2\|x-y\|} \geq \frac{1}{2K}.$$

■

### A.2 Proof of Proposition 6

Let us first define the notion of approximation of unity subordinated to an atlas.

**Definition 23** *A collection of $C^\infty$ functions $(\rho_i)_{i=1,...,m}$ on $\mathcal{M}$ is said to be an approximation of unity subordinated to $(\varphi_i^{-1}(B^d(0,\tau)))_i$ if for all $i \in \{1,...,m\}$,*

- $0 \leq \rho_i \leq 1$ *and* $\sum_{i=1}^{m} \rho_i(x) = 1$ *for* $x \in \mathcal{M}$,
- $supp(\rho_i) \subset \varphi_i^{-1}(B^d(0,\tau))$.

It is well known that there exists an approximation of unity subordinated to any atlas on a closed manifold (Jost, 2008). Using this approximation of unity we can now give the proof of Proposition 6.

**Proof** As $\mu$ has a density $f_\mu$ with respect to the volume measure on $\mathcal{M}$, then by change of variable we have that the measure $\varphi_{i\#\mu}$ admits a density $f_\mu^i$ with respect to the Lebesgue measure on $B^d(0,\tau)$ given by

$$f_\mu^i(u) = f_\mu(\varphi_i^{-1}(u))\det((\nabla\varphi_i^{-1}(u))^\top \nabla\varphi_i^{-1}(u))^{1/2}.$$

Furthermore, as $\mu$ verifies the $(\beta, K)$-regularity density condition, we have by the Faa di Bruno formula that $f_\mu^i \in \mathcal{H}_C^\beta(B^d(0,\tau),\mathbb{R})$ and $f_\mu^i$ is bounded below by $\frac{1}{2K}$. Write $\lambda_i^d$ the Lebesgue measure on $B^d(0,\tau)$ normalized by $C_i^{-1} = \int f_\mu^i d\lambda^d / \int_{B^d(0,\tau)} d\lambda^d$. Then by Caffarelli's regularity (Theorem 12.50, Villani (2009)) we have that there exists $T_i \in \mathcal{H}_C^{\beta+1}(B^d(0,\tau), B^d(0,\tau))$ such that $T_{i\#\lambda_i^d} = \varphi_{i\#\mu}$. Furthermore, $T_i$ is solution to the equation

$$\det(\nabla T_i)^{-1} = f_\mu^i \circ T_i$$

so in particular $\det(\nabla T_i) \geq C^{-1}$. As $\lambda_{\max}(\nabla T_i) \leq C$ we deduce that $\lambda_{\min}(\nabla T_i) \geq C^{-1}$. Therefore we have that $\lambda_{\min}(\nabla(\varphi_i^{-1} \circ T_i)) \geq C^{-1}$. Using Theorem 31, we can extend $\varphi_i^{-1} \circ T_i$ to a map $\phi_i \in \mathcal{H}_C^{\beta+1}(\mathbb{R}^d, \mathbb{R}^p)$.

Let $h : \mathbb{R}^p \to \mathbb{R}$ a bounded continuous function, then for $\rho_i$ an approximation of unity subordinated to $(\varphi_i^{-1}(B^d(0,\tau)))_i$, we have

$$\begin{aligned}\mathbb{E}_\mu[h(X)] &= \sum_{i=1}^m \mathbb{E}_\mu[h(X)\rho_i(X)] = \sum_{i=1}^m \int_{B^d(0,\tau)} h(\varphi_i^{-1}(u))\rho_i(\varphi_i^{-1}(u))d\varphi_{i\#\mu}(u)\\ &= \sum_{i=1}^m \frac{1}{C_i}\int_{B^d(0,\tau)} h(\varphi_i^{-1}\circ T_i(u))\rho_i(\varphi_i^{-1}\circ T_i(u))d\lambda_i^d(u)\\ &= \sum_{i=1}^m \frac{1}{C_i}\int_{\mathbb{R}^d} h(\phi_i(u))\rho_i(\phi_i(u))\mathbb{1}_{\{u\in B^d(0,\tau)\}}d\lambda_i^d(u).\end{aligned}$$

Let $\zeta_i : \mathbb{R}^d \to \mathbb{R}$ defined by

$$\zeta_i(u) = \rho_i(\phi_i(u))\mathbb{1}_{\{u\in B^d(0,\tau)\}}.$$

Then as $\rho_i$ is subordinated to $(\varphi_i^{-1}(B^d(0,\tau)))_i$, we have that $\zeta_i \in \mathcal{H}_C^{\beta+1}(\mathbb{R}^d,\mathbb{R})$. Finally,

$$\mathbb{E}_\mu[h(X)] = \sum_{i=1}^m \frac{1}{C_i}\int_{\mathbb{R}^d} h(\phi_i(u))\zeta_i(u)d\lambda_i^d(u),$$

which concludes the proof. ■

## Appendix B. Proof of the Existence of a Diffeomorphism Between the Manifolds (Section 3.2.2)

Let us first define an operator that allows to change the regularity of a function by modifying its wavelet coefficients.

**Definition 24** *For a tempered distribution $f \in \mathcal{B}^{s,b}_{\infty,\infty}(\mathbb{R}^p,\mathbb{R})$, define the tempered distribution $\Gamma^{\gamma,c}(f) \in \mathcal{B}^{s+\gamma,b+c}_{\infty,\infty}(\mathbb{R}^p,\mathbb{R})$ by its wavelets coefficients*

$$\langle \Gamma^{\gamma,c}(f), \psi_{jlz}\rangle = 2^{j\gamma}(1+j)^c\langle f, \psi_{jlz}\rangle.$$

The tempered distribution $\Gamma^{\gamma,c}(f)$ is a "regularization" of $f$ if $\gamma < 0$ or $\gamma = 0$ and $c < 0$. For ease of notation, we will write $\Gamma^\gamma$ instead of $\Gamma^{\gamma,0}$. Let us recall two results from Stéphanovitch et al. (2023) :

**Proposition 25** *Let $h_1 \in \mathcal{B}^{s_1,b_1}_{\infty,\infty}(\mathbb{R}^m,\mathbb{R}^p,1)$ , $h_2 \in \mathcal{B}^{s_2,b_2}_{\infty,\infty}(\mathbb{R}^m,\mathbb{R}^p,1)$ for $s_1,s_2,b_1,b_2 \in \mathbb{R}$. Then for $\tau \in \mathbb{R}$, $t,r \in [0,1]$, $q > 1$ and $1/q + 1/q^\star = 1$ we have*

$$\Big\langle h_1, h_2\Big\rangle_{L^2} \le \Big\langle \tilde{\Gamma}^{t\tau}(h_1), \Gamma^{(1-t)\tau}(h_2)\Big\rangle_{L^2}^{\frac{1}{q}} \Big\langle \tilde{\Gamma}^{-r\frac{q^\star}{q}\tau}(h_1), \Gamma^{-(1-r)\frac{q^\star}{q}\tau}(h_2)\Big\rangle_{L^2}^{\frac{1}{q^\star}}$$

*for $\tilde{\Gamma}^{t\tau}(h_1) \in \mathcal{B}^{s_1+t\tau,b_1}_{\infty,\infty}$ and $\tilde{\Gamma}^{-r\frac{q^\star}{q}\tau}(h_1) \in \mathcal{B}^{s_1-r\frac{q^\star}{q},b_1}_{\infty,\infty}$ such that*

$$\langle \tilde{\Gamma}^{t\tau}(h_1)_i, \psi_{jlz}\rangle = S(j,l,w)_i\langle \Gamma^{t\tau}(h_1)_i, \psi_{jlz}\rangle$$

*and*

$$\langle \tilde{\Gamma}^{-t\frac{q^\star}{q}}\tau(h_1)_i, \psi_{jlz}\rangle = S(j,l,w)_i \langle \Gamma^{-r\frac{q^\star}{q}}\tau(h_1)_i, \psi_{jlz}\rangle$$

*with* $S(j,l,w)_i \in \{-1,1\}$ *the sign of* $\langle h_{1_i}, \psi_{jlz}\rangle\langle h_{2_i}, \psi_{jlz}\rangle$, $i = 1, ..., p$.

This proposition is a key result that we will use numerous times along the proof of Theorem 12. It is the classical interpolation inequality between Besov spaces of Section 3.1 in a more general form. An important Corollary is its applications to Hölder IPMs.

**Corollary 26** *Let* $\mu, \mu^\star$ *two probability measures with compact support in* $\mathbb{R}^p$. *Then for any* $\theta, \theta_1, \theta_2 > 0$ *such that* $\theta_1 < \theta < \theta_2$, *we have for all* $\epsilon > 0$

$$d_{\mathcal{H}_1^\theta}(\mu, \mu^\star) \leq C \log(\epsilon^{-1})^{2\delta_{\theta_1,\theta_2}} d_{\mathcal{H}_1^{\theta_1}}(\mu, \mu^\star)^{\frac{\theta_2-\theta}{\theta_2-\theta_1}} d_{\mathcal{H}_1^{\theta_2}}(\mu, \mu^\star)^{\frac{\theta-\theta_1}{\theta_2-\theta_1}} + \epsilon\delta_{\theta_1,\theta_2},$$

*with* $\delta_{\theta_1,\theta_2} = 1$ *if* $\theta_1$ *or* $\theta_2$ *is an integer and* $\delta_{\theta_1,\theta_2} = 0$ *otherwise.*

Let us define the notion of approximate Jacobian (Definition 2.10 in Federer (1959)).

**Definition 27** *Let* $\mathcal{M}$ *be a* $d$*-dimensional submanifold of* $\mathbb{R}^p$ *and a map* $T : \mathbb{R}^p \to \mathcal{M}$ *of regularity* $C^1$. *For* $x \in \mathbb{R}^p$, *using the matrix of* $\nabla T(x)$ *with respect to orthonormal basis of* $\mathbb{R}^p$ *and* $\mathcal{T}_{T(x)}(\mathcal{M})$, *the approximate Jacobian of* $T$, *noted* $ap_d(\nabla T(x))$, *is equal to the square root of the sum of the squares of the determinants of the* $d$ *by* $d$ *minors of this matrix.*

We will use multiple times the next proposition.

**Proposition 28** *(Theorem 3.2.22, Federer (2014)) For* $t > 0$, $T : \mathcal{M}^{\star t} \to \mathcal{M}^\star$ *of regularity* $C^1$ *and* $D : \mathcal{M}^{\star t} \to \mathbb{R}$, *we have*

$$\int_{\mathcal{M}^{\star t}} D(x) ap_d(\nabla T(x)) d\lambda^p(x) = \int_{\mathcal{M}^\star} \int_{T^{-1}(\{z\})} D(x) d\lambda^{p-d}(x) d\lambda_{\mathcal{M}^\star}(z),$$

for $\lambda_{\mathcal{M}^\star}$ the volume measure on the submanifold $\mathcal{M}^\star$.

### B.1 Proof of Lemma 8

**Proof** If $\alpha$ is not an integer, using Corollary 26 with $\theta = 1$, $\theta_1 = 1/2$ and $\theta_2 = \alpha$, we have

$$\begin{aligned} W_1(\mu, \mu^\star) &\leq 2K d_{\mathcal{H}_1^1}(\mu, \mu^\star) \\ &\leq C d_{\mathcal{H}_1^{1/2}}(\mu, \mu^\star)^{\frac{\alpha-1}{\alpha-1/2}} d_{\mathcal{H}_1^\alpha}(\mu, \mu^\star)^{\frac{1}{2\alpha-1}} \\ &\leq C d_{\mathcal{H}_1^\alpha}(\mu, \mu^\star)^{\frac{1}{2\alpha-1}} \end{aligned}$$

If now $\alpha$ is an integer, then using Corollary 26 with $\theta = 1$, $\theta_1 = \frac{\alpha-3/2}{2(\alpha-1)}$ and $\theta_2 = \alpha + 1/2$, we get the same result. Then by hypothesis we have

$$W_1(\mu, \mu^\star) \leq Ct^{d+1}.$$

Let $\delta > 0$ and suppose that $\mathbb{H}(\mathcal{M}, \mathcal{M}^\star) \geq \delta$. For $x \in \mathcal{M}$ such that $d(x, \mathcal{M}^\star) \geq \delta$, write $\nu = \mu(\cdot \cap B^p(x, \delta/2))$. Then has $\mu$ has a density with respect to the volume measure of $\mathcal{M}$ bounded below by $K^{-1}$, we have that

$$\nu(B^p(x, \delta/2) \cap \mathcal{M}) \geq K^{-1} \lambda_{\mathcal{M}}(\mathcal{M} \cap B^p(x, \delta/2) \geq C\delta^d.$$

Using the dual formulation of the 1-Wasserstein distance, we have that

$$\begin{aligned} W_1(\mu, \mu^\star) &= \sup_{D \in \mathrm{Lip}(1)} \mathbb{E}_{\substack{X \sim \mu^\star \\ Y \sim \mu}}[D(X) - D(Y)] \\ &\geq \int_{B^p(x, \delta/2)} d(y, \mathcal{M}^\star) d\mu(y) \geq \frac{\delta}{2} \mu(B^p(x, \delta/2) \cap \mathcal{M}) \\ &\geq C\delta^{d+1}. \end{aligned}$$

Therefore as $W_1(\mu, \mu^\star) \leq Ct^{d+1}$, we can deduce that

$$\mathbb{H}(\mathcal{M}, \mathcal{M}^\star) \leq Ct.$$

■

### B.2 Proof of Theorem 9

Let us first use Proposition 3.1 in Fefferman et al. (2020) to show that there exists a $C^\infty$ manifold close to $\mathcal{M}^\star$.

As every charts $\phi_i$ (given by Proposition 6) of $\mathcal{M}^\star$ belongs to $\mathcal{H}^2_K(B^d(0, \tau), \mathcal{M})$, we have in particular that for all $x \in \mathcal{M}^\star$ and $0 < s < \tau$,

$$\mathbb{H}(B^p(x, s) \cap \mathcal{M}^\star, T_x(\mathcal{M})) \leq Ks^2.$$

Then applying Proposition 3.1 in Fefferman et al. (2020) we have that there exists a $C^\infty$ closed manifold $\mathcal{M}_s$ such that $\mathbb{H}(\mathcal{M}^\star, \mathcal{M}_s) \leq 5s^2$ with charts $\varphi^s_x \in \mathcal{H}^\eta_{C_\eta}(B^d(0, s), \mathcal{M}_s)$ $\forall \eta > 0$, and its reach $r_s$ verifies $r_s \geq C^{-1}s$. We are going to use the orthogonal projection $\pi_s$ onto the submanifold $\mathcal{M}_s$ to build our map.

From Leobacher and Steinicke (2021) we know that the canonical projection $\pi$ onto a submanifold $\mathcal{M}$ with strictly positive reach $r$, verifies

$$\nabla \pi(x) = \left(\mathrm{Id}_{\mathcal{T}_{\pi(x)}(M)} - \|x - \pi(x)\| L_{\pi(x), v}\right)^{-1} P_{\mathcal{T}_{\pi(x)}(M)}. \tag{16}$$

for $L_{\pi(x),v}$ the shape operator in the direction $v = \frac{x - \pi(x)}{\|x - \pi(x)\|}$. From Corollary 3 in Leobacher and Steinicke (2021), we also have that

$$\| \left(\mathrm{Id}_{\mathcal{T}_{\pi(x)}(\mathcal{M})} - \|x - \pi(x)\| L_{\pi(x), v}\right)^{-1} \| \leq (1 - \|x - \pi(x)\|/r)^{-1}. \tag{17}$$

Furthermore if $\mathcal{M}$ is of regularity $\eta$, there exists a constant $C_\eta > 0$ depending on the $\mathcal{H}^\eta$ norm of the charts such that

$$\|\pi_{|_{\mathcal{M}^{r/2}}}\|_{\mathcal{H}^{\eta - 1}} \leq C_\eta. \tag{18}$$

Using these results, let us show that the restriction of $\pi_s$ to the manifold $\mathcal{M}^\star$ is a diffeomorphism.

**Lemma 29** *Let $\mathcal{M}_s$ and $\mathcal{M}^\star$ satisfy the $(\beta+1, K)$-manifold condition, with $\mathbb{H}(\mathcal{M}_s, \mathcal{M}^\star) \leq 5s^2$. For sufficiently small $s > 0$, the map $\pi_{s|_{\mathcal{M}^\star}} : \mathcal{M}^\star \to \mathcal{M}_s$, i.e. the restriction of the projection $\pi_s$ to the manifold $\mathcal{M}^\star$, is a diffeomorphism.*

**Proof** Let us first show that $\pi_{s|_{\mathcal{M}^\star}}$ is a local diffeomorphism, by showing that if there existed $x \in \mathcal{M}$ such that $\lambda_{\min}(\nabla(\pi_s \circ \varphi_x^{-1})(0)^\top \nabla(\pi_s \circ \varphi_x^{-1})(0))$ was too small, then the Hausdorff distance between $\mathcal{M}_s$ and $\mathcal{M}^\star$ would be too large.
Let $x \in \mathcal{M}^\star$ and $h \in (0, \frac{\tau}{2K})$. For $v \in \mathbb{R}^d$ with $\|v\| = 1$ such that $\langle \pi_s(\varphi_x^{-1}(0)) - \varphi_x^{-1}(0), \nabla(\pi_s \circ \varphi_x^{-1} - \varphi_x^{-1})(0)v\rangle \geq 0$, using a Taylor expansion we have

$$\begin{aligned}
&\|\pi_s(\varphi_x^{-1}(hv)) - \varphi_x^{-1}(hv)\| \\
&\geq \|\pi_s(\varphi_x^{-1}(0)) - \varphi_x^{-1}(0) + \nabla(\pi_s \circ \varphi_x^{-1} - \varphi_x^{-1})(0)hv\| - Ch^2 \\
&= \Big(\|\pi_s(\varphi_x^{-1}(0)) - \varphi_x^{-1}(0)\|^2 + 2\langle \pi_s(\varphi_x^{-1}(0)) - \varphi_x^{-1}(0), \nabla(\pi_s \circ \varphi_x^{-1} - \varphi_x^{-1})(0)hv\rangle \\
&\quad + \|\nabla(\pi_s \circ \varphi_x^{-1} - \varphi_x^{-1})(0)hv\|^2\Big)^{1/2} - Ch^2 \\
&\geq h\|\nabla(\pi_s \circ \varphi_x^{-1} - \varphi_x^{-1})(0)v\| - Ch^2 \\
&\geq h(\|\nabla\varphi_x^{-1}(0)v\| - \|\nabla(\pi_s \circ \varphi_x^{-1})(0)v\|) - Ch^2 \\
&\geq h(1/2 - \|\nabla(\pi_s \circ \varphi_x^{-1})(0)v\|) - Ch^2.
\end{aligned}$$

As $\mathbb{H}(\mathcal{M}_s, \mathcal{M}^\star) \leq 5s^2$ we deduce with $h = \frac{\tau(1/2 - \|\nabla(\pi_s \circ \varphi_x^{-1})(0)v\|)}{2C}$ that

$$\frac{\tau^2(1/2 - \|\nabla(\pi_s \circ \varphi_x^{-1})(0)v\|)^2}{4C} \leq 5s^2$$

so

$$1/2 - \sqrt{20C}\frac{s}{\tau} \leq \|\nabla(\pi_s \circ \varphi_x^{-1})(0)v\|$$

which gives for $s$ small enough

$$\lambda_{\min}(\nabla(\pi_s \circ \varphi_x^{-1})(0)^\top \nabla(\pi_s \circ \varphi_x^{-1})(0))^{1/2} \geq \frac{1}{4}. \tag{19}$$

In particular $\pi_s \circ \varphi_x^{-1}$ is an immersion and therefore $\pi_{s|_{\mathcal{M}^\star}} : \mathcal{M}^\star \to \mathcal{M}_s$ is a local diffeomorphism.

Suppose there exists $x, y \in \mathcal{M}^\star$ different such that $\pi_s(x) = \pi_s(y)$. Then

$$\|x - y\| \leq \|x - \pi_s(x)\| + \|y - \pi_s(y)\| \leq 2\mathbb{H}(\mathcal{M}^\star, \mathcal{M}_s) \leq 10s^2.$$

Then taking $10s^2 \leq \frac{1}{K}$, we have that $y \in \varphi_x^{-1}(B(0, 10s^2))$. But from (19) and the fact that $\|\nabla^2(\pi_s \circ \varphi_x^{-1})\| \leq C$ we have that for $s$ small enough, $\pi_s \circ \varphi_x^{-1}$ is a diffeomorphism on $B^d(0, 10s^2)$ and therefore $\pi_s(x) \neq \pi_s(y)$.

As $\mathcal{M}^\star$ is compact without boundary, $\pi_s(\mathcal{M}^\star)$ is a closed submanifold of $\mathcal{M}_s$ so being of same dimension, they are equal. We can conclude that $\pi_{s|_{\mathcal{M}^\star}}$ is a diffeomorphism. ■

We can now define the map $T : \mathcal{M}^{\star r_s/4} \to \mathcal{M}^\star$ by

$$T(x) = (\pi_{s|_{\mathcal{M}^\star}})^{-1} \circ \pi_s(x)$$

which is well defined for $s$ small enough as $\mathbb{H}(\mathcal{M}_s, \mathcal{M}^\star) \leq 5s^2 < r_s/4$. Let us now state that $T$ is the map we are interested in.

**Proposition 30** *The map $T : \mathcal{M}^{\star r_s/4} \to \mathcal{M}^\star$ defined by $T(x) = (\pi_{s|_{\mathcal{M}^\star}})^{-1} \circ \pi_s(x)$, is $(\mathcal{M}, \mathcal{M}^\star)_{K_T}$-compatible with radius $t = C_\alpha^{-1}$ and $K_T = C_\alpha$.*

**Proof** We have

$$\nabla T(x) = \nabla(\pi_{s|_{\mathcal{M}^\star}})^{-1}(\pi_s(x)) \circ \nabla \pi_s(x). \tag{20}$$

From (19) and the fact that for all $x \in \mathcal{M}^\star$, $\varphi_x^{-1}$ is 2-Lipschitz on $B^d(0,\tau)$, we have for all $h \in \mathcal{T}_x(\mathcal{M}^\star)$,

$$\|\nabla \pi_{s|_{\mathcal{M}^\star}} h\| \geq \frac{1}{8}\|h\|. \tag{21}$$

Then, from (20) and the Faa di Bruno formula, we deduce that $T \in \mathcal{H}^{\beta+1}_{C_s}(\mathcal{M}_s^{r_s/4}, \mathcal{M}^\star)$.

Supposing $d_{\mathcal{H}^\alpha_1}(\mu, \mu^\star) \leq s^{2(d+1)(2\alpha-1)}$ we have from Lemma 8 that $\mathbb{H}(\mathcal{M}, \mathcal{M}^\star) \leq Cs^2$ so $\mathbb{H}(\mathcal{M}_s, \mathcal{M}) \leq \mathbb{H}(\mathcal{M}, \mathcal{M}^\star) + \mathbb{H}(\mathcal{M}_s, \mathcal{M}^\star) \leq Cs^2$. We show the same way as we did with $\mathcal{M}^\star$, that for all $x \in \mathcal{M}$, $\pi_s \circ \phi_x^{-1}$ is a local diffeomorphism for $\phi_x^{-1}$ the charts of $\mathcal{M}$ and that $\pi_s$ is injective on $\mathcal{M}$, so $\pi_{s|\mathcal{M}}$ is a diffeomorphism. Therefore, we deduce that $T_{|\mathcal{M}}$ is also a diffeomorphism.

Doing the same derivation as for $\mathcal{M}^\star$ to obtain (19), we get for $\mathcal{M}$ that

$$\lambda_{\min}(\nabla(\pi_s \circ \phi_x^{-1})(0)^\top \nabla(\pi_s \circ \phi_x^{-1})(0))^{1/2} \geq C^{-1}.$$

Furthermore, as for all $y \in \mathcal{M}^\star$ and $h \in \mathcal{T}_y(\mathcal{M}^\star)$, we have

$$\|\nabla \pi_{s|_{\mathcal{M}^\star}}(y) h\| \leq 2\|h\|,$$

we deduce that

$$\lambda_{\min}\big(\nabla(T \circ \phi_x^{-1})(0)^\top \nabla(T \circ \phi_x^{-1})(0)\big)^{1/2} \geq C^{-1}. \tag{22}$$

Then as $\phi_x^{-1}$ is $C$-Lipschitz, we have that

$$\lambda_{\min}\big(\nabla T_{|\mathcal{M}}(X)^\top \nabla T_{|\mathcal{M}}(X)\big)^{1/2} \geq C^{-1},$$

so we deduce that $T^{-1}_{|\mathcal{M}} \in \mathcal{H}^{\beta+1}_{C_s}(\mathcal{M}^\star, \mathcal{M})$. Therefore point $i)$ is verified.

From Theorem 4.8 in Federer (1959) we have that $\pi_s : \mathcal{M}_s^{r_s} \to \mathcal{M}_s$ the projection onto $\mathcal{M}_s$ verifies point $ii)$ of Definition 7 for a radius $t = r_s$, so as $\mathbb{H}(\mathcal{M}^\star, \mathcal{M}_s) \leq Cs^2 < r_s$, we deduce that $T$ verifies $ii)$ for a radius $r_s/2$.

Let us prove point $iii)$. Let $f \in \mathcal{H}^{\beta+1}_{C_s}(\mathcal{M}_s, (\mathbb{R}^p)^{p-d})$ such that $(f_1(x), ..., f_{p-d}(x))$ is an orthonormal basis of $\mathcal{T}_x(\mathcal{M}_s)^\top$. Let $R_x : \mathbb{R}^{p-d} \to \mathbb{R}^p$ a linear map such that $R_x(e_i) = f_i(x)$ for $(e_1, ..., e_{p-d})$ the canonical basis of $\mathbb{R}^{p-d}$. Then for $h \in \mathcal{H}^\eta_1(\mathcal{M}^{\star t}, \mathbb{R})$ and $x \in \mathcal{M}^\star$, we have

$$\begin{aligned}\int_{T^{-1}(\{x\})} h(y) d\lambda^{p-d}_{E_x}(y) &= \int_{\mathcal{T}_{\pi_s(x)}(\mathcal{M}_s)^\perp \cap B(x,t)} h(y) d\lambda^{p-d}_{\mathcal{T}_{\pi_s(x)}(\mathcal{M}_s)^\perp}(y) \\ &= \int_{\mathbb{R}^{p-d} \cap B(0,t)} h(x + R_x(z)) d\lambda^{p-d}(z).\end{aligned}$$

Then as $\forall z \in \mathbb{R}^{p-d}$, we have that the map $x \mapsto x + R_x(z)$ belongs to $\mathcal{H}^{\beta+1}_{C_s}(\mathcal{M}^\star, \mathbb{R}^p)$, so $x \mapsto h(x + R_x(z))$ belongs to $\mathcal{H}^\eta_{C_s}(\mathcal{M}^\star, \mathbb{R}^p)$ and therefore $T$ verifies $iii)$.

Let us prove *iv*). Using (19) and the fact that $\pi_s$ is 2-Lipschitz on $\mathcal{M}^\star$, we have that for all $x \in \mathcal{M}^{\star t}$,

$$\inf_{z \in \mathcal{T}_{T(x)}(\mathcal{M}^\star)} \|\nabla T(x) \frac{z}{\|z\|}\| \geq C^{-1}.$$

Therefore, we deduce that the approximate Jacobian of $T$ verifies

$$\mathrm{ap}_d(\nabla T(x)) \geq C^{-1},$$

so $T$ verifies the point *iv*) for $s > 0$ small enough. ■

## Appendix C. Proofs of the Interpolation Inequalities in the Manifold Case (Section 3.2.3) and Additional Results

### C.1 Extension Lemma

In this section we prove a general result allowing to extend maps that are defined on the neighborhood of a submanifold. This result is a corollary of the following version of Whitney's extension Theorem (Theorem A in Fefferman (2005) and Theorem 4 in Chapter 6 of Stein (1970)).

**Theorem 31** *Given $\eta, p \geq 1$, there exists $k \in \mathbb{N}$ depending only on $p$ and $\beta$, for which the following holds. Let $f : E \to \mathbb{R}$ a closed subset of $\mathbb{R}^p$. Suppose that for any $k$ distinct points $x_1, ..., x_k \in E$, there exists $\lfloor \eta \rfloor$ degree polynomials $P_1, ..., P_k$ on $\mathbb{R}^p$, satisfying*

*a) $P_i(x_i) = f(x_i)$ for $i = 1, ..., k$;*

*b) $\|\nabla^\alpha P_i(x_i)\| \leq M$ for $i = 1, ..., k$ and $|\alpha| \leq \lfloor \eta \rfloor$ and*

*c) $\|\nabla^\alpha (P_i - P_j)(x_i)\| \leq M \|x_i - x_j\|^{\eta - \alpha}$ for $i, j = 0, ..., k$ and $|\alpha| \leq \lfloor \eta \rfloor$ with $M$ independent of $x_1, ..., x_k$*

*Then $f$ extends to $\mathcal{H}^\eta_{C_M}(\mathbb{R}^p, \mathbb{R})$.*

Using this theorem we obtain the following extension result.

**Proposition 32** *Let $E \subset \mathbb{R}^p$ a closed set and $\mathcal{M}^\star$ a closed submanifold covered by an atlas $(\varphi_i^{-1}(B^d(0, K^{-1})))_{\{i=1,...m\}}$ with $\varphi_i^{-1} \in \mathcal{H}^1_K$ and with reach $r^\star \geq K^{-1}$. Suppose that*

*i) $d(E, \mathcal{M}^\star) < r^\star/4$,*

*ii) $\mathcal{M}^\star \subset E$,*

*iii) there exists a map $\pi \in \mathcal{H}^0_C(E, \mathcal{M}^\star)$ such that for all $x \in E$ and $t \in [0, 1]$, $x + t(\pi(x) - x) \in E$,*

*iv) there exists a map $D \in \mathcal{H}^1_C(\mathcal{M}^{\star r^\star/2}, E)$ such that $D_{|E} = Id$.*

*Then for all maps $H \in \mathcal{H}^\eta_M(E, \mathbb{R})$, $H$ can be extended into a map in $\mathcal{H}^\eta_{C_M}(\mathbb{R}^p, \mathbb{R})$.*

**Proof** Let $x_1,...,x_k \in E$ and define for all $r \in \{1,...,k\}$

$$P_r(x) = \sum_{\alpha=0}^{\lfloor \eta \rfloor} \nabla^\alpha H(x_r) \frac{(x-x_r)^\alpha}{\alpha!}.$$

with the notations $\nabla^\alpha H(x_r)$ the $\alpha$-differential of $H$ and

$$\nabla^\alpha H(x_r)(x-x_r)^\alpha = \nabla^\alpha H(x_r)\big(x-x_r, x-x_r, ..., x-x_r\big).$$

Then the polynomials $P_r$ check conditions a) and b) of Theorem 31 for $M = \|H\|_{\mathcal{H}^\eta(E,\mathbb{R})}$ . Let us show that they also verify condition c).

First, for $\alpha = \lfloor \eta \rfloor$, we have

$$\|\nabla^{\lfloor \eta \rfloor} P_r(x_r) - \nabla^{\lfloor \eta \rfloor} P_j(x_r)\| = \|\nabla^{\lfloor \eta \rfloor} H(x_r) - \nabla^{\lfloor \eta \rfloor} H(x_j)\| \leq C\|x_r - x_j\|^{\eta - \lfloor \eta \rfloor}.$$

Let $\alpha \in \{0,...,\lfloor \eta \rfloor - 1\}$ and take a path $\gamma \in \mathcal{H}^1([0,1],E)$ such that $\gamma(0) = x_j$ and $\gamma(1) = x_r$. Applying Taylor expansions recursively we obtain

$$\begin{aligned}
\nabla^\alpha P_r(x_r) = \nabla^\alpha H(x_r) &= \nabla^\alpha H(x_j) + \int_0^1 \nabla^{\alpha+1} H(\gamma(t_1))\dot{\gamma}(t_1)dt_1 \\
&= \nabla^\alpha H(x_j) + \int_0^1 \left( \nabla^{\alpha+1} H(x_j) + \int_0^1 \nabla^{\alpha+2} H(\gamma(t_2 t_1)) t_1 \dot{\gamma}(t_2 t_1) d_{t_2} \right) \dot{\gamma}(t_1) dt_1 \\
&= \nabla^\alpha H(x_j) \\
&\quad + \sum_{l=1}^{\lfloor \eta \rfloor - \alpha - 1} \int_0^1 ... \int_0^1 \nabla^{\alpha+l} H(x_j) \dot{\gamma}(t_l t_{l-1} ... t_1) d_{t_l} \dot{\gamma}(t_{l-1} t_{l-2} ... t_1) t_{l-1} d_{t_{l-1}} ... \dot{\gamma}(t_1) t_1^{l-1} dt_1 \\
&\quad + \int_0^1 ... \int_0^1 \nabla^{\lfloor \eta \rfloor} H(\gamma(t_{\lfloor \eta \rfloor - \alpha} ... t_1) \dot{\gamma}(t_{\lfloor \eta \rfloor - \alpha} ... t_1) d_{t_{\lfloor \eta \rfloor - \alpha}} \\
&\qquad \times \dot{\gamma}(t_{\lfloor \eta \rfloor - \alpha - 1} ... t_1) t_{t_{\lfloor \eta \rfloor - \alpha - 1}} d_{t_{\lfloor \eta \rfloor - \alpha - 1}} ... \dot{\gamma}(t_1) t_1^{\lfloor \eta \rfloor - \alpha - 1} dt_1 \\
&= \nabla^\alpha H(x_j) \\
&\quad + \sum_{l=1}^{\lfloor \eta \rfloor - \alpha} \int_0^1 ... \int_0^1 \nabla^{\alpha+l} H(x_j) \dot{\gamma}(t_l t_{l-1} ... t_1) d_{t_l} \dot{\gamma}(t_{l-1} t_{l-2} ... t_1) t_{l-1} d_{t_{l-1}} ... \dot{\gamma}(t_1) t_1^{l-1} dt_1 \\
&\quad + \int_0^1 ... \int_0^1 \big( \nabla^{\lfloor \eta \rfloor} H(\gamma(t_{\lfloor \eta \rfloor - \alpha} ... t_1) - \nabla^{\lfloor \eta \rfloor} H(x_j) \big) \dot{\gamma}(t_{\lfloor \eta \rfloor - \alpha} ... t_1) d_{t_{\lfloor \eta \rfloor - \alpha}} ... \dot{\gamma}(t_1) t_1^{\lfloor \eta \rfloor - \alpha - 1} dt_1.
\end{aligned}$$

Using integration by part one can show that

$$\begin{aligned}
&\sum_{l=1}^{\lfloor \eta \rfloor - \alpha} \int_0^1 ... \int_0^1 \nabla^{\alpha+l} H(x_j) \dot{\gamma}(t_l t_{l-1} ... t_1) d_{t_l} \dot{\gamma}(t_{l-1} t_{l-2} ... t_1) t_{l-1} d_{t_{l-1}} ... \dot{\gamma}(t_1) t_1^{l-1} dt_1 \\
&\qquad = \sum_{l=1}^{\lfloor \eta \rfloor - \alpha} \nabla^{\alpha+l} H(x_j) \frac{1}{l!} (x_r - x_j)^l,
\end{aligned}$$

so from the previous derivation we get

$$\begin{aligned}
&\|\nabla^\alpha P_r(x_r) - \nabla^\alpha P_j(x_r)\| \\
&= \| \int_0^1 ... \int_0^1 \left(\nabla^{\lfloor \eta \rfloor} H(\gamma(t_{\lfloor \eta \rfloor - \alpha} ... t_1) - \nabla^{\lfloor \eta \rfloor} H(x_j)\right) \dot{\gamma}(t_{\lfloor \eta \rfloor - \alpha} ... t_1) d_{t_{\lfloor \eta \rfloor - \alpha}} ... \dot{\gamma}(t_1) t_1^{\lfloor \eta \rfloor - \alpha - 1} dt_1 \| \\
&\leq \int_0^1 ... \int_0^1 C \|\gamma(t_{\lfloor \eta \rfloor - \alpha} ... t_1) - x_j\|^{\beta - \lfloor \beta \rfloor} \|\dot{\gamma}(t_{\lfloor \eta \rfloor - \alpha} ... t_1)\| d_{t_{\lfloor \eta \rfloor - \alpha}} ... \|\dot{\gamma}(t_1)\| t_1^{\lfloor \eta \rfloor - \alpha - 1} dt_1.
\end{aligned}$$

Suppose that $\|x_r - x_j\| \leq r^\star/4$ and for $D$ the map of assumption iv), define

$$\gamma(s) = D(x_r + s(x_j - x_r))$$

which is well defined as for $s \in [0,1]$,

$$d(x_r + s(x_j - x_r), \mathcal{M}^\star) \leq \|x_r + s(x_j - x_r) - x_r\| + d(x_r, \mathcal{M}^\star) < r^\star/2.$$

We have $\|\dot{\gamma}(s)\| \leq C\|x_r - x_j\|$ so we deduce that

$$\begin{aligned}
&\|\nabla^\alpha P_r(x_r) - \nabla^\alpha P_j(x_r)\| \\
&\quad \leq \int_0^1 ... \int_0^1 C\|x_r - x_j\|^{\beta - \lfloor \beta \rfloor} 4C\|x_r - x_j\| d_{t_{\lfloor \eta \rfloor}} 4C\|x_r - x_j\| d_{t_{\lfloor \beta \rfloor}} ... 4C\|x_r - x_j\| dt_\alpha \\
&\quad \leq C\|x_r - x_j\|^{\beta + 1 - \alpha}.
\end{aligned}$$

Suppose now that $\|x_r - x_j\| \geq r^\star/4$ and define $\delta \in \mathcal{H}^1([0,1], \mathcal{M}^\star)$ a geodesic on $\mathcal{M}^\star$ from $x_j$ to $x_r$. Let

$$\gamma(s) = \begin{cases} x_r + 3s(\pi(x_r) - x_r) & \text{if } s \in [0, 1/3] \\ \delta(3(t - 1/3)) & \text{if } s \in (1/3, 2/3) \\ \pi(x_j) + 3(s - 2/3)(x_j - \pi(x_j)) & \text{if } s \in [2/3, 1] \end{cases}$$

be the path from $x_r$ to $x_j$ that starts by projecting $x_r$ onto $\mathcal{M}^\star$, then circulates along $\mathcal{M}^\star$ until it reaches the projection of $x_j$, and finally go to $x_j$ in a straight line.

For $s \in [0, 1/3] \cup [2/3, 1]$ we have

$$\|\dot{\gamma}(s)\| \leq 3 \max(\|\pi(x_r) - x_r\|, \|\pi(x_j) - x_j\|) \leq C \leq C r^{\star - 1} \|x_r - x_j\|.$$

As $\mathcal{M}^\star$ can be covered by an atlas $(\varphi_i^{-1}(B^d(0,\tau)))_{\{i=1,...m\}}$ with $\varphi_i^{-1} \in \mathcal{H}^1_K$, we deduce that

$$\int_0^1 \|\dot{\delta}(t)\| dt \leq mK\tau.$$

Therefore, for $\delta$ with constant speed we have

$$\|\dot{\delta}(t)\| \leq mK\tau \leq C r^{\star - 1} \|x_r - x_j\|.$$

We then get as in the case where $\|x_r - x_j\| \leq t/4$, that

$$\|\nabla^\alpha P_r(x_r) - \nabla^\alpha P_j(x_r)\| \leq C(r^{\star - 1}\|x_r - x_j\|)^{\beta + 1 - \alpha}.$$

Therefore using the sharp form of Whitney's extension (Theorem 31), $H$ can be extended to a map in $\mathcal{H}^\eta_{C_M}(\mathbb{R}^p, \mathbb{R})$. ■

### C.2 Proof of Lemma 10

Let us first recall a result from Stéphanovitch et al. (2023) showing that we can gain some weak Besov regularity by paying a logarithmic term. For $1 \leq p \leq \infty$, we let $L^p(\mathbb{R}^d, \mathbb{R}, C)$ be the set of functions $f : \mathbb{R}^d \to \mathbb{R}$ with $p$-norm $\|f\|_{L^p} := \left(\int \|f\|^p d\lambda^d\right)^{1/p}$ bounded by $C > 0$.

**Proposition 33** *Let $f \in L^1(\mathbb{R}^d, \mathbb{R}, 1) \cap L^2(\mathbb{R}^d, \mathbb{R})$ and $g \in \mathcal{H}_1^\gamma(\mathbb{R}^d, \mathbb{R}, 1)$ with $\gamma > 0$. Then for all $\tau > 0$, $\epsilon \in (0,1)$ we have*

$$\int_{\mathbb{R}^d} f(x)g(x)d\lambda_{\mathbb{R}^d}(x) \leq C \log(\epsilon^{-1})^\tau \int_{\mathbb{R}^d} f(x)\tilde{\Gamma}_\epsilon^{0,-\tau}(g)(x)d\lambda_{\mathbb{R}^d}(x) + C\epsilon$$

*for*

$$\tilde{\Gamma}_\epsilon^{0,-\tau}(g)(x) = \sum_{j=0}^{\log(\epsilon^{-1})} \sum_{l=1}^{2^d} \sum_{w \in \mathbb{Z}^d} (1+j)^{-\tau} \alpha_g(j,l,w) \frac{\alpha_g(j,l,w)\alpha_f(j,l,w)}{|\alpha_g(j,l,w)\alpha_f(j,l,z)|} \psi_{j,l,z}(x).$$

Using Proposition 32, we can extend $T$ into a map belonging to $\mathcal{H}_C^{\beta+1}(\mathbb{R}^p, \mathbb{R}^p)$. Let $h \in \mathcal{H}_1^1(\mathbb{R}^p, \mathbb{R})$, using Proposition 6 we have

$$\begin{aligned}\int (h(x) - h(T(x))f_\mu(x)d\lambda_{\mathcal{M}}(x) &\leq \int \|x - T(x)\| f_\mu(x)d\lambda_{\mathcal{M}}(x) \\ &= \sum_{i=1}^m \int_{\mathbb{R}^d} \frac{1}{C_i} \|\phi_i(u) - T(\phi_i(u))\| \zeta_i(u) d\lambda^d(u).\end{aligned}$$

Let $\xi \in (0,1)$ and for all $i \in \{1, ..., m\}$, define

$$X_i(u) = \phi_i(u) - T(\phi_i(u))$$

and $L_i : \mathbb{R}^d \to \mathbb{R}^p$ by

$$L_i(u) = \left( \frac{X_i(u)}{\xi} \mathbb{1}_{\{\|X_i(u)\| < \xi\}} + \frac{X_i(u)}{\|X_i(u)\|} \mathbb{1}_{\{\|X_i(u)\| \geq \xi\}} \right) \mathbb{1}_{\{u \in B^d(0,2\tau)\}}. \tag{23}$$

We have

$$\begin{aligned}&\int_{\mathbb{R}^d} \|\phi_i(u) - T(\phi_i(u))\| \zeta_i(u) d\lambda^d(u) \\ &\quad = \int_{\mathbb{R}^d} \langle \frac{\phi_i(u) - T(\phi_i(u))}{\|\phi_i(u) - T(\phi_i(u))\|}, \phi_i(u) - T(\phi_i(u)) \rangle \zeta_i(u) d\lambda^d(u) \\ &\quad \leq \int_{\mathbb{R}^d} \langle L_i(u), X_i(u) \rangle \zeta_i(u) d\lambda^d(u) + \xi.\end{aligned}$$

Let us write $K_X = \max_{i \in \{1,...,m\}} \|X_i\|_{\mathcal{H}^{\beta+1}} \vee 1 \leq C$ and $\delta_u^i = \|X_i(u)\|$ for $u \in B^d(0, 2\tau)$. We have that $L_i \in \mathcal{H}_1^0(B^d(0,2\tau), \mathbb{R}^p)$, let us fix $\eta \in (1, \beta+1]$ and show that $L_i$ has some additional Hölder regularity.

**Proposition 34** *For all $i \in \{1, ..., m\}$ and $L_i : \mathbb{R}^d \to \mathbb{R}^p$ defined in (23), we have for all $u, v \in B^d(0, 2\tau)$,*

$$\|L_i(u) - L_i(v)\| \leq 8K_X \min(\delta_u^i, \delta_v^i)^{-(1-(\eta-\lfloor\eta\rfloor))}\|u - v\|^{1-(\eta-\lfloor\eta\rfloor)}.$$

**Proof** Suppose first that $\|u - v\| \geq \min(\delta_u^i, \delta_v^i)/(4K_X)$. Then

$$\begin{aligned}\|L_i(u) - L_i(v)\| &\leq 2 = 2\|u - v\|^{-(1-(\eta-\lfloor\eta\rfloor))}\|u - v\|^{1-(\eta-\lfloor\eta\rfloor)} \\ &\leq 8K_X \min(\delta_u^i, \delta_v^i)^{-(1-(\eta-\lfloor\eta\rfloor))}\|u - v\|^{1-(\eta-\lfloor\eta\rfloor)}.\end{aligned}$$

Now if $\|u - v\| \leq \min(\delta_u^i, \delta_v^i)/(4K_X)$, then for $u_s = u + s(v - u)$ with $s \in [0, 1]$, we have

$$\begin{aligned}\|X(u_s)\| &\geq \|X_i(u)\| - sK_X\|u - v\| \\ &\geq \delta_u^i - \frac{1}{4}\min(\delta_u^i, \delta_v^i) \\ &\geq \frac{1}{2}\min(\delta_u^i, \delta_v^i).\end{aligned}$$

Furthermore,

$$\nabla L_i(u) = \frac{\mathbb{1}_{\{\|X_i(u)\| \geq \xi\}}}{\|X_i(u)\|}\Big(\nabla X_i(u) - \frac{X_i(u)}{\|X_i(u)\|}\frac{(X_i(u))^\top}{\|X_i(u)\|}\nabla X_i(u)\Big) + \frac{\mathbb{1}_{\{\|X_i(u)\| < \xi\}}}{\xi}\nabla X_i(u),$$

so $\|\nabla L_i(u)\| \leq K_X\|X_i(u)\|^{-1} = K_X(\delta_u^i)^{-1}$. Then

$$\begin{aligned}\|L_i(u) - L_i(v)\| &= \|\int_0^1 \nabla L_i(u + s(v - u))(v - u)ds\| \\ &\leq \|v - u\|\int_0^1 \|\nabla L_i(u + s(v - u))\|ds \leq 2K_X \min(\delta_u^i, \delta_v^i)^{-1}\|v - u\| \\ &\leq 2K_X \min(\delta_u^i, \delta_v^i)^{-(1-(\eta-\lfloor\eta\rfloor))}\|u - v\|^{1-(\eta-\lfloor\eta\rfloor)}.\end{aligned}$$

■

We have that $\phi_{i|B^d(0,\tau)}$ the restriction of $\phi_i$ to $B^d(0, \tau)$, is a diffeomorphism on its image. For $x \in \phi_i(B^d(0, \tau))$, we will write $\phi_i^{-1}(x)$ for the inverse application of $\phi_{i|B^d(0,\tau)}$.

Let us define $F_i : T^{-1}(\phi_i(B^d(0, \tau))) \to B^d(0, \tau)$ by

$$F_i := \phi_i^{-1} \circ T_{|\mathcal{M}}^{-1} \circ T,$$

$$A := \{x \in \mathcal{M}^{\star t} |\ \|x - T(x)\| \leq \|T_{|\mathcal{M}}^{-1} \circ T(x) - T(x)\|\} \tag{24}$$

and $H_i : A \to \mathbb{R}$ by

$$H_i(x) := \Big\langle \tilde{\Gamma}_\epsilon^{-\eta+1,-2}(L_i) \circ F_i(x)\ ,\ x - T(x)\Big\rangle \zeta_i(F_i(x))\mathbb{1}_{\{T_{|\mathcal{M}}^{-1} \circ T(x) \in \phi_i(B^d(0,\tau))\}}. \tag{25}$$

We have the following result on $L_i, X_i$ and $H_i$.

**Proposition 35** *For all $i \in \{1,...,m\}$ and $\epsilon \in (0,1)$ we have*

$$\int_{\mathbb{R}^d} \langle L_i(u), X_i(u)\rangle \zeta_i(u) d\lambda^d(u) \leq C \log(\epsilon^{-1})^2 \left( \int_{\mathcal{M}} H_i(x) d\mu(x) \right)^{\frac{\beta+1}{\beta+\eta}} + C\epsilon.$$

**Proof** For $r \in \{1,...,p\}$, Let

$$\tilde{\Gamma}_\epsilon^{0,-2}(L_{i_r})(u) = \sum_{j=0}^{\log(\lfloor \epsilon^{-1} \rfloor)} \sum_{l=1}^{2^d} \sum_{z \in \mathbb{Z}^d} (1+j)^{-2} \alpha_{L_{i_r}}(j,l,z) S(j,l,w)_{ir} \psi_{jlz}(u)$$

with $S(j,l,w)_{ir} = \frac{\alpha_{L_{i_r}}(j,l,z)\alpha_{X_i\zeta_i}(j,l,z)_r}{|\alpha_{L_{i_r}}(j,l,z)\alpha_{X_i\zeta_i}(j,l,z)_r|}$. Then using Proposition 33, we have

$$\int_{\mathbb{R}^d} \langle L_i(u), X_i(u)\rangle \zeta_i(u) d\lambda^d(u) \leq C \log(\epsilon^{-1})^2 \int_{\mathbb{R}^d} \langle \tilde{\Gamma}_\epsilon^{0,-2}(L_i)(u), X_i(u)\rangle \zeta_i(u) d\lambda^d(u) + C\epsilon.$$

Applying Proposition 25 for $h_1 = \tilde{\Gamma}_\epsilon^{0,-2}(L_i)$, $h_2 = X_i\zeta_i$, $s_1 = 0$, $b_1 = -2$, $s_2 = \beta+1$, $b_2 = 0$, $\tau = -\eta+1$, $t = 1$, $r = 0$ and $q = \frac{\beta+\eta}{\beta+1}$ we get

$$\begin{aligned}
&\int_{\mathbb{R}^d} \langle \tilde{\Gamma}_\epsilon^{0,-2}(L_i)(u), X_i(u)\rangle \zeta_i(u) d\lambda^d(u) \\
&\quad \leq \left\langle \tilde{\Gamma}^{-\eta+1}(\tilde{\Gamma}_\epsilon^{0,-2}(L_i)), X_i\zeta_i \right\rangle_{L^2}^{\frac{\beta+1}{\beta+\eta}} \left\langle \tilde{\Gamma}^0(\tilde{\Gamma}_\epsilon^{0,-2}(L_i)), \Gamma^{\beta+1}(X_i\zeta_i) \right\rangle_{L^2}^{\frac{\eta-1}{\beta+\eta}} \\
&\quad \leq \left\langle \tilde{\Gamma}_\epsilon^{-\eta+1,-2}(L_i), X_i\zeta_i \right\rangle_{L^2}^{\frac{\beta+1}{\beta+\eta}} \left\langle \tilde{\Gamma}_\epsilon^{0,-2}(L_i), \Gamma^{\beta+1}(X_i\zeta_i) \right\rangle_{L^2}^{\frac{\eta-1}{\beta+\eta}}.
\end{aligned}$$

We have that $\Gamma^{\beta+1}(X_i\zeta_i) \in \mathcal{B}^0_{\infty,\infty}(C)$ and $\tilde{\Gamma}_\epsilon^{0,-2}(L_i) \in \mathcal{B}^{0,2}_{\infty,\infty}(C)$ so

$$\begin{aligned}
&\sum_{r=1}^{p} \sum_{j=0}^{\infty} \sum_{l=1}^{2^d} \sum_{z \in \mathbb{Z}^d} \alpha_{\tilde{\Gamma}_\epsilon^{0,-2}(L_i)}(j,l,z)_r \alpha_{\Gamma^{\beta+1}(X_i\zeta_i)}(j,l,z)_r \\
&\qquad \leq \sum_{r=1}^{p} \sum_{j=0}^{\infty} \sum_{l=1}^{2^d} \sum_{z \in \mathbb{Z}^d} (1+j)^{-2} C 2^{-jd} \mathbb{1}_{\{z \in B^d(0,C)\}} \\
&\qquad \leq C \sum_{r=1}^{p} \sum_{j=0}^{\infty} (1+j)^{-2} \leq C.
\end{aligned}$$

Therefore, we have

$$\left\langle \tilde{\Gamma}_\epsilon^{0,-2}(L_i), \Gamma^{\beta+1}(X_i\zeta_i) \right\rangle_{L^2}^{\frac{\eta-1}{\beta+\eta}} \leq C.$$

Furthermore, recalling the proof of Proposition 6, we have

$$(\phi_i)_{\#\lambda^d_{|B^d(0,\tau)}} = \mu_{|\phi_i(B^d(0,\tau))}.$$

Therefore,

$$\begin{aligned}\left\langle \tilde{\Gamma}_\epsilon^{-\eta+1,-2}(L_i), X_i\zeta_i \right\rangle_{L^2} &= \int_{\mathbb{R}^d} \langle \tilde{\Gamma}_\epsilon^{-\eta+1,-2}(L_i)(u), X_i(u)\rangle \zeta_i(u) d\lambda^d(u) \\ &= \int_{\mathbb{R}^d} \langle \tilde{\Gamma}_\epsilon^{-\eta+1,-2}(L_i)(\phi_i^{-1}(x)), x - T(x)\rangle \zeta_i(\phi_i^{-1}(x)) d(\phi_i)_{\#\lambda^d_{\lfloor B^d(0,\tau)}}(x) \\ &= \int_{\mathcal{M}} H_i(x) d\mu(x).\end{aligned}$$

We can then conclude that

$$\int_{\mathbb{R}^d} \langle L_i(u), X_i(u)\rangle \zeta_i(u) d\lambda^d(u) \leq C \log(\epsilon^{-1})^2 \left( \int_{\mathcal{M}} H_i(x) d\mu(x) \right)^{\frac{\beta+1}{\beta+\eta}} + C\epsilon.$$

■

Let us now show that the maps $H_i$ from (25) are of regularity $\eta$. To that end, let us first show that $\nabla^{\lfloor \eta \rfloor} \tilde{\Gamma}_\epsilon^{-\eta+1,-2}(L_i)$ exists and has some Hölder regularity.

**Lemma 36** *For all $i \in \{1, ..., m\}$, $i_1, ..., i_{\lfloor \eta \rfloor} \in \{1, ..., d\}$ and $u \in B^d(0, \tau)$, we have*

$$\| \partial^{\lfloor \eta \rfloor}_{i_1,...,i_{\lfloor \eta \rfloor}} \tilde{\Gamma}_\epsilon^{-\eta+1,-2}(L_i)(u) \| \leq C K_X (\delta_u^i)^{-(\lfloor \eta \rfloor + 1 - \eta)},$$

*(with $\delta_u^i = \|X_i(u)\|$) and for $u, v \in B^d(0, 2\tau)$ we have*

$$\| \partial^{\lfloor \eta \rfloor}_{i_1,...,i_{\lfloor \eta \rfloor}} \tilde{\Gamma}_\epsilon^{-\eta+1,-2}(L_i)(u) - \partial^{\lfloor \eta \rfloor}_{i_1,...,i_{\lfloor \eta \rfloor}} \tilde{\Gamma}_\epsilon^{-\eta+1,-2}(L_i)(v) \| \leq C K_X (\delta_u^i \wedge \delta_v^i)^{-1} \|u - v\|^{\eta - \lfloor \eta \rfloor}.$$

**Proof** For $u \in B^d(0, \tau)$, we shall distinguish whether we have $\delta_u^i > \xi$ or not.

Let us first consider the case $\delta_u^i > \xi$. Define $L_{ui} : B^d(u, \frac{\delta_u^i}{4K_X}) \to \mathbb{R}^p$ being equal to $L_i$. We then extend $L_{ui}$ to $\mathbb{R}^d$ by

$$\overline{L}_{ui}(v) = L_{ui}(\pi_{B_u}(v))(0 \vee (1 - 3\|v - \pi_{B_u}(v)\|)), \tag{26}$$

for $\pi_{B_u}$ the projection on $B^d(u, \frac{\delta_u^i}{4K_X})$. For all $v \in B^d(u, \frac{\delta_u^i}{4K_X})$, we have $\delta_v^i \geq \delta_u^i/2$ so from Proposition 34 we get

$$L_{ui} \in \mathcal{H}_1^0(B^d(u, \frac{\delta_u^i}{4K_X}), \mathbb{R}^p) \cap \mathcal{H}^{1-(\eta - \lfloor \eta \rfloor)}_{16K_X(\delta_u^i)^{-(1-(\eta-\lfloor \eta \rfloor))}}(B^d(u, \frac{\delta_u^i}{4K_X}), \mathbb{R}^p).$$

Therefore, we also have that

$$\overline{L}_{ui} \in \mathcal{H}_1^0(\mathbb{R}^d, \mathbb{R}^p) \cap \mathcal{H}^{1-(\eta - \lfloor \eta \rfloor)}_{16K_X(\delta_u^i)^{-(1-(\eta-\lfloor \eta \rfloor))}}(\mathbb{R}^d, \mathbb{R}^p).$$

For $r \in \{1, ..., p\}$, let us write $(\alpha_{\overline{L}_{ui}}(j, l, z)_r)_{(j,l,z)}$ the wavelet coefficients of $(\overline{L}_{ui})_r$. As $\text{support}(\psi_{jlz}) \subset B^d(2^{-j}z, C2^{-j})$, consider any $(j, l, z)$ such that

$$j \geq \lfloor \log_2(C K_X (\delta_u^i)^{-1}) \rfloor + 1,$$
$$\text{support}(\psi_{jlz}) \cap B^d(u, \frac{\delta_u^i}{8K_X}) \neq \varnothing.$$

Then $\alpha_{\overline{L}_{ui}}(j,l,z)_i = \alpha_L(j,l,z)_i$. Let us note

$$\xi_u^i = \lfloor \log_2(CK_X(\delta_u^i)^{-1}) \rfloor + 1$$

and

$$\tilde{\Gamma}_\epsilon^{-\eta+1,-2}(\overline{L}_{ui})_r(v) = \sum_{j=0}^{\log_2(\epsilon^{-1})} \sum_{l=1}^{2^d} \sum_{z\in\mathbb{Z}^d} 2^{-j(\eta-1)}(1+j)^{-2} S(j,l,z)_{ir} \alpha_{\overline{L}_{ui}}(j,l,z)_r \psi_{j,l,z}(v),$$

the $(\eta-1,2)$ wavelet regularization of $\overline{L}_{u_i}$. In particular we have

$$\tilde{\Gamma}_\epsilon^{-\eta+1,-2}(\overline{L}_{ui}) \in \mathcal{H}_C^{\eta-1}(\mathbb{R}^d,\mathbb{R}^p) \cap \mathcal{H}^{\lfloor\eta\rfloor}_{CK_X(\delta_u^i)^{-(1-(\eta-\lfloor\eta\rfloor))}}(\mathbb{R}^d,\mathbb{R}^p).$$

Let $i_1,...,i_{\lfloor\eta\rfloor} \in \{1,...,d\}$, we have

$$\begin{aligned}
&|\partial^{\lfloor\eta\rfloor}_{i_1,...,i_{\lfloor\eta\rfloor}} \tilde{\Gamma}_\epsilon^{-\eta+1,-2}(L_i)_r(u)| \\
=&|\partial^{\lfloor\eta\rfloor}_{i_1,...,i_{\lfloor\eta\rfloor}} \sum_{j=0}^{\log_2(\epsilon^{-1})} \sum_{l=1}^{2^d} \sum_{z\in\mathbb{Z}^d} \alpha_{\tilde{\Gamma}^{-\eta+1,-2}(L_i)}(j,l,z)_r \psi_{j,l,z}(u)| \\
\leq&|\partial^{\lfloor\eta\rfloor}_{i_1,...,i_{\lfloor\eta\rfloor}} \sum_{j=0}^{\xi_u^i\wedge\log_2(\epsilon^{-1})} \sum_{l=1}^{2^d} \sum_{z\in\mathbb{Z}^d} (\alpha_{\tilde{\Gamma}^{-\eta+1,-2}(L_i)}(j,l,z)_r - \alpha_{\tilde{\Gamma}^{-\eta+1,-2}(\overline{L}_{ui})}(j,l,z)_r)\psi_{j,l,z}(u)| \\
&+ |\partial^{\lfloor\eta\rfloor}_{i_1,...,i_{\lfloor\eta\rfloor}} \sum_{j=0}^{\xi_u^i\wedge\log_2(\epsilon^{-1})} \sum_{l=1}^{2^d} \sum_{z\in\mathbb{Z}^d} \alpha_{\tilde{\Gamma}^{-\eta+1,-2}(\overline{L}_{ui})}(j,l,z)_r \psi_{j,l,z}(u) \\
&+ \partial^{\lfloor\eta\rfloor}_{i_1,...,i_{\lfloor\eta\rfloor}} \sum_{j=\xi_u^i\wedge\log_2(\epsilon^{-1})+1}^{\log_2(\epsilon^{-1})} \sum_{l=1}^{2^d} \sum_{z\in\mathbb{Z}^d} \alpha_{\tilde{\Gamma}^{-\eta+1,-2}(L_i)}(j,l,z)_r \psi_{j,l,z}(u)|. \qquad (27)
\end{aligned}$$

For $k \in \{1, ..., d\}$, let us write $\theta_k = \sum_{r=1}^{\lfloor \eta \rfloor} \mathbb{1}_{\{i_r = k\}}$. For the first term of (27) we have

$$
\begin{aligned}
&|\partial^{\lfloor \eta \rfloor}_{i_1,...,i_{\lfloor \eta \rfloor}} \sum_{j=0}^{\xi^i_u \wedge \log_2(\epsilon^{-1})} \sum_{l=1}^{2^d} \sum_{z \in \mathbb{Z}^d} (\alpha_{\tilde{\Gamma}^{-\eta+1,-2}(L_i)}(j,l,z)_r - \alpha_{\tilde{\Gamma}^{-\eta+1,-2}(\overline{L}_{ui})}(j,l,z)_r)\psi_{j,l,z}(u)| \\
&\quad \leq \sum_{j=0}^{\xi^i_u \wedge \log_2(\epsilon^{-1})} \sum_{l=1}^{2^d} \sum_{z \in \mathbb{Z}^d} |(\alpha_{\tilde{\Gamma}^{-\eta+1,-2}(L_i)}(j,l,z)_r - \alpha_{\tilde{\Gamma}^{-\eta+1,-2}(\overline{L}_{ui})}(j,l,z)_r) \\
&\qquad \times 2^{j(\lfloor \eta \rfloor + d/2)} \prod_{k=0}^{d} \partial^{\theta_k} \psi_{l_k}(2^j u_k - z_k)| \\
&\quad \leq C \sum_{j=0}^{\xi^i_u \wedge \log_2(\epsilon^{-1})} \sum_{z \in \mathbb{Z}^d} 2^{j(\lfloor \eta \rfloor + 1 - \eta)} | \prod_{k=0}^{d} \partial^{\theta_k} \psi_{l_k}(2^j u_k - z_k)| \\
&\quad \leq C \sum_{j=0}^{\xi^i_u \wedge \log_2(\epsilon^{-1})} \sum_{z \in \mathbb{Z}^d} 2^{j(\lfloor \eta \rfloor + 1 - \eta)} C \mathbb{1}_{\{u \in supp(\psi_{jlz})\}} \\
&\quad \leq C \sum_{j=0}^{\xi^i_u \wedge \log_2(\epsilon^{-1})} 2^{j(\lfloor \eta \rfloor + 1 - \eta)} \leq C 2^{\xi^i_u(\lfloor \eta \rfloor + 1 - \eta)} \leq C K_X (\delta^i_u)^{-(\lfloor \eta \rfloor + 1 - \eta)}.
\end{aligned}
$$

If $\xi^i_u < \log_2(\epsilon^{-1})$, for the second term we have

$$
\begin{aligned}
&|\partial^{\lfloor \eta \rfloor}_{i_1,...,i_{\lfloor \eta \rfloor}} \sum_{j=0}^{\xi^i_u} \sum_{l=1}^{2^d} \sum_{z \in \mathbb{Z}^d} \alpha_{\tilde{\Gamma}^{-\eta+1,-2}(\overline{L}_{ui})}(j,l,z)_r \psi_{j,l,z}(u) \\
&\qquad + \partial^{\lfloor \eta \rfloor}_{i_1,...,i_{\lfloor \eta \rfloor}} \sum_{j=\xi^i_u+1}^{\log_2(\epsilon^{-1})} \sum_{l=1}^{2^d} \sum_{z \in \mathbb{Z}^d} \alpha_{\tilde{\Gamma}^{-\eta+1,-2}(L_i)}(j,l,z)_r \psi_{j,l,z}(u)| \\
&\quad = |\partial^{\lfloor \eta \rfloor}_{i_1,...,i_{\lfloor \eta \rfloor}} \sum_{j=0}^{\log_2(\epsilon^{-1})} \sum_{l=1}^{2^d} \sum_{z \in \mathbb{Z}^d} \alpha_{\tilde{\Gamma}^{-\eta+1,-2}(\overline{L}_{ui})}(j,l,z)_r \psi_{j,l,z}(u)| \\
&\quad \leq \|\partial^{\lfloor \eta \rfloor}_{i_1,...,i_{\lfloor \eta \rfloor}} \tilde{\Gamma}^{-\eta+1,-2}(\overline{L}_{ui})_r\|_\infty \leq C K_X (\delta^i_u)^{-(\lfloor \eta \rfloor + 1 - \eta)}.
\end{aligned}
$$

On the other hand, if $\xi^i_u \geq \log_2(\epsilon^{-1})$ we have that the second term is equal to

$$
\begin{aligned}
&|\partial^{\lfloor \eta \rfloor}_{i_1,...,i_{\lfloor \eta \rfloor}} \sum_{j=0}^{\log_2(\epsilon^{-1})} \sum_{l=1}^{2^d} \sum_{z \in \mathbb{Z}^d} \alpha_{\tilde{\Gamma}^{-\eta+1,-2}(\overline{L}_{ui})}(j,l,z)_r \psi_{j,l,z}(u)| \\
&\leq \|\partial^{\lfloor \eta \rfloor}_{i_1,...,i_{\lfloor \eta \rfloor}} \tilde{\Gamma}^{-\eta+1,-2}(\overline{L}_{ui})_r\|_\infty \leq C K_X (\delta^i_u)^{-(\lfloor \eta \rfloor + 1 - \eta)}.
\end{aligned}
$$

Now that we have shown that

$$
|\partial^{\lfloor \eta \rfloor}_{i_1,...,i_{\lfloor \eta \rfloor}} \tilde{\Gamma}^{-\eta+1,-2}_\epsilon(L_i)_r(u)| \leq C K_X (\delta^i_u)^{-(\lfloor \eta \rfloor + 1 - \eta)}, \tag{28}
$$

for $u \in B^d(0,\tau)$ satisfying $\delta_u^i > \xi$, in the case $\delta_u^i \leq \xi$ we have $L_i(u) = X_i(u)$ so it can be treated the same way. Defining $L_{ui}$ coinciding with $L_i$ on $B^d(u, \frac{\xi}{4K_X})$ and doing the same derivations, we also obtain $|\partial^{\lfloor\eta\rfloor}_{i_1,...,i_{\lfloor\eta\rfloor}} \tilde{\Gamma}_\epsilon^{-\eta+1,-2}(L_i)(u)_r| \leq CK_X(\delta_u^i)^{-(\lfloor\eta\rfloor+1-\eta)}$.

Let us now show that for $u, v \in B^d(0,2\tau)$ we have

$$\begin{aligned}|\partial^{\lfloor\eta\rfloor}_{i_1,...,i_{\lfloor\eta\rfloor}} \tilde{\Gamma}_\epsilon^{-\eta+1,-2}(L_i)_r(u) - \partial^{\lfloor\eta\rfloor}_{i_1,...,i_{\lfloor\eta\rfloor}} \tilde{\Gamma}_\epsilon^{-\eta+1,-2}(L_i)_r(v)| \\ \leq CK_X((\delta_u^i)^{-1} \vee (\delta_v^i)^{-1})\|u-v\|^{\eta-\lfloor\eta\rfloor}. \end{aligned} \tag{29}$$

If $\|u-v\| < \frac{\delta_u^i \wedge \delta_v^i}{8K_X}$, define $L_{i(u,v)}$ coinciding with $L$ on $B^d(\underset{w\in\{u,v\}}{\arg\max}\, \delta_w, \frac{\delta_u^i \vee \delta_v^i}{8K_X})$ and extend it into $\overline{L}_{i(u,v)} : B^d(0,\tau) \to \mathbb{R}^p$ by projection as we did with $L_{ui}$ in (26). Otherwise, if $\|u-v\| \geq \frac{\delta_u^i \wedge \delta_v^i}{8K_X}$, define $L_{i(u,v)}$ coinciding with $L$ on $B^d(u, \frac{\delta_u^i \wedge \delta_v^i}{32K_X}) \cup B^d(v, \frac{\delta_u^i \wedge \delta_v^i}{32K_X})$. Then, we have $L_{i(u,v)} \in \mathcal{H}^1_{CK_X(\delta_u^i)^{-1}\vee(\delta_v^i)^{-1}}$ and we can extend it by projection on the balls as we did with $L_{ui}$ in (26):

$$\begin{aligned}\overline{L}_{i(u,v)}(w) =& L_{i(u,v)}(\pi_{B_u}(w))(0 \vee (1 - 32K_X(\delta_u^i)^{-1}\|w - \pi_{B_u}(w)\|)) \\ &+ L_{i(u,v)}(\pi_{B_v}(w))(0 \vee (1 - 32K_X(\delta_v^i)^{-1}\|w - \pi_{B_v}(w)\|)).\end{aligned}$$

We also have $\overline{L}_{i(u,v)} \in \mathcal{H}^1_{C(\delta_u^i)^{-1}\vee(\delta_v^i)^{-1}}$. Then by splitting the wavelets coefficients of $\partial^{\lfloor\eta\rfloor}_{i_1,...,i_{\lfloor\eta\rfloor}} \tilde{\Gamma}_\epsilon^{-\eta+1,-2}(L_i)_r(u) - \partial^{\lfloor\eta\rfloor}_{i_1,...,i_{\lfloor\eta\rfloor}} \tilde{\Gamma}_\epsilon^{-\eta+1,-2}(L_i)_r(v)$ as in (27), we obtain (29). ■

Using Lemma 36, we obtain that the maps $H_i$ from (25) are of regularity $\eta$.

**Proposition 37** *For all $i \in \{1,...,m\}$, the application $H_i$ belongs to $\mathcal{H}^\eta_C(A,\mathbb{R})$.*

**Proof** Define

$$F_i := \phi_i^{-1} \circ T^{-1}_{|\mathcal{M}} \circ T \quad \text{and} \quad \sigma_i := \mathbb{1}_{\{T^{-1}_{|\mathcal{M}} \circ T(\cdot) \in \phi_{i|B^d(0,\tau)}\}} \zeta_i \circ F_i.$$

We have $F_i \in \mathcal{H}^{\beta+1}_C(A,\mathbb{R}^d)$. Since $\text{supp}(\zeta_i) \subset B^d(0,\tau)$, the map $\sigma_i$ belongs to $\mathcal{H}^{\beta+1}_C(A,\mathbb{R}^p)$. Furthermore, $(\text{Id} - T) \in \mathcal{H}^{\beta+1}_C(A,\mathbb{R}^p)$, and

$$\tilde{\Gamma}_\epsilon^{-\eta+1,-2}(L_i) \in \mathcal{H}^{\eta-1}_C(B^d(0,\tau),\mathbb{R}^p).$$

Thus, by the Faa di Bruno formula, $H_i \in \mathcal{H}^{\eta-1}_C(A,\mathbb{R})$. Let $i_1,...,i_{\lfloor\eta\rfloor} \in \{1,...,p\}$ and write $P(\{i_1,...,i_{\lfloor\eta\rfloor}\})$ the set of subsets of $\{i_1,...,i_{\lfloor\eta\rfloor}\}$. For $x \in A$, we have

$$\partial^{\lfloor\eta\rfloor}_{i_1,...,i_{\lfloor\eta\rfloor}} H_i(x) = \sum_{S \in P(\{i_1,...,i_{\lfloor\eta\rfloor}\})} \Big\langle \partial^{|S|}_{i\in S}(\tilde{\Gamma}_\epsilon^{-\eta+1,-2}(L_i) \circ F_i)(x)\ ,\ \partial^{\lfloor\eta\rfloor-|S|}_{i\notin S}((\text{Id} - T)\sigma_i)(x)\Big\rangle.$$

For all $S \in P(\{i_1,...,i_{\lfloor\eta\rfloor}\})$ with $|S| < \lfloor\eta\rfloor$, we have that

$$x \mapsto \Big\langle \partial^{|S|}_{i\in S}(\tilde{\Gamma}_\epsilon^{-\eta+1,-2}(L_i) \circ F_i)(x)\ ,\ \partial^{\lfloor\eta\rfloor-|S|}_{i\notin S}(\text{Id} - T)\sigma_i(x)\Big\rangle \in \mathcal{H}^{\eta-\lfloor\eta\rfloor}_C(A,\mathbb{R}).$$

From the previous calculations, we know that

$$\begin{aligned}|\partial^{\lfloor\eta\rfloor}_{i_1,\dots,i_{\lfloor\eta\rfloor}}\tilde{\Gamma}^{-\eta+1,-2}_{\epsilon}(L_i)(F_i(x))| &\leq CK_X\delta^{i-(\lfloor\eta\rfloor+1-\eta)}_{F_i(x)}\\ &= CK_X\|T^{-1}_{|\mathcal{M}}\circ T(x) - T(T^{-1}_{|\mathcal{M}}\circ T(x))\|^{-(\lfloor\eta\rfloor+1-\eta)}\\ &= CK_X\|T^{-1}_{|\mathcal{M}}\circ T(x) - T(x)\|^{-(\lfloor\eta\rfloor+1-\eta)}\\ &\leq C\|x-T(x)\|^{-(\lfloor\eta\rfloor+1-\eta)},\end{aligned}$$

so

$$|\Big\langle \partial^{\lfloor\eta\rfloor}_{i_1,\dots,i_{\lfloor\eta\rfloor}}(\tilde{\Gamma}^{-\eta+1,-2}_{\epsilon}(L_i)\circ F_i)(x)\ , (\mathrm{Id}-T)\sigma_i(x)\Big\rangle| \leq C\|x-T(x)\|^{\eta-\lfloor\eta\rfloor}\sigma_i(x). \tag{30}$$

Let $x, y \in A$ with $T^{-1}_{|\mathcal{M}}\circ T(x), T^{-1}_{|\mathcal{M}}\circ T(y) \in \phi_i(B^d(0,\tau))$ and $\|x-T(x)\| \leq \|y-T(y)\|$, from (29) we get

$$\begin{aligned}&|\Big\langle \partial^{\lfloor\eta\rfloor}_{i_1,\dots,i_{\lfloor\eta\rfloor}}(\tilde{\Gamma}^{-\eta+1,-2}_{\epsilon}(L_i)\circ F_i)(x)\ , (\mathrm{Id}-T)\sigma_i(x)\Big\rangle\\ &\quad - \Big\langle \partial^{\lfloor\eta\rfloor}_{i_1,\dots,i_{\lfloor\eta\rfloor}}(\tilde{\Gamma}^{-\eta+1,-2}_{\epsilon}(L_i)\circ F_i)(y)\ , (\mathrm{Id}-T)\sigma_i(y)\Big\rangle|\\ &\leq |\Big\langle \partial^{\lfloor\eta\rfloor}_{i_1,\dots,i_{\lfloor\eta\rfloor}}(\tilde{\Gamma}^{-\eta+1,-2}_{\epsilon}(L_i)\circ F_i)(x) - \partial^{\lfloor\eta\rfloor}_{i_1,\dots,i_{\lfloor\eta\rfloor}}(\tilde{\Gamma}^{-\eta+1,-2}_{\epsilon}(L_i)\circ F_i)(y)\ , (\mathrm{Id}-T)\sigma_i(x)\Big\rangle|\\ &\quad + |\Big\langle \partial^{\lfloor\eta\rfloor}_{i_1,\dots,i_{\lfloor\eta\rfloor}}(\tilde{\Gamma}^{-\eta+1,-2}_{\epsilon}(L_i)\circ F_i)(y)\ , (\mathrm{Id}-T)\sigma_i(x) - (\mathrm{Id}-T)\sigma_i(y)\Big\rangle|\\ &\leq C\min(\|x-T(x)\|, \|y-T(y)\|)^{-1}\|x-y\|^{\eta-\lfloor\eta\rfloor}\|x-T(x)\|\\ &\quad + C\|T^{-1}_{|\mathcal{M}}\circ T(y) - T(y)\|^{-(1-(\eta-\lfloor\eta\rfloor))}\|(\mathrm{Id}-T)\sigma_i(x) - (\mathrm{Id}-T)\sigma_i(y)\|\\ &\leq C\|x-y\|^{\eta-\lfloor\eta\rfloor} + C\|T^{-1}_{|\mathcal{M}}\circ T(y) - T(y)\|^{-(1-(\eta-\lfloor\eta\rfloor))}\|(\mathrm{Id}-T)\sigma_i(x) - (\mathrm{Id}-T)\sigma_i(y)\|.\end{aligned}$$

Furthermore,

$$\begin{aligned}&\|(\mathrm{Id}-T)\sigma_i(x) - (\mathrm{Id}-T)\sigma_i(y)\|\\ &\quad = \|(\mathrm{Id}-T)\sigma_i(x) - (\mathrm{Id}-T)\sigma_i(y)\|^{1-(\eta-\lfloor\eta\rfloor)}\|(\mathrm{Id}-T)\sigma_i(x) - (\mathrm{Id}-T)\sigma_i(y)\|^{\eta-\lfloor\eta\rfloor}\\ &\quad \leq (\|(\mathrm{Id}-T)(x)\| + \|(\mathrm{Id}-T)(y)\|)^{1-(\eta-\lfloor\eta\rfloor)}\|x-y\|^{\eta-\lfloor\eta\rfloor}\\ &\quad \leq 2\|(\mathrm{Id}-T)(y)\|^{1-(\eta-\lfloor\eta\rfloor)}\|x-y\|^{\eta-\lfloor\eta\rfloor}\\ &\quad \leq 2\|T^{-1}_{|\mathcal{M}}\circ T(y) - T(y)\|^{1-(\eta-\lfloor\eta\rfloor)}\|x-y\|^{\eta-\lfloor\eta\rfloor}.\end{aligned}$$

Now if $T^{-1}_{|\mathcal{M}}\circ T(x) \in \phi_i(B^d(0,\tau))$ and $T^{-1}_{|\mathcal{M}}\circ T(y) \notin \phi_i(B^d(0,\tau))$, we have from (30),

$$\begin{aligned}|\Big\langle \partial^{\lfloor\eta\rfloor}_{i_1,\dots,i_{\lfloor\eta\rfloor}}(\tilde{\Gamma}^{-\eta+1,-2}_{\epsilon}(L_i)\circ F_i)(x)\ , (\mathrm{Id}-T)\sigma_i(x)\Big\rangle| &\leq C\|x-T(x)\|^{\eta-\lfloor\eta\rfloor}\sigma_i(x)\\ &\leq C\sigma_i(x)\\ &= C(\sigma_i(x)-\sigma_i(y))\\ &\leq C\|x-y\|.\end{aligned}$$

Therefore we have finally that $H_i \in \mathcal{H}^{\eta}_C(A,\mathbb{R})$. ■

Let us now prove that we can extend $H$ to the whole $\mathbb{R}^p$ space using our extension result for

maps defined on the neighborhood of a submanifold (Proposition 32). Define $D : \mathcal{M}^{\star t} \to A$ by

$$D(x) = \begin{cases} x & \text{if } x \in A \\ \frac{x-T(x)}{\|x-T(x)\|}\|T^{-1}_{|\mathcal{M}} \circ T(x) - T(x)\| + T(x) & \text{if } x \notin A. \end{cases}$$

For $x \notin A$ we have

$$\begin{aligned}\nabla D(x) =& \frac{\|T^{-1}_{|\mathcal{M}} \circ T(x) - T(x)\|}{\|x - T(x)\|}\left(\nabla(\mathrm{Id} - T)(x) - \frac{x - T(x)}{\|x - T(x)\|}\frac{(x - T(x))^\top}{\|x - T(x)\|}\nabla(\mathrm{Id} - T)(x)\right) \\ &+ \frac{x - T(x)}{\|x - T(x)\|}\frac{(x - T(x))^\top}{\|x - T(x)\|}\nabla(T^{-1}_{|\mathcal{M}} \circ T - T)(x) + \nabla T(x),\end{aligned}$$

so $\|\nabla D(x)\| \leq C$. Then using Proposition 32, for all $i \in \{1, ..., m\}$, $H_i$ can be extended to a map in $\mathcal{H}^\eta_C(\mathbb{R}^p, \mathbb{R})$.

We can now conclude the proof of Lemma 10 using Proposition 35 as

$$\int_{\mathcal{M}} H_i(x)d\mu(x) \leq C \sup_{f \in \mathcal{H}^\eta_1,\ f_{|\mathcal{M}^\star}=0} \int_{\mathcal{M}} f(x)d\mu(x).$$

### C.3 Proof of Lemma 11

**Proof** Let $T$ be the $(\mathcal{M}, \mathcal{M}^\star)$-compatible map given by Theorem 9. As $T^{-1}_{|\mathcal{M}} \in \mathcal{H}^{\beta+1}_C(\mathcal{M}^\star, \mathcal{M})$, we have that $T_{\#\mu}$ admits a density $f_T \in \mathcal{H}^\beta_C(\mathcal{M}^\star, \mathbb{R})$ with respect to the volume measure on $\mathcal{M}^\star$. For $t \geq C^{-1}$ the radius of compatibility of the map $T$ (see Definition 7) define the $t$-envelope of $f_T$ and $f_{\mu^\star}$ as:

$$f^t_T(x) = \frac{f_T(T(x))\Theta(4\|x - T(x)\|^2/t^2)|\mathrm{ap}_d(\nabla T(x))|}{\int_{\mathbb{R}^{p-d}} \Theta(4\|y\|^2/t^2)d\lambda^{p-d}(y)}$$

and

$$f^t_{\mu^\star}(x) = \frac{f_{\mu^\star}(T(x))\Theta(4\|x - T(x)\|^2/t^2)|\mathrm{ap}_d(\nabla T(x))|}{\int_{\mathbb{R}^{p-d}} \Theta(4\|y\|^2/t^2)d\lambda^{p-d}(y)}$$

for $\mathrm{ap}_d$ the approximate jacobian defined in Definition 27 and $\Theta \in \mathcal{H}^\beta_C(\mathbb{R}, \mathbb{R}_+)$ such that $\Theta(x) = \Theta(-x)$, $\Theta(0) = 1$ and $\Theta_{|(1,\infty)} = 0$.

As $\nabla T \in \mathcal{H}^\beta_K(\mathbb{R}^p, L(\mathbb{R}^p, \mathbb{R}^p))$ and $\mathrm{ap}_d(\nabla T(x))$ is bounded below by $K_T^{-1}$ ($T$ verifes point (iv) of Definition 7), then from the Faa di Bruno formula, we deduce that $\mathrm{ap}_d(\nabla T)$ belongs to $\mathcal{H}^\beta_C(\mathcal{M}^{\star t}, \mathbb{R})$. Let $h \in \mathcal{H}^\gamma_K(\mathbb{R}^p, \mathbb{R})$ and define $\overline{h} : \mathcal{M}^{\star t} \to \mathbb{R}$ by

$$\overline{h}(x) = h(T(x))\kappa(2\|x - T(x)\|^2/t^2),$$

for $\kappa \in \mathcal{H}^\gamma_C(\mathbb{R}, \mathbb{R})$ such that $\kappa_{|(-\infty,1/2)} = 1$ and $\kappa_{|(1,\infty)} = 0$. Then $\overline{h} \in \mathcal{H}^\gamma_C(\mathbb{R}^p, \mathbb{R})$ and from Proposition 28 we have,

$$\begin{aligned}
&\int_{\mathcal{M}} h(T(x))d\mu(x) - \int_{\mathcal{M}^\star} h(x)d\mu^\star(x) = \int_{\mathcal{M}^\star} h(x)(f_T(x) - f_{\mu^\star}(x))d\lambda_{\mathcal{M}^\star}(x) \\
&= \int_{\mathcal{M}^\star} \frac{h(x)(f_T(x) - f_{\mu^\star}(x))}{\int_{\mathbb{R}^{p-d}} \Theta(4\|y\|^2/t^2)d\lambda^{p-d}(y)} \int_{T^{-1}(\{x\})} \Theta(4\|z - T(z)\|^2/t^2)d\lambda^{p-d}(z)d\lambda_{\mathcal{M}^\star}(x) \\
&= \int_{\mathcal{M}^\star} \int_{T^{-1}(\{x\})} \frac{h(T(z))(f_T(T(z)) - f_{\mu^\star}(T(z)))\Theta(4\|z - T(z)\|^2/t^2)}{\int_{\mathbb{R}^{p-d}} \Theta(4\|y\|^2/t^2)d\lambda^{p-d}(y)} d\lambda^{p-d}(z)d\lambda_{\mathcal{M}^\star}(x) \\
&= \int_{\mathbb{R}^p} h(T(x))(f_T^t(T(x)) - f_{\mu^\star}^t(T(x)))d\lambda^p(x) \\
&= \int_{\mathbb{R}^p} \overline{h}(x)(f_T^t(x) - f_{\mu^\star}^t(x))d\lambda^p(x).
\end{aligned}$$

Let

$$\tilde{\Gamma}_\epsilon^{0,-2}(\overline{h})(x) = \sum_{j=0}^{\log(\lfloor \epsilon^{-1} \rfloor)} \sum_{l=1}^{2^d} \sum_{w \in \mathbb{Z}^p} (1+j)^{-2} S(j,l,w) \alpha_{\overline{h}}(j,l,w) \psi_{jlw}(x)$$

for $S(j,l,w) = \frac{\alpha_{\overline{h}}(j,l,w)(\alpha_{f_T^t}(j,l,w) - \alpha_{f_{\mu^\star}^t}(j,l,w))}{\alpha_{\overline{h}}(j,l,w)(\alpha_{f_T^t}(j,l,w) - \alpha_{f_{\mu^\star}^t}(j,l,w))|}$ . Then using Proposition 33, we have

$$\int_{\mathbb{R}^p} \overline{h}(x)(f_T^t(x) - f_{\mu^\star}^t(x))d\lambda^p(x) \leq C \log(\epsilon^{-1})^2 \int_{\mathbb{R}^p} \tilde{\Gamma}_\epsilon^{0,-2}(\overline{h})(f_T^t(x) - f_{\mu^\star}^t(x))d\lambda^p(x) + C\epsilon.$$

Applying Proposition 25 for $h_1 = \tilde{\Gamma}_\epsilon^{0,-2}(\overline{h})$, $h_2 = f_T^t - f_{\mu^\star}^t$, $s_1 = \gamma$, $b_1 = 2$, $s_2 = \beta$, $b_2 = 0$, $\tau = \gamma - \eta$, $t = 1$, $r = \frac{\gamma}{\beta+\gamma}$ and $q = \frac{\beta+\eta}{\beta+\gamma}$ we get

$$\begin{aligned}
&\int_{\mathbb{R}^p} \overline{h}^\epsilon(x)(f_T^t(x) - f_{\mu^\star}^t(x))d\lambda^p(x) \\
&\quad \leq \left\langle \tilde{\Gamma}_\epsilon^{\gamma-\eta,-2}(\overline{h}), f_T^t - f_{\mu^\star}^t \right\rangle_{L^2(\mathbb{R}^p)}^{\frac{\beta+\gamma}{\beta+\eta}} \left\langle \tilde{\Gamma}_\epsilon^{\gamma,-2}(\overline{h}), \Gamma^\beta(f_T^t - f_{\mu^\star}^t) \right\rangle_{L^2(\mathbb{R}^p)}^{\frac{\eta-\gamma}{\beta+\eta}}
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{j=0}^{\infty} \sum_{l=1}^{2^p} \sum_{w \in \mathbb{Z}^p} 2^{j(\beta+\gamma)} |\alpha_{\tilde{\Gamma}_\epsilon^{0,-2}(\overline{h})}(j,l,w)||\alpha_{f_T}(j,l,w) - \alpha_{f_{\mu^\star}}(j,l,w)| \\
&\qquad \leq \sum_{j=0}^{\infty} \sum_{l=1}^{2^d} \sum_{w \in \mathbb{Z}^p} (1+j)^{-2} C 2^{-jp} \mathbb{1}_{\{supp(\psi_{jlw} \cap B^p(0,K)) \neq \varnothing\}} \\
&\qquad \leq C \sum_{j=0}^{\infty} (1+j)^{-2} \leq C.
\end{aligned}$$

Define $F_h : \mathcal{M}^\star \to \mathbb{R}$ by

$$F_h(x) = \int_{T^{-1}(\{x\})} \tilde{\Gamma}_\epsilon^{\gamma-\eta,-2}(\overline{h})(z) \ \frac{\Theta(4\|z - T(z)\|^2/t^2)}{\int_{\mathbb{R}^{p-d}} \Theta(4\|y\|^2/t^2)d\lambda^{p-d}(y)} d\lambda_{T^{-1}(\{x\})}(z).$$

Then

$$\begin{aligned}
&\int_{\mathbb{R}^p} \tilde{\Gamma}_\epsilon^{\gamma-\eta,-2}(\overline{h})(x)(f_T^t(x)-f_{\mu^\star}^t(x))d\lambda^p(x)\\
&=\int_{\mathcal{M}^\star}\int_{T^{-1}(\{x\})} \tilde{\Gamma}_\epsilon^{\gamma-\eta,-2}(\overline{h})(z)(f_T^t(z)-f_{\mu^\star}^t(z))|\mathrm{ap}_d(\nabla T(z))|^{-1}d\lambda^{p-d}(z)d\lambda_{\mathcal{M}^\star}(x)\\
&=\int_{\mathcal{M}^\star}\int_{T^{-1}(\{x\})} \tilde{\Gamma}_\epsilon^{\gamma-\eta,-2}(\overline{h})(z)\frac{(f_T(T(z))-f_{\mu^\star}(T(z)))\Theta(4\frac{\|z-T(z)\|^2}{t^2})}{\int_{\mathbb{R}^{p-d}}\Theta(4\|y\|^2/t^2)d\lambda^{p-d}(y)}d\lambda^{p-d}(z)d\lambda_{\mathcal{M}^\star}(x)\\
&=\int_{\mathcal{M}^\star} F_h(x)(f_T(x)-f_{\mu^\star}(x))d\lambda_{\mathcal{M}^\star}(x).
\end{aligned}$$

As $T$ verifies point iii) of Definition 7, we have $F_h \in \mathcal{H}_C^\eta(\mathcal{M}^\star,\mathbb{R})$ so it can be extended using Proposition 32 to a map in $\mathcal{H}_C^\eta(\mathbb{R}^p,\mathbb{R})$. Then, we can conclude that

$$\int_{\mathcal{M}^\star} h(x)(f_T(x)-f_{\mu^\star}(x))d\lambda_{\mathcal{M}^\star}(x) \leq C\sup_{h\in\mathcal{H}_1^\eta}\left(\int_{\mathcal{M}^\star} h(x)(f_T(x)-f_{\mu^\star}(x))d\lambda_{\mathcal{M}^\star}(x)\right)^{\frac{\beta+\gamma}{\beta+\eta}}.$$

■

### C.4 Proof of Theorem 12

**Proof** Let $C_\eta$ the constant given by Theorem 9 such that if

$$d_{\mathcal{H}_1^\eta}(\mu,\mu^\star) \leq C_\eta^{-1},$$

then there exists a map $T$ being $(\mathcal{M},\mathcal{M}^\star)$-compatible.

Suppose first that $d_{\mathcal{H}_1^\eta}(\mu,\mu^\star) > C_\eta^{-1}$, then

$$d_{\mathcal{H}_1^\gamma}(\mu,\mu^\star) \leq 2K \leq 2KC_\eta d_{\mathcal{H}_1^\eta}(\mu,\mu^\star) \leq C d_{\mathcal{H}_1^\eta}(\mu,\mu^\star)^{\frac{\beta+\gamma}{\beta+\eta}}.$$

Suppose now that $d_{\mathcal{H}_1^\eta}(\mu,\mu^\star) \leq C_\eta^{-1}$. Using Corollary 26 we have for all $\epsilon\in(0,1)$

$$d_{\mathcal{H}_1^\gamma}(\mu,\mu^\star) \leq C\log(\epsilon^{-1})^2 d_{\mathcal{H}_1^1}(\mu,\mu^\star)^{\frac{\eta-\gamma}{\eta-1}} d_{\mathcal{H}_1^\eta}(\mu,\mu^\star)^{\frac{\gamma-1}{\eta-1}} + \epsilon. \tag{31}$$

Suppose first that $\eta \leq \beta+1$. Using lemmas 10 and 11 we have

$$d_{\mathcal{H}_1^1}(\mu,\mu^\star) \leq C\log(\epsilon^{-1})^4\left(\sup_{\substack{h\in\mathcal{H}_1^\eta\\ h_{|\mathcal{M}^\star}=0}}\left(\int_{\mathcal{M}} h(x)d\mu(x)\right)^{\frac{\beta+1}{\beta+\eta}} + d_{\mathcal{H}_1^\eta}(T_{\#\mu},\mu^\star)^{\frac{\beta+1}{\beta+\eta}}\right)+\epsilon.$$

On one hand we have

$$\sup_{\substack{h\in\mathcal{H}_1^\eta\\ h_{|\mathcal{M}^\star}=0}}\int_{\mathcal{M}} h(x)d\mu(x) \leq \sup_{h\in\mathcal{H}_1^\eta}\int_{\mathcal{M}} h(x)d\mu(x) - \int_{\mathcal{M}^\star} h(x)d\mu^\star(x) = d_{\mathcal{H}_1^\eta}(\mu,\mu^\star),$$

on the other hand

$$d_{\mathcal{H}_1^\eta}(T_{\#\mu},\mu^\star)=\sup_{h\in\mathcal{H}_1^\eta}\int_{\mathcal{M}}h(T(x))d\mu(x)-\int_{\mathcal{M}^\star}h(T(x))d\mu^\star(x)\leq Cd_{\mathcal{H}_1^\eta}(\mu,\mu^\star),$$

as any map $h\circ T\in\mathcal{H}_C^\eta(\mathcal{M}^{\star t/4},\mathbb{R})$ can be extended to a map in $\mathcal{H}_C^\eta(\mathbb{R}^p,\mathbb{R})$ using Proposition 32. Therefore, we have

$$d_{\mathcal{H}_1^1}(\mu,\mu^\star)\leq C\log(\epsilon^{-1})^4 d_{\mathcal{H}_1^\eta}(\mu,\mu^\star)^{\frac{\beta+1}{\beta+\eta}}+\epsilon.$$

Then plugging this in (31) we obtain

$$\begin{aligned}d_{\mathcal{H}_1^\gamma}(\mu,\mu^\star)&\leq C\log(\epsilon^{-1})^6 d_{\mathcal{H}_1^\eta}(\mu,\mu^\star)^{\frac{\beta+1}{\beta+\eta}\frac{\eta-\gamma}{\eta-1}+\frac{\gamma-1}{\eta-1}}+\epsilon\\&=C\log(\epsilon^{-1})^6 d_{\mathcal{H}_1^\eta}(\mu,\mu^\star)^{\frac{\beta+\gamma}{\beta+\eta}}+\epsilon\end{aligned}$$

so taking $\epsilon=d_{\mathcal{H}_1^\eta}(\mu,\mu^\star)^{\frac{\beta+\gamma}{\beta+\eta}}$, we get the result.

Now if $\eta>\beta+1$, using lemmas 10 and 11 we have again

$$\begin{aligned}d_{\mathcal{H}_1^1}(\mu,\mu^\star)&\leq d_{\mathcal{H}_1^1}(\mu,T_{\#\mu})+d_{\mathcal{H}_1^1}(T_{\#\mu},\mu^\star)\\&\leq C\log(\epsilon^{-1})^4 d_{\mathcal{H}_1^{\beta+1}}(\mu,\mu^\star)^{\frac{\beta+1}{2\beta+1}}+\epsilon.\end{aligned}$$

And using again Corollary 26 we have

$$d_{\mathcal{H}_1^{\beta+1}}(\mu,\mu^\star)\leq C\log(\epsilon^{-1})^2 d_{\mathcal{H}_1^1}(\mu,\mu^\star)^{\frac{\eta-(\beta+1)}{\eta-1}}d_{\mathcal{H}_1^\eta}(\mu,\mu^\star)^{\frac{\beta}{\eta-1}}+\epsilon$$

so

$$d_{\mathcal{H}_1^1}(\mu,\mu^\star)\leq C\log(\epsilon^{-1})^6 d_{\mathcal{H}_1^1}(\mu,\mu^\star)^{\frac{\eta-(\beta+1)}{\eta-1}\frac{\beta+1}{2\beta+1}}d_{\mathcal{H}_1^\eta}(\mu,\mu^\star)^{\frac{\beta}{\eta-1}\frac{\beta+1}{2\beta+1}}+\epsilon.$$

Then, taking $\epsilon=d_{\mathcal{H}_1^1}(\mu,\mu^\star)d_{\mathcal{H}_1^\eta}(\mu,\mu^\star)$, we have

$$d_{\mathcal{H}_1^1}(\mu,\mu^\star)\leq C\log(d_{\mathcal{H}_1^\eta}(\mu,\mu^\star)^{-1})^6 d_{\mathcal{H}_1^1}(\mu,\mu^\star)^{\frac{\eta-(\beta+1)}{\eta-1}\frac{\beta+1}{2\beta+1}}d_{\mathcal{H}_1^\eta}(\mu,\mu^\star)^{\frac{\beta}{\eta-1}\frac{\beta+1}{2\beta+1}},$$

which gives

$$d_{\mathcal{H}_1^1}(\mu,\mu^\star)^{(\eta-1)(2\beta+1)-(\eta-1-\beta)(\beta+1)}\leq C\log(d_{\mathcal{H}_1^\eta}(\mu,\mu^\star)^{-1})^{6(\eta-1)(2\beta+1)}d_{\mathcal{H}_1^\eta}(\mu,\mu^\star)^{\beta(\beta+1)}.$$

As $(\eta-1)(2\beta+1)-(\eta-1-\beta)(\beta+1))=\beta(\eta+\beta)$ we obtain

$$d_{\mathcal{H}_1^1}(\mu,\mu^\star)\leq C\log(d_{\mathcal{H}_1^\eta}(\mu,\mu^\star)^{-1})^{6\frac{(\eta-1)(2\beta+1)}{\beta(\eta+\beta)}}d_{\mathcal{H}_1^\eta}(\mu,\mu^\star)^{\frac{\beta+1}{\beta+\eta}}.$$

Then plugging this result in (31), we obtain for $\epsilon=d_{\mathcal{H}_1^\eta}(\mu,\mu^\star)$,

$$\begin{aligned}d_{\mathcal{H}_1^\gamma}(\mu,\mu^\star)&\leq C\log(d_{\mathcal{H}_1^\eta}(\mu,\mu^\star)^{-1})^{C_2}d_{\mathcal{H}_1^\eta}(\mu,\mu^\star)^{\frac{\beta+1}{\beta+\eta}\frac{\eta-\gamma}{\eta-1}+\frac{\gamma-1}{\eta-1}}\\&=C\log(d_{\mathcal{H}_1^\eta}(\mu,\mu^\star)^{-1})^{C_2}d_{\mathcal{H}_1^\eta}(\mu,\mu^\star)^{\frac{\beta+\gamma}{\beta+\eta}}.\end{aligned}$$

■

### C.5 Proof of Proposition 13

**Proof** As explained at the beginning of the proof of Theorem 12, we can suppose that the distance $d_{\mathcal{H}_1^\eta}(\mu, \mu^\star)$ is small enough so that there exists a $(\mathcal{M}, \mathcal{M}^\star)$-compatible map $T$. We have

$$d_{\mathcal{H}_1^\gamma}(\mu, \mu^\star) \leq d_{\mathcal{H}_1^\gamma}(\mu, T_{\#\mu}) + d_{\mathcal{H}_1^\gamma}(T_{\#\mu}, \mu^\star)$$

Using Lemma 11 we obtain

$$\begin{aligned} d_{\mathcal{H}_1^\gamma}(T_{\#\mu}, \mu^\star) &\leq C \log(\epsilon^{-1})^2 d_{\mathcal{H}_1^\eta}(T_{\#\mu}, \mu^\star)^{\frac{\beta+\gamma}{\beta+\eta}} + \epsilon \\ &\leq C d_{\mathcal{H}_1^\eta}(T_{\#\mu}, \mu^\star)^{\frac{\gamma}{\eta}} \end{aligned}$$

for $\epsilon = d_{\mathcal{H}_1^\eta}(T_{\#\mu}, \mu^\star)^{\frac{\gamma}{\eta}}$. On the other hand,

$$\begin{aligned} d_{\mathcal{H}_1^\gamma}(\mu, T_{\#\mu}) &= \sup_{h \in \mathcal{H}_1^\gamma} \int_{\mathcal{M}} (h(x) - h(T(x))) f_\mu(x) d\lambda_{\mathcal{M}}(x) \\ &\leq \int_{\mathcal{M}} \|x - T(x)\|^\gamma f_\mu(x) d\lambda_{\mathcal{M}}(x) \\ &\leq \left( \int_{\mathcal{M}} \|x - T(x)\|^\eta f_\mu(x) d\lambda_{\mathcal{M}}(x) \right)^{\frac{\gamma}{\eta}} \end{aligned}$$

using Jensen's inequality. The function $H : \mathcal{M}^{\star t} \to \mathbb{R}$ defined by

$$H(x) = \|x - T(x)\|^\eta,$$

belongs to $\mathcal{H}_C^\eta(\mathcal{M}^{\star t}, \mathbb{R})$ and can therefore be extend to a map in $\mathcal{H}_C^\eta(\mathbb{R}^p, \mathbb{R})$ using Proposition 32.

■

### C.6 Proof of Proposition 15

**Proof** Let $s \in (0, r_{\mathcal{M}^\star})$, we are going to show that $W_1(\mu, \mu^\star)$ is bounded below which will give us that $d_{\mathcal{H}_1^\eta}(\mu, \mu^\star)$ is also bounded below by Lemma 8.

Suppose first that $\mathbb{H}(\mathcal{M}, \mathcal{M}^\star) > s$ and for $y \in \mathcal{M}^\star$ such that $d(y, \mathcal{M}) \geq s$, define $D_s^y : \mathbb{R}^p \to \mathbb{R}$ as

$$D_s^y(x) = (s - \|x - y\|) \vee 0.$$

As $D_s$ is 1-Lipschitz, we have

$$\begin{aligned} W_1(\mu, \mu^\star) &\geq \left( \int_{\mathcal{M}^\star} D_s^y(x) f_{\mu^\star}(x) d\lambda_{\mathcal{M}^\star}(x) - \int_{\mathcal{M}} D_s^y(x) f_\mu(x) d\lambda_{\mathcal{M}}(x) \right) \\ &= \int_{\mathcal{M}^\star \cap B^p(y,s)} (s - \|x - y\|) f_{\mu^\star}(x) d\lambda_{\mathcal{M}^\star} \\ &\geq C s^d K^{-1} \end{aligned}$$

which gives the result. Now if $\mathbb{H}(\mathcal{M}, \mathcal{M}^\star) \leq s$, by Lemma 29 we have that for $s > 0$ small enough, $\pi$ the canonical projection onto $\mathcal{M}^\star$, is a diffeomorphism from $\mathcal{M}$ to $\mathcal{M}^\star$. Le be $z \in \mathcal{M}$ such that $f_\mu(z) \leq s$. Define $D_s^z : \mathbb{R}^p \to \mathbb{R}$ as

$$D_s(x) = (\frac{s}{2} \wedge (s - \|x - \pi(z)\|)) \vee 0.$$

As $D_s$ is 1-Lipschitz, we have

$$W_1(\mu, \mu^\star) \geq \int_{\mathcal{M}^\star} D_s^z(x) f_{\mu^\star}(x) d\lambda_{\mathcal{M}^\star}(x) - \int_{\mathcal{M}} D_s^z(x) f_\mu(x) d\lambda_{\mathcal{M}}(x)$$

and

$$\begin{aligned}
\int_{\mathcal{M}^\star} D_s^z(x) f_{\mu^\star}(x) d\lambda_{\mathcal{M}^\star}(x) &= \int_{\mathcal{M}^\star \cap B^p(\pi(z), s)} \frac{s}{2} \wedge (s - \|x - \pi(z)\|) f_{\mu^\star}(x) d\lambda_{\mathcal{M}^\star}(x) \\
&\geq \int_{\mathcal{M}^\star \cap B^p(\pi(z), s/2)} \frac{s}{2} f_{\mu^\star}(x) d\lambda_{\mathcal{M}^\star}(x) \\
&\geq C s^{d+1} K^{-1}.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\int_{\mathcal{M}} D_s^z(x) f_\mu(x) d\lambda_{\mathcal{M}}(x) &= \int_{\mathcal{M} \cap B^p(\pi(z), s)} \frac{s}{2} \wedge (s - \|x - \pi(z)\|) f_\mu(x) d\lambda_{\mathcal{M}}(x) \\
&\leq \int_{\mathcal{M} \cap B^p(\pi(z), s)} s f_\mu(x) d\lambda_{\mathcal{M}}(x) \\
&= s \int_{\varphi_x(\mathcal{M} \cap B^p(\pi(z), s))} f_\mu(\varphi_x^{-1}(u)) d\varphi_{x\#\lambda_{\mathcal{M}}}(u) \\
&\leq s \int_{B^d(\varphi_x(z), 2s)} f_\mu(\varphi_x^{-1}(u)) K^d d\lambda^d(u) \\
&\leq Cs \int_{B^d(\varphi_x(\pi(z)), 2s)} (f_\mu(z) + K\|\varphi_x^{-1}(z) - u\|) d\lambda^d(u) \\
&\leq Cs \int_{B^d(\varphi_x(\pi(z)), 2s)} (1 + 2K) s d\lambda^d(u) \\
&\leq C s^{d+2}.
\end{aligned}$$

Therefore we have

$$W_1(\mu, \mu^\star) \geq C s^{d+1}(1 - C_2 s),$$

so for $s$ small enough we get the result. ■

## C.7 Additional Proofs from Section 3.1

### C.7.1 Proof of Proposition 4

**Proof** Take two densities $f, f^\star \in L^2(\mathbb{R}^p, \mathbb{R})$ and a potential $h \in \mathcal{B}^s_{\infty,\infty}(\mathbb{R}^p, \mathbb{R})$ to compare them. The family of wavelets functions $(\psi_{jlw})_{jlw}$ being an orthonormal basis of $L^2$, we can

write for $\tau > 0$ and $q \in (0,1), q^\star = (1-1/q)^{-1}$,

$$\int_{\mathbb{R}^p} h(x)(f(x)-f^\star(x))d\lambda^p(x) = \sum_{j=0}^{\infty}\sum_{l=1}^{2^p}\sum_{w\in\mathbb{Z}^p} \alpha_h(j,l,w)(\alpha_f(j,l,w)-\alpha_{f^\star}(j,l,w))$$
$$\leq \sum_{j=0}^{\infty}\sum_{l=1}^{2^p}\sum_{w\in\mathbb{Z}^p} 2^{-j\frac{\tau}{q}+j\frac{\tau}{q}}|\alpha_h(j,l,w)(\alpha_f(j,l,w)-\alpha_{f^\star}(j,l,w))|^{1/q+1/q^\star}.$$

Then, applying Hölder's inequality we get

$$\int_{\mathbb{R}^p} h(x)(f(x)-f^\star(x))d\lambda^p(x)$$
$$\leq \left(\sum_{j=0}^{\infty}\sum_{l=1}^{2^p}\sum_{w\in\mathbb{Z}^p} 2^{-j\tau}|\alpha_h(j,l,w)(\alpha_f(j,l,w)-\alpha_{f^\star}(j,l,w))|\right)^{\frac{1}{q}}$$
$$\times \left(\sum_{j=0}^{\infty}\sum_{l=1}^{2^p}\sum_{w\in\mathbb{Z}^p} 2^{j\tau\frac{q^\star}{q}}|\alpha_h(j,l,w)(\alpha_f(j,l,w)-\alpha_{f^\star}(j,l,w))|\right)^{\frac{1}{q^\star}}. \tag{32}$$

Let us take $\tilde{\Gamma}^{-\tau}(h) \in \mathcal{B}^{s+\tau}_{\infty,\infty}$ and $\tilde{\Gamma}^{\frac{q^\star}{q}\tau}(h) \in \mathcal{B}^{s-\frac{q^\star}{q}\tau}_{\infty,\infty}$ defined through their wavelet coefficients:

$$\alpha_{\tilde{\Gamma}^{-\tau}(h)}(j,l,w) = S(j,l,w)\alpha_{\Gamma^{-\tau}(h)}(j,l,w)$$

and

$$\alpha_{\tilde{\Gamma}^{\frac{q^\star}{q}\tau}(h)}(j,l,w) = S(j,l,w)\alpha_{\Gamma^{\frac{q^\star}{q}\tau}(h)}(j,l,w)$$

with $S(j,l,w) \in \{-1,1\}$ the sign of $\alpha_h(j,l,w)(\alpha_f(j,l,w)-\alpha_{f^\star}(j,l,w))$. Then, from (32) we obtain

$$\int_{\mathbb{R}^p} h(x)(f(x)-f^\star(x))d\lambda^p(x)$$
$$\leq \left(\int_{\mathbb{R}^p} \tilde{\Gamma}^{-\tau}(h)(x)(f(x)-f^\star(x))d\lambda^p(x)\right)^{\frac{1}{q}} \left(\int_{\mathbb{R}^p} \tilde{\Gamma}^{\frac{q^\star}{q}\tau}(h)(x)(f(x)-f^\star(x))d\lambda^p(x)\right)^{\frac{1}{q^\star}},$$

which gives a first interpolation inequality between Besov IPMs:

$$d_{\mathcal{B}^s_{\infty,\infty}}(f,f^\star) \leq d_{\mathcal{B}^{s+\tau}_{\infty,\infty}}(f,f^\star)^{\frac{1}{q}}\, d_{\mathcal{B}^{s-\frac{q^\star}{q}\tau}_{\infty,\infty}}(f,f^\star)^{\frac{1}{q^\star}}. \tag{33}$$

Using (33), we are looking for the smallest $q \in (0,1)$ such that $d_{\mathcal{B}^{s-\frac{q^\star}{q}\tau}_{\infty,\infty}}(f,f^\star)$ is finite. Taking $f, f^\star \in \mathcal{B}^{\beta,2}_{\infty,\infty}$ compactly supported in $B^p(0,K)$ we have

$$
\begin{aligned}
&d_{\mathcal{B}_{\infty,\infty}^{s-\frac{q^\star}{q}\tau}}(f, f^\star) \\
&= \sup_{h \in \mathcal{B}_{\infty,\infty}^{s-\frac{q^\star}{q}\tau}(1)} \sum_{j=0}^{\infty} \sum_{l=1}^{2^p} \sum_{w \in \mathbb{Z}^p} \alpha_h(j,l,w)(\alpha_f(j,l,w) - \alpha_{f^\star}(j,l,w)) \\
&= \sum_{j=0}^{\infty} \sum_{l=1}^{2^p} \sum_{w \in \mathbb{Z}^p} 2^{-j(s-\frac{q^\star}{q}\tau+p/2)} \alpha_h(j,l,w) |\alpha_f(j,l,w) - \alpha_{f^\star}(j,l,w)| \\
&\leq \sum_{j=0}^{\infty} \sum_{l=1}^{2^p} \sum_{w \in \mathbb{Z}^p} 2^{-j(s-\frac{q^\star}{q}\tau+p/2)} 2^{-j(\beta+p/2)} (1+j)^{-2} \|f - f^\star\|_{\mathcal{B}_{\infty,\infty}^{\beta,2}} \mathbb{1}_{\{supp(\psi_{jlw}) \cap B^p(0,K) \neq \varnothing\}} \\
&\leq C \|f - f^\star\|_{\mathcal{B}_{\infty,\infty}^{\beta,2}} \sum_{j=0}^{\infty} 2^{-j(s-\frac{q^\star}{q}\tau+\beta)} (1+j)^{-2}.
\end{aligned}
$$

The last quantity is finite iff $\beta \geq \frac{q^\star}{q}\tau - s$ which gives $\frac{1}{q} \leq \frac{\beta+s}{\beta+s+\tau}$. ■

### C.7.2 Proof of Theorem 5

**Proof** Using lemma 3 we have that $f - f^\star \in \mathcal{B}_{\infty,\infty}^{\beta}(C)$ and $h \in \mathcal{B}_{\infty,\infty}^{\gamma}(C)$, then for $\epsilon \in (0,1)$ we have

$$
\begin{aligned}
&\int_{\mathbb{R}^p} h(x)(f(x) - f^\star(x)) d\lambda^p(x) \\
&= \sum_{j=0}^{\infty} \sum_{l=1}^{2^p} \sum_{w \in \mathbb{Z}^p} \alpha_h(j,l,w)(\alpha_f(j,l,w) - \alpha_{f^\star}(j,l,w)) \\
&\leq \sum_{j=0}^{\log_2(\epsilon^{-1})} \sum_{l=1}^{2^p} \sum_{w \in \mathbb{Z}^p} \alpha_h(j,l,w)(\alpha_f(j,l,w) - \alpha_{f^\star}(j,l,w)) \\
&\quad + \sum_{j=\log_2(\epsilon^{-1})+1}^{\infty} \sum_{l=1}^{2^p} \sum_{w \in \mathbb{Z}^p} 2^{-j(\gamma+\beta+p)} \mathbb{1}_{\{supp(\psi_{jlw}) \cap B^p(0,K) \neq \varnothing\}} \\
&\leq \log_2(\epsilon^{-1})^2 \sum_{j=0}^{\log_2(\epsilon^{-1})} \sum_{l=1}^{2^p} \sum_{w \in \mathbb{Z}^p} (1+j)^{-2} |\alpha_h(j,l,w)(\alpha_f(j,l,w) - \alpha_{f^\star}(j,l,w))| \\
&\quad + C\epsilon^{\gamma+\beta} \\
&\leq C \log_2(\epsilon^{-1})^2 d_{\mathcal{B}_{\infty,\infty}^{\gamma}(1)}(\Gamma^{0,-2}(f), \Gamma^{0,-2}(f^\star)) + C\epsilon^{\gamma+\beta}.
\end{aligned}
$$

As $f, f^\star \in \mathcal{B}^{\beta,2}_{\infty,\infty}(C)$, we can now use Proposition 4 and obtain

$$\begin{aligned} d_{\mathcal{B}^{\gamma}_{\infty,\infty}(1)}(\Gamma^{0,-2}(f),\Gamma^{0,-2}(f^\star)) &\leq C d_{\mathcal{B}^{\alpha}_{\infty,\infty}(1)}(\Gamma^{0,-2}(f),\Gamma^{0,-2}(f^\star))^{\frac{\beta+\gamma}{\beta+\alpha}} \\ &= d_{\mathcal{B}^{\gamma,2}_{\infty,\infty}(1)}(f,f^\star)^{\frac{\beta+\gamma}{\beta+\alpha}} \\ &\leq C d_{\mathcal{H}^{\gamma}_{C}}(f,f^\star)^{\frac{\beta+\gamma}{\beta+\alpha}}, \end{aligned}$$

where we used Lemma 3 for the last inequality. Putting everything together and taking $\epsilon = d_{\mathcal{H}^{\alpha}_1}(f,f^\star) \wedge 1/2$ we obtain the result. ■

### C.8 Proof of Theorem 17

#### C.8.1 Definition of the Manifolds and Measures

Let $n, \beta \in \mathbb{N}$ and $g_n, g^\star : [0,1] \to \mathbb{R}^2$ defined by

$$g^\star(t) := (\cos(2\pi t), \sin(2\pi t)) \text{ and } g_n(t) := (1 + (2\pi n)^{-(\beta+1)} \sin(2\pi n t)) g^\star(t). \tag{34}$$

Define the two submanifolds $\mathcal{M}_n := g_n([0,1]), \mathcal{M}^\star := g^\star([0,1])$ and $\mu_n$ the uniform measure on $\mathcal{M}_n$. The manifolds $\mathcal{M}_n$ and $\mathcal{M}^\star$ are represented in Figure 1. We choose the $(\mathcal{M}_n, \mathcal{M}^\star)$-compatible map $T$ to be the projection onto $\mathcal{M}^\star$ for simplicity of the derivations.

We have that $\mathcal{M}_n$ verifies the $(\beta+1, C)$-manifold condition and for all $\alpha > \beta + 1$, we have $\|g_n\|_{\mathcal{H}^\alpha} \geq C^{-1} n^{\alpha-(\beta+1)}$. Therefore this example is well suited to represent the case of $\beta+1$ regular manifolds.

#### C.8.2 Expected Shape of Optimal Potentials

To get some intuition let us analyze the behavior of $d_{\mathcal{H}^{\gamma}_1}(\mu_n, T_{\#\mu_n})$ for $\gamma = 1$. We have

$$\begin{aligned} d_{\mathcal{H}^1_1}(\mu_n, T_{\#\mu_n}) &= \max_{h \in \mathcal{H}^1_1} \int_{\mathcal{M}_n} (h(x) - h(T(x)) f_{\mu_n}(x) d\lambda_{\mathcal{M}_n}(x) \\ &\leq \int_{\mathcal{M}_n} \|x - T(x)\| f_{\mu_n}(x) d\lambda_{\mathcal{M}_n}(x) \\ &= \int_{\mathcal{M}^\star} \|y - T^{-1}_{|\mathcal{M}_n}(y)\| f_{\mu_n}(T^{-1}_{|\mathcal{M}_n}(y)) \mathrm{ap}_d(\nabla T(T^{-1}_{|\mathcal{M}_n}(y)))^{-1} d\lambda_{\mathcal{M}^\star}(y) \\ &\leq C \int_{\mathcal{M}^\star} \|y - T^{-1}_{|\mathcal{M}_n}(y)\| d\lambda_{\mathcal{M}^\star}(y) \\ &\leq C \int_0^1 n^{-(\beta+1)} |\sin(2\pi n t)| dt \\ &\leq C n^{-(\beta+1)}. \end{aligned}$$

Note that using the same derivations, we can show that $d_{\mathcal{H}^1_1}(\mu_n, T_{\#\mu_n}) \geq C^{-1} n^{-(\beta+1)}$ so $d_{\mathcal{H}^1_1}(\mu_n, T_{\#\mu_n})$ behaves like $O(n^{-(\beta+1)})$. To build an optimal potential in $\mathcal{H}^\eta_1$ with $\eta \geq 1$, let us look at how to regularize the optimal Lipschitz potential

$$L^\star_n := \|x - T(x)\| \in \mathrm{Lip}_1$$

such that it belongs to $\mathcal{H}^\eta$ for $\eta \geq 1$. A simple way could be to use Jensen's inequality:

$$\int_{\mathcal{M}_n} \|x - T(x)\| f_{\mu_n}(x) d\lambda_{\mathcal{M}_n}(x) \leq \left( \int_{\mathcal{M}_n} \|x - T(x)\|^2 f_{\mu_n}(x) d\lambda_{\mathcal{M}_n}(x) \right)^{\frac{1}{2}}. \tag{35}$$

We have that $\overline{L_n^\star}(x) := \|x - T(x)\|^2$ belongs to $\mathcal{H}^{\beta+1}$ but is not an optimal potential as its cost is too low. Indeed, we obtain an exponent $1/2$ in (35), which is strictly smaller than $\frac{\beta+1}{2\beta+1}$, the one of Lemma 10. The reason why $\overline{L_n^\star}$ is not optimal is that it does not use the fact that $\mathcal{M}_n$ is of regularity $\beta + 1$. Let us study what should be the shape of an optimal potential. For $h \in \mathcal{H}^\eta$, we have

$$\begin{aligned}
&\int_{\mathcal{M}_n} (h(x) - h(T(x))) f_{\mu_n}(x) d\lambda_{\mathcal{M}_n}(x) \\
&\quad = \int_{\mathcal{M}_n} \langle \int_0^1 \nabla h(T(x) + t(x - T(x))) dt, x - T(x) \rangle f_{\mu_n}(x) d\lambda_{\mathcal{M}_n}(x) \\
&\quad = \int_{\mathcal{M}_n} \Big( \langle \nabla h(T(x)), x - T(x) \rangle f_{\mu_n}(x) + O(\|x - T(x)\|^2) \Big) d\lambda_{\mathcal{M}_n}(x) \\
&\quad \leq C \int_{\mathcal{M}_n} \langle \nabla h(T(x)), x - T(x) \rangle d\lambda_{\mathcal{M}_n}(x) + C n^{-2(\beta+1)} \\
&\quad \leq C \int_0^1 \langle \nabla h(g^\star(t)), g^\star(t) n^{-(\beta+1)} \sin(2\pi n t) \rangle dt + C n^{-2(\beta+1)},
\end{aligned} \tag{36}$$

with $g^\star$ defined in (34). We see that in order to determine a good potential $h \in \mathcal{H}^\eta$, it is enough to determine the function $H : [0,1] \to \mathbb{R}^2$ defined by

$$H(t) := \nabla h(g^\star(t)). \tag{37}$$

When $n$ is very large, $\mathcal{M}_n$ and $\mathcal{M}^\star$ behave locally like the images of

$$\overline{g}^\star(t) = (t, 0) \text{ and } \overline{g}_n(t) = (t, n^{-(\beta+1)} \sin(2\pi n t)), \tag{38}$$

as represented in Figure 2. Let us first study the local case where we are looking for an optimal potential between $\overline{\mathcal{M}}_n$ and $\overline{\mathcal{M}}^\star$ defined as the images of $\overline{g}^\star$ and $\overline{g}_n$ in (38). Recall from (36) that we want to maximize with respect to $H \in \mathcal{H}_1^{\eta-1}([0,1], \mathbb{R}^2)$ the quantity

$$\int_0^1 \langle H(t), g^\star(t) n^{-(\beta+1)} \sin(2\pi n t) \rangle dt,$$

which in the local case (38) corresponds to

$$n^{-(\beta+1)} \int_0^1 H_1(t) \sin(2\pi n t) dt, \tag{39}$$

for $H_1(t) := \langle H(t), g^\star(t) \rangle$.

To maximise (39), we are going to take $H_1$ having the same sign as the function $x \mapsto \sin(2\pi n x)$. Therefore we will take $H_1(t) = 0$ when $\sin(2\pi n t) = 0$ (i.e. $t \in \{0, \frac{1}{2n}, ..., \frac{2n-1}{2n}\}$),

and for $t \in [\frac{k}{2n}, \frac{k+1}{2n}]$, $H_1(t)$ will be increasing in the direction of $\frac{\sin(\frac{k+1/2}{2n})}{|\sin(\frac{k+1/2}{2n})|}$ as $t$ gets closer to the middle point $\frac{k+1/2}{2n}$. This is illustrated in Figure 2.

The idea behind this construction is to be able to utilize that $\mathcal{M}_n$ is of regularity $\beta+1$. It implies that we can quantify the distance between two points of intersections of the manifolds $\mathcal{M}_n$ and $\mathcal{M}^\star$ with respect to the distance between the manifolds. Indeed for two critical points $\frac{k}{2n}, \frac{k+1}{2n}$, we have

$$\frac{k+1}{2n} - \frac{k}{2n} = \frac{1}{2n} = O(n^{-(\beta+1)})^{\frac{1}{\beta+1}} = \left( \sup_{t \in [\frac{k}{2n}, \frac{k+1}{2n}]} n^{-\beta+1} |\sin(2\pi n t)| \right)^{\frac{1}{\beta+1}}$$
$$= \left( \sup_{x \in \Theta_n(k)} \|x - T(x)\| \right)^{\frac{1}{\beta+1}}, \tag{40}$$

with $\Theta_n(k) = \{x \in \mathcal{M}_n |\ \exists \lambda \in [0,1], T(x) = g^\star(\lambda \frac{k}{2n} + (1-\lambda)\frac{k+1}{2n})\}$. Therefore as $H_1$ is increasing between critical points $\frac{k}{2n}, \frac{k+1}{2n}$, we will be able to relate the value $H_1(\frac{k+1/2}{2n})$ to the distance $\left( \sup_{x \in \Theta_n(k)} \|x - T(x)\| \right)^{\frac{1}{\beta+1}}$.

### C.8.3 Explicit Construction of an Optimal Potential in the Case $\eta \geq 1$

We first address the case $\eta, \gamma \geq 1$, while the remaining cases follow directly by interpolation. Let us build $H_1 \in \mathcal{H}_1^{\eta-1}$ that maximizes (up to a multiplicative constant) the objective (39). Let $\lambda \in \mathcal{H}^\eta([0,1],[0,1])$ such that for all $k \in \{0, ..., \eta-1\}$, $\nabla^k \lambda(0) = \nabla^k \lambda(1) = 0$ and for all $x \in [1/4, 3/4]$, $\lambda(x) \geq 1/2$. We define the optimal potential by

$$H_1(t) := \frac{(-1)^{\lfloor 2nt \rfloor}}{n^{\eta-1}} \lambda\big(2n(t - \frac{\lfloor 2nt \rfloor}{2n})\big).$$

The cost (39) of $H_1$ can be lower bounded as follow.

$$n^{-(\beta+1)} \int_0^1 H_1(t) \sin(2\pi n t) dt = 2n^{-\beta} \int_0^{\frac{1}{2n}} H_1(t) \sin(2\pi n t) dt \geq 2n^{-\beta} \int_{\frac{1}{8n}}^{\frac{3}{8n}} H_1(t) \sin(2\pi n t) dt$$

and

$$2n^{-\beta} \int_{\frac{1}{8n}}^{\frac{3}{8n}} H_1(t) \sin(2\pi n t) dt \geq 2n^{-\beta} \int_{\frac{1}{8n}}^{\frac{3}{8n}} \frac{1}{2n^{\eta-1}} \sin(2\pi n t) dt \geq C^{-1} n^{-(\beta+\eta)}.$$

Let us now show that this potential is optimal (up to a multiplicative constant) by giving an upper bound on (39). For $f : \mathbb{R}/\mathbb{Z} \to \mathbb{C}$, let $(c_f(k))_{k \in \mathbb{Z}}$ denote its Fourier coefficients and let

$$\mathcal{W}_1^{\eta-1,2} = \{f \in L^2(\mathbb{R}/\mathbb{Z}, \mathbb{R}) |\ \sum_{k \in \mathbb{Z}} |c_f(k)|^2 |k|^{2(\eta-1)} \leq 1\}$$

be the unit ball of the $L^2$ Sobolev space of regularity $\eta-1$. As $\mathcal{H}^{\eta-1} \hookrightarrow \mathcal{W}^{\eta-1,2}$, we have

$$\begin{aligned}\sup_{f\in\mathcal{H}_1^{\eta-1}} \int_0^1 f(t)\sin(2\pi nt)dt &\leq C \sup_{f\in\mathcal{W}_1^{\eta-1,2}} \int_0^1 f(t)\sin(2\pi nt)dt \\ &= C \sup_{f\in\mathcal{W}_1^{\eta-1,2}} \sum_{k\in\mathbb{Z}} c_f(k)c_{\sin(2\pi n\cdot)}(k) \\ &= C \sup_{f\in\mathcal{W}^{\eta-1,2}} -\frac{i}{2}c_f(n) + \frac{i}{2}c_f(-n) \\ &\leq Cn^{-(\eta-1)}.\end{aligned}$$

Therefore,

$$n^{-(\beta+1)} \sup_{f\in\mathcal{H}_1^{\eta-1}} \int_0^1 f(t)\sin(2\pi nt)dt \leq Cn^{-(\beta+\eta)} \leq C_2 n^{-(\beta+1)} \int_0^1 H_1(t)\sin(2\pi nt)dt,$$

so $H_1$ is indeed optimal.

Coming back to the non local case, for $\Theta : \mathcal{M}^\star \to \mathbb{R}^2$ such that $\Theta(x)$ is the unitary normal vector pointing outward the manifold $\mathcal{M}^\star$ at the point $x$, defining

$$h(x) = H_1 \circ g^{\star-1} \circ T(x)\langle \Theta \circ T(x), x - T(x)\rangle,$$

we have $h \in \mathcal{H}_C^\eta(\mathcal{M}^{\star n^{-(\beta+1)}}, \mathbb{R})$ so using Proposition 32 it can be extended into a map belonging to $\mathcal{H}_C^\eta(\mathbb{R}^2, \mathbb{R})$. Furthermore, we have

$$\begin{aligned}\int_{\mathcal{M}_n} (h(x) - h(T(x)))f_{\mu_n}(x)d\lambda_{\mathcal{M}_n}(x) &\geq C^{-1} \int_0^1 \langle H_1(t)g^\star(t)), g^\star(t)n^{-(\beta+1)}\sin(2\pi nt)\rangle dt \\ &\geq C^{-1}n^{-(\beta+\eta)},\end{aligned}$$

and for all $h' \in \mathcal{H}_1^\eta(\mathbb{R}^2, \mathbb{R})$ with $h'_{|\mathcal{M}_{g^\star}=0}$, we have

$$\begin{aligned}\int_{\mathcal{M}_n} h'(x)f_{\mu_n}(x)d\lambda_{\mathcal{M}_n}(x) &\leq C \int_0^1 \langle \nabla h'(g^\star(t)), g^\star(t)n^{-(\beta+1)}\sin(2\pi nt)\rangle dt + Cn^{-2(\beta+1)} \\ &\leq n^{-(\beta+1)} \sup_{f\in\mathcal{H}_1^{\eta-1}} \int_0^1 f(t)\sin(2\pi nt)dt + Cn^{-2(\beta+1)} \\ &\leq Cn^{-(\beta+\eta)} \\ &\leq C \int_{\mathcal{M}_n} h(x)f_{\mu_n}(x)d\lambda_{\mathcal{M}_n}(x).\end{aligned}$$

We conclude that $h$ is an optimal potential for the non local case which gives

$$C^{-1}n^{-(\beta+\eta)} \leq \int h(x)d\mu_n(x) \leq Cd_{\mathcal{H}_1^\eta}(\mu_n, T_{\#\mu_n}) \leq C_2 \int h(x)d\mu_n(x) \leq C_3 n^{-(\beta+\eta)}.$$

In particular, we obtain the sought-after inequality for all $1 \leq \gamma < \eta$:

$$d_{\mathcal{H}_1^\eta}(\mu_n, T_{\#\mu_n})^{\frac{\beta+\gamma}{\beta+\eta}} \leq \left(Cn^{-(\beta+\eta)}\right)^{\frac{\beta+\gamma}{\beta+\eta}} \leq Cn^{-\beta+\gamma} \leq C_2 d_{\mathcal{H}_1^\gamma}(\mu_n, T_{\#\mu_n}).$$

Finally note that this example also allows to show that for $\gamma \in (0,1)$ and $\eta \geq 1$ we have

$$d_{\mathcal{H}_1^\gamma}(\mu_n, T_{\#\mu_n}) = O(n^{-\gamma(\beta+1)}) = O(n^{-\beta+\eta})^{\frac{\beta\gamma+\gamma}{\beta+\eta}} = O(d_{\mathcal{H}_1^\eta}(\mu_n, T_{\#\mu_n})^{\frac{\beta\gamma+\gamma}{\beta+\eta}}). \tag{41}$$

### C.8.4 Further Remarks on the Example

We see that in the case $\eta = \beta + 1$, the exponent $\frac{\beta+1}{2\beta+1}$ is an improvement over the exponent $1/2$ given by Jensen's inequality (35). In contrary to $x \mapsto \|x - T(x)\|^2$, the potential $h$ uses the fact that $\mathcal{M}_n$ is of regularity $\beta + 1$. Instead of being equal to 0 on $\mathcal{M}^\star$, $\nabla h$ increases between the intersections points of $\mathcal{M}^\star$ and $\mathcal{M}_n$ which corresponds to the irregularities of the function $x \mapsto \|x - T(x)\|$. As the distance between these points is proportional to the distance between $\mathcal{M}^\star$ and $\mathcal{M}_n$ to the power $\frac{1}{\beta+1}$, we have that $\|\nabla h(T(x))\|$ is proportional to $\|x - T(x)\|^{\frac{\eta-1}{\beta+1}} = \|x - T(x)\|^{\frac{\beta}{\beta+1}}$ for $\eta = \beta + 1$. This explains the improvement over the potential $x \mapsto \|x - T(x)\|^2$ as the norm of its gradient is proportional to $\|x - T(x)\|$.

The construction of $h$ is similar to the one used in the proof of Lemma 10. In both cases we take

$$h(x) = \langle F(x), x - T(x) \rangle,$$

with $F$ a regularization of the map $x \mapsto \frac{x-T(x)}{\|x-T(x)\|}$ belonging to $\mathcal{H}_C^{\eta-1}$ and such that $\|\nabla^{\lfloor \eta \rfloor} F(x)\| \leq \|x - T(x)\|^{-1+\eta-\lfloor \eta \rfloor}$. The difference lies in the regularization of the map $x \mapsto \frac{x-T(x)}{\|x-T(x)\|}$ where here we did a geometrical construction which is already quite complex even in the simple setting of this example. We believe that Lemma 10 could be proven using a construction of this type but it might be very complicated to describe it in the general case. In the proof of Lemma 10, we simply regularize the function $x \mapsto \frac{x-T(x)}{\|x-T(x)\|}$ by lowering the size of its wavelet coefficients. This trick allows us to obtain a good potential of regularity $\eta - 1$ without really knowing its shape. In particular, this outlines the efficiency of the wavelet tool which allows to easily modify the regularity of functions while controlling their potential action.

## C.9 Existence of the Minimizer (11)

**Proposition 38** *The minimizer*

$$\hat{\mu} \in \operatorname*{arg\,min}_{\mu \in \mathcal{F}} d_{\mathcal{H}_1^{d/2}}(\mu, \mu_n^\star),$$

*is well defined.*

**Proof** As each $\mu \in \mathcal{F}$ is supported in $B_p(0, K)$, we may have that $\mathcal{F} \subset \mathcal{P}(B_p(0, K))$, the space of Borel probability measures on the compact metric space $B_p(0, K)$. As $\mathcal{P}(B_p(0, K))$ is compact for the weak topology (Prokhorov's theorem), any sequence in $\mathcal{F}$ is tight and admits weakly convergent subsequences in $\mathcal{P}(B_p(0, K))$. We then need to show that $\mathcal{F}$ is *closed* in $\mathcal{P}(B_p(0, K))$, so that it is itself compact.

Let $(\mu_k)_{k\geq 1} \subset \mathcal{F}$ with $\mu_k \rightharpoonup \mu$ weakly in $\mathcal{P}(B_p(0, K))$. Write that each $\mu_k$ is supported on some $(\beta+1, K)$-manifold $\mathcal{M}_k$ and has a $(\beta, K)$-density $f_k$ bounded below by $K^{-1}$ (Definition 2). For each $k$, applying Proposition 4 to the pair $(\mathcal{M}_k, \mu_k)$ we obtain that there exist some maps $\varphi_{i,k} \in \mathcal{H}_C^{\beta+1}(\mathbb{R}^d, \mathbb{R}^p)$ and weights $\zeta_{i,k} \in \mathcal{H}_C^{\beta+1}(\mathbb{R}^d, \mathbb{R})$ with $\operatorname{supp}(\zeta_{i,k}) \subset B_d(0, \tau)$, such that for every bounded continuous $h : \mathbb{R}^p \to \mathbb{R}$,

$$\int h \, \mathrm{d}\mu_k = \sum_{i=1}^{m} \frac{1}{C_{i,k}} \int_{\mathbb{R}^d} h(\varphi_{i,k}(u)) \, \zeta_{i,k}(u) \, \mathrm{d}\lambda^d(u).$$

By Arzelà–Ascoli, after extracting subsequences (by diagonal argument over $i$), we may assume $\varphi_{i,k} \to \varphi_i$ and $\zeta_{i,k} \to \zeta_i$ uniformly together with all derivatives controlled by the Hölder norms; the Jacobian lower bound passes to the limit, so each $\varphi_i$ is a $C^{\beta+1}$-immersion with a uniform metric lower bound. Setting

$$\mathcal{M} := \bigcup_{i=1}^{m_0} \varphi_i(B_d(0,\tau)) \subset B_p(0,K), \qquad \text{and} \qquad \int h \,\mathrm{d}\mu := \sum_{i=1}^{m_0} \frac{1}{C_i} \int h(\varphi_i(u))\zeta_i(u) \,\mathrm{d}\lambda^d(u),$$

we get that $\mu$ is supported on a $d$-dimensional $(\beta+1, \tilde{K})$-manifold $\mathcal{M}$ and admits a $(\beta, \tilde{K})$-density bounded below by $\tilde{K}^{-1}$.

Now for $x \in \mathcal{M}$, we can take a sequence of points $x_k \in \mathcal{M}_k$ with $x_k \to x$ and as $\mathcal{M}_k$ satisfies the $(\beta+1, K)$-manifold condition, we have that the inverse of the projection into the tangent plane at $x_k$ satisfies

$$(\varphi_x)_k^{-1} \in \mathcal{H}_K^{\beta+1}((\varphi_x)_k((U_x)_k), (U_x)_k).$$

Passing to the limit, we obtain that $\mathcal{M}$ satisfies the $(\beta+1, K)$-manifold condition and doing the same for densities we obtain that $\mu$ satisfies the $(\beta, K)$-density condition and therefore belongs to $\mathcal{F}$. By continuity of the distance $d_{\mathcal{H}_1^{d/2}}$, we finally obtain that $\mu$ is a minimizer over $\mathcal{F}$ of the function $\nu \mapsto d_{\mathcal{H}_1^{d/2}}(\nu, \mu_n^\star)$. ■

## References


M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein generative adversarial networks. In D. Precup and Y.W. Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pages 214–223. PMLR, 2017.

Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006. ISBN 0387310738.

Saptarshi Chakraborty and Peter L Bartlett. On the statistical properties of generative adversarial models for low intrinsic data dimension. *arXiv preprint arXiv:2401.15801*, 2024.

Frédéric Chazal, David Cohen-Steiner, and Quentin Mérigot. Geometric inference for probability measures. *Foundations of Computational Mathematics*, 11:733–751, 2011.

Ingrid Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on pure and applied mathematics*, 41(7):909–996, 1988.

Valentin De Bortoli. Convergence of denoising diffusion models under the manifold hypothesis. *arXiv preprint arXiv:2208.05314*, 2022.

Vincent Divol. Measure estimation on manifolds: an optimal transport approach. *Probability Theory and Related Fields*, 183(1-2):581–647, 2022.

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

Herbert Federer. *Geometric measure theory*. Springer, 2014.

Charles Fefferman, Sergei Ivanov, Yaroslav Kurylev, Matti Lassas, and Hariharan Narayanan. Reconstruction and interpolation of manifolds. i: The geometric Whitney problem. *Foundations of Computational Mathematics*, 20(5):1035–1133, 2020.

Charles L Fefferman. A sharp form of Whitney's extension theorem. *Annals of mathematics*, 161(1):509–577, 2005.

Evarist Giné and Richard Nickl. *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2015. doi: 10.1017/CBO9781107337862.

Hichem Hajaiej, Luc Molinet, Tohru Ozawa, and Baoxiang Wang. Sufficient and necessary conditions for the fractional gagliardo-nirenberg inequalities and applications to navier-stokes and generalized boson equations. *arXiv preprint arXiv:1004.4287*, 2010.

Dorothee D Haroske. *Envelopes and sharp embeddings of function spaces*. CRC Press, 2006.

Jürgen Jost. *Riemannian Geometry and Geometric Analysis*. Universitext. Springer, 5 edition, 2008. doi: 10.1007/978-3-540-77341-2.

Mario Köppen. The curse of dimensionality. In *5th online world conference on soft computing in industrial applications (WSC5)*, volume 1, pages 4–8, 2000.

Gunther Leobacher and Alexander Steinicke. Existence, uniqueness and regularity of the projection onto differentiable manifolds. *Annals of Global Analysis and Geometry*, 60(3): 559–587, 2021. doi: 10.1007/s10455-021-09788-z.

Tengyuan Liang. How well generative adversarial networks learn distributions. *Journal of Machine Learning Research*, 22(228):1–41, 2021.

Alessandra Lunardi. *Interpolation theory*, volume 16. Springer, 2018.

A. Müller. Integral probability metrics and their generating classes of functions. *Advances in Applied Probability*, 29:429–443, 1997.

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

Lars Ruthotto and Eldad Haber. An introduction to deep generative modeling. *GAMM-Mitteilungen*, 44(2):e202100008, 2021.

Filippo Santambrogio. *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*, volume 87 of *Progress in Nonlinear Differential Equations and Their Applications*. Birkhäuser, 2015. doi: 10.1007/978-3-319-20828-2.

Nicolas Schreuder, Victor-Emmanuel Brunel, and Arnak Dalalyan. Statistical guarantees for generative models without domination. In *Algorithmic Learning Theory*, pages 1051–1071. PMLR, 2021.

Jian Shen, Han Zhao, Weinan Zhang, and Yong Yu. Model-based policy optimization with unsupervised model adaptation. *Advances in Neural Information Processing Systems*, 33: 2823–2834, 2020.

Alexander J Smola, A Gretton, and K Borgwardt. Maximum mean discrepancy. In *13th international conference, ICONIP*, pages 3–6, 2006.

ELIAS M. Stein. *Singular Integrals and Differentiability Properties of Functions (PMS-30)*. Princeton University Press, 1970. ISBN 9780691080796. URL `http://www.jstor.org/stable/j.ctt1bpmb07`.

Arthur Stéphanovitch. Generative model for optimal density estimation on unknown manifold. *arXiv preprint arXiv:2506.19587*, 2025.

Arthur Stéphanovitch, Eddie Aamari, and Clément Levrard. Wasserstein gans are minimax optimal distribution estimators. *arXiv preprint arXiv:2311.18613*, 2023.

Csaba Szepesvári. *Algorithms for reinforcement learning*. Springer Nature, 2022.

Rong Tang and Yun Yang. Minimax rate of distribution estimation on unknown submanifolds under adversarial losses. *The Annals of Statistics*, 51(3):1282–1308, 2023.

Hans Triebel. *Theory of Function Spaces II*. Modern Birkhäuser Classics. Birkhäuser Basel, 2010. doi: 10.1007/978-3-0346-0419-2.

Sergio Verdú. Total variation distance and the distribution of relative information. In *2014 Information Theory and Applications Workshop (ITA)*, pages 1–3. IEEE, 2014.

Cédric Villani. *Optimal transport: old and new*, volume 338. Springer, 2009.

Cédric Villani. *Topics in optimal transportation*, volume 58. American Mathematical Soc., 2021.

Jie Wang, Minshuo Chen, Tuo Zhao, Wenjing Liao, and Yao Xie. A manifold two-sample test study: integral probability metric with neural networks. *Information and Inference: A Journal of the IMA*, 12(3):1867–1897, 2023.