École Doctorale Mathématiques et STIC

THÈSE

Kernel smoothing and diffusion processes on manifolds

Présentée pour l'obtention du grade de DOCTEUR
de l'Université Gustave Eiffel
et de l'Université du Québec à Montréal

Spécialité

Mathématiques Appliquées

par

NGUYEN Dinh Toàn

Soutenue le mardi 10 juin 2025 devant le jury composé de

Jurgen Angst	MCF	Université Rennes 1	Rapporteur
Thomas Bonis	MCF	Université Gustave Eiffel	Examinateur
Nicolas Fournier	Pr	Paris Sorbonne Université	Examinateur
Hélène Guérin	Pr	Université du Québec à Montréal	Directrice
Claire Lacour	Pr	Université Gustave Eiffel	Examinatrice
Frédéric Rochon	Pr	Université du Québec à Montréal	Examinateur
Laurent Saloff-Coste	Pr	Cornell University	Rapporteur
Viet Chi Tran	DR	Centre Inria de l'Univ. de Lille	Directeur

Remerciements

This PhD thesis has been funded by an international cotutelle doctoral program and was jointly supervised by the Université Gustave Eiffel in France and Université du Québec à Montréal (UQAM) in Canada.
The works presented here have been supported by the Mathematics for Public Health program funded by the Natural Sciences and Engineering Research Council of Canada (NSERC), by Labex Bézout (ANR-10-LABX-58) and by GdR GeoSto 3477.

Productions scientifiques liées à la thèse

Several papers were written during the course of this PhD thesis, including one that is not presented here but stemmed from my master thesis.

Articles

- Dinh-Toàn Nguyen, Scaling limit of the collision measures of multiple random walks. *ALEA, Lat. Am. J. Probab. Math. Stat.* 20, 1385–1410 (2023). `doi.org/10.30757/ALEA.v20-52` [89]
- Hélène Guérin, Dinh-Toàn Nguyen, Viet Chi Tran, Strong uniform convergence of Laplacians of random geometric and directed kNN graphs on compact manifolds. Submitted (2024). `doi.org/10.48550/arXiv.2212.1028` [61]
- Vincent Divol, Hélène Guérin, Dinh-Toàn Nguyen, Viet Chi Tran. Measure estimation on a manifold explored by a diffusion process. Submitted (2024).`doi.org/10.48550/arXiv.2410.11777` [38]

Article en cours de rédaction

- 1-Wasserstein minimax estimation for general smooth probability densities, D.-T. Nguyen, 2025+, in preparation.

Contents

Chapter 1

Introduction

Contents

1.1 Introduction

My thesis mainly lies at the interface between probability theory and differential geometry. From my perspective, these two beautiful fields of mathematics have always been closely connected. The interplay of these two fields has led to significant advances, enabling the modeling of sophisticated systems in physics, finance, machine learning, and beyond. For example, the geometric perspective allows probabilists to leverage concepts like curvature [12], Riemannian metrics [8, 9], and geometric structures [5] to gain deeper insights into probabilistic phenomena such as the theory of concentration of measure, with applications in algorithm optimization and machine learning (e.g. [58, 85]). Of particular interest for us are the considerations of stochastic processes in curved spaces. We refer to [43, 67] for the definition of Brownian motions and stochastic differential equations (SDEs) on manifolds.

The manifold hypothesis has become a foundational concept in modern machine learning and topological data analysis. It posits that high-dimensional data often lie on a low-dimensional manifold embedded within a higher-dimensional space [84, 103, 114]. This paradigm underpins the efficiency of nonparametric methods in high-dimensional statistical models and has spurred extensive research into manifold learning and analysis, e.g. [52, 91, 99]. For example, [37] recently applied optimal transport to measure estimation on manifolds in this framework, providing a robust nonparametric framework that accurately recovers distributions on complex, high-dimensional data spaces with theoretical guarantees of efficiency and optimality. Nevertheless, previous studies were based on independent and identically distributed (i.i.d.) samples, we are interested in this thesis with data stemming from stochastic processes exploring the manifolds. The exploration of complex structures in view of understanding their geometric and topological properties is an old idea, think of the PageRank algorithm for example [94]. Therefore, building on the foundational connections between probability theory and differential geometry, we embark on our doctoral journey with an exploration of random walks and diffusion processes on (random geometric graphs) on manifolds. This study opens up new perspectives for us, revealing exciting ways to reconstruct manifold information through empirical observations of trajectories, explore minimax optimality in density estimation, and understand the intricate dynamics of stochastic processes on manifolds such as their long-time behavior or their contractivity properties.

As we navigate between these layers of complexity, our work gradually focuses on three central themes: the convergence properties of random operators on manifolds [61], the estimation of invariant measures associated with diffusion processes [38], and the minimax convergence rates in density estimation on manifolds. More precisely, this document is organized as follows.

In this introductory chapter, we introduce various preliminary mathematical concepts necessary to lay the groundwork for this thesis. Then, a summary of the thesis's contributions will be provided in Section 1.6.

In Chapter 2, we investigate the convergence of random operators associated with points sampled on smooth compact and connected manifolds, when their number tends to infinity. A particular emphasis is put on the convergence of graph Laplacians built from these points. This question was already much considered in the literature [55, 115, 24, 25]. Here, we extend the existing results by weakening the assumptions on the kernel functions used in the graph construction. This provides a uniform convergence rate for wider ranges of kernel-induced random operators, including the exact generator of the k-nearest neighbor (k-NN) random walk. We use these limits to establish the functional convergence, when the sample size grows to infinity, of the random walks on the graph to stochastic diffusive processes on the manifold.

Then, in Chapter 3, we focus on the convergence properties in long time of occupation measures associated with diffusion processes $(X_t)_{t\geq 0}$ on smooth compact connected manifolds, that can be the limiting processes obtained in the previous chapter. We study these convergences in Wasserstein distance, when the occupation measures are smoothed by the convolution with an

appropriate kernel. The limits are the invariant measures of the diffusions. Thanks to the regularization with the convolution, the convergence rates that we obtain improve those in [121]. We establish the optimality of the convergence rates in the minimax sense. From the perspective of manifold estimation, our work serves as a counterpart to the study conducted by Divol in [37], with the key distinction that our data comprises trajectories of stochastic processes rather than independent and identically distributed (i.i.d.) samples. Specifically, we consider the estimation problem where the observed data are continuous sample paths of a diffusion process evolving on a manifold.
Finally, in Chapter 4, we revisit the problem of density estimation on manifolds introduced in [37], putting aside the stochastic process point of view for the moment. We extend these results as well as those in related works [118, 90]. Specifically, we will demonstrate that the minimax convergence rate established in [37, 90] holds for a broader class of density functions, not only the one that are bounded below by a positive constant. Additionally, this provides a convergence theorem generalizing the result in [118] in the case of density functions with possibly unbounded support in $\mathbb{R}^m$. An almost sure convergence result is also provided for the compact manifold setting.

We start with presenting the fundamental concepts and basic results across various fields, including Stochastic Differential Equations (Section 1.2), Differential Geometry (Sections 1.3 and 1.4), and Operator Theory (Section 1.5) in the purpose of studying stochastic differential equations on manifolds. We then present the main results of the thesis in Section 1.6.

1.2 Fundamentals of Stochastic Differential Equations

Stochastic differential equations (SDEs) are fundamental tools in modeling systems that evolve over time under the influence of both deterministic and random forces. They have widespread applications in fields such as physics, finance, biology, and engineering. In this section, we summarize some basic concepts related to SDEs in the Euclidean space $\mathbb{R}^m$, which will later serve as a foundation for developing the theory of SDEs on manifolds. The interested reader is referred to [68, 48, 101].
Let us consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. An SDE in the Euclidean framework is characterized by the following components:

- A **drift** coefficient vector $b : \mathbb{R}^m \to \mathbb{R}^m$, representing the deterministic part of the system's evolution.
- A **volatility** (or **diffusion**) coefficient matrix $\sigma = (\sigma^i_k)_{1 \leqslant i \leqslant m,\, 1 \leqslant k \leqslant l}$, where $\sigma : \mathbb{R}^m \to \mathbb{R}^{m \times l}$, capturing the random fluctuations.
- A driving l-dimensional Brownian motion process $B = (B^1_t, B^2_t, \ldots, B^l_t)_{t \geqslant 0}$, which introduces the stochasticity into the system. The Brownian motion B is adapted to a filtration $(\mathcal{F}_t)_{t \geqslant 0}$ that satisfies the usual hypotheses: it is right-continuous and complete.
- An $\mathcal{F}_0$-measurable random variable X_0 independent from B that will be the **initial condition.**

To focus on the essential ideas without unnecessary complications, we assume that the functions b and σ are **continuous**.
Let us now formalize our notations and go deeper into the properties of SDEs.

1.2.1 Strong Solutions, Uniqueness, and Explosion

In the study of SDEs, it is crucial to understand the concept of *strong solutions*, the *uniqueness*, and the *explosion* of SDE solutions.

Definition 1.2.1 (Strong Solution of an SDE)**.** *[73, p. 336], [68, p. 163], [67, p. 7] Let* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$ *be a filtered probability space, and consider the filtration, Brownian motion, initial condition, drift and volatility functions introduced above. Let* τ *be a* $(\mathcal{F}_t)$*-stopping time, and let* $X = \{X_t : 0 \leqslant t < \tau\}$ *be a* $(\mathcal{F}_t)$*-adapted continuous semimartingale defined up to time* τ*. We say that* X *is a* ***strong solution*** *of the stochastic differential equation starting from* X_0

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}B_t, \tag{1.1}$$

if, for all $t \geqslant 0$*, the following integral equation holds:*

$$X_t = X_0 + \int_0^{t\wedge\tau} b(X_s)\,\mathrm{d}s + \int_0^{t\wedge\tau} \sigma(X_s)\,\mathrm{d}B_s, \tag{1.2}$$

where $\int \cdot \mathrm{d}B_s$ *denotes the Itô integral [68, Chapter II].*

Notice that the definition of a strong solution can be extended to SDEs driven by other stochastic processes, such as fractional Brownian motion or Lévy processes. In such cases, the integrals are interpreted in the sense appropriate to the driving process (e.g., the Riemann-Stieltjes integral or the Itô integral for Lévy processes).

A fundamental question in the theory of SDEs is under what conditions does a unique strong solution exist. The following theorem addresses this question.

Theorem 1.2.2 (Existence and Uniqueness of Strong Solutions)**.** *[75, pp. 287–289], [67, Thm. 1.1.8]*
Suppose that the functions b *and* σ *are* ***locally Lipschitz continuous****. That is, for every compact set* $K \subset \mathbb{R}^m$*, there exists a constant* $L_K > 0$ *such that for all* $x, y \in K$*,*

$$\|b(x) - b(y)\| + \|\sigma(x) - \sigma(y)\| \leqslant L_K\|x - y\|.$$

Then, for any $\mathcal{F}_0$*-measurable initial condition* X_0*, the SDE* (1.1) *has a unique strong solution* $X = \{X_t : 0 \leqslant t < \tau\}$ *up to an explosion time* τ*. In other words, for any other solution* $Y = \{Y_t : 0 \leqslant t < \tau'\}$ *of the same SDE with the same initial condition, we have* $\tau' \leqslant \tau$ *and*

$$Y_t = X_t \quad \text{for all } t < \tau'.$$

Furthermore, on the event $\{\tau < +\infty\}$*, we have almost surely that* $\lim_{t\to\tau} \|X_t\| = \infty$*, meaning that* τ *is the* ***explosion time*** *of* X*.*

This theorem assures us that under the local Lipschitz condition, the SDE has a unique maximal strong solution. However, it does not guarantee that the solution exists for all time. Solutions may explode, becoming unbounded in finite time.

Example 1.2.3. *Consider the one-dimensional SDE:*

$$\mathrm{d}X_t = X_t^3\,\mathrm{d}t + X_t^2\mathrm{d}B_t, \quad X_0 = 1.$$

The unique solution for this SDE is $X_t = \frac{1}{1-B_t}$*. We can see that even though the coefficients are continuous, the solution for the above SDE explodes in finite time. More precisely, the explosion time for this solution is* $\tau_1 := \inf\{t : B_t \geq 1\}$*.*

A natural question arises: under what conditions can we ensure that solutions to an SDE do not explode? The following theorem provides a sufficient condition that is weaker than Lipschitz continuity.

Theorem 1.2.4 (Non-Explosiveness Criterion). *[68, Thm. 2.4, p. 177]*
Suppose that the functions b and σ are ***linearly bounded****; that is, there exists a constant $C > 0$ such that for all $x \in \mathbb{R}^m$,*

$$\|b(x)\| + \|\sigma(x)\| \leqslant C(1 + \|x\|).$$

Then, for any initial condition X_0, the solution X_t of the SDE (1.1) exists for all $t \geqslant 0$ and does not explode. Moreover, if $\mathbb{E}[\|X_0\|^2] < \infty$, then

$$\mathbb{E}[\|X_t\|^2] < \infty \quad \text{for all } t \geqslant 0.$$

This theorem indicates that if the coefficients of the SDE grow at most linearly, the solution remains finite for all time, and its second moment is finite, provided the initial condition has a finite second moment.

1.2.2 Weak Solutions

In some situations, it is important to consider the distribution of solutions rather than their specific sample paths. This leads to the concept of **weak solutions**, where the focus is on the existence of a probability space and processes satisfying the SDE in distribution.

Definition 1.2.5 (Weak Solution of an SDE). *A tuple $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P}, ; B, X_0, X)$ is called a* ***weak solution*** *of the SDE (1.1) if:*

- *$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$ is a filtered probability space satisfying the usual hypotheses.*
- *$B = (B_t)_{t\geqslant 0}$ is an l-dimensional Brownian motion adapted to the filtration $(\mathcal{F}_t)_{t\geqslant 0}$.*
- *A $\mathcal{F}_0$-measurable random variable X_0.*
- *$X = (X_t)_{t\geqslant 0}$ is an $\mathbb{R}^m$-valued continuous adapted process satisfying the SDE (1.1) with initial condition X_0.*

In a weak solution, the probability space, the Brownian motion and the initial condition are part of the solution, rather than being prescribed in advance.
An important relationship exists between the uniqueness of strong solutions and weak solutions.

Theorem 1.2.6 (Strong Uniqueness Implies Weak Uniqueness). *[68, p. 166]*
If the SDE (1.1) has a unique strong solution for every initial condition $X_0 = x_0 \in \mathbb{R}^m$, then it also has a unique weak solution. Specifically, for any two weak solutions $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P}; B, x_0, X)$ and $\left(\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t), \tilde{\mathbb{P}}; \tilde{B}, x_0, \tilde{X}\right)$, the laws of X and $\tilde{X}$ are identical:

$$\mathcal{L}(X) = \mathcal{L}(\tilde{X}).$$

Notation 1.2.7. *In what follows, we will denote by $\mathbb{P}_{x_0}$ the distribution associated to the weak solution with initial position x_0 and $\mathbb{E}_{x_0}$ will be the corresponding expectation.*

This theorem implies that the distribution of the solution is uniquely determined by the coefficients of the SDE and the initial condition when strong uniqueness holds.

Corollary 1.2.8. *If the SDE (1.1) has a unique strong solution, then the solution X is* ***adapted to the natural filtration*** *generated by the driving Brownian motion B.*

1.2.3 Semi-group and generators

Now, suppose that for each $x_0 \in \mathbb{R}^m$, the weak solution of the SDE (1.1) started from $X_0 = x_0$ exists, is unique, and does not explode (e.g., when b and σ are globally Lipschitz continuous). We can then define a family of operators that describe the evolution of the system.

Definition 1.2.9 (Semigroup of an SDE)**.** *For any $t \geqslant 0$ and any bounded measurable function $f : \mathbb{R}^m \to \mathbb{R}$, define*

$$P_t f(x_0) = \mathbb{E}_{x_0}[f(X_t)],$$

where (X_t) is a weak solution of the SDE (1.1) *starting from x_0. The family of operators $(P_t)_{t\geqslant 0}$ is called the **semigroup** of the SDE* (1.1)*.*

The semigroup $(P_t)_{t\geqslant 0}$ satisfies the following property that reflects the Markovian nature of the stochastic process X_t:

Proposition 1.2.10 (Semigroup Property)**.** *For any $s, t \geqslant 0$, and any bounded measurable function f, we have*

$$P_s(P_t f) = P_{s+t} f.$$

The **generator** of the semigroup provides information about the infinitesimal behavior of the process. It is defined as follows.

Definition 1.2.11 (Generator of a semigroup)**.** *Given a semigroup $(P_t)_{t\geqslant 0}$ acting on a Banach space $\mathbf{B}$, the **generator** $\mathcal{A}$ is defined on the domain*

$$\mathcal{D}(\mathcal{A}) = \left\{ f \in \mathbf{B} \,\Big|\, \lim_{t\to 0^+} \frac{P_t f - f}{t} \text{ exists in } \mathbf{B} \right\},$$

and for $f \in \mathcal{D}(\mathcal{A})$, the generator is given by

$$\mathcal{A} f = \lim_{t\to 0^+} \frac{P_t f - f}{t}.$$

In the context of SDEs, the generator can often be explicitly calculated.

Proposition 1.2.12 (Generator of an SDE)**.** *[68, Thm 6.1]*
Consider the semigroup $(P_t)_{t\geqslant 0}$ associated with the SDE (1.1)*. For any function $f \in \mathcal{C}_c^2(\mathbb{R}^m)$ (twice continuously differentiable with compact support), we have*

$$\lim_{t\to 0^+} \frac{P_t f - f}{t} = \mathcal{A} f \quad \text{in } L^\infty,$$

where the generator $\mathcal{A}$ is given by

$$\mathcal{A} f(x) = \frac{1}{2} \sum_{i,j=1}^{m} a^{ij}(x) \frac{\partial^2 f}{\partial x_i \partial x_j}(x) + \sum_{i=1}^{m} b^i(x) \frac{\partial f}{\partial x_i}(x), \tag{1.3}$$

and $a^{ij}(x)$ are the entries of the matrix $a(x) = \sigma(x)\sigma(x)^\top$.

This result shows that the generator of the semigroup associated with an SDE is a second-order differential operator.
An essential tool in stochastic calculus is Itô's formula, which allows us to compute the differential of a function of a stochastic process. In the particular case of SDE, we have the following formula for solutions of SDE:

Proposition 1.2.13 (Itô's Formula)**.** *Let $X = \{X_t : 0 \leqslant t < \tau\}$ be a strong solution of the SDE* (1.1)*, and let $f \in \mathcal{C}_c^2(\mathbb{R}^m)$. Then, for any $t \geqslant 0$, we have*

$$f(X_{t\wedge\tau}) = f(X_0) + \int_0^{t\wedge\tau} \mathcal{A} f(X_s)\,\mathrm{d}s + \sum_{k=1}^{l} \int_0^{t\wedge\tau} \left(\sigma^\top(X_s)\nabla f(X_s)\right)_k \mathrm{d}B_s^k,$$

where $\mathcal{A}$ is the generator defined in (1.3)*, and ∇f denotes the gradient of f.*

1.2.4 Girsanov's Theorem

Girsanov's theorem is a fundamental result that allows for a change of measure, effectively transforming one SDE into another by adjusting the drift term. This has proved particularly useful in financial mathematics and stochastic control, but not only in these fields.

Theorem 1.2.14 (Girsanov's Theorem)**.** *[93, Thm. 8.6.4]*
Let $X = (X_t)_{t\geqslant 0}$ be a strong solution of the SDE, for a given Brownian motion B and a initial condition X_0,

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}B_t.$$

Suppose that there exist functions $a : \mathbb{R}^m \to \mathbb{R}^m$ and $u : \mathbb{R}^m \to \mathbb{R}^l$ such that:

- *u is continuous.*
- *For all $x \in \mathbb{R}^m$, $b(x) - a(x) = \sigma(x)u(x)$.*
- *The process*
$$\mathcal{E}_t = \exp\left(-\int_0^t u(X_s)^\top\,\mathrm{d}B_s - \frac{1}{2}\int_0^t \|u(X_s)\|^2\,\mathrm{d}s\right)$$
is a martingale satisfying $\mathbb{E}[\mathcal{E}_T] = 1$ for some $T > 0$ fixed.

Then, under the new probability measure $\mathbb{Q}$ defined by $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \mathcal{E}_T$, the process X is a strong solution of the SDE

$$\mathrm{d}X_t = a(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}\hat{B}_t \quad \forall 0 < t \le T,$$

starting from X_0 and where $\hat{B}_t = B_t + \int_0^t u(X_s)\,\mathrm{d}s$ is a Brownian motion under $\mathbb{Q}$.

Girsanov's theorem essentially allows us to shift the drift of an SDE by changing the underlying probability measure, provided that the Radon-Nikodym derivative $\mathcal{E}_T$ is a martingale with expectation one. The process $\hat{B}_t$ adjusts for the change in drift, ensuring that it remains a Brownian motion under the new measure $\mathbb{Q}$.

Our aim is to extend the notion of stochastic differential equations to manifolds, and therefore, we now provide some fundamentals of Differential Geometry. We will recall in Section 1.3 the bases of Differential Geometry, namely the notions of topological manifolds and atlases. Then, smooth manifolds and submanifolds are defined, allowing to introduce tangent spaces. All these notations serve in Section 1.4 to introduce the concepts of Riemaniann geometry that will be needed to extend the results of SDEs to manifolds. The theory of SDEs on manifolds is strongly based on the notion of infinitesimal generator and requires results in operator theory that are recalled in Section 1.5. The SDEs on manifold are introduced in Section 1.5.4 in particular.

1.3 Fundamentals of differential geometry I: topological manifolds and submanifolds of $\mathbb{R}^d$

In this Section, we give a concise introduction to most of fundamental geometric notions used in this document. First, we define coordinate charts and atlases of manifolds. We also discuss on vector fields, tangent spaces, curves and maps defined on manifolds. The main references for this Section are [78, 40], and[59].

1.3.1 Topological manifolds, coordinate charts, change of coordinates

Let $\mathcal{M}$ be a topological space. Recall that $\mathcal{M}$ is called *Hausdorff* if for any two distinct points $p, q \in \mathcal{M}$, there exist two disjoint open neighborhoods U, V of p and q, respectively.
The simplest manifolds are topological manifolds. In simple terms, a d-dimensional topological manifold $\mathcal{M}$ is a topological space that locally look like open subsets of $\mathbb{R}^d$.

Definition 1.3.1 (Coordinate charts)**.** *A d-dimensional **coordinate chart** (or just **chart**) on $\mathcal{M}$ is a pair (U, φ) where U is an open subset of $\mathcal{M}$ and $\varphi : U \to \widehat{U}$ is a homeomorphism from U to an open subset $\widehat{U}$ of $\mathbb{R}^d$. [78, p.4]*

Definition 1.3.2 (Topological manifold)**.** *[78, p.3] A topological manifold of dimension d is a Hausdorff topological space $\mathcal{M}$ with a countable base such that any point of $\mathcal{M}$ belongs to a d-dimensional chart on $\mathcal{M}$.*

Example 1.3.3. *Circles and triangles (see Fig. 1.1) are then 1-dimensional manifolds because locally, these shapes can be continuously mapped to open intervals of $\mathbb{R}$.*

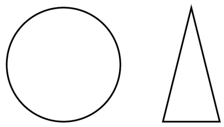


Figure 1.1: Examples of 1-dimensional manifolds.

Remark 1.3.4 (Boundary)**.** *Manifolds given by the definitions above are called 'manifolds without boundary'. Another category of manifolds are 'manifolds with boundary', where some points (the boundary points) have no neighborhood comparable with $\mathbb{R}^d_+$ but rather locally homeomorphic to closed half spaces $\mathbb{H}^d := \{x \in \mathbb{R}^d : x_1 \geq 0\}$. For the sake of simplicity, we do not treat this category of manifold here.*

Definition 1.3.5 (Local coordinate system)**.** *A local coordinate system $x^1, x^2, \ldots, x^d$ is a d-uple of functions from an open subset U of $\mathcal{M}$ to $\mathbb{R}$ such that $(U, (x^1, x^2, \ldots, x^d))$ is a chart of $\mathcal{M}$.*

Notation 1.3.6. *Within the scope of this section, the letters p and q will be used to denote points on the manifold $\mathcal{M}$, while x and y will represent the corresponding local coordinates of these points (when the choice of local chart is clear). However, in some other sections, when the number of mathematical objects under consideration increases considerably, the symbols p and q will be saved for other mathematical objects, while x and y will be used to describe points on the manifolds as well.*

Normally, for convenience, given a chart (U, φ), we identify the region U with its image $\varphi(U)$. Hence in practice, when the choice of local chart is clear, we use interchangeably a point p and its local coordinates $x^1(p), x^2(p), \ldots, x^d(p)$ to denote a same object.

Given two charts (U, ϕ) and (V, ψ) of a d-dimensional topological manifold, then on their intersection $U \cap V$, one can define two different local coordinate systems, say $x^1, x^2, \ldots, x^d$ and $y^1, y^2, \ldots, y^d$ corresponding respectively to ϕ and ψ. Thus, a point $p \in U \cap V$ can be represented by either of two tuples of local coordinates $(x^1(p), x^2(p), \ldots, x^d(p))$ and $(y^1(p), y^2(p), \ldots, y^d(p))$. The **change of the coordinates** (or **transition map**) from the local coordinate system (x^i) to (y^i) is given then by the continuous function $\psi \circ \phi^{-1}$.

1.3.2 Smooth structures, smooth manifolds, smooth functions

Although topological manifolds have their own appeal, many important applications of manifold theory require us to move beyond topology into the realm where calculus is the primary language.

To achieve this, it is not hard to see that a concept of 'smoothness' on manifolds is necessary. In what follows, we will define a *smooth manifold* as a topological manifold endowed with an additional *smooth structure* which can be understood as a structure that determines the manifold smoothness.

Definition 1.3.7 (Smooth structure)**.** *[78, p.4, p.13] A* ***smooth atlas*** $\mathscr{A}$ *on* $\mathcal{M}$ *is a collection of charts on* $\mathcal{M}$*, whose domains cover* $\mathcal{M}$*, such that any two charts in* $\mathscr{A}$ *are smoothly compatible, i.e., the transition map* $\psi \circ \varphi^{-1}$ *is smooth for any pair of charts* (ψ, φ) *with intersecting domains.*
A ***smooth structure*** $\mathscr{A}$ *on a topological manifold* $\mathcal{M}$ *is a maximal smooth atlas on* $\mathcal{M}$*, i.e., any chart that is smoothly compatible with all charts in* $\mathscr{A}$ *must also be a chart in* $\mathscr{A}$*.*

Definition 1.3.8 (Smooth manifold)**.** *[78, p.13] A* ***smooth manifold*** *is then a pair* $(\mathcal{M}, \mathscr{A})$ *with* $\mathscr{A}$ *being a smooth structure on a topological manifold* $\mathcal{M}$*. In other words, a smooth manifold is a topological manifold endowed with a* $\mathcal{C}^\infty$*-structure.*

By a *smooth chart* on a smooth manifold, we will always mean a chart from its $\mathcal{C}^\infty$ atlas. In the sequel, since we will only deal with smooth manifolds and smooth charts, when there is no ambiguity, the term *"chart"* will be used as a synonymous term for *"smooth chart"* and the term *"manifold"* will be used as a synonymous term for *"smooth manifold"*.

Remark 1.3.9. *Smooth structure of a manifold is not uniquely determined by its topological structure as smoothness is not invariant under homeomorphism. Indeed, Milnor has shown in [87] that one can construct two different smooth structures on the* 7*-dimensional sphere* $\mathbb{S}^7$*.*

On smooth manifolds, smooth functions are defined as

Definition 1.3.10 (Smooth function, derivation)**.** *[78, p.32]*
Given a smooth manifold $(\mathcal{M}, \mathscr{A})$*, a function* $f : \mathcal{M} \to \mathbb{R}$ *is said to be* ***smooth*** *if every* ***local coordinate representation*** *of* f*, i.e.* $f \circ \varphi^{-1}$ *with* $\varphi \in \mathscr{A}$*, is smooth. The space of smooth functions on* $\mathcal{M}$ *is denoted by* $\mathcal{C}^\infty(M)$*.*

We see that smooth functions on manifolds are mainly described locally via local charts. The following theorem is an important property of smooth manifolds that provides a tool to globalize local properties of manifolds by ‘gluing’ the local charts.

Theorem 1.3.11 (Partition of unity)**.** *[78, Th. 2.23] Let* $\{\Omega_\alpha\}_{\alpha \in A}$ *be an abitrary open cover of a smooth manifold* $\mathcal{M}$*. Then there exists a family* $\{\psi_\alpha\}_{\alpha \in A}$ *of functions of* $\mathcal{C}^\infty(\mathcal{M})$ *such that:*

(i) $0 \le \psi_\alpha \le 1$ *for all* $\alpha \in A$*.*

(ii) supp$(\psi_\alpha) \subset \Omega_\alpha$*.*

(iii) The family of supports $\{supp(\psi_\alpha)\}_{\alpha \in A}$ *is locally finite, i.e., any point in* $\mathcal{M}$ *has a neighborhood that intersects with supp*(ψ_α) *for only finitely values of* A*.*

(iv) $\sum_{\alpha \in A} \psi_\alpha(p) = 1$ *for all point* p *in* $\mathcal{M}$*.*

A such family of functions is called a ***partition of unity*** *of* $\mathcal{M}$ *subordinate to the open cover* $\{\Omega_\alpha\}_{\alpha \in A}$*.*

A particular case of the above theorem is when the open cover is chosen to be $\{U, \mathcal{M} \setminus K\}$ where K is a closed set contained in U. In this case, we imply the existence of a smooth function $0 \le \phi \le 1$ such that $\phi = 1$ on K and $\phi = 0$ outside U. Such a function ϕ is called a *cutoff function* of K in U.

1.3.3 Tangent vectors, vector fields, differentials, covectors

Let $\mathcal{M}$ be a smooth manifold. In order to make sense of calculus on manifolds, we need to define the tangent space at a given point on a manifold. For example, recall that in accordance with the laws of mechanics, a curve can be obtained by specifying the velocity vector along the displacement. Velocity vectors are tangent vectors to the curve. When $\mathcal{M}$ is a lower-dimensional subspace of $\mathbb{R}^m$, each tangent space of $\mathcal{M}$ can be understood as the set of vectors in $\mathbb{R}^m$ passing through the point under consideration on $\mathcal{M}$ and tangent to $\mathcal{M}$. This definition is often referred to as the geometric definition of the tangent space [78, p.51]. However, such a definition depends on the choice of the ambient space of $\mathcal{M}$. To better capture the intrinsic nature of tangent vectors, we have the following definition, which is commonly known as the algebraic definition of tangent vectors.

Definition 1.3.12 (Derivation, Tangent vectors, Tangent space). *A linear mapping $v : \mathcal{C}^\infty(\mathcal{M}) \to \mathbb{R}$ is called a **derivation** at a point $p \in \mathcal{M}$ if it satisfies the following product rule:*

$$v(fg) = g(p)v(f) + f(p)v(g), \tag{1.4}$$

for all $f, g \in \mathcal{C}^\infty(\mathcal{M})$.
*The set of all derivations at a point p is called the **tangent space** of $\mathcal{M}$ at point p, denoted by $T_p\mathcal{M}$. Each element of $T_p\mathcal{M}$ (that is, derivations at z), is also called **tangent vectors** of $\mathcal{M}$ at p.*

Remark 1.3.13 (Tangent spaces of $\mathbb{R}^m$). *For $\mathcal{M} = \mathbb{R}^m$ and any point $p \in \mathbb{R}^m$, the tangent space $T_p\mathbb{R}^m$ can be identified with $\mathbb{R}^m$ by the identification:*

$$\begin{aligned} \mathbb{R}^m &\longrightarrow T_p\mathbb{R}^m \\ v &\longmapsto D_v\big|_p, \end{aligned}$$

where $D_v\big|_p$ is the directional derivative at p with respect to vector v.
Indeed, this mapping is a linear injection between two tangent spaces with the same dimension (c.f. Theorem 1.3.14).

It is easy to check that $T_p\mathcal{M}$ is vector space over $\mathbb{R}$ and that even though $\mathcal{C}^\infty(\mathcal{M})$ is an infinite-dimensional vector space, the space $T_p\mathcal{M}$ is finite-dimensional.

Theorem 1.3.14. *[78, Prop. 3.10] If $\mathcal{M}$ is a smooth manifold of dimension d, for all $p \in \mathcal{M}$, $T_p\mathcal{M}$ is a vector space of the same dimension d.*

The proof of Theorem 1.3.14 is essentially based on a Taylor-like expansion and the remark that for all derivation v at p and smooth functions f, g such $f(p) = g(p) = 0$, we have $v(gf) = 0$.

Example 1.3.15 (Basis of tangent space $T_p\mathcal{M}$). *Fix a local chart (U, φ) on $\mathcal{M}$ and a point $p \in U$. For any smooth function f on $\mathcal{M}$, the local representation of f under (U, φ) is just a multivariable function on $\mathbb{R}^d$(Note that $\varphi(U)$ is an open subset of $\mathbb{R}^d$). Therefore, the usual partial derivative along the i-th coordinate of the local representation, i.e.,*

$$f \mapsto \frac{\partial\big(f \circ \varphi^{-1}\big)}{\partial x^i}\bigg|_{\varphi(p)},$$

defines a derivation of $\mathcal{M}$ at p. In the sequel, by abuse of notation, this derivation is denoted by $\dfrac{\partial}{\partial x^i}\bigg|_p$.

Proposition 1.3.16 (Basis of tangent spaces). *[78, p.60] If $\mathcal{M}$ is a smooth manifold of dimension d, for any point p and any local coordinate system (x^i) around p, the tangent vectors $\dfrac{\partial}{\partial x^1}\bigg|_p, \dfrac{\partial}{\partial x^2}\bigg|_p, \dots, \dfrac{\partial}{\partial x^d}\bigg|_p$ form a basis of $T_p\mathcal{M}$.*

Definition 1.3.17 (Vector field). *[78, p.174] A vector field V on $\mathcal{M}$ is a family $\{V_p\}_{p\in\mathcal{M}}$ of tangent vectors of $\mathcal{M}$ such that associates each point $p \in \mathcal{M}$ with a tangent vector $V_p \in T_p\mathcal{M}$ at this point.*

By Proposition 1.3.16, in a local coordinates $(x^1, x^2, \dots, x^d)$, the vector field V can be represented in the form:

$$V_p = \sum_{i=1}^{d} V^i(p) \frac{\partial}{\partial x^i}\bigg|_p, \tag{1.5}$$

for some real functions $V^1, V^2, \dots, V^n$ on the domain of the local coordinate system.

Definition 1.3.18 (Smooth vector field). *[78, p.175] The vector field V is said to be smooth if for any smooth local coordinate system, the corresponding functions $V^1, V^2, \dots, V^d$ are smooth.*

Now, let p be a point in $\mathcal{M}$ and f be a smooth function on $\mathcal{M}$. The mapping $T_p\mathcal{M} \ni v \mapsto v(f)$ defines a linear mapping on $T_p\mathcal{M}$. This mapping is called **differential of f at p.**

Definition 1.3.19 (Differential). *[78, p.62] [59, p.56] Fix a point p in $\mathcal{M}$ and let f be a smooth function in a neighborhood of p. The differential $\mathrm{d}f_p$ at p is a linear function on $T_p\mathcal{M}$ given by*

$$\langle \mathrm{d}f_p, v\rangle = v(f) \text{ for any } v \in T_p\mathcal{M}. \tag{1.6}$$

Thus, $\mathrm{d}f_p$ is an element of the dual space $T_p^*\mathcal{M}$ of $T_p\mathcal{M}$, which is also called a **cotangent space**. The elements of $T_p^*\mathcal{M}$ are then called **covectors**.
Any basis $\{e_1, e_2, \dots, e_d\}$ in $T_p\mathcal{M}$ has a dual basis $\{e^1, e^2, \dots, e^d\}$ in the dual space $T_p^*\mathcal{M}$, which is defined by:

$$\langle e^i, e_j\rangle = \delta^i_j := \begin{cases} 1, & i = j, \\ 0, & \text{otherwise.} \end{cases}$$

For example, the basis $\left\{\frac{\partial}{\partial x^i}\big|_p\right\}_{1\le i\le d}$ has the dual $\{\mathrm{d}(x^i)_p\}_{1\le i\le d}$ because

$$\langle \mathrm{d}(x^i)_p, \frac{\partial}{\partial x^j}\big|_p\rangle = \frac{\partial}{\partial x^j} x^i = \delta^i_j.$$

Hence, the covector $\mathrm{d}f_p$ can be represented in the basis $\{\mathrm{d}x^i\}_{1\le i\le d}$ as follows:

$$\mathrm{d}f_p = \sum_{i=1}^{d} \frac{\partial f}{\partial x^i}\bigg|_p \mathrm{d}x^i_p. \tag{1.7}$$

Indeed, for any $j = 1, \dots, n$.

$$\sum_{i=1}^{d} \langle \frac{\partial f}{\partial x^i} \mathrm{d}x^i_p, \frac{\partial}{\partial x^j}\rangle = \sum_{i=1}^{d} \frac{\partial f}{\partial x^i} \langle \mathrm{d}x^i_p, \frac{\partial}{\partial x^j}\rangle = \sum_{i=1}^{d} \frac{\partial f}{\partial x^i} \delta^i_j = \frac{\partial f}{\partial x^j} = \langle \mathrm{d}f, \frac{\partial}{\partial x^j}\rangle.$$

Notice that in the general framework of smooth mappings between two manifolds, the differential has a more general definition, say, as a linear mapping between tangent spaces, but we will not explain this aspect here for the sake of simplicity.

1.3.4 Smooth map, smooth curve

So far, we have focused on real-valued functions on a manifold $\mathcal{M}$, meaning functions that map $\mathcal{M}$ to $\mathbb{R}$. Now that smooth manifolds are well-defined, we can extend this concept to mappings between two manifolds. For example, to define a local representation of a mapping F between manifolds, we use local charts not only to describe the domain of F but also to describe its codomain.

Definition 1.3.20 (Local representation of maps between manifolds, smooth maps, diffeomorphism)**.** *[78, p.34] Given two manifolds $\mathcal{M}$ and $\mathcal{N}$, and a mapping $F : \mathcal{M} \to \mathcal{N}$.*
*A **local representation** of F is a function of form $\psi^{-1} \circ F \circ \varphi$, where (U, φ) and (V, ψ) are respectively local charts of $\mathcal{M}$ and $\mathcal{N}$ such that, for the sake of well-definedness, the image $F(U)$ is included in V (see Figure 1.2).*
*The map $F : \mathcal{M} \to \mathcal{N}$ is said to be **smooth** if every local representation of F is smooth.*

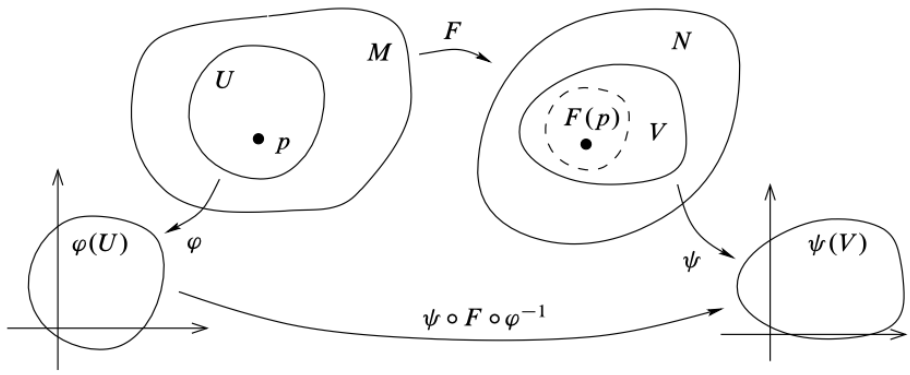
For smooth maps, their differentials are defined to be linear mappings between tangent spaces.

Definition 1.3.21 (Differential of smooth maps)**.** *[78, p.55] Given two manifolds $\mathcal{M}$ and $\mathcal{N}$, and a mapping $F : \mathcal{M} \to \mathcal{N}$. For any point $p \in \mathcal{M}$, the **differential** $\mathrm{d}F_p$ of F at point p is the linear mapping $\mathrm{d}F_p : T_p\mathcal{M} \to T_{F(p)}\mathcal{N}$ such that:*

$$\langle \mathrm{d}F_p(v), g \rangle = v(F \circ g) \quad \forall g \in \mathcal{C}^\infty(\mathcal{M}).$$

Remark 1.3.22. *When $\mathcal{N} \equiv \mathbb{R}$, the smooth map F is indeed a real function. In this case, notice that the tangent spaces of $\mathbb{R}$ can be identified with $\mathbb{R}$ (c.f. Remark 1.3.13), the above definition is therefore consistent with the one given in Definition 1.3.19.*

Figure 1.2: Definition of smooth maps (source: [78, Fig. 2.2]).



One of the smooth mappings we often encounter is a **smooth curve**.

Definition 1.3.23 (Smooth curve)**.** *A smooth curve is a smooth map of the form $\gamma : I \to \mathcal{M}$, where I is an interval of $\mathbb{R}$. In this case, the derivative of γ at t, denoted by $\gamma'(t)$, is a vector (i.e. derivation) in $T_{\gamma(t)}\mathcal{M}$ defined by:*

$$\left\langle \gamma'(y), f \right\rangle = \lim_{s \to t} \frac{f(\gamma(t)) - f(\gamma(s)}{t - s} = \frac{\partial (f \circ \gamma)}{\partial t}\Big|_t. \tag{1.8}$$

Note that when viewed as a manifold, a closed interval like $[0, 1]$ falls into the category of manifolds with boundary [78, p.25]. For simplicity, the definition of such manifolds is not covered in this document. However, specific cases, including closed intervals and half-open intervals, will be used when studying smooth curves.

Remark 1.3.24. *If $\mathcal{M}$ is a manifold without boundary, any given smooth curve $\gamma : I \to \mathcal{M}$ can always be extended to another smooth curve $\tilde{\gamma} : \tilde{I} \to \mathcal{M}$, where $\tilde{I}$ is an open interval.*

1.3.5 Smooth embedding, submanifolds

While abstract manifolds are the primary focus of differential geometry, substructures like circles and spheres within a Euclidean space $\mathbb{R}^m$ are visually more intuitive to perceive. These substructures are called **submanifolds** of $\mathbb{R}^m$. The main interest of this section is to define this mathematical object.

Definition 1.3.25 (Smooth immersion, smooth embedding, embedded smooth manifold)**.** *[78, p. 78, 98] Given two manifolds $\mathcal{M}$ and $\mathcal{N}$.*
*A mapping $F : \mathcal{M} \to \mathcal{N}$ is called a **smooth immersion** if $\mathrm{d}F_p : T_p\mathcal{M} \to T_{F(p)}\mathcal{N}$ is injective for all $p \in \mathcal{M}$. This map F is further called a **smooth embedding** if F is also a topological embedding, i.e., F is a homeomorphism between $\mathcal{M}$ and $F(\mathcal{M})$.*
*An **embedded submanifold** of $\mathcal{M}$ is a subset $\mathcal{S}$ of $\mathcal{M}$ which is a topological manifold in the subspace topology, and endowed with a smooth structure with respect to which the inclusion map $\iota : \mathcal{S} \hookrightarrow \mathcal{M}$ is a smooth embedding.*

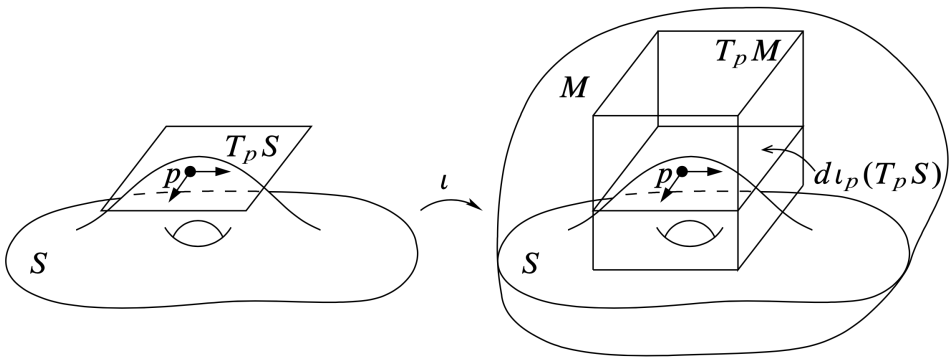
Example 1.3.26.

(a) The unit circle $\mathbb{S}^1$ is an embedded submanifold $\mathbb{R}^2$. The torus $\mathbb{T}^2$ is an embedded submanifold of $\mathbb{R}^3$.

(b) If W is an open set in $\mathbb{R}^n$ and H is any affine subspace of $\mathbb{R}^n$ that cuts through W, then $W \cap H$ is a smooth embedded submanifold of $\mathbb{R}^n$.

Example 1.3.26(b) provides a trivial way to construct submanifolds of Eulidean spaces. Indeed, any embedded submanifold locally has this type of relation with its ambient space, up to diffeomorphism.

Figure 1.3: The tangent space to an embedded submanifold (source: [78, Fig. 5.12]).



Remark 1.3.27 (Tangent spaces of submanifolds are linear subspaces of the tangent spaces of its ambient manifolds)**.** *Given an embedded submanifold $\mathcal{S}$ of $\mathcal{M}$ and the inclusion map $\iota : \mathcal{S} \hookrightarrow \mathcal{M}$, by definition, for all $p \in \mathcal{S}$, the differential $d\,\iota_p : T_p\mathcal{M} \to T_p\mathcal{M}$ is an injective linear mapping. In other words, each tangent vector v in $T\mathcal{S}$ is associated with a unique tangent vector space $d\,\iota_p(v)$ in $T\mathcal{M}$. Thus, from here one, we adopt the convention of identifying $T_p\mathcal{S}$ with its image under this map $d\,\iota$, thereby thinking of $T_p\mathcal{S}$ as a certain linear subspace of $T_p\mathcal{M}$ (see Figure 1.3).*

Definition 1.3.28 (Vector fields tangent to submanifolds)**.** *Let $\mathcal{S}$ be an embedded submanifold of $\mathcal{M}$, and V be a smooth vector field on $\mathcal{M}$. V is said to be tangent to $\mathcal{S}$ if $V_p \in T_p\mathcal{S}$ for all $p \in \mathcal{S}$.*

1.4 Fundamentals of differential geometry II: Riemannian geometry

In order to make sense of distances and volumes on a smooth manifold $\mathcal{M}$, a metric on $\mathcal{M}$ has to be conceptualized. Riemannian geometry studies this aspect and its implications. The main references for this section are *"Introduction to Riemannian manifolds"* by Lee [79] and *"Riemannian Geometry"* by do Carmo [41]. We will next see how this metric structure allows to define the gradient, ∇ and divergence operators that are the bases to introduce the Laplace-Beltrami operator and later the operators that will be the infinitesimal generators of diffusions on manifolds.

1.4.1 Riemannian metric, gradient operator

Let $\mathcal{M}$ be a smooth d-dimensional manifold. A *Riemannian metric* (or a Riemannian *metric tensor*) on $\mathcal{M}$ is a family $\mathbf{g} = \{\mathbf{g}(p)\}_{p\in\mathcal{M}}$ such that for each p, $\mathbf{g}(p)$ is a symmetric, positive definite, bilinear form on the tangent space $T_p\mathcal{M}$, smoothly depending on $p \in \mathcal{M}$.
Using the metric tensor, one can define an inner product $\langle\cdot,\cdot\rangle_{\mathbf{g}}$ in any tangent space $T_p\mathcal{M}$ by

$$\langle\eta,\xi\rangle_{\mathbf{g}} \equiv \mathbf{g}(p)(\eta,\xi),$$

for all tangent vectors $\eta, \xi \in T_p\mathcal{M}$. Hence, $T_p\mathcal{M}$ becomes a Euclidean space and the length of any tangent vector ξ is defined as

$$|\xi|_{\mathbf{g}} = \sqrt{\langle\xi,\xi\rangle_{\mathbf{g}}}.$$

Given a local coordinate system $x^1, x^2, \ldots, x^d$, the above inner product of $T_p\mathcal{M}$ has the form

$$\langle\eta,\xi\rangle_{\mathbf{g}} = \sum_{i,j=1}^{d} g_{ij}(p)\eta^i\xi^j,$$

where $(g_{ij}(p))_{i,j=1}^d$ is a symmetric positive definite $n\times n$ matrix, and $(\eta^i), (\xi^i)$ are coordinates of η, ξ in the basis $\left\{\frac{\partial}{\partial x^i}\big|_p\right\}_{1\le i\le d}$(see Equation (1.5)). The condition that *"$\mathbf{g}(p)$ smoothly depends on p"* means that all the components g_{ij} are $\mathcal{C}^\infty$-functions in the corresponding charts.

Definition 1.4.1. *[40, p.38] A Riemannian manifold is a couple $(\mathcal{M},\mathbf{g})$ where $\mathbf{g}$ is a Riemannian metric on a smooth manifold $\mathcal{M}$.*

Let $(\mathcal{M},\mathbf{g})$ be a Riemannian manifold. The metric tensor $\mathbf{g}$ provides a canonical way to identify the tangent space $T_p\mathcal{M}$ with the cotangent space $T_p^*\mathcal{M}$. Indeed, for any vector $\eta \in T_p\mathcal{M}$, denote by $\widehat{\mathbf{g}}(p)\eta$ a covector that is defined by the identity

$$\langle\widehat{\mathbf{g}}(p)\eta,\xi\rangle = \langle\eta,\xi\rangle_{\mathbf{g}}. \tag{1.9}$$

It is easy to check that the mapping $\widehat{\mathbf{g}}(p) : T_p\mathcal{M} \to T_p^*\mathcal{M}$ is linear and that given a local coordinate system $x^1, x^2, \ldots, x^d$, the matrix representation of $\widehat{\mathbf{g}}(p)$ with respect to the basis $\left\{\frac{\partial}{\partial x^i}\big|_p\right\}_{1\le i\le d}$ of $T_p\mathcal{M}$ and the basis $\left\{\mathrm{d}x^i\right\}_{1\le i\le d}$ of $T_p^*\mathcal{M}$ is $(g_{ij})_{i,j=1}^d$.
Besides, if $\xi \neq 0$, $\mathbf{g}(p)\xi$ is also non-zero because $\langle\mathbf{g}(p)\xi,\xi\rangle = \langle\xi,\xi\rangle_{\mathbf{g}} > 0$. Thus, $\mathbf{g}(p)$ is inversible with the inverse mapping

$$\widehat{\mathbf{g}}^{-1}(p) : T_p^*\mathcal{M} \to T_p\mathcal{M}$$

whose the matrix representation in the above bases denoted by $(g^{ij})_{i,j=1}^d$ that satisfies:

$$(g^{ij})_{i,j=1}^d = \left[(g_{ij})_{i,j=1}^d\right]^{-1}. \tag{1.10}$$

For any smooth function f on $\mathcal{M}$, we define its **gradient** $\nabla f(p)$ at a point $p \in \mathcal{M}$ by:

$$\nabla f(p) := \widehat{\mathbf{g}}^{-1}(p)(\mathrm{d}f_p), \tag{1.11}$$

that is, $\nabla f(p)$ is a tangent vector version of $\mathrm{d}f_p$. By the above remark on the matrix representation of $\widehat{\mathbf{g}}^{-1}$ and Equation (1.7), given a local coordinate system, the component of the gradient ∇f can be calculated explicitly as follows:

$$(\nabla f)^i := \sum_{j=1}^{d} g^{ij}\frac{\partial f}{\partial x^j}, \tag{1.12}$$

that is, $\nabla f = \sum_{i,j=1}^{d} g^{ij} \frac{\partial f}{\partial x^j} \frac{\partial}{\partial x^i}$. Now, if h is another smooth function on $\mathcal{M}$, by using (1.12) and the fact that (g^{ij}) is the inverse of (g_{ij}), we imply that:

$$\langle \nabla f, \nabla h \rangle_{\mathbf{g}} = \sum_{i,j=1}^{d} g^{ij} \frac{\partial f}{\partial x^i} \frac{\partial h}{\partial x^j}. \tag{1.13}$$

We end this Section by showing for all smooth functions f, h, we have:

$$\nabla(fh) = h\nabla(f) + f\nabla(h). \tag{1.14}$$

Indeed, for any tangent vector v, by definition,

$$\langle \mathrm{d}(fg), v \rangle = v(fh) = fv(h) + hv(f) = f\langle \mathrm{d}h, v \rangle + h\langle \mathrm{d}f, v \rangle.$$

Thus, $\mathrm{d}(fh) = f\mathrm{d}h + h\mathrm{d}f$. Hence, $\widehat{\mathbf{g}}^{-1}(p)\mathrm{d}(fg) = \widehat{\mathbf{g}}^{-1}(p)\left(f(p)\mathrm{d}h + h(p)\mathrm{d}f\right)$. Hence, we imply (1.14).

1.4.2 Length, distance and volume

One of the main purposes of Riemannian metric is to rigorously define curve lengths, distances and volumes on manifolds.
Let $\gamma : I \to \mathcal{M}$ be a smooth curve on the Riemannian manifold $(\mathcal{M}, \mathbf{g})$. The **length** of γ is defined as, [79, p.34],

$$\mathrm{Length}(\gamma) = \int_I |\gamma'(t)|_{\mathbf{g}} \mathrm{d}t, \tag{1.15}$$

where $\gamma'(t)$ has been defined in (1.8). This definition of curve length allows us to construct the distance between points on a Riemannian manifold.

Definition 1.4.2 (Geodesic distance)**.** *A smooth curve $\gamma : I \to \mathcal{M}$ is a geodesic if for all $t \in I$, $\nabla\gamma'(t) = 0$.*
The geodesic distance between any two points p, q is the shortest length among all curves with the same endpoints p and q:

$$d_{\mathcal{M}}(p, q) = \inf\{\mathrm{Length}(\gamma) \mid \gamma : I = [a, b] \to \mathcal{M},\ \gamma(a) = p,\ \gamma(b) = q\}.$$

For geodesic curves, the acceleration $\nabla\gamma'(t)$ vanishes for any $t \in I$, which means that the velocity $|\gamma'(t)|_{\mathbf{g}}$ is constant and the length of the curve between any two of its points is extremal, since it has a zero derivative.

Denote by $\mathcal{B}(\mathcal{M})$ the smallest σ-algebra containing all open subsets of $\mathcal{M}$. The purpose of this section is to show that there is a canonical volume measure μ of $\mathcal{M}$ defined on the measurable space $(\mathcal{M}, \mathcal{B}(\mathcal{M}))$, which is called the *Riemannian measure (or volume)* of $\mathcal{M}$.

Theorem 1.4.3. *[59, p.59] For any Riemannian manifold $(\mathcal{M}, \mathbf{g})$, there is a unique measure μ on $\mathcal{B}(\mathcal{M})$ such that in any chart (U, ϕ),*

$$\mathrm{d}\mu = \sqrt{\det g}\, \mathrm{d}\lambda, \tag{1.16}$$

where $g = (g_{ij})$ is the matrix representation of the Riemannian metric $\mathbf{g}$ in the local chart U, and λ is the Lebesque measure of $\mathbb{R}^d$.

There are various proofs for this theorem. Here, we propose one.

A proof of Theorem 1.4.3. Because $\mathcal{M}$ is second countable and locally compact, there is a countable family of smooth charts $\{(U_i, \phi_i)\}_{i\in\mathbb{N}}$ that covers M and $\overline{U_i}$ is compact for each i. Let $\{\psi_i\}$ be a partition of unity of M subordinate to $\{U_i\}_{i\in\mathbb{N}}$. Let $\tilde{g}^i$ denote the representation of $\mathbf{g}$ in the local chart (U_i, ϕ_i). We consider the following linear mapping $T : \mathcal{C}_c(\mathcal{M}) \to \mathbb{R}$

$$T(f) := \sum_{i=0}^{\infty} \int_{\phi_i(U_i)} (f \circ \phi_i)(\psi_i \circ \phi_i)\sqrt{\det \tilde{g}^i}\, \mathrm{d}\lambda. \tag{1.17}$$

Roughly speaking, $T(f)$ is the integration of f against μ and each term on the right-hand side is the integration of $f\psi_i$ on U_i. Note that (1.17) is well defined because the family of the supports of $\{\psi_i\}$ are locally finite, hence the compact set supp$f$ intersects with finitely many supp(ψ_i). Clearly, T is a positive linear mapping. Hence, by Riesz's representation theorem [104, Thm 2.1] and the fact that $\mathcal{M}$ is a locally compact Hausdorff space, we imply that there is a measure μ on $(\mathcal{M}, \mathcal{B}(\mathcal{M}))$ such that $\int f\mathrm{d}\mu = T(f)$.
To prove that μ is indeed the Riemannian measure we are looking for, it is clearly sufficient to show that given any chart (U, ϕ),

$$\mu(U) = \int_{\phi(U)} \sqrt{\det h}\mathrm{d}\lambda$$

where h is the representation of $\mathbf{g}$ in the local chart (U, ϕ). This is then reduced to prove that for all $i \in \mathbb{N}$,

$$\int_{\phi_i(U\cap U_i)} (\psi_i \circ \phi_i)\sqrt{\det \tilde{g}^i}\, \mathrm{d}\lambda = \int_{\phi(U\cap U_i)} (\psi_i \circ \phi)\sqrt{\det h}\, \mathrm{d}\lambda,$$

which is just a formula of change of coordinates (see Section 1.3.1, or [40, p.44]). □

1.4.3 Divergence theorem, Laplace operator, Green formula

For any smooth vector field X on a Riemannian manifold $(\mathcal{M}, \mathbf{g})$, its *divergence* div$X$ is a smooth function on $\mathcal{M}$, defined via the following theorem

Theorem 1.4.4 (A particular case of Theorem 16.32 [79])**.** *For any smooth vector field X on a Riemannian manifold $(\mathcal{M}, \mathbf{g})$, there is a unique smooth function on $\mathcal{M}$, denoted by* div$X$, *such that the following identity holds*

$$\int_{\mathcal{M}} (\mathrm{div}X)u\mathrm{d}\mu = -\int_{\mathcal{M}} \langle X, \nabla u\rangle_{\mathbf{g}}\mathrm{d}\mu, \tag{1.18}$$

for all $u \in \mathcal{C}_c^\infty(\mathcal{M})$.

Before explaining the proof of this theorem, let us take a smooth chart (U, ϕ) of $\mathcal{M}$, by (1.11) and (1.9), we obtain for any function $u \in \mathcal{C}_c^\infty(U)$,

$$\begin{aligned}
\int_U \langle \nabla u, X\rangle_{\mathbf{g}}\mathrm{d}\mu &= \int_U \langle \mathbf{g}^{-1}(\mathrm{d}u), X\rangle_{\mathbf{g}}\mathrm{d}\mu \\
&= \int_U \langle \mathrm{d}u, X\rangle\mathrm{d}\mu \\
&= \sum_{k=1}^{d} \int_U \frac{\partial u}{\partial x^k} X^k \sqrt{\det g}\mathrm{d}\lambda \\
&= -\sum_{k=1}^{d} \int_U u\frac{\partial}{\partial x^k}(X^k\sqrt{\det g})\mathrm{d}\lambda \\
&= -\sum_{k=1}^{d} \int_U u\frac{1}{\sqrt{\det g}}\frac{\partial}{\partial x^k}(X^k\sqrt{\det g})\mathrm{d}\lambda
\end{aligned}$$

Thus by comparing to Equation (1.18), we see that the divergence of X on (U, ϕ) can be chosen as:

$$\operatorname{div} X = \sum_{k=1}^{d} \frac{1}{\sqrt{\det g}} \frac{\partial}{\partial x^k} (X^k \sqrt{\det g}). \tag{1.19}$$

Clearly, the uniqueness in the above theorem implies that the formula in Equation (1.19) does not depend the choice of ϕ in (U, ϕ).

Sketch of the proof of Theorem 1.4.4. We begin with proving the uniqueness and it is easy to check that the uniqueness follows from the fact that if a continuous function g satisfies that $\int_{\mathcal{M}} g f d\mu = 0$ for all $f \in \mathcal{C}_c^\infty(\mathcal{M})$, g must be identical to 0.
The uniqueness then shows that the formula provided in Equation (1.19) does not depend our choice of local chart. Hence, Equation (1.19) can be extended to define a smooth function $\operatorname{div} X$ on $\mathcal{M}$. Furthermore, this function $\operatorname{div} X$ satisfies that for any local chart (U, ϕ) and any function $u \in \mathcal{C}_c^\infty(U)$, we have that:

$$\int_{\mathcal{M}} (\operatorname{div} X) u \mathrm{d}\mu = - \int_{\mathcal{M}} \langle X, \nabla u \rangle \mathrm{d}\mu.$$

It is easy to show this property is still true if $u \in \mathcal{C}_c^\infty(\mathcal{M})$ by using a partition of unity of $\mathcal{M}$ and noting that $\nabla(uv) = u\nabla v + v\nabla u$ and $\nabla(1) = 0$. □

Having defined gradient and divergence, we can now define the *Laplace operator* (called also the *Laplace-Beltrami* operator) on any Riemannian manifold $(\mathcal{M}, \mathbf{g})$ as follows:

$$\Delta = \operatorname{div} \circ \nabla. \tag{1.20}$$

That is, for any smooth function f on $\mathcal{M}$, $\Delta f = \operatorname{div}(\nabla f)$, so Δf is a smooth function on $\mathcal{M}$. This can be formulated in terms of local coordinates.

Definition 1.4.5. *[79, Prop 2.46] The* ***Laplace–Beltrami operator*** *on the manifold $\mathcal{M}$ is the unique linear operator $\Delta : \mathcal{C}^\infty(\mathcal{M}) \to \mathcal{C}^\infty(\mathcal{M})$ such that for any smooth function f and any smooth local coordinates (x^i) on an open set $U \subseteq \mathcal{M}$:*

$$\Delta f = \frac{1}{\sqrt{\det \mathbf{g}}} \sum_{i=1}^{d} \frac{\partial}{\partial x^i} \left(\sum_{j=1}^{d} \mathbf{g}^{ij} \sqrt{\det \mathbf{g}} \frac{\partial f}{\partial x^j} \right), \tag{1.21}$$

where $\det \mathbf{g} = \det(\mathbf{g}_{ij})$ is the determinant of the component matrix of $\mathbf{g}$ in these coordinates, and $(\mathbf{g}^{ij})$ are coefficients of the inverse matrix of $(\mathbf{g}_{ij})_{1 \leq i,j \leq d}$.

If $\mathcal{M}$ is a Euclidean space, i.e., $\mathcal{M} = \mathbb{R}^m$, its Laplacian Δ can easily be expressed as a finite sum of second derivatives, and Definition 1.4.5 can be seen as a generalization of Laplacian operators in Euclidean spaces $\mathbb{R}^m$ to manifolds. These operators play a central role in the study of both heat equation on manifolds and Brownian motion on manifolds, as will be seen in what follows. In the general case, when not in $\mathbb{R}^m$, the explicit expression for Δ is more complicated. Fortunately, with additional assumptions about the embedding of $\mathcal{M}$, such a formulation can be achieved.

Theorem 1.4.6 (Hörmander formulation for Laplacian)**.** *[67, Thm 3.1.4]*
Suppose that $\mathcal{M}$ is a submanifold of the Euclidean space $\mathbb{R}^m$ with induced metric. Let $\{\xi_\alpha\}_{1 \leqslant \alpha \leqslant m}$ be the standard orthonormal basis on $\mathbb{R}^m$. For each $x \in \mathcal{M}$, let $P_\alpha(x)$ be the orthogonal projection of ξ_α to $T_x\mathcal{M}$. Then, we have:

$$\Delta = \sum_{\alpha=1}^{m} P_\alpha^2, \tag{1.22}$$

In other words, for any smooth function f, $\Delta f = \sum_{\alpha=1}^{m} P_\alpha(P_\alpha(f))$.

Remark 1.4.7. *The number of vector fields P_i here is m, which is always greater than the actual dimension of $\mathcal{M}$.*

Theorem 1.4.8. *If u, v are two smooth functions on a Riemannian manifold $(\mathcal{M}, \mathbf{g})$ and one of them has a compact support then :*

$$\int_{\mathcal{M}} u\Delta v \mathrm{d}\mu = -\int_{\mathcal{M}} \langle \nabla u, \nabla v \rangle_{\mathbf{g}} \mathrm{d}\mu = \int_{\mathcal{M}} v\Delta u \mathrm{d}\mu. \tag{1.23}$$

Sketch of the proof of Theorem 1.4.8. If both functions have compact supports, the above theorem follows directly from the divergence theorem (Theorem 1.4.4). When either one of them does not have a compact support, we can use a cutoff function based on the support of the other function to revert the problem to the previous situation. □

From this point to the rest of this section, we assume that readers are familiar with most of the standard notions in theory of smooth manifolds, even with the notions we did not mention, say, *smooth mappings*, *tangent space* $T\mathcal{M}$, *flow* .

1.4.4 Normal coordinates

Given a Riemannian manifold $(\mathcal{M}, \mathbf{g})$, a *normal coordinate chart* of $\mathcal{M}$ is a kind of local charts of $\mathcal{M}$ that is specifically made to bear many convenient geometric properties. For its definition, existence and uniqueness, we refer the interested readers to [79, p 131-132]. In this section, we will only recap the properties in which we are interested.

Theorem 1.4.9. *[Derivatives of Riemannian metrics in normal coordinate charts][79, Prop. 5.24] Let $\Phi : \mathcal{M} \supset U \to \mathbb{R}^d$ be a normal coordinate chart at a point x in $\mathcal{M}$ such that $\Phi(x) = 0$ and $(g_{ij}; 1 \leq i, j \leq d)$ be the local representation of the Riemannian metric of $\mathcal{M}$ in the coordinate chart Φ. We have that for all i, j,*

$$g_{ij}(0) = \delta_{ij}, \quad g'_{ij}(0) = 0. \tag{1.24}$$

To illustrate the benefits of using normal coordinates in computations, we consider the problem of calculating the Laplacian and the gradient at a point x of smooth functions $f, h : \mathcal{M} \to \mathbb{R}$.

Proposition 1.4.10. *Suppose that $\Phi : \mathcal{M} \supset U \to \mathbb{R}^d$ be a normal coordinate chart at a point x in $\mathcal{M}$ such that $\Phi(x) = 0$, then:*

i. $\Delta f(p) = \Delta \hat{f}(0)$,

ii. $\langle \nabla f(p), \nabla h(p) \rangle = \langle \nabla \hat{f}(0), \nabla \hat{h}(0) \rangle$,

where $\hat{f}, \hat{h}$ are the local representation of f and h in the local chart Φ.

Notice that in Equality *i.* of Proposition 1.4.10, the symbol Δ on the left hand side stands for the Laplace-Beltrami operator while the symbol Δ on the right hand side stands for the usual Laplacian in $\mathbb{R}^d$. The same remark applies for Equality *ii.*.

Proof for Proposition 1.4.10. Let (g_{ij}) be the local representation of the Riemannian metric $\mathbf{g}$ in the local coordinate chart Φ. By Theorem 1.4.9, we know that $g_{ij}(0) = \delta_{ij}$, thus, $g^{ij} = \delta_{ij}$. (Recall that (g^{ij}) is defined as the inverse matrix of (g_{ij})). Hence, by comparing with Equation 1.13, we have Equality *ii.*.
For Equality *i.*, we will use Equation (1.21) at $x = 0$. By noticing that $\det g(0) = 1$ and that every term involving derivatives of g_{ij} and g^{ij} in the above formula vanishes, we have the conclusion. □

1.4.5 Tensor, differential forms, exterior derivative

We can upgrade the notions of tangent spaces and gradient to higher dimensions. Hessian operators, volume measures and integration theory are extended by tensors and differential k-forms. This section can be skipped in a first reading. One of its main result is that the Laplace Beltrami operator can be expressed in terms of these objects.

The **tensor bundle** and **exterior derivative** on manifolds are additional concepts in differential geometry, built on top of the notion of the tangent space. These concepts are inspired by their counterparts in tensor algebra and introduce an additional (algebraic) structure to the existing differential objects.

Definition 1.4.11 ((k,l)-Tensor Bundle)**.** *[78, chapter 12]*
Let V be a finite-dimensional vector space. The ***space of*** (k,l)***-tensors*** *on V is defined as:*

1. *$T^{(k,l)}V = T^k(V) \otimes T^l(V^*)$, where $T^k(V)$ denotes the k-fold tensor product of V, and $T^l(V^*)$ denotes the l-fold tensor product of the dual space V^*.*

For a smooth manifold $\mathcal{M}$, we construct the corresponding tensor bundles as follows:

2. *The* ***bundle of covariant*** k***-tensors*** *on $\mathcal{M}$ is given by:*

$$T^k(T^*\mathcal{M}) := \coprod_{x\in\mathcal{M}} T^k(T_x^*\mathcal{M}),$$

where $T_x^*\mathcal{M}$ is the cotangent space at $x \in \mathcal{M}$, and $\coprod$ denotes the disjoint union over all points in $\mathcal{M}$. This bundle collects all covariant k-tensors at each point of the manifold.

3. *The (k,l)**-tensor bundle** over $\mathcal{M}$ is defined as:*

$$T^{(k,l)}T\mathcal{M} := \coprod_{x\in\mathcal{M}} T^{(k,l)}(T_x\mathcal{M}),$$

where $T^{(k,l)}(T_x\mathcal{M})$ is the space of (k,l)-tensors at the point $x \in \mathcal{M}$, constructed from the tangent space $T_x\mathcal{M}$ and its dual $T_x^*\mathcal{M}$.

In this definition, the (k,l)-tensor bundle encompasses all possible tensors that are contravariant of order k and covariant of order l at each point of the manifold $\mathcal{M}$.

Definition 1.4.12. *[78, p.360] The* ***space of all smooth differential*** k***-forms*** *on $\mathcal{M}$ is denoted by $\Omega^k(\mathcal{M})$ and is defined as:*

$$\Omega^k(\mathcal{M}) = \Gamma\left(\Lambda^k T^*\mathcal{M}\right).$$

Here, Γ denotes the space of smooth sections of a vector bundle, and $\Lambda^k T^*\mathcal{M}$ represents the k-th exterior power of the cotangent bundle $T^*\mathcal{M}$.
Then, we define

$$\Omega^*(\mathcal{M}) = \bigoplus_{i=0}^{k} \Omega^k(\mathcal{M}),$$

Remark 1.4.13. *The wedge product on Ω^k for each k turns $\Omega^*(\mathcal{M})$ into an associative, anti-commutative graded algebra.*

Differential forms are antisymmetric tensor fields that play a crucial role in calculus on manifolds. They generalize the concepts of functions and vector fields and are fundamental in the formulation of integration on manifolds, Stokes' theorem, and de Rham cohomology.

Lemma 1.4.14 (Characterization of Differential Forms). *[78, p. 318]*
A map

$$\mathcal{A} : \underbrace{\mathcal{C}^\infty(\mathcal{M}) \times \cdots \times \mathcal{C}^\infty(\mathcal{M})}_{k \text{ copies}} \to \mathcal{C}^\infty(\mathcal{M})$$

is induced by a smooth covariant k-tensor if and only if it is multilinear over $\mathcal{C}^\infty(\mathcal{M})$. Moreover, if $\mathcal{A}$ is alternating (i.e., it changes sign upon swapping any two of its arguments), then it is induced by a differential k-form.

This lemma provides a characterization of tensor fields and differential forms in terms of their action on smooth vector fields.

Definition 1.4.15 (Pullback of Differential Forms). *[78, p.284]*
*Given a smooth map $F : \mathcal{M} \to \mathcal{N}$ between smooth manifolds, the **pullback** $F^* : \Omega(\mathcal{N}) \to \Omega(\mathcal{M})$ is defined by:*

$$F^*(\omega)(X_1, \ldots, X_k) = \omega\left(F_* X_1, \ldots, F_* X_k\right),$$

for all $k \in \mathbb{N}$, $\omega \in \Omega^k(\mathcal{N})$, and $X_1, \ldots, X_k \in \mathcal{C}^\infty(\mathcal{M})$.
Here, $F_* : T_x\mathcal{M} \to T_{F(x)}\mathcal{N}$ is the differential (pushforward) of F at $x \in \mathcal{M}$. The pullback F^* allows us to transfer differential forms from $\mathcal{N}$ back to $\mathcal{M}$ via the map F, enabling the comparison and manipulation of forms on different manifolds.

The pullback operation is fundamental in differential geometry, as it preserves the differential structure when mapping forms between manifolds. It ensures that the integral of a form over a manifold corresponds to the integral of its pullback over the preimage under F, aligning with the change of variables in integration.

Definition 1.4.16 (Exterior Derivative). *[78, Thm 14.24]*
*There exists a unique graded map $\mathrm{d} : \Omega^*(\mathcal{M}) \to \Omega^*(\mathcal{M})$, called the **exterior derivative** on $\mathcal{M}$, satisfying the following properties:*

1. ***Linearity:*** *d is linear over $\mathbb{R}$.*

2. ***Action on Functions:*** *For all $f \in \mathcal{C}^\infty(\mathcal{M}) = \Omega^0(\mathcal{M})$ and $X \in \mathfrak{X}(\mathcal{M})$, the exterior derivative acts as:*

$$\mathrm{d}f(X) = Xf,$$

   where Xf denotes the derivative of f in the direction of X.

3. ***Nilpotency:*** *The exterior derivative satisfies $\mathrm{d} \circ \mathrm{d} = 0$, meaning that applying it twice yields zero.*

4. ***Leibniz Rule (Graded Product Rule):*** *For all $\omega \in \Omega^k(\mathcal{M})$ and $\eta \in \Omega^l(\mathcal{M})$,*

$$\mathrm{d}(\omega \wedge \eta) = (\mathrm{d}\omega) \wedge \eta + (-1)^k \omega \wedge \mathrm{d}\eta.$$

The exterior derivative generalizes the concept of differentiation to higher-degree differential forms. It is a crucial operator in differential geometry and topology, playing a central role in de Rham cohomology, Stokes' theorem, and the theory of differential equations on manifolds. The property $\mathrm{d} \circ \mathrm{d} = 0$ leads to the concept of closed and exact forms, which are essential in the study of the topology of manifolds.

Notation 1.4.17. *(Notation conflict) When d has multiple meanings, such as representing a differential in a differential equation, we use $\mathbf{d}$ instead of d to denote the exterior derivative.*

The following lemma highlights the **naturality** of the exterior derivative with respect to smooth maps between manifolds. The pullback operation preserves the algebraic and differential structures of forms, making it a homomorphism of differential graded algebras.

Lemma 1.4.18 (Naturality of the Exterior Derivative)**.** *[78, p. 366]*
Given a smooth map $F : \mathcal{M} \to \mathcal{N}$ between smooth manifolds, the following properties hold:

1. ***Linearity of Pullback:*** *The pullback $F^* : \Omega^*(\mathcal{N}) \to \Omega^*(\mathcal{M})$ is linear over $\mathbb{R}$.*
2. ***Compatibility with Wedge Product:*** *For all $\omega, \eta \in \Omega^*(\mathcal{N})$,*
$$F^*(\omega \wedge \eta) = F^*(\omega) \wedge F^*(\eta).$$
3. ***Commutation with Exterior Derivative:*** *The pullback commutes with the exterior derivative, i.e.,*
$$F^*(\mathrm{d}\omega) = \mathrm{d}F^*(\omega),$$
for all $\omega \in \Omega^*(\mathcal{N})$.

Proposition 1.4.19. *The Laplace–Beltrami operator on the manifold $\mathcal{M}$ defined in* (1.21) *can be rewritten using the exterior derivative and its pullback:*

$$\Delta u = \mathrm{d} \otimes \mathrm{d}^*.$$

1.4.6 Linear connection and covariant derivatives

When exploring classical results or developing new ones on manifolds, we often begin by translating fundamental concepts from well-understood spaces such as the Euclidean space $\mathbb{R}^d$ into their corresponding, more general versions on curved spaces like manifolds. The primary focus of this section is to introduce one such translation. In particular, we will examine how to define higher-order derivatives of functions on manifolds. To achieve this, we first introduce the concept of a linear connection. Essentially, a linear connection is a rule that enables us to differentiate vector fields along curves on a manifold; this operation is also known as the covariant derivative. Once the covariant derivative is well-defined for vector fields, it can be naturally and uniquely extended to more advanced objects, such as tensor fields.
This section is organized into three main parts. In the first part (Section 1.4.6.1), we revisit the standard concept of covariant derivatives for multivariable functions in $\mathbb{R}^d$. Next, in Section 1.4.6.2, we define the linear connection and covariant derivatives of functions on manifolds. Finally, as an illustrative example, we present a coordinate-free Taylor expansion formula for functions on manifolds in Section 1.4.6.3.
The main references for this section are chapter 5 and chapter 6 in [79].

1.4.6.1 Covariant derivatives in $\mathbb{R}^d$

Given a multivariable function $p : \mathbb{R}^d \to \mathbb{R}$, the kth-order covariant derivative of p at a point x, denoted by $\nabla^k p(x)$ (and occasionally by $p^{(k)}(x)$ or $\nabla^k p\big|_x$ in this thesis), is defined as the multilinear map

$$\nabla^k p(x) : \underbrace{\mathbb{R}^d \times \cdots \times \mathbb{R}^d}_{k \text{ times}} \to \mathbb{R},$$

given by

$$\nabla^k p(x)(v_1, \ldots, v_k) = \sum_{(\alpha_1,\ldots,\alpha_k)\in\{1,\ldots,d\}^k} v_1^{\alpha_1} \cdots v_k^{\alpha_k} \frac{\partial^k p}{\partial x^{\alpha_1} \cdots \partial x^{\alpha_k}}(x).$$

Although this definition may appear cumbersome, it is merely a generalization of the standard derivative. For example, when $k = 1$, $\nabla^1 p(x)$ coincides with the gradient $\nabla p(x)$ since, for every vector $v \in \mathbb{R}^d$,

$$\langle \nabla p(x), v \rangle_{\mathbb{R}^d} = \nabla^1 p(x)(v).$$

Similarly, when $k = 2$, $\nabla^2 p(x)$ is exactly the Hessian matrix $\mathbf{H}p(x)$ of p, in the sense that for all $v_1, v_2 \in \mathbb{R}^d$,

$$\langle v_1, \mathbf{H}p(x)\, v_2\rangle_{\mathbb{R}^d} = \nabla^2 p(x)(v_1, v_2).$$

Using this objects, the Taylor expansion of $p(x)$ about 0 up to order k can be expressed as follows:

$$p(x) = \sum_{i=0}^{k} \frac{1}{i!}\nabla^i p(0)(x^{\times k}) + \frac{1}{k!}\int_0^1 (1-s)^k \nabla^{k+1} p(sx)(x^{\times (k+1)})\mathrm{d}s.$$

where $x^{\times i}$ denotes the element $\underbrace{(x, ..., x)}_{i \text{ times}}$ in $(\mathbb{R}^d)^i$. Note that, this formulation of Taylor expansion is of minimal usefulness in multivariable calculus in practice. It only serves as an illustrative example for our use case.

1.4.6.2 Linear connection and covariant derivatives of functions on manifolds

One significant advantage of $\mathbb{R}^d$ over a general manifold in performing derivative computations is that all tangent spaces can be naturally identified with a single vector space. For example, given a tangent vector $v \in T_x\mathbb{R}^d$, for any other point y different form x in $\mathbb{R}^d$ the tangent space $T_y\mathbb{R}^d$ is canonically identified with $\mathbb{R}^d$, so that the corresponding tangent vector can also be regarded as v. This clear correspondence does not hold on a general manifold. In fact, on a general manifold there exist multiple ways to establish such identifications via different linear connections. The process of establishing these identifications is known as *parallel transport*, which depends entirely on the chosen linear connection and the paths connecting the points. We will not discuss parallel transport here, and will focus solely on the concept of a linear connection.

Definition 1.4.20 (Linear connection). *[79, p.91] A linear connection on $\mathcal{M}$ is a map*

$$\nabla \colon \mathfrak{X}(\mathcal{M}) \times \mathfrak{X}(\mathcal{M}) \longrightarrow \mathfrak{X}(\mathcal{M}),$$

written $(X, Y) \mapsto \nabla_X Y$, satisfying the following properties:

(i) $\nabla_X Y$ is linear over $\mathcal{C}^\infty(\mathcal{M})$ in X. In other words, for all $f_1, f_2 \in \mathcal{C}^\infty(\mathcal{M})$ and $X_1, X_2 \in \mathfrak{X}(\mathcal{M})$,

$$\nabla_{f_1 X_1 + f_2 X_2} Y = f_1 \nabla_{X_1} Y + f_2 \nabla_{X_2} Y.$$

(ii) $\nabla_X Y$ is linear over $\mathbb{R}$ in Y. That is, for all $a_1, a_2 \in \mathbb{R}$ and $Y_1, Y_2 \in \mathfrak{X}(\mathcal{M})$,

$$\nabla_X \big(a_1 Y_1 + a_2 Y_2\big) = a_1 \nabla_X Y_1 + a_2 \nabla_X Y_2.$$

(iii) $\nabla_X(fY) = f\nabla_X Y + \big(Xf\big) Y$ for all $f \in \mathcal{C}^\infty(\mathcal{M})$,

where $\mathcal{M}$ denotes the space of all smooth vector fields on $\mathcal{M}$.

Informally, for any point $x \in \mathcal{M}$, $\nabla_X Y\big|_x$ is the "derivative" the vector field Y at x in the direction $X(x)$. Indeed, the "locality" of the directional derivative is still reserved for linear connection. That is,

Proposition 1.4.21 (Locality of linear connection). *([79, Proposition 4.5])$\nabla_X Y\big|_x$ depends only on the values of Y in a neighborhood of x and on the value of X at x.*

In the context of Riemannian geometry, though we will not go into the full details here, every Riemannian metric on a manifold defines a unique linear connection known as the Levi-Civita connection(cf. [79, Thm 5.10]). Consequently, unless otherwise specified, any mention of a linear

connection or related concepts on a Riemannian manifold should be understood as referring to this canonical Levi-Civita connection.
We now discuss the extensibility of linear connection. Note that no Riemannian structure is assumed in the subsequent discussion.

Covariant derivatives of tensor fields

Every linear connection on a manifold defines a unique procedure for differentiating smooth tensor fields.

Proposition 1.4.22. *([79, Prop. 4.15]). Let $\mathcal{M}$ be a smooth manifold with or without boundary, and let ∇ be a connection in $T\mathcal{M}$. Then ∇ extends uniquely to each tensor bundle $T^{(k,l)}(\mathcal{M})$, also denoted by ∇, so that the following conditions are satisfied:*

(i) In $T^{(1,0)}(\mathcal{M}) = T\mathcal{M}$, ∇ agrees with the given connection.

(ii) In $T^{(0,0)}(\mathcal{M}) = \mathcal{M} \times \mathbb{R}$, ∇ is given by ordinary differentiation of functions:

$$\nabla_X f \;=\; Xf.$$

(iii) For any tensor fields F, G of appropriate types,

$$\nabla_X\big(F \otimes G\big) \;=\; \big(\nabla_X F\big) \otimes G \;+\; F \otimes \big(\nabla_X G\big).$$

(iv) ∇ commutes with all contractions: if "tr" denotes a trace on any pair of indices, one covariant and one contravariant, then

$$\nabla_X\big(\operatorname{tr} r\big) \;=\; \operatorname{tr}\big(\nabla_X r\big).$$

We are now ready to discuss covariant derivatives of a smooth function.

First order derivative, problem of notations

For a smooth function $p\colon \mathcal{M} \to \mathbb{R}$, as discussed in previous sections, ∇p is defined as the mapping:

$$X \in \mathfrak{X}(\mathcal{M}) \mapsto X(p).$$

In other words, ∇p is simply the 1-form $\mathrm{d}p$.
Note that the gradient operator ∇ does not exist on a general manifold; it arises only in the presence of a Riemannian metric, which provides a canonical correspondence between 1-forms and tangent vectors.
Furthermore, as one should have observed by now, the notation ∇ is used in various contexts within differential geometry. Its precise meaning should be inferred from the surrounding geometric framework.

Second derivative and symmetry

The second order covariant derivative $\nabla^2 p$ of p, which is also called covariant Hessian, is then defined by as:

$$\nabla^2 p(Y, X) = X(Yp) - (\nabla_X Y)p, \tag{1.25}$$

where X and Y are smooth vector fields on $\mathcal{M}$. In general, there is no symmetry in X and Y for a general covariant Hessian. Nonetheless, within the framework of Riemannian metric, this symmetric is valid.

Higher order covariant derivatives

For any $k \geq 1$, the k-th covariant derivative $\nabla^k p$ of p is then the $\mathcal{C}^\infty(\mathcal{M})$ multi-linear mapping

$$\nabla^k p : \underbrace{\mathfrak{X}(\mathcal{M}) \times \cdots \times \mathfrak{X}(\mathcal{M})}_{k \text{ times}} \to \mathcal{C}^\infty(\mathcal{M})$$

defined by the recursive relation:

$$\nabla^k p := \nabla(\nabla^{k-1} p). \tag{1.26}$$

Note that these objects do have local representations using Christoffel symbols. However, they are not in the scope of this introduction.

1.4.6.3 A Taylor expansion formula for functions on manifolds

In this section, to illustrate a concrete application of the above constructions, we present a formulation for the Taylor expansion of smooth functions on manifolds. A key advantage of this approach is that it is coordinate-free, meaning it does not depend on any local charts.

Definition 1.4.23. *([79, p.103]) A smooth curve γ on a smooth manifold $\mathcal{M}$ is called geodesic with respect to a linear connection ∇ if:*

$$\nabla_{\dot{\gamma}}\dot{\gamma} = 0.$$

Theorem 1.4.24. *Given a manifold $\mathcal{M}$ with a linear connection $\nabla : \mathcal{M} \times \mathcal{M} \to \mathcal{M}$. For any smooth function p on $\mathcal{M}$, two points x and y on $\mathcal{M}$, a smooth geodesic curve $\gamma : [0,1] \to \mathcal{M}$ connecting x and y, that is $\gamma(0) = x$ and $\gamma(1) = y$. For any positive integer k, we have that:*

$$p(y) - p(x) = \sum_{i=1}^{k} \frac{1}{i!} \nabla^i p\big|_x (\underbrace{\dot{\gamma}(0), \dot{\gamma}(0), ..., \dot{\gamma}(0)}_{i \ times}) + \frac{1}{k!} \int_0^1 (1-s)^k \nabla^{k+1} p\big|_{\gamma(s)} (\underbrace{\dot{\gamma}(s), \dot{\gamma}(s), ..., \dot{\gamma}(s)}_{k+1 \ times}) \mathrm{d}s.$$

Proof. Consider $F(t) = p(\gamma(t))$. It is sufficient to show that for any k and $s \in (0,1)$

$$F^{(k)}(s) = \nabla^k p\big|_{\gamma(s)} (\underbrace{\dot{\gamma}(s), \dot{\gamma}(s), ..., \dot{\gamma}(s)}_{k \text{ times}}).$$

Clearly, this is true for $k = 0$. Suppose this is true for l, we will prove that it is true for $k := l+1$. Indeed, firstly, we have:

$$\frac{\mathrm{d}F^{(l)}}{\mathrm{d}s} = \nabla_{\dot{\gamma}} \left(\nabla^l p\big|_{\gamma} (\underbrace{\dot{\gamma}, \dot{\gamma}, ..., \dot{\gamma}}_{l \text{ times}}) \right).$$

Then, after Eq 4.12 in [79, p. 96] and the fact that $\nabla_{\dot{\gamma}}\dot{\gamma} = 0$, we see that:

$$\nabla_{\dot{\gamma}} \left(\nabla^l p\big|_{\gamma} (\underbrace{\dot{\gamma}, \dot{\gamma}, ..., \dot{\gamma}}_{l \text{ times}}) \right) = \left(\nabla_{\dot{\gamma}} \nabla^l p \right) (\underbrace{\dot{\gamma}, \dot{\gamma}, ..., \dot{\gamma}}_{l \text{ times}}).$$

Finally, after Eq 4.14 [79, p. 97], we have:

$$\left(\nabla_{\dot{\gamma}} \nabla^l p \right) (\underbrace{\dot{\gamma}, \dot{\gamma}, ..., \dot{\gamma}}_{l \text{ times}}) = \nabla^{l+1} p (\underbrace{\dot{\gamma}, \dot{\gamma}, ..., \dot{\gamma}}_{l+1 \text{ times}}).$$

Therefore, we have the desired conclusion. □

1.4.7 Frame bundle and horizontal lift

It is not always easy and beneficial to analyze a Riemannian manifold through its ambient space. In this section, we will analyze the lifting of our initial manifold $\mathcal{M}$ to a more abstract manifold called the **frame bundle**. This lift is the central concept of the Eells-Elworthy-Malliavin construction of Brownian motion and remains the standard approach to study Brownian motion and its variants on manifolds.

Our primary reference for this section is Chapter 2 of [67]. Our modest contribution lies in refining the regularity aspects of various statements. We also suppose in this subsection that readers are fairly familiar with Riemannian geometry.

Definition 1.4.25. *A frame at $x \in \mathcal{M}$ is an $\mathbb{R}$-vector space isomorphism $u : \mathbb{R}^d \to T_x\mathcal{M}$ between $\mathbb{R}^d$ and $T_x\mathcal{M}$.*

In other words, suppose that $\{e_i\}_{1\leqslant i\leqslant d}$ is the standard orthonormal basis on $\mathbb{R}^d$, then $ue_1, ue_2, ..., ue_d$ make up a basis (equivalently, a frame) for $T_x\mathcal{M}$.

Definition 1.4.26 (Frame bundle)**.** *We use* $\mathrm{F}(\mathcal{M})_x$ *to denote the space of all frames at* x*. Then the frame bundle of* $\mathcal{M}$ *is the disjoint union:*

$$\mathrm{F}(\mathcal{M}) := \coprod_{x\in\mathcal{M}} \mathrm{F}(\mathcal{M})_x.$$

Hence, the ***projection*** $\mathbf{p} : \mathrm{F}(\mathcal{M}) \to \mathcal{M}$ *is a bundle over* $\mathcal{M}$*. This bundle* $\mathcal{M}$ *naturally makes into a smooth bundle of dimension* $d + d^2$*.*

Suppose $\mathcal{M}$ is equipped with a connection $\nabla : \Gamma(T\mathcal{M}) \times \Gamma(T\mathcal{M}) \to \Gamma(T\mathcal{M})$.

Definition 1.4.27 (Horizontal curve)**.** *Let* I *be an open interval in* $\mathbb{R}$*. A horizontal curve* $(u_t)_{t\in I}$ *in* $\mathcal{M}$ *is a* $\mathcal{C}^1$*- choice of frames in* $\mathrm{F}(\mathcal{M})$ *such that for every vector* $v \in \mathbb{R}^d$*,* $(u_t v)$ *is parallel along* $(\mathbf{p}u_t)$*:.*

On the one hand, a horizontal curve in $\mathrm{F}\mathcal{M}$ clearly defines a unique $\mathcal{C}^1$-curve in $\mathcal{M}$. The vice versa is also true, that is,

Proposition 1.4.28. *Given a* $\mathcal{C}^1$*-curve* (γ_t) *in* $\mathcal{M}$ *and a frame* $u \in \mathrm{F}(\mathcal{M})_x$ *at* $x \in \mathcal{M}$*, there is a unique horizontal curve* (u_t) *along* (γ_t) *such that* $u_0 = u$*. Moreover,* $\gamma \in \mathcal{C}^k$ *if and only if* $u \in \mathcal{C}^k$*.*

Then a tangent vector of the framebundle $\mathrm{F}(\mathcal{M})$ is defined to be *horizontal* as follows:

Definition 1.4.29 (Horizontal vector)**.** *For* $u \in \mathrm{F}(\mathcal{M})$*, a tangent vector* $X \in T_u\mathrm{F}(\mathcal{M})$ *is said to be horizontal if it is tangent to a horizontal curve.*

In other words, X is horizontal if there is a horizontal curve (u_t) such that $u_0 = u$ and for all $F \in \mathcal{C}^1(\mathrm{F}(\mathcal{M}))$,

$$\frac{\mathrm{d}}{\mathrm{d}t}F(u_t) = XF.$$

Having defined the concept *'horizontal'*, we can now introduce the main concept of this section: the **horizontal lift**.

Definition 1.4.30 (Horizontal lift of a tangent vector of $\mathcal{M}$)**.** *Given a tangent vector* $z \in T_x\mathcal{M}$*, for* $x \in \mathcal{M}$*, the horizontal lift* $\mathbf{H}_z$ *of* z *at* $u \in \mathrm{F}(\mathcal{M})_x$ *is the tangent vector at* x *of any horizontal curve* (u_t) *in* $\mathrm{F}(\mathcal{M})$ *such that* $u_0 = u$ *and* $\dot{x}(0) = z \in T_{x(0)}\mathcal{M}$ *where* $x(t) := \mathbf{p}u_t$*.*

We use $\mathrm{H}_u\mathrm{F}(\mathcal{M})$ and $\mathrm{HF}(\mathcal{M})$ to denote respectively the space of all horizontal vector at u, and the horizontal bundle of $\mathrm{F}(\mathcal{M})$. Indeeds, $\mathrm{HF}(\mathcal{M}) \to \mathrm{F}(\mathcal{M})$ is a vector bundle of rank d [67, p.38].

Definition 1.4.31. *For any vector* $v \in \mathbb{R}^d$*, the horizontal lift* H_v *of* v *is the vector field* $u \mapsto \mathbf{H}_{uz} \in T_u\mathrm{F}(\mathcal{M})$*.*

Hence, for any vector field $Z \in \Gamma(T\mathcal{M})$, $\mathbf{H}_Z$ can be rewritten as $u \mapsto H_{u^{-1}Z}(u)$. Therefore, $\mathbf{H}_Z$ is $\mathcal{C}^k$ if and only if $\mathbf{H}_Z \in \mathcal{C}^k$.
We use H_i to denote H_{e_i} for each i.

Remark 1.4.32. $\mathbf{H}_v(f \circ \mathbf{p}) = (vf) \circ \mathbf{p}$ *for all* $f \in \mathcal{C}^1(\mathcal{M})$*,* $v \in T\mathcal{M}$*.*

1.5 Fundamentals of Operator theory

In this section, we lay out foundational concepts in operator theory, particularly focusing on unbounded operators in Hilbert spaces. We have explored the definitions and properties of densely defined, closed, symmetric, and self-adjoint operators, which are essential in understanding the behavior of differential operators like the Laplacian. Additionally, we have introduced weighted Riemannian manifolds and the weighted Laplacian, providing insights into how weight functions influence spectral properties and the analysis on manifolds. These concepts will serve as building blocks for further exploration of functional analysis and its applications in mathematical physics and differential geometry.
In this section, we introduce essential concepts related to unbounded operators on Hilbert spaces, which are particularly important when dealing with differential operators like the Laplacian.
Let H be a separable Hilbert space, with inner product denoted by $\langle \cdot, \cdot \rangle$. Let $\mathcal{A} : D \to H$ be an unbounded linear operator on H, where $D \subset H$ is the domain of $\mathcal{A}$. The operator $\mathcal{A}$ maps elements from its domain D back into H.
We begin by defining several fundamental properties of unbounded operators, which will be crucial in our subsequent discussions.

Definition 1.5.1 (Densely defined operators, Closed operators, Symmetric operators, Positive pperators)**.** *[105, p.347-348]*
We say:

- $\mathcal{A}$ *is* ***densely defined*** *if its domain D is dense in H; that is, for every $u \in H$, there exists a sequence $(u_n) \subset D$ such that $u_n \to u$ in H.*
- $\mathcal{A}$ *is* ***symmetric*** *if*
$$\langle \mathcal{A}u, v \rangle = \langle u, \mathcal{A}v \rangle \quad \textit{for all } u, v \in D,$$
This means that $\mathcal{A}$ equals its adjoint on D.
- $\mathcal{A}$ *is* ***positive*** *if*
$$\langle \mathcal{A}u, u \rangle \geqslant 0 \quad \textit{for all } u \in D.$$
This property ensures that the operator does not decrease the "energy" associated with u.
- $\mathcal{A}$ *is* ***closed*** *if its graph is a closed subset of $H \times H$. Specifically,*
$$\mathrm{Graph}(\mathcal{A}) := \{(u, \mathcal{A}u) \in H \times H : u \in D\}$$
is closed in the Hilbert space $H \times H$ equipped with the product topology.

Understanding these properties is essential because they determine the behavior of the operator and its suitability for analysis in various contexts.

Remark 1.5.2. *An alternative and often practical characterization of a closed operator $\mathcal{A}$ is the following: $\mathcal{A}$ is closed if and only if, for every pair $(u, v) \in H \times H$ and any sequence $(u_n) \subset D$ satisfying*
$$u_n \xrightarrow{H} u \quad \textit{and} \quad \mathcal{A}u_n \xrightarrow{H} v,$$
we have $u \in D$ and $\mathcal{A}u = v$. This means that if both u_n and $\mathcal{A}u_n$ converge in H, then the limit point u belongs to the domain D, and the operator $\mathcal{A}$ acts continuously at u.

Remark 1.5.3. *It is important to note that the domain D of $\mathcal{A}$ does not need to be closed in H, and not every unbounded operator is closed. Some operators can be extended to closed operators, known as their closures, while others cannot.*

In certain cases, we can extend an unbounded operator to a closed operator, which is essential for defining self-adjoint operators and studying their spectral properties.

Proposition 1.5.4 (Closure of a symmetric operator)**.** *If $\mathcal{A}$ is a densely defined symmetric operator on H, then there exists a unique extension $\overline{\mathcal{A}} : \overline{D} \to H$ of $\mathcal{A}$ such that the graph of $\overline{\mathcal{A}}$ is the closure of the graph of $\mathcal{A}$ in $H \times H$.*
Moreover, we have:

1. *$\overline{\mathcal{A}}$ is symmetric.*
2. *The domain $\overline{D}$ of $\overline{\mathcal{A}}$ is the completion of D with respect to the graph norm:*

$$\|u\|_{\mathcal{A}} := \|u\|_H + \|\mathcal{A}u\|_H.$$

Additionally, if $\mathcal{A}$ is positive, then so is $\overline{\mathcal{A}}$.

Proof. It is sufficient to show that the closure of Graph($\mathcal{A}$) in $H \times H$ is indeed a graph of an unbounded operator.
Suppose the otherwise, i.e., there is a sequence $(f_n) \subset D$ and a non-null element $u \in H$ such that with respect to the norm in H,

$$\lim_{n\to\infty} f_n = 0, \quad \lim_{n\to\infty} \mathcal{A}f_n = u$$

Thus, for every element $g \in D$, by symmetry of $\mathcal{A}$, we have:

$$\langle u, g\rangle = \lim_{n\to\infty} \langle \mathcal{A}f_n, g\rangle = \lim_{n\to\infty} \langle f_n, \mathcal{A}g\rangle = 0$$

Besides, D is dense in H. Hence, u must be 0, which is a contradiction. Hence, the conclusion. □

Notation 1.5.5 (Closure)**.** *In the rest of this text, the operator $\overline{\mathcal{A}}$ defined in Proposition 1.5.4 is called the **closure** of $\mathcal{A}$.*

Remark 1.5.6. *The second point in Proposition 1.5.4 also implies that $u \in \overline{D}$ if and only if there exists a sequence $(u_n) \subset D$ such that $u_n \to u$ in H and $(\mathcal{A}u_n)$ is a Cauchy sequence in H. This characterization is particularly useful when dealing with unbounded operators, as it allows us to understand their domains through limits of sequences in D.*

Next, we introduce the concept of the adjoint of an unbounded operator, which generalizes the notion of the transpose of a matrix to infinite-dimensional spaces.

Definition 1.5.7 (Adjoint of an unbounded operator)**.** *[105, p.348] Suppose that $\mathcal{A}$ is densely defined. The **adjoint** $\mathcal{A}^* : D^* \to H$ of $\mathcal{A}$ is an unbounded operator defined by:*

- *The domain D^* consists of all $g \in H$ such that the linear functional*

$$D \ni f \mapsto \langle \mathcal{A}f, g\rangle$$

  is bounded on D with respect to the norm of H.

- *For each $g \in D^*$, there exists a unique element $\mathcal{A}^*g \in H$ such that*

$$\langle \mathcal{A}^*g, f\rangle = \langle g, \mathcal{A}f\rangle \quad \text{for all } f \in D.$$

Remark 1.5.8. *Informally, the adjoint $\mathcal{A}^*$ is the maximal extension of $\mathcal{A}$ that satisfies the relation*

$$\langle \mathcal{A}f, g\rangle = \langle f, \mathcal{A}^*g\rangle \quad \text{for all } f \in D,\ g \in D^*.$$

It captures how $\mathcal{A}$ interacts with the inner product structure of H.

Several important properties relate an operator to its adjoint.

Proposition 1.5.9. *[105, Thm 13.9, Thm 13.11] Suppose that $\mathcal{A}$ is densely defined. Then:*

- *If $\mathcal{A}$ is symmetric, then $\mathcal{A}^*$ extends $\mathcal{A}$.*
- *The graph of $\mathcal{A}^*$ is closed, so $\mathcal{A}^*$ is a closed operator.*
- *If $\mathcal{A}$ is symmetric, then the double adjoint $\mathcal{A}^{**}$ equals the closure of $\mathcal{A}$, that is, $\mathcal{A}^{**} = \overline{\mathcal{A}}$.*

These results highlight the significance of the adjoint operator in understanding the closure and extensions of $\mathcal{A}$.

Definition 1.5.10 (Self-adjoint and Essentially self-adjoint operators)**.** *Suppose $\mathcal{A}$ is densely defined. We say:*

- *$\mathcal{A}$ is **self-adjoint** if $\mathcal{A}^* = \mathcal{A}$; that is, $\mathcal{A}$ coincides with its adjoint.*
- *$\mathcal{A}$ is **essentially self-adjoint** if $\mathcal{A}$ is symmetric and its closure $\overline{\mathcal{A}}$ is self-adjoint; equivalently, $\mathcal{A}^* = \overline{\mathcal{A}}$.*

Self-adjoint operators are crucial in quantum mechanics and spectral theory because they guarantee real eigenvalues and a complete set of eigenfunctions, which are necessary for the physical interpretation of observables.

Proposition 1.5.11. *If $\mathcal{A}$ is densely defined and symmetric, then $\mathcal{A}$ is self-adjoint if and only if its adjoint $\mathcal{A}^*$ is symmetric.*

Proof. If $\mathcal{A}$ is self-adjoint, then $\mathcal{A}^* = \mathcal{A}$ is symmetric by definition. Conversely, if $\mathcal{A}^*$ is symmetric, since $\mathcal{A}^*$ extends $\mathcal{A}$ and both are symmetric, it follows that $\mathcal{A}^* = \mathcal{A}$, so $\mathcal{A}$ is self-adjoint. □

To illustrate these concepts, let us consider an example involving the Laplacian operator.

Example 1.5.12 (The Laplacian on $L^2((0,1))$)**.** *Consider the Laplacian $\mathcal{A} = -\frac{\mathrm{d}^2}{\mathrm{d}x^2}$ defined on the dense subspace $D := \mathcal{C}_c^\infty(0,1)$ of the Hilbert space $H = L^2((0,1))$. Here, $\mathcal{C}_c^\infty(0,1)$ denotes the set of infinitely differentiable functions with compact support in $(0,1)$.*
The closure $\overline{\mathcal{A}}$ of $\mathcal{A}$ is defined on the Sobolev space $H_0^2((0,1))$, which is the completion of $\mathcal{C}_0^\infty(0,1)$ under the Sobolev norm

$$\|f\|_{H_0^2(0,1)} := \|f\|_{L^2(0,1)} + \|\mathcal{A}f\|_{L^2(0,1)}.$$

This space consists of functions in $L^2((0,1))$ whose weak derivatives up to second order are in $L^2((0,1))$ and vanish at the boundary.
Although $\mathcal{A}$ is symmetric and positive on D, its adjoint $\mathcal{A}^*$ is not symmetric and not necessarily positive. For instance, consider the function $f(x) = e^{kx}$ with $k \in \mathbb{C}$. This function may not belong to D but can belong to D^*. We find that

$$\mathcal{A}^* f = -k^2 e^{kx} = -k^2 f,$$

where $-k$ is not necessarily real or positive, depending on k. This example demonstrates that an operator's adjoint can have different properties from the original operator, emphasizing the importance of understanding the domains and closures of operators.

1.5.1 Weighted Riemannian Manifolds and Weighted Laplacians

In differential geometry and analysis, the concept of a weighted manifold extends the classical notion of a Riemannian manifold by incorporating a measure with a smooth density function. This framework is useful in various applications, including probability theory, geometric analysis, and the study of heat kernels. We can refer the interested readers to [59].
Let $(\mathcal{M}, \mathbf{g})$ be a Riemannian manifold, where $\mathbf{g}$ is the Riemannian metric. Let μ be a measure on $\mathcal{M}$.

Definition 1.5.13 (Weighted Riemannian manifold)**.** *The triple $(\mathcal{M}, \mathbf{g}, \mu)$ is called a **weighted Riemannian manifold** if μ is a measure on $\mathcal{M}$ with a smooth positive density p with respect to the Riemannian volume measure $\mathrm{vol}_{\mathcal{M}}$; that is,*

$$\mu = p \,\mathrm{dvol}_{\mathcal{M}}.$$

*For any smooth vector field X on $\mathcal{M}$, the **weighted divergence** $\mathrm{div}_\mu(X)$ is defined by*

$$\mathrm{div}_\mu(X) := \frac{1}{p}\,\mathrm{div}(pX),$$

where div *denotes the usual divergence operator associated with the metric* $\mathbf{g}$*.*

The weighted divergence adjusts the classical divergence to account for the measure μ, incorporating the effect of the density function p.

Definition 1.5.14 (Weighted Laplacian)**.** *For any smooth function f on $\mathcal{M}$, the **weighted Laplacian** $\Delta_\mu f$ is defined as*

$$\Delta_\mu f := \mathrm{div}_\mu(\nabla f),$$

where ∇f is the gradient of f with respect to the metric $\mathbf{g}$*, see Section 1.4.1.*

The weighted Laplacian generalizes the standard Laplace-Beltrami operator by including the weight function p, making it essential in studying diffusion processes and heat flow on weighted manifolds. As an example, for two positive functions $p,\ q \in \mathcal{C}^2$, the operators $\mathcal{A}_{pq}$ given for any test function f of class $\mathcal{C}^2$ on $\mathcal{M}$ by

$$\mathcal{A}_{pq}(f) := q\Delta f + \langle q\nabla \ln(pq), \nabla f\rangle, \tag{1.27}$$

are weighted Laplacians associated with the weighted Riemannian metric $\widetilde{\mathbf{g}} = q\,\mathbf{g}$ and with the measure $\mathrm{d}\mu = p\ \mathrm{dvol}_{\mathcal{M}}$. When we take $q = \frac{p}{2}$, we recover the generator

$$\frac{p}{2}\Delta f + \langle \nabla p, \nabla f\rangle$$

studied in [25, 55, 61]. When $q = 1$, we recover a Langevin diffusion studied in [121] and whose generator $\mathcal{L}$ is defined for any test function f of class $\mathcal{C}^2$ on $\mathcal{M}$ by

$$\mathcal{L}(f) := \Delta f + \langle \nabla \ln p, \nabla f\rangle = \Delta f + \left\langle \frac{\nabla p}{p}, \nabla f \right\rangle. \tag{1.28}$$

Remark 1.5.15. *By applying Green's identity (a generalization of integration by parts), we find that Δ_μ is a symmetric operator on $L^2(\mu)$. Specifically, for any $f, g \in \mathcal{C}^\infty(\mathcal{M})$, we have*

$$\begin{aligned}\int_{\mathcal{M}} (\Delta_\mu f)\, g \,\mathrm{d}\mu &= \int_{\mathcal{M}} \mathrm{div}_\mu(\nabla f)\, g \,\mathrm{d}\mu \\ &= -\int_{\mathcal{M}} \langle \nabla f, \nabla g\rangle_{\mathbf{g}} \,\mathrm{d}\mu \\ &= \int_{\mathcal{M}} f\,(\Delta_\mu g)\,\mathrm{d}\mu.\end{aligned}$$

Here, $\langle\cdot,\cdot\rangle_{\mathbf{g}}$ denotes the inner product induced by the Riemannian metric $\mathbf{g}$*. This symmetry is crucial for spectral analysis and ensures that the operator has real eigenvalues.*

The spectral properties of the weighted Laplacian are of significant interest. Under appropriate conditions, its spectrum consists of a discrete set of eigenvalues that can be compared to those of the standard Laplacian.
Let

$$0 = \lambda_0^\mu < \lambda_1^\mu \leqslant \lambda_2^\mu \leqslant \cdots$$

denote the eigenvalues of $-\Delta_\mu$ (counted with multiplicity). Similarly, let

$$0 = \lambda_0 < \lambda_1 \leqslant \lambda_2 \leqslant \cdots$$

denote the eigenvalues of the usual Laplacian $-\Delta_{\mathcal{M}}$.
To study the behavior of λ_k^μ, we recall the following variational characterization known as the minimax principle.

Theorem 1.5.16 (Minimax Principle)**.** *[59, Theorem 10.18] The eigenvalues of $-\Delta_\mu$ satisfy*

$$\lambda_k^\mu = \sup_{\dim E = k-1} \inf_{f \in E^\perp \setminus \{0\}} \mathcal{R}_\mu(f),$$

where:

- *The supremum is taken over all $(k-1)$-dimensional subspaces E of $W^1(\mathcal{M},\mu)$, the Sobolev space of functions with square-integrable first derivatives.*
- *$E^\perp$ denotes the orthogonal complement of E in $L^2(\mu)$.*
- *The Rayleigh quotient $\mathcal{R}_\mu(f)$ is defined by*

$$\mathcal{R}_\mu(f) := \frac{\int_{\mathcal{M}} |\nabla f|^2 \, \mathrm{d}\mu}{\int_{\mathcal{M}} f^2 \, \mathrm{d}\mu}.$$

Using this principle, we can compare the eigenvalues of the weighted and unweighted Laplacians.

Lemma 1.5.17. *There exists a constant $C > 0$ such that for all $k \geqslant 1$,*

$$C^{-1}\lambda_k \leqslant \lambda_k^\mu \leqslant C\lambda_k.$$

In other words, λ_k^μ and λ_k are comparable up to a constant factor that depends on the weight function p but not on k.

Proof. Let $\mathcal{R}(f)$ denote the Rayleigh quotient with respect to the unweighted measure:

$$\mathcal{R}(f) = \frac{\int_{\mathcal{M}} |\nabla f|^2 \, \mathrm{dvol}_{\mathbf{g}}}{\int_{\mathcal{M}} f^2 \, \mathrm{dvol}_{\mathbf{g}}}.$$

Since p is smooth and positive on $\mathcal{M}$, there exist constants $p_{\min}, p_{\max} > 0$ such that

$$0 < p_{\min} \leqslant p(x) \leqslant p_{\max} \quad \text{for all } x \in \mathcal{M}.$$

Therefore, for any $f \in W^1(\mathcal{M},\mu)$,

$$p_{\min} \int_{\mathcal{M}} |\nabla f|^2 \, \mathrm{dvol}_{\mathbf{g}} \leqslant \int_{\mathcal{M}} |\nabla f|^2 \, \mathrm{d}\mu \leqslant p_{\max} \int_{\mathcal{M}} |\nabla f|^2 \, \mathrm{dvol}_{\mathbf{g}},$$

and similarly for the denominator involving f^2.
Using the minimax principle and these inequalities, we can relate the eigenvalues λ_k^μ and λ_k through the constants $p_{\min}$ and $p_{\max}$, yielding the desired comparison. □

By applying Weyl's asymptotic formula, which describes the behavior of eigenvalues of elliptic operators on compact manifolds, we obtain the following corollary.

Corollary 1.5.18. *For a fixed measure μ, there exists a constant $\kappa > 1$ such that*

$$\kappa^{-1} k^{2/d} \leqslant \lambda_k^\mu \leqslant \kappa k^{2/d},$$

where d is the dimension of the manifold $\mathcal{M}$.

Remark 1.5.19. *An alternative approach to obtaining this result is to use spectral theory for elliptic operators. Specifically, as Δ_μ is an essentially self-adjoint elliptic differential operator on the weighted manifold $(\mathcal{M}, \mathbf{g}, \mu)$, one can apply results from spectral theory, such as those found in [107, Problem 15.4, p. 131], to derive the asymptotic behavior of its eigenvalues.*

These results demonstrate that the introduction of a smooth positive weight function does not drastically alter the spectral properties of the Laplacian. The eigenvalues remain comparable to those of the standard Laplacian, preserving the essential analytical and geometrical features of the manifold.

1.5.2 Semigroup theory

In this section, we will explore some basic concepts of semigroup theory, thereby laying the groundwork for further analysis of SDEs through an approach from an analytical perspective. Our main reference for this section is the book [72] by Engel and Nagel.

Let $(\mathbf{B}, \|\cdot\|)$ be a Banach space. Keep it mind that in most of our cases, $\mathbf{B}$ is L^p-spaces with $1 \leq p \leq \infty$.

Definition 1.5.20. *[72, p.14,p.36,p.40] A family $(P_t)_{t\geq 0}$ of bounded linear operators on $\mathbf{B}$ is called a* ***semigroup*** *on $\mathbf{B}$ if*

$$\begin{cases} P_s P_t = P_{s+t} \text{ for all } s, t \geq 0, \\ P_0 = Id. \end{cases}$$

Then, this semigroup $(P_t)_{t\geq 0}$ is said to be

- ***strongly continuous*** *if $t \mapsto P_t f$ is continuous for all $f \in \mathbf{B}$.*
- ***contractive*** *if $\|P_t\| \leq 1$ for all t.*

Given a strongly continuous contraction semigroup (P_t), its **generator** is defined as:

Definition 1.5.21. *The* ***generator*** *of a semigroup $(P_t)_{t\geq 0}$ is defined as the unbounded operator $\mathcal{A} : \mathcal{D}(\mathcal{A}) \to \mathbf{B}$ such that $\mathcal{D}(\mathcal{A}) = \{f \in \mathbf{B} : \lim_{t\to 0^+} \frac{T_t f - f}{t} \text{ exists}\}$ and*

$$\mathcal{A} f = \lim_{t\to 0^+} \frac{T_t f - f}{t}.$$

Not all operators can be generator of semigroup. Generators of semigroup have many specific properties. For example,

Theorem 1.5.22. *[72, p.51] The generator of a strongly continuous semigroup is closed and densely defined. In addition, it determines the semigroup uniquely.*

More precisely, the following theorem is due to Hille and Yoshida gives the necesssary and sufficient conditions for an unbouned operator $\mathcal{A}$ to be able to generate a strongly continuous contraction semigroup.

Theorem 1.5.23 (Hille-Yoshida). *[72, p.73] An unbounded linear operator $\mathcal{A}$ on* $\mathbf{B}$ *generates a strongly continuous contraction semigroup if and only if:*

- *$\mathcal{A}$ is densely-defined and closed,*
- $\mathrm{Spec}(\mathcal{A}) \subset (-\infty, 0]$*, and*
- $\|(\lambda\mathbf{Id} - \mathcal{A})^{-1}\| \leq \frac{1}{\lambda}$ *for all* $\lambda > 0$*,*

where the spectrum $\mathrm{Spec}(\mathcal{A}) := \{\lambda \in \mathbb{C} : \lambda\mathrm{Id} - \mathcal{A}$ *is not bijective*$\}$.

Remark 1.5.24. *When $\lambda\mathrm{Id} - \mathcal{A}$ is bijective, by closed graph theorem, its inverse $(\lambda\mathbf{Id} - \mathcal{A})^{-1}$ is a continuous closed operator. Therefore, $\|(\lambda\mathbf{Id} - \mathcal{A})^{-1}\| < \infty$.*

Note that in practice, we usually only determine the value of $\mathcal{A}f$ for f within a small, well-behaved subspace D of $\mathcal{D}(\mathcal{A})$, rather than for the entire domain $\mathcal{D}(\mathcal{A})$. In fact, $\mathcal{D}(\mathcal{A})$ is rarely expressed in a simple manner, and $\mathcal{A}f$ in many cases must be defined indirectly, which can introduce unnecessary complexities into the reasoning process. Thus,

Definition 1.5.25. *We also say that an unbouned operator $(\mathcal{A}, D)$ generates a semigroup $(P_t)_{t\geq 0}$ if the closure of $(\mathcal{A}, D)$ is the generator of $(P_t)_{t\geq 0}$.*

1.5.3 Contractivity of semigroups

In this thesis (see Chapter 3), the contractivity properties of the semigroups will be use, as they are related with the long-time behavior or the associated stochastic processes. Let us first introduce this notion (see e.g. [120, Section 2.6.3]).
For p and $q \in [1, +\infty]$, and for an operator P, we define the operator norm:

$$\|P\|_{p\to q} = \sup\big\{\|Pf\|_q : f \in L^p,\ \|f\|_p \leqslant 1\big\}. \tag{1.29}$$

Definition 1.5.26. *A semigroup $(P_t)_{t\geqslant 0}$ is called :*

- *ultracontractive if $\|P_t\|_{2\to\infty} < +\infty$ for any $t > 0$.*
- *supercontractive if $\|P_t\|_{2\to 4} < +\infty$ for any $t > 0$.*
- *hypercontractive if $\|P_t\|_{2\to 4} \leqslant 1$ for some $t > 0$.*

Many criteria associated with measure concentration and functional inequalities exist in the literature and we refer to [120] for a complete exposition.

Lemma 1.5.27. *[120, Theorem 3.5.5.] The semigroup (P_t) associated to the operator $\mathcal{A}$ is ultracontractive. In other words, for each $t > 0$, there is a minimal positive value $u_t > 0$, such that for any bounded measurable function f, we have*

$$\|P_t f\|_\infty \leq u_t \|f\|_{L^1(\mu)}. \tag{1.30}$$

1.5.4 Stochastic differential equations and diffusions on manifolds

In this section, we now explain what are semi-martingales on manifolds and extend the notions of stochastic differential equations seen in Section 1.2 to manifolds. We will revisit several foundational results in the theory of SDEs on manifolds, and our primary focus will be on relaxing the regularity conditions typically imposed on the coefficient functions of SDEs. The main source guiding this exploration is Chapter 1 of Hsu's book [67] where the smoothness of coefficient functions is assumed in general. We can also refer to [68].

1.5.4.1 Semimartingales on $\mathcal{M}$

Definition 1.5.28 ($\mathcal{M}$-valued semimartingale)**.** *Let τ be a stopping time. A continuous, $\mathcal{M}$-valued process X defined on $[0,\tau)$ is called an $\mathcal{M}$-valued semimartingale if f(X) is a real-valued semimartingale on $[0,\tau)$ for all $f \in \mathcal{C}^2(\mathcal{M})$.*

Continuous real-valued martingales can be written as solutions of SDEs driven by Brownian motions (see [69, Section II.2]). On manifold, the stochastic integrals are expressed using the Stratonovitch integral (see [98, page 82]) rather than the Itô integral, that leads to less convenient formulas when working on manifolds. The Stratonovitch integral is defined as follows:

Definition 1.5.29 (Stratonovitch integral)**.** *Let X, Y be two continuous real-valued semimartingales. The Stratonovitch integral of Y with respect to X, denoted by $\int_0^t Y_s \circ \mathrm{d}X_s$, is defined by*

$$\int_0^t Y_s \circ \mathrm{d}X_s := \int_0^t Y_s \mathrm{d}X_s + \frac{1}{2}\langle Y, X\rangle_t,$$

where the first term is the Itô integral of Y with respect to X and $\langle .,.\rangle$ is the bracket process (also known as the quadratic covariation process).

With the Stratonovitch integral, the classical Itô formula can then be written is the following way (see [98, Theorems 20-21, pages 277-278], [68, Th. III.1.3 p.101]), for a continuous d-dimensional semimartingale X and a function $f : \mathbb{R}^d \to \mathbb{R}$ of class $\mathcal{C}^2$: $f(X)$ is a semimartingale and

$$f(X_t) - f(X_0) = \sum_{i=1}^d \int_0^t \frac{\partial f}{\partial x^i}(X_s) \circ \mathrm{d}X_s^i. \tag{1.31}$$

Because of this chain rule that holds under a more convenient form as in the ordinary calculus, the Stratonovitch integral is better fitted to differential calculus on manifolds than the usual Itô integral.

We recall that a vector field V on a manifold $\mathcal{M}$ is a family $\{V(x)\}_{x\in\mathcal{M}}$ such that $\forall x \in \mathcal{M}$, $V(x) \in T_x\mathcal{M}$ (see for e.g. [77, Chapter 4]). In local coordinates $(x^1, x^2, ..., x^d)$, a smooth vector field V can be represented as

$$V(x) = \sum_{i=1}^d V^i(x) \left.\frac{\partial}{\partial x^i}\right|_x,$$

where $V^1, \dots, V^d$ are real smooth functions on the domain of the local coordinate system, and where $\left\{\frac{\partial}{\partial x^i}\right\}_{1\le i\le d}$ denotes a basis of $T_x\mathcal{M}$.

Proposition 1.5.30 (Theorem 1.2.9 in [67])**.** *Let $l \geq 1$. Consider the Stratonovich SDE*

$$\mathrm{d}X_t = \sum_{\alpha=1}^l V_\alpha(X_t) \circ \mathrm{d}B_t^\alpha + V_0(X_t)\mathrm{d}t \tag{1.32}$$

where $(V_\alpha)_{0\le\alpha\le l}$ are $\mathcal{C}^2$ vector fields on $\mathcal{M}$ and $B = (B^\alpha)_{1\le\alpha\le l}$ is the standard l-dimensional Brownian motion. Then, there exists a unique strong solution to (1.32) *(up to explosion time) whose infinitesimal generator is*

$$\mathcal{A}f(x) = \frac{1}{2}\sum_{\alpha=1}^l \left(V_\alpha^2 f\right)(x) + (V_0 f)(x),$$

where $\left(V_\alpha^2 f\right)(x) := (V_\alpha(V_\alpha f))(x)$, and whose carré du champ operator is given by

$$\Gamma(f,g) = \frac{1}{2}\sum_{\alpha=1}^l V_\alpha(f)V_\alpha(g),$$

where by definition

$$\Gamma(f,g) = \frac{1}{2}\left(\mathcal{A}(fg) - f\mathcal{A}(g) - g\mathcal{A}(f)\right).$$

Notice that (1.32) is a SDE on $\mathcal{M}$ because the V_α's are vector fields on $\mathcal{M}$. Extending these fields to the whole ambient space allows to solve the SDE in $\mathbb{R}^m$ using Picard's iterations. Provided the initial condition lies in $\mathcal{M}$, then the solution remains on the manifold, [67, Prop. 1.2.8].

From the Itô formula, remark that the distribution of the solution of the SDE (1.32) is also characterized by the fact that for all $f \in \mathcal{C}^\infty(\mathcal{M})$,

$$f(X_t) = f(X_0) + \int_0^t \sum_{\alpha=1}^{l} V_\alpha f(X_s) \circ dB_s^\alpha + \int_0^t V_0 f(X_s)\, ds.$$

As for Euclidean semimartingales, we can show that all continuous semimartingales on $\mathcal{M}$ solve a SDE of the form (1.32).

Proposition 1.5.31 (Chapter 2 [67])**.** *If (X_t) is a continuous semimartingale on $\mathcal{M}$, then there are m smooth vector fields $V_1, ..., V_m$ and a $\mathbb{R}^m$-valued semimartingale $W = (W^i)_{1\leqslant i\leqslant m}$ such that:*

$$\mathrm{d}X_t = \sum_{i=1}^{m} V_i(X_t) \circ \mathrm{d}W_t^i,$$

where $\int \cdot \circ \mathrm{d}W_t^i$ denotes the Stratonovich stochastic integral with respect to the Brownian motion W^i (see [68, Chapter III]).

1.5.4.2 Brownian motion on $\mathcal{M}$

We are now in position to define the Brownian motion on the manifold $\mathcal{M}$. Let $\{\xi_\alpha\}_{1\leq\alpha\leq m}$ be an orthonormal basis on $\mathbb{R}^m$ of which $\mathcal{M}$ is here considered to be a submanifold. For each $x \in \mathcal{M}$, we consider $P_\alpha(x)$ the orthogonal projection of e_α to $T_x\mathcal{M}$. Let us note that P_α is a vector field on $\mathcal{M}$. In a local coordinate system $(x^1, x^2, ..., x^d)$,

$$P_\alpha(x) = \sum_{i=1}^{d} P_\alpha^i(x) \left.\frac{\partial}{\partial x^i}\right|_x .$$

Recall Equation (1.22) from Theorem 1.4.6, that the Laplace-Beltrami operator satisfies $\Delta = \sum_{\alpha=1}^m P_\alpha^2$ and remark that for two real-valued functions of class $\mathcal{C}^2$ on $\mathcal{M}$, we have:

$$\langle \nabla f, \nabla h \rangle = \sum_{\alpha=1}^{m} (P_\alpha f)(P_\alpha h). \tag{1.33}$$

Then, as an application of Proposition 1.5.30, we have the following result.

Proposition 1.5.32. *There exists a unique strong solution starting at $x \in \mathcal{M}$ to the following SDE*

$$\mathrm{d}X_t = \sqrt{2}\sum_{\alpha=1}^{m} P_\alpha(X_t) \circ \mathrm{d}B_t^\alpha, \tag{1.34}$$

that has the infinitesimal generator Δ. This solution is defined as the **Brownian motion on the manifold** *$\mathcal{M}$.*

Example 1.5.33. *Consider $\mathcal{M} = \mathbb{S}^m$ the unit sphere of $\mathbb{R}^{m+1}$. The projection to the tangent sphere at $x \in \mathbb{S}^m$ is*

$$P(x)\xi = \xi - \langle \xi, x\rangle x, \tag{1.35}$$

for any $\xi \in \mathbb{R}^{m+1}$. From this, we deduce that (1.38) *becomes:*

$$X_t^i = X_0^i + \int_0^t \sum_{\alpha=1}^m \left(\mathbf{1}_{\alpha=i} - X_s^\alpha X_s^i\right) \circ dB_s^\alpha, \quad 1 \leqslant i \leqslant m+1, \quad X_0 = (X_0^1, \dots X_0^{m+1}) \in \mathbb{S}^m. \tag{1.36}$$

This is known as Stroock's representation of the spherical Brownian motion.

1.5.4.3 Other diffusions on $\mathcal{M}$

Another application of Proposition 1.5.30 is that the SDE corresponding to the operator $\mathcal{A}_{pq}$ defined in (1.27) is:

$$\mathrm{d}X_t = \sum_{\alpha=1}^m \sqrt{2q(X_t)} P_\alpha(X_t) \circ \mathrm{d}B_t^\alpha + \sum_{\alpha=1}^m \left(\frac{1}{2}(P_\alpha q)(X_t) + q(P_\alpha(\ln p))(X_t)\right)(P_\alpha f)(X_t) \tag{1.37}$$

for $(B^\alpha)_{1\le\alpha\le m}$ independent Euclidean 1-dimensional Brownian motions. Taking $q \equiv 1$, we deduce that the unique solution to the SDE

$$\mathrm{d}X_t = \sqrt{2}\sum_{\alpha=1}^m P_\alpha(X_t) \circ \mathrm{d}B_t^\alpha + \sum_{\alpha=1}^m P_\alpha(\ln p)(X_t) P_\alpha(X_t)\mathrm{d}t, \tag{1.38}$$

has the infinitesimal generator $\mathcal{L}f = \Delta f + \langle \nabla \ln p, \nabla f\rangle$, which reduced to Δ and (1.38) if we have additionally that $p \equiv 1$.

1.5.5 Non-smooth non-symmetric elliptic operators on manifolds

In this section, we investigate second-order differential operators on manifolds, particularly focusing on those that may lack smoothness or symmetry (with respect to a certain measure). Our objective is to establish a theoretical foundation that ensures the rigor of analyses related to diffusion processes generated on manifolds. Specifically, we consider operators of the form:

$$\mathcal{A}f(x) = \sum_{i=1}^d b^i(x)\frac{\partial f}{\partial x^i}(x) + \sum_{i,j=1}^d a^{ij}(x)\frac{\partial^2 f}{\partial x^i \partial x^j}(x), \tag{1.39}$$

where x is a point on a manifold $\mathcal{M}$, f is a twice differentiable function on $\mathcal{M}$, and $(x^1, x^2, \dots, x^d)$ are local coordinates around x. The coefficients $b^i(x)$ and $a^{ij}(x)$ are assumed to be twice differentiable functions.

Amongst other classic results, the main result of this section is the establishment of conditions under which a non-smooth, non-symmetric elliptic operator $\mathcal{A}$ on a manifold $\mathcal{M}$ generates a unique Feller semigroup and is essentially self-adjoint when considered as an operator on $L^2(\mu)$. This result provides a foundation for our later analysis on manifolds.

For readers interested in the special case of smooth symmetric elliptic operators in $\mathbb{R}^d$, we recommend the discussion presented by Baudoin in Chapter 4 of [14].

Assumption 1. *Throughout this section, we adopt the following notations and assumptions:*

- *The operator $\mathcal{A} : \mathcal{C}_c^2(\mathcal{M}) \to \mathcal{C}_0(\mathcal{M})$ is an elliptic differential operator. In any local coordinate system $(x^1, x^2, \dots, x^d)$ on $\mathcal{M}$, the corresponding coefficient functions $a^{ij}(x)$ and $b^i(x)$ are twice differentiable.*

- *The manifold $\mathcal{M}$ is a d-dimensional embedded submanifold without boundary of $\mathbb{R}^m$.*
- *The measure μ is a probability Borel measure on $\mathcal{M}$.*
- *The space $W(\mathcal{M})$ denotes the set of all continuous processes taking values in $\mathcal{M}$.*

We now present the results concerning the operator $\mathcal{A}$ and its associated semigroups.

Proposition 1.5.34. *Suppose there exist functions $\tilde{b} : \mathbb{R}^m \to \mathbb{R}^m$ and $\tilde{\sigma} : \mathbb{R}^m \to \mathbb{R}^{m\times m}$ satisfying the following conditions:*
(i) The functions $\tilde{b}$ and $\tilde{\sigma}$ are bounded and locally Lipschitz continuous.
(ii) The operator $\tilde{\mathcal{A}}$, defined by

$$\tilde{\mathcal{A}}f = \frac{1}{2}\sum_{i,j=1}^{m} \tilde{a}^{ij}\frac{\partial^2 f}{\partial x^i \partial x^j} + \sum_{i=1}^{m} \tilde{b}^i \frac{\partial f}{\partial x^i}, \quad \text{with } \tilde{a} = \tilde{\sigma}\tilde{\sigma}^\top,$$

extends the operator $\mathcal{A}$; that is, $\tilde{\mathcal{A}}f = \mathcal{A}f$ for all $f \in \mathcal{C}_c^2(\mathcal{M})$.
Then, the operator $\mathcal{A}$ generates a unique semigroup $(P_t)_{t\geqslant 0}$ on $\mathcal{C}_0(\mathcal{M})$. Moreover, this semigroup is a contraction, strongly continuous, and has the Feller property.

Proof. Since $\tilde{b}$ and $\tilde{\sigma}$ are bounded and locally Lipschitz continuous, the stochastic differential equation (SDE)

$$\mathrm{d}\tilde{X}_t = \tilde{b}(\tilde{X}_t)\,\mathrm{d}t + \tilde{\sigma}(\tilde{X}_t)\,\mathrm{d}B_t$$

has a unique strong solution for any initial condition $\tilde{X}_0 = x \in \mathbb{R}^m$, where B_t is a standard Brownian motion in $\mathbb{R}^m$. This uniqueness is ensured by standard results on SDEs with Lipschitz coefficients, see Theorem 1.2.2.
The associated semigroup $(\tilde{P}_t)_{t\geqslant 0}$ on $\mathbb{R}^m$ is defined by

$$\tilde{P}_t\tilde{f}(x) = \mathbb{E}_x\left[\tilde{f}(\tilde{X}_t)\right],$$

for $\tilde{f} \in \mathcal{C}_0(\mathbb{R}^m)$. This semigroup is strongly continuous and satisfies the Feller property, meaning it maps continuous functions vanishing at infinity into themselves and preserves the maximum norm. The boundedness of the coefficients ensures that $\tilde{P}_t$ is a contraction semigroup.
Since $\tilde{\mathcal{A}}$ extends $\mathcal{A}$, for any $f \in \mathcal{C}_c^2(\mathcal{M})$, we have $\tilde{\mathcal{A}}f = \mathcal{A}f$. Therefore, any $\mathcal{A}$-diffusion process X^x starting at $x \in \mathcal{M}$ is also a $\tilde{\mathcal{A}}$-diffusion process when considered in $\mathbb{R}^m$. By the uniqueness of solutions to the martingale problem associated with $\tilde{\mathcal{A}}$, the processes X^x and $\tilde{X}^x$ have the same law on the path space $W(\mathbb{R}^m)$ when started at $x \in \mathcal{M}$.
Therefore, for all $f \in \mathcal{C}_c^2(\mathcal{M})$, extended to $\tilde{f} \in \mathcal{C}_c^2(\mathbb{R}^m)$, and for all $x \in \mathcal{M}$ and $t > 0$, we have

$$P_t f(x) = \tilde{P}_t\tilde{f}(x).$$

This equality shows that the semigroup $(P_t)_{t\geqslant 0}$ on $\mathcal{C}_0(\mathcal{M})$ inherits the contraction, strong continuity, and Feller properties from $(\tilde{P}_t)_{t\geqslant 0}$. The uniqueness of (P_t) follows from the uniqueness of $(\tilde{P}_t)$. □

Remark 1.5.35. *The boundedness of the coefficients $\tilde{b}$ and $\tilde{\sigma}$ is crucial for ensuring the strong continuity of the associated semigroup. Relaxing this condition may lead to semigroups that are not strongly continuous, which would complicate the analysis.*

Remark 1.5.36. *Since $\mathcal{M}$ is an embedded submanifold, under the initial assumptions on $\mathcal{A}$, there always exists a second-order elliptic differential operator $\tilde{\mathcal{A}}$ on $\mathbb{R}^m$ that extends $\mathcal{A}$. The challenge lies in ensuring that $\tilde{\mathcal{A}}$ satisfies the boundedness and Lipschitz conditions required in Proposition 1.5.34. Constructing such an extension may involve extending the coefficients $b^i(x)$ and $a^{ij}(x)$ to functions on $\mathbb{R}^m$ with the desired properties.*

As a direct consequence, we have the following corollary.

Corollary 1.5.37. *If $\mathcal{M}$ is compact, then $\mathcal{A}$ generates a unique semigroup $(P_t)_{t\geqslant 0}$ on $\mathcal{C}_0(\mathcal{M})$. Moreover, this semigroup is a contraction, strongly continuous, and has the Feller property.*

Proof. On a compact manifold, continuous functions are automatically bounded, and locally Lipschitz functions are globally Lipschitz. Therefore, the conditions of Proposition 1.5.34 are satisfied without the need for further adjustments. The conclusion follows directly. □

We next consider the extension of the semigroup to Lebesgue spaces L^s.

Theorem 1.5.38. *Suppose that $\mathcal{M}$ is compact, μ a positive measure on $\mathcal{M}$, and that for all $f \in \mathcal{C}_c^2(\mathcal{M})$, we have*

$$\int_{\mathcal{M}} \mathcal{A}f \,\mathrm{d}\mu = 0.$$

Then, for each $s \in [1,\infty)$ and $t > 0$, there exists a unique continuous operator $P_t^{(s)} : L^s(\mu) \to L^s(\mu)$ such that

$$P_t^{(s)} f = P_t f \quad \text{for all } f \in \mathcal{C}_c(\mathcal{M}).$$

Proof. Let $\overline{\mathcal{A}} : \mathcal{D}(\overline{\mathcal{A}}) \to \mathcal{C}_0(\mathcal{M})$ denote the closure of $\mathcal{A}$ in $\mathcal{C}_0(\mathcal{M})$. Since $\mathcal{A}$ generates a contraction semigroup on $\mathcal{C}_0(\mathcal{M})$, for all $f \in \mathcal{C}_c^2(\mathcal{M})$ and $t > 0$, we have $P_t f \in \mathcal{D}(\overline{\mathcal{A}})$.
For such f and t, there exists a sequence $(g_n)_{n\in\mathbb{N}} \subset \mathcal{D}(\mathcal{A})$ such that

$$\mathcal{A}P_t f = \lim_{n\to\infty} \mathcal{A}g_n \quad \text{in } \mathcal{C}_0(\mathcal{M}).$$

Therefore,

$$\int_{\mathcal{M}} \mathcal{A}P_t f \,\mathrm{d}\mu = \lim_{n\to\infty} \int_{\mathcal{M}} \mathcal{A}g_n \,\mathrm{d}\mu = 0.$$

This implies that the mapping $t \mapsto \int_{\mathcal{M}} P_t f \,\mathrm{d}\mu$ is constant in t. Since at $t = 0$ we have $\int_{\mathcal{M}} P_0 f \,\mathrm{d}\mu = \int_{\mathcal{M}} f \,\mathrm{d}\mu$, it follows that

$$\int_{\mathcal{M}} P_t f \,\mathrm{d}\mu = \int_{\mathcal{M}} f \,\mathrm{d}\mu \quad \text{for all } t > 0.$$

Consequently, for $f \in \mathcal{C}_0(\mathcal{M})$,

$$\|P_t f\|_{L^1(\mu)} = \|f\|_{L^1(\mu)} \quad \text{and} \quad \|P_t f\|_{L^\infty(\mu)} \leqslant \|f\|_{L^\infty(\mu)}.$$

By the Riesz–Thorin interpolation theorem, P_t extends uniquely to a contraction on $L^s(\mu)$ for each $s \in [1,\infty)$. The density of $\mathcal{C}_c(\mathcal{M})$ in $L^s(\mu)$ ensures that this extension is unique and continuous. We denote this extension by $P_t^{(s)}$. □

We now address the essential self-adjointness of $\mathcal{A}$.

Theorem 1.5.39. *If $\mathcal{A} : \mathcal{C}_c^2(\mathcal{M}) \to \mathcal{C}_0(\mathcal{M}) \subset L^2(\mu)$ is symmetric and the measure μ has a strictly positive density on $\mathcal{M}$, then $\mathcal{A}$ is essentially self-adjoint when considered as an unbounded operator on $L^2(\mu)$.*

Proof. Since $\mathcal{A}$ is symmetric on the Hilbert space $H = L^2(\mu)$, we have

$$\int_{\mathcal{M}} \mathcal{A}f \,\mathrm{d}\mu = 0 \quad \text{for all } f \in \mathcal{C}_c^2(\mathcal{M}).$$

By Theorem 1.5.38 with $s = 2$, the closure $\overline{\mathcal{A}} : \mathcal{D}(\overline{\mathcal{A}}) \to H$ of $\mathcal{A}$ generates a semigroup $(P_t^{(2)})_{t\geqslant 0}$ on H, satisfying $P_t^{(2)} f = P_t f$ for all $f \in \mathcal{C}_c(\mathcal{M})$.

The operator $\overline{\mathcal{A}}$ is symmetric, being the closure of a symmetric operator. Moreover, since $\overline{\mathcal{A}}$ generates $(P_t^{(2)})_{t\geqslant 0}$, we have $P_t^{(2)} f \in \mathcal{D}(\overline{\mathcal{A}})$ for all $t > 0$ and $f \in \mathcal{C}_c(\mathcal{M})$.
Consider the function

$$F(s) = \langle P_s^{(2)} f, P_{t-s}^{(2)} g\rangle_{L^2(\mu)},$$

for fixed $t > 0$ and $f, g \in \mathcal{C}_c(\mathcal{M})$. This function is constant in $s \in [0, t]$. Indeed, for $0 < s < t$, we compute

$$\frac{\mathrm{d}}{\mathrm{d}s} F(s) = \langle \overline{\mathcal{A}} P_s^{(2)} f, P_{t-s}^{(2)} g\rangle_{L^2(\mu)} - \langle P_s^{(2)} f, \overline{\mathcal{A}} P_{t-s}^{(2)} g\rangle_{L^2(\mu)} = 0,$$

since $\overline{\mathcal{A}}$ is symmetric. This implies that

$$F(s) = F(0) = \langle f, P_t^{(2)} g\rangle_{L^2(\mu)} \quad \text{for all } s \in [0, t].$$

Therefore, the semigroup $(P_t^{(2)})$ is self-adjoint, meaning that $(P_t^{(2)})^* = P_t^{(2)}$. Consequently, the generator $\overline{\mathcal{A}}$ is self-adjoint, and since $\mathcal{A}$ is densely defined, it follows that $\mathcal{A}$ is essentially self-adjoint. □

Remark 1.5.40. *Although $\mathcal{A}$ is symmetric, this does not guarantee that its adjoint $\mathcal{A}^*$ is symmetric unless $\mathcal{A}$ is self-adjoint. The essential self-adjointness of $\mathcal{A}$ means that its closure $\overline{\mathcal{A}}$ is self-adjoint, ensuring that the operator has a unique self-adjoint extension. This property is significant in applications, as it allows the use of spectral theory to analyze the operator and the semigroup it generates. One approach to understanding this relationship is to analyze the semigroup $(P_t^{(2)})$ and its connection to the associated stochastic differential equation. For further discussion, see Section 4.5.1 in [14], where a similar problem is considered in the Euclidean setting.*

1.5.6 Wasserstein distance

The Wasserstein distance (also known as the Earth Mover's distance) is a metric that measures the distance between two probability distributions on a given metric space. It is commonly used in optimal transport theory. The most frequently used version is the Wasserstein-1 distance and the more general Wasserstein-q distance.

Definition 1.5.41. *[12, p.436] Let $(\mathcal{M}, \rho)$ be a metric space, and let μ and ν be two probability measures on $\mathcal{M}$. The Wasserstein-q distance between μ and ν, for $q > 0$, is defined as:*

$$\mathcal{W}_q(\mu, \nu) = \left(\inf_{\gamma \in \Pi(\mu,\nu)} \int_{\mathcal{M}\times\mathcal{M}} \rho(x, y)^q \, d\gamma(x, y) \right)^{\frac{1}{q}},$$

where:

- *$\rho(x, y)$ is the metric (or distance) between two points x and y in the space $\mathcal{M}$,*
- *$\Pi(\mu, \nu)$ is the set of all couplings (or transport plans) of μ and ν, meaning the set of all probability measures γ on $\mathcal{M} \times \mathcal{M}$ with marginals μ and ν.*

The Wasserstein distance measures the "cost" of transforming one probability distribution into another, where the "cost" is defined by the metric $\rho(x, y)$, which measures the distance between points x and y, and the optimal transport plan γ, which minimizes the total cost of this transformation.
In particular, for the Wasserstein-1 distance (with $q = 1$), the distance can be interpreted as the minimum amount of "work" required to move probability mass from the distribution μ to the distribution ν.

Example 1.5.42 (1-Wasserstein distance on $\mathbb{R}$)**.** *For probability distributions μ and ν on the real line $\mathbb{R}$, the Wasserstein-1 distance can be simplified to:*

$$\mathcal{W}_1(\mu,\nu) = \int_0^1 |F_\mu^{-1}(t) - F_\nu^{-1}(t)|\, dt,$$

where F_μ^{-1} and F_ν^{-1} are the quantile functions (inverse cumulative distribution functions) of μ and ν, respectively.
This gives a way to compute the Wasserstein distance between two distributions in 1D based on their quantiles.

It can also be shown that Definition 1.5.41 for $q = 1$ is also equivalent to:

$$\mathcal{W}_1(\mu,\nu) = \sup_{f \text{ is 1-Lipschitz}} \left(\int_{\mathcal{M}} f \mathrm{d}\mu - \int_{\mathcal{M}} f \mathrm{d}\nu \right), \tag{1.40}$$

see [45, Proposition 2.6.6].

Intuitively speaking, the Wasserstein distance provides a way to compare probability distributions in terms of how much "effort" is needed to morph one distribution into another based on the underlying metric space. Thus, the nature of the underlying also a special role in Wasserstein distance analysis.
For example, when $\mathcal{M}$ is a Riemannian manifold, beyond focusing on measure coupling to estimate the corresponding Wasserstein distances, Peyre [96] proposed the following estimate, providing an upper bound for the Wasserstein distance using an analytic norm.

Theorem 1.5.43. *Given a compact Riemannian manifold (with or without border) $\mathcal{M}$ Then, for all probability measures μ and ν on $\mathcal{M}$, we have:*

$$\mathcal{W}_2(\mu,\nu) \leq 2\|\mu - \nu\|_{H^{-1}(\mu)},$$

where

$$\|\mu - \nu\|_{H^{-1}(\mu)} := \sup_{g : \int_{\mathcal{M}} |\nabla g|^2 \mathrm{d}\mu \leq 1} \left(\int_{\mathcal{M}} g \mathrm{d}(\mu - \nu) \right).$$

Moreover,

$$\mathcal{W}_2(\mu,\nu) \leq 2 \sup_{\substack{g \in Lip(\mathcal{M}): \\ \int_{\mathcal{M}} |\nabla g|^2 \mathrm{d}\mu \leq 1}} \left(\int_{\mathcal{M}} g \mathrm{d}(\mu - \nu) \right), \tag{1.41}$$

where $Lip(\mathcal{M})$ denotes the space of all Lipschitz continuous functions on $\mathcal{M}$.

Proof of Theorem 1.5.43. The result is due to [96] but we provide a new proof, as the one provided by Peyre holds only under additional assumptions that were not detailed in his paper and that are not necessary in this proof. Consider the Hamilton-Jacobi semigroup $(Q_t)_{t>0}$ on $\mathrm{Lip}(\mathcal{M})$:

$$Q_t\phi(x) := \inf_{y \in \mathcal{M}} \left\{ \phi(y) + \frac{1}{2t}\rho(y,x)^2 \right\}, \quad t > 0,\ \phi \in \mathrm{Lip}(\mathcal{M}),$$

where ρ is the geodesic distance of $\mathcal{M}$.
From [81, Theorem 2.5], for any ϕ continuous bounded, $Q_0\phi := \lim_{t\to 0} Q_t\phi = \phi$, $\|\nabla Q_t\phi\|_\infty$ is bounded for all $t > 0$, and $Q_t\phi$ solves the Hamilton-Jacobi equation:

$$\frac{\mathrm{d}}{\mathrm{d}t} Q_t\phi = -\frac{1}{2}|\nabla Q_t\phi|^2, \quad t > 0. \tag{1.42}$$

As used in [120], the Kantorovich dual formula implies that (see [119, Theorem 2.10]):

$$\frac{1}{2}\mathcal{W}_2^2(\mu,\nu) = \sup_{\phi \in \mathrm{Lip}(\mathcal{M}) \text{ bounded}} \{\nu(Q_1\phi) - \mu(\phi)\}.$$

Consider the following curve on the space of measures $(\mu_t)_{0\le t\le 1}$ defined by

$$\mu_t := (1-t)\mu + t\nu.$$

By taking the derivative along t, we have that for all ϕ continuous and bounded:

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mu_t(Q_t\phi)) = -\frac{1}{2}\left(\int_{\mathcal{M}} |\nabla Q_t\phi|^2 \mathrm{d}\mu_t\right) + \int_{\mathcal{M}} Q_t\phi \mathrm{d}(\nu-\mu).$$

We now analyze the above term. Let t be in $(0,1)$.
If $\int_{\mathcal{M}} |\nabla Q_t\phi|^2 \mathrm{d}\mu_t > 0$, by the fact that $-\frac{1}{2}a^2 + ab \le \frac{1}{2}b^2$ for all a, b real, we have:

$$\begin{aligned} &-\frac{1}{2}\left(\int_{\mathcal{M}} |\nabla Q_t\phi|^2 \mathrm{d}\mu_t\right) + \int_{\mathcal{M}} Q_t\phi \mathrm{d}(\nu-\mu) \\ &\qquad \le \frac{1}{2}\frac{\left(\int_{\mathcal{M}} Q_t\phi \mathrm{d}(\nu-\mu)\right)^2}{\int_{\mathcal{M}} |\nabla Q_t\phi|^2 \mathrm{d}\mu_t} \le \frac{1}{2} \sup_{\substack{g\in \mathrm{Lip}(\mathcal{M}):\\ \int_{\mathcal{M}} |\nabla g|^2 \mathrm{d}\mu_t \le 1}} \left(\int_{\mathcal{M}} g\mathrm{d}(\mu-\nu)\right)^2 \end{aligned} \tag{1.43}$$

If $\int_{\mathcal{M}} |\nabla Q_t\phi|^2 \mathrm{d}\mu_t = 0$ and $\int_{\mathcal{M}} Q_t\phi \mathrm{d}(\nu-\mu) = 0$, clearly we have

$$-\frac{1}{2}\left(\int_{\mathcal{M}} |\nabla Q_t\phi|^2 \mathrm{d}\mu_t\right) + \int_{\mathcal{M}} Q_t\phi \mathrm{d}(\nu-\mu) \le \frac{1}{2} \sup_{\substack{g\in \mathrm{Lip}(\mathcal{M}):\\ \int_{\mathcal{M}} |\nabla g|^2 \mathrm{d}\mu_t \le 1}} \left(\int_{\mathcal{M}} g\mathrm{d}(\mu-\nu)\right)^2. \tag{1.44}$$

If $\int_{\mathcal{M}} |\nabla Q_t\phi|^2 \mathrm{d}\mu_t = 0$ and $\int_{\mathcal{M}} Q_t\phi \mathrm{d}(\nu-\mu) \neq 0$, using the Lipschitz function $\alpha \times Q_t\phi$ on $\mathcal{M}$ for any arbitrary $\alpha \in \mathbb{R}\setminus\{0\}$, we remark that

$$\sup_{\substack{g\in \mathrm{Lip}(\mathcal{M}):\\ \int_{\mathcal{M}} |\nabla g|^2 \mathrm{d}\mu_t \le 1}} \left(\int_{\mathcal{M}} g\mathrm{d}(\mu-\nu)\right)^2 = \infty, \tag{1.45}$$

and clearly

$$-\frac{1}{2}\left(\int_{\mathcal{M}} |\nabla Q_t\phi|^2 \mathrm{d}\mu_t\right) + \int_{\mathcal{M}} Q_t\phi \mathrm{d}(\nu-\mu) \le \frac{1}{2} \sup_{\substack{g\in \mathrm{Lip}(\mathcal{M}):\\ \int_{\mathcal{M}} |\nabla g|^2 \mathrm{d}\mu_t \le 1}} \left(\int_{\mathcal{M}} g\mathrm{d}(\mu-\nu)\right)^2. \tag{1.46}$$

Hence, for all $t \in (0,1)$,

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mu_t(Q_t\phi)) \le \frac{1}{2} \sup_{\substack{g\in \mathrm{Lip}(\mathcal{M}):\\ \int_{\mathcal{M}} |\nabla g|^2 \mathrm{d}\mu_t \le 1}} \left(\int_{\mathcal{M}} g\mathrm{d}(\mu-\nu)\right)^2,$$

which implies

$$\mathcal{W}_2^2(\mu,\nu) \le \int_0^1 \sup_{\substack{g\in \mathrm{Lip}(\mathcal{M}):\\ \int_{\mathcal{M}} |\nabla g|^2 \mathrm{d}\mu_t \le 1}} \left(\int_{\mathcal{M}} g\mathrm{d}(\mu-\nu)\right)^2 dt.$$

Besides, $\mu_t \geq (1-t)\mu$ for all $t \in (0,1)$. Thus, for all g with $\int_{\mathcal{M}} |\nabla g|^2 \mathrm{d}\mu_t \leq 1$, we have $(1-t)\int_{\mathcal{M}} |\nabla g|^2 \mathrm{d}\mu \leq 1$. This leads to the fact that for all $0 < t_0 < t_1 < 1$, the following holds

$$\mathcal{W}_2^2(\mu_{t_0}, \mu_{t_1}) \leq \int_0^1 \sup_{\substack{g \in \mathrm{Lip}(\mathcal{M}): \\ (1-t_1)\int_{\mathcal{M}} |\nabla g|^2 \mathrm{d}\mu \leq 1}} \left((t_1 - t_0) \int_{\mathcal{M}} g \mathrm{d}(\mu - \nu) \right)^2 \mathrm{d}t$$
$$\leq \frac{(t_1-t_0)^2}{1-t_1} \sup_{\substack{g \in \mathrm{Lip}(\mathcal{M}): \\ \int_{\mathcal{M}} |\nabla g|^2 \mathrm{d}\mu \leq 1}} \left(\int_{\mathcal{M}} g \mathrm{d}(\mu - \nu) \right)^2.$$

Hence, for any $n \in \mathbb{N}^*$ and $0 < t_0 < t_1 < ... < t_n < 1$, we have:

$$\mathcal{W}_2(\mu_{t_0}, \mu_{t_n}) \leq \sum_{i=1}^n \mathcal{W}_2(\mu_{t_{i-1}}, \mu_{t_i}) \leq \sum_{i=1}^n \frac{t_i - t_{i-1}}{\sqrt{1-t_i}} \sup_{\substack{g \in \mathrm{Lip}(\mathcal{M}): \\ \int_{\mathcal{M}} |\nabla g|^2 \mathrm{d}\mu \leq 1}} \left(\int_{\mathcal{M}} g \mathrm{d}(\mu - \nu) \right)$$

Thus, for all $0 < t_0 < t_1 < 1$, by convergence of the Riemann sum,

$$\begin{aligned}\mathcal{W}_2(\mu_{t_0}, \mu_{t_1}) \leq & \left(\int_{t_0}^{t_1} \frac{1}{\sqrt{1-t}} \mathrm{d}t \right) \sup_{\substack{g \in \mathrm{Lip}(\mathcal{M}): \\ \int_{\mathcal{M}} |\nabla g|^2 \mathrm{d}\mu \leq 1}} \left(\int_{\mathcal{M}} g \mathrm{d}(\mu - \nu) \right) \\ = & \, 2(\sqrt{1-t_0} - \sqrt{1-t_1}) \sup_{\substack{g \in \mathrm{Lip}(\mathcal{M}): \\ \int_{\mathcal{M}} |\nabla g|^2 \mathrm{d}\mu \leq 1}} \left(\int_{\mathcal{M}} g \mathrm{d}(\mu - \nu) \right).\end{aligned}$$

By taking $t_0 \to 0$ and $t_1 \to 1$, we obtain that:

$$\mathcal{W}_2(\mu, \nu) \leq 2 \sup_{\substack{g \in \mathrm{Lip}(\mathcal{M}): \\ \int_{\mathcal{M}} |\nabla g|^2 \mathrm{d}\mu \leq 1}} \left(\int_{\mathcal{M}} g \mathrm{d}(\mu - \nu) \right),$$

which is our desired conclusion. □

1.6 Thesis contributions

In this section, I present a summary of the results obtained through my research and collaborations from the following papers:

- Strong uniform convergence of Laplacians of random geometric and directed k-NN graphs on compact manifolds, Guérin, H., Nguyen, D.-T., and Tran, V., 2022, `doi.org/10.48550/arXiv.2212.1028`, [61]

- Measure estimation on a manifold explored by a diffusion process, Divol, V., Guérin, H., Nguyen, D.-T., and Tran, V. C., 2024, `doi.org/10.48550/arXiv.2410.11777`, [38]

- 1-Wasserstein minimax estimation for general smooth probability densities, D.-T. Nguyen, 2024+, in preparation.

The organization of this section is as follows: Section 1.6.1 discusses the convergence of graph Laplacians based on [61], Section 1.6.2 presents the results from [38], and Section 1.6.3 introduces our initial findings from the forthcoming paper.

1.6.1 Convergence of graph Laplacians

In this work, [61], we investigate the convergence of random operators on compact smooth manifolds, with particular emphasis on the convergence of graph Laplacians built from random samples.
More precisely, its goal is to study the uniform convergence speed of the following random operators (1.47)(1.48) defined on random points, and to derive probabilistic bounds on their deviations.

1.6.1.1 Analysis Framework

We work with a compact smooth d-dimensional submanifold $\mathcal{M}$ of $\mathbb{R}^m$, endowed with the Riemannian structure induced by the ambient space $\mathbb{R}^m$. Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of independent and identically distributed (i.i.d.) points sampled from the probability measure $p(x)\mu(\mathrm{d}x)$ on $\mathcal{M}$, where $p \in \mathcal{C}^2(\mathcal{M})$ is a continuous density function with respect to the volume measure μ of $\mathcal{M}$.
These are two principal types of random measure we will work with in this paper. The first one is the graph Laplacian induced by a fixed kernel $K : \mathbb{R}_+ \to \mathbb{R}_+$ and a deterministic sequence of bandwidths (h_n):

$$\mathcal{A}_{h_n,n}(f)(x) := \frac{1}{nh_n^{d+2}} \sum_{i=1}^{n} K\left(\frac{\|x - X_i\|_2}{h_n}\right)(f(X_i) - f(x)), \tag{1.47}$$

where K satisfies the following assumption.

Assumption 2. *The kernel $K : \mathbb{R}_+ \to \mathbb{R}_+$ is a measurable function with $K(\infty) = 0$ and bounded variation H, such that:*

$$\int_0^\infty a^{d+3}\, \mathrm{d}H(a) < \infty.$$

The second is the graph Laplacian induced from a family from k-nearest neighbor graphs corresponding to a sequence of integers $(k_n)_{n\geq 0}$:

$$\mathcal{A}_n^{kNN}(f)(x) := \frac{1}{nr_n(x)^{d+2}} \sum_{i=1}^{n} K\left(\frac{\|x - X_i\|_2}{r_n(x)}\right)(f(X_i) - f(x)), \tag{1.48}$$

where $r_n(x)$ is the shortest radius such that there exists at least k_n points X_i with $1 \leq i \leq d$ that lie in the ball $B_{\mathbb{R}^m}(x, r_n(x))$, i.e.,

$$r_n(x) := \min\left\{r : \text{there are at least } k_n \text{ index } i \in [1, n] \text{ such that: } \|x - X_i\| \leq r\right\}.$$

1.6.1.2 Previous Work & Contributions

The convergence of graph Laplacians constructed from random samples has been extensively studied in the literature [115, 55]. Giné and Koltchinskii [55] considered Gaussian kernels and established convergence results under strong smoothness assumptions on the kernel K. Their work provided foundational results for the statistical analysis of graph-based methods.
Calder and García Trillos [25] obtained deviation inequalities for random operators similar to those we study, but their results did not include uniformity over classes of test functions. Additionally, their analysis required stronger assumptions on the regularity of the kernel and the underlying manifold.

Our work generalizes these previous results by weakening the assumptions on the kernel function and providing uniform convergence results over a class of functions. This allows for non-continuous or non-smooth kernels, such as indicator functions, broadening the applicability of our results to a wider range of graph structures used in practice, including ε-geometric and k-nearest neighbor (kNN) graphs

1.6.1.3 Main Theorems

Theorem 1.6.1 (Uniform Convergence of Random Operators)**.** *Under Assumption 2 on the kernel K, where $h_n \to 0$ and $\frac{\log h_n^{-1}}{nh_n^{d+2}} \to 0$ as $n \to \infty$, we have that, almost surely, for every function $f \in \mathcal{C}^3(\mathcal{M})$,*

$$\sup_{x\in\mathcal{M}} |\mathcal{A}_{h_n,n}(f)(x) - \mathcal{A}(f)(x)| = O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}} + h_n\right), \tag{1.49}$$

where $\mathcal{A}$ is the second-order differential operator on $\mathcal{M}$ defined as:

$$\mathcal{A}(f)(x) = c_0\left(\langle\nabla p(x), \nabla f(x)\rangle + \frac{1}{2}p(x)\Delta f(x)\right),$$

with ∇ and Δ being the gradient and Laplace-Beltrami operators on $\mathcal{M}$, respectively, and

$$c_0 = \frac{1}{d}\int_{\mathbb{R}^d} K\left(\|v\|_2\right)\|v\|_2^2\,\mathrm{d}v.$$

Remark 1.6.2. *The constant facteur the convergence speed estimation* (1.49) *depends only on the embedding $\mathcal{M} \subset \mathbb{R}^m$, K and $\|f\|_{\mathcal{C}^3}$.*

Here, to obtain results on uniform convergence, we apply the use of the Vapnik-Chervonenkis theory from the [55] paper with our generalized kernel. This result broadens the scope of investigation in Ting's work [115], where uniform convergence was considered only for a finite number of points.
This proof also provides a deviation inequality:

Theorem 1.6.3 (Deviation inequality)**.** *There exists a constant $C' > 0$ such that for all n, h, δ satisfying: $h \vee \sqrt{\frac{\log h^{-1}}{nh^{d+2}}} \leqslant \delta \leqslant 1$,*

$$\mathbb{P}\left(\sup_{f\in\mathcal{F}}\sup_{x\in\mathcal{M}} |\mathcal{A}_{h,n}(f)(x) - \mathcal{A}(f)(x)| > C'\delta\right) \leqslant \exp\left(-nh^{d+2}\delta^2\right).$$

where $\mathcal{F} := \{f \in \mathcal{C}^3(\mathcal{M}) : \|f\|_{\mathcal{C}^3} \leq 1\}$.

On top of that, we observe that the bandwidths used to define $\mathcal{A}_{h_n,n}$ are not necessarily deterministic or pre-determined for the convergence result to be valid. Indeed, only certain suitable convergence conditions for (h_n) are required.
This observation leads to our extension of our analysis to the convergence of Laplacians constructed from k-nearest neighbor (kNN) graphs.

Theorem 1.6.4. *Under Assumption 1 on K and the condition $0 < p_{\min} \leqslant p(x) \leqslant p_{\max}$ for all $x \in \mathcal{M}$, along with the following conditions on k_n, the number of nearest neightbors to consider,*

$$\lim_{n\to\infty}\frac{k_n}{n} = 0, \quad \textit{and} \quad \lim_{n\to\infty}\frac{1}{n}\left(\frac{k_n}{n}\right)^{-1-2/d}\log\left(\frac{k_n}{n}\right) = 0,$$

we have that, almost surely,

$$\sup_{x\in\mathcal{M}}\left|\mathcal{A}_n^{kNN}(f)(x)-\mathcal{A}(f)(x)\right| = O\left(\sqrt{\log\left(\frac{n}{k_n}\right)}\frac{1}{\sqrt{k_n}}\left(\frac{n}{k_n}\right)^{1/d}+\left(\frac{k_n}{n}\right)^{1/d}\right).$$

Here, $\mathcal{A}_n^{kNN}$ is a normalized graph Laplacian constructed from the kNN graph with k_n-nearest neighbors for the first n random points $(X_i)_{1\leq i\leq n}$, see Equation (1.48).

1.6.2 Convergence in Wasserstein distance of occupation measure with convolution

The manifold hypothesis has become ubiquitous in modern machine learning, explaining the efficiency of nonparametric methods in high-dimensional statistical models [23]. This paradigm has motivated statisticians to study inference problems under manifold constraints [91, 52, 2, 36, 99]. Given n i.i.d. samples from a distribution μ supported on a d-dimensional manifold $\mathcal{M}$, the task of estimating either μ or geometric quantities related to $\mathcal{M}$ naturally arises. In this section, μ is a probability measure on $\mathcal{M}$, while the volume measure is denoted by $\mathrm{d}x$.
However, when leaving the i.i.d. setting, the literature is less abundant. A natural framework arises when data is generated through an exploration process, such as a random walk on a graph approximating the manifold (e.g., the PageRank algorithm [94]). In the limit, this random walk converges to a continuous-time diffusion exploring the manifold.
Our goal is to propose reconstruction methods for the measure μ based on the observation of the sample path $(X_t)_{t\in[0,T]}$. In this work, we introduce a kernel-based estimator for the invariant measure μ of a diffusion process on a manifold $\mathcal{M}$, providing improved convergence rates in the Wasserstein distance compared to previous methods. By smoothing the occupation measure μ_T, we achieve minimax optimal rates under mild regularity conditions on the density p and the diffusion generator $\mathcal{A}$. These results contribute to the understanding of statistical inference in non-i.i.d. settings on manifolds and have potential applications in machine learning and data analysis involving manifold-valued data.

1.6.2.1 Analysis Framework

Consider a diffusion process $(X_t)_{t\in[0,T]}$ on a compact submanifold without boundary $\mathcal{M}\subseteq\mathbb{R}^m$, generated by a uniformly elliptic $\mathcal{C}^2$-differential operator $\mathcal{A}$, symmetric with respect to some invariant measure μ.
The general framework includes operators of the form $\mathcal{A}_{pq}$, defined by (1.27) in Section 1.5.1.

1.6.2.2 Previous Work & Contributions

In $\mathbb{R}^m$, the estimation of the invariant measure of a diffusion has been extensively studied [30, 42, 95, 102]. For manifold-valued data, the problem of reconstructing the stationary measure μ from a sample path was first addressed by Wang and Zhu [121] for the generator $\mathcal{L}$. They considered the occupation measure μ_T, defined for every bounded measurable test function f by

$$\int_{\mathcal{M}} f(x)\mu_T(\mathrm{d}x)=\frac{1}{T}\int_0^T f(X_s)\,\mathrm{d}s. \tag{1.50}$$

They showed that for the process with generator $\mathcal{L}$,

$$\mathbb{E}_x\left[\mathcal{W}_2^2(\mu_T,\mu)\right]\lesssim\begin{cases}T^{-1} & \text{if } d\leqslant 3,\\ T^{-1}\ln(1+T) & \text{if } d=4,\\ T^{-2/(d-2)} & \text{if } d\geqslant 5,\end{cases} \tag{1.51}$$

where $\mathbb{E}_x$ denotes the expectation taken from the diffusion process starting at $x \in \mathcal{M}$, and $\mathcal{W}_2$ is the 2-Wasserstein distance with the geodesic distance ρ on $\mathcal{M}$.
We extend these results beyond the i.i.d. setting by studying the convergence properties of an estimator $\widehat{\mu}_{T,h}$ of μ, obtained by smoothing the occupation measure μ_T with a kernel K of bandwidth $h > 0$. When $d \geqslant 5$ and for an appropriate choice of h, we obtain the rate of convergence

$$\mathbb{E}_x\left[\mathcal{W}_2^2(\widehat{\mu}_{T,h}, \mu)\right] \lesssim T^{-\frac{2\ell+2}{2\ell+d-2}}, \tag{1.52}$$

where μ has a density of regularity $\ell \geqslant 2$. This rate not only holds for the Langevin diffusion with generator $\mathcal{L}$ but for all diffusion paths $(X_t)_{t\in[0,T]}$ whose generator $\mathcal{A}$ is a uniformly elliptic $\mathcal{C}^2$-differential operator, symmetric with respect to μ.
Furthermore, we show that these rates cannot be improved by providing minimax rates of convergence for this problem.

1.6.2.3 Main Theorem

The first result we obtained is an indirect estimate for the the convergence of $\widehat{\mu}_{T,h}$.

Theorem 1.6.5 (Estimation from a Diffusion with Generator $\mathcal{A}$)**.** *Let $d \geqslant 1$ and p be a positive $\mathcal{C}^2$ density function with associated measure μ. Let $(X_t)_{t\geqslant 0}$ be a diffusion with generator $\mathcal{A}$ which is a uniformly elliptic $\mathcal{C}^2$-differential operator, symmetric with respect to μ.*
Let $0 < h \leqslant h_0$ for some constant h_0 depending on $\mathcal{M}$ and K. Assume that either K is nonnegative or that $d \geqslant 4$ and $Th^d \geqslant c\ln(T)$ (in which case, h_0 additionally depends on the $\mathcal{C}^1$-norm of p). Then,

$$\sup_{x\in\mathcal{M}} \mathbb{E}_x\left[\mathcal{W}_2^2(\widehat{\mu}_{T,h}, \mu_h)\right] \leqslant c_0 \frac{u_{\mathcal{A}}\, p_{\max}^2}{p_{\min}^2} \|K\|_\infty^2 \begin{cases} \dfrac{h^{4-d}}{T} & \text{if } d \geqslant 5, \\ \dfrac{\ln(1/h)}{T} & \text{if } d = 4, \\ \dfrac{1}{T} & \text{if } d \leqslant 3, \end{cases} \tag{1.53}$$

where:

- *μ_h is the convolution of μ with K.*
- *c_0 depends on $\mathcal{M}$.*
- *$u_{\mathcal{A}}$ is the ultracontractivity constant of $\mathcal{A}$ (see Section 1.5.3).*
- *$p_{\min}$ and $p_{\max}$ are bounds on p.*

Remark 1.6.6. *The role of the ultracontractivity constant $\mu_{\mathcal{A}}$ is to control the additional latency in the convergence speed due to the deviation from μ, the invariant measure, of the initial measure of X.*
In general, ultracontractivity refers to a property of certain semigroups (e.g., those generated by diffusion operators), which ensures that the semigroup maps L^2 functions to L^∞ functions in a controlled way. This property can be characterized by an ultracontractivity constant that quantifies this mapping strength.

Thus, by performing a bias-variance decomposition for the initially considered Wasserstein distance, along with an optimal bandwidth selection, we obtain the following result, which is the desired convergence rate:

Theorem 1.6.7. *Under a suitable choice for the bandwidth $h = h(T)$, when T converges to infinity, we have:*

$$\sup_{x\in\mathcal{M}} \mathbb{E}_x\left[\mathcal{W}_2^2(\widehat{\mu}_{T,h}, \mu)\right] \lesssim T^{-\frac{2\ell+2}{2\ell+d-2}}. \tag{1.54}$$

We then analyze the minimax optimality of this result and prove that these rates are minimax optimal.

Proposition 1.6.8. *Let $\ell \geqslant 2$ be an integer. Then, for $\kappa_{\min}$, $p_{\min}$ small enough and $p_{\max}$, $u_{\max}$, L large enough,*

$$\mathcal{R}(\mathcal{P}_{T,\ell}) \gtrsim \begin{cases} T^{-1/2} & \text{if } d \leqslant 4, \\ T^{-\frac{\ell+1}{2\ell+d-2}} & \text{if } d \geqslant 5, \end{cases} \tag{1.55}$$

where $\mathcal{R}(\mathcal{P}_{T,\ell})$ is the minimax rate over the class $\mathcal{P}_{T,\ell}$ of diffusion processes with generators satisfying the conditions specified above.

This matches the rates achieved by our estimator $\widehat{\mu}_{T,h}$, confirming its optimality.

1.6.3 1-Wasserstein minimax estimation for general smooth probability densities

In Chapter 4, we revisit the problem of approximating probability measures with smooth density under the Wasserstein metric, a topic extensively studied in recent literature [118, 90, 37]. Specifically, given a sample consisting of n independent and identically distributed (i.i.d.) random variables drawn from an unknown probability measure μ, our objective is to construct from this sample an estimator $\tilde{\mu}_n$ for μ that attains optimal asymptotic convergence rates with respect to the Wasserstein metric $\mathcal{W}_1(\tilde{\mu}_n, \mu)$ when the sample size n goes to infinity.

1.6.3.1 Analysis framework

Recognizing that convergence rates of empirical measures in Wasserstein distance crucially depend on the dimensional characteristics of the underlying space [46, 90, 37], we divide our analysis into two distinct scenarios. The first scenario addresses the case where the measure μ is absolutely continuous with respect to the Lebesgue measure on the Euclidean space $\mathbb{R}^d$. The second scenario considers the setting in which μ is supported on a low-dimensional space $\mathcal{M}$, which we assume to be a compact d-dimensional manifold without boundary, smoothly embedded in a high-dimensional Euclidean space $\mathbb{R}^m$ $(m > d)$.
Besides, in statistics, the regularity control on density functions is usually expressed in terms of Besov norms [57]. Nevertheless, Besov norms are interpolations of Sobolev norms $H^s_q(\mathcal{M})$ [80, p.152,153]:

$$\|f\|_{H^s_q(\mathbb{R}^d)} = \left(\int_{\mathbb{R}^d} \max_{1\le i\le s} \|\nabla^i p\|^q_{\mathrm{op}}(x)\,\mathrm{d}x\right)^{1/q} \quad \text{with } s \in \mathbb{N} \text{ and } f \in \mathcal{C}^\infty(\mathbb{R}^d),$$

$$\|f\|_{H^s_q(\mathcal{M})} = \left(\int_{\mathcal{M}} \max_{1\le i\le s} \|\nabla^i p\|^q_{\mathrm{op}}(x)\,\mathrm{d}x\right)^{1/q} \quad \text{with } s \in \mathbb{N} \text{ and } f \in \mathcal{C}^\infty(\mathcal{M}).$$

For the sake of simplicity, in this chapter, Sobolev norms are the only measure of regularity for density functions we will use.

Case of $\mathbb{R}^d$.
Let us begin by examining the Euclidean scenarios $(\mathbb{R}^d, \|\cdot\|_2)$.
Let μ be a probability measure on $\mathbb{R}^d$ with density p with respect to the Lebesgue measure of $\mathbb{R}^d$ and $X_1, X_2, ..., X_n$ be a sample of n i.i.d. random elements sampled from μ. In this setting, we analyze the asymptotic behavior of the kernel measure estimator $\mu_{n,h}$ as the sample size n tends to infinity and the smoothing bandwidth h decreases at an appropriate speed to zero. This kernel estimator is explicitly defined by:

$$\widehat{\mu}_{n,h}(\mathrm{d}y) = \frac{1}{n}\sum_{i=1}^{n} h^{-d} K\left(\frac{\|X_i - y\|_2}{h}\right)\mathrm{d}y, \tag{1.56}$$

where h is the smoothing parameter, and the kernel function $K : \mathbb{R}_+ \to \mathbb{R}$ is bounded, measurable, and supported on $[0, 1]$, satisfying the normalization condition:

$$\int_{\mathbb{R}^d} K(\|x\|_2)\,\mathrm{d}x = 1. \tag{1.57}$$

This method of estimator construction is called kernel smoothing [63, Chapter 6]. Note that, the normalization condition Eq (4.5) implies that

$$\widehat{\mu}_{n,h}(\mathbb{R}^d) = 1, \tag{1.58}$$

regardless of the choice of n and h.
Then, we formalize the definition of "k-vanishing kernel" used previously:

Definition 1.6.9 (k-vanishing kernel)**.** *Let k be a positive integer. A kernel function $K : \mathbb{R}_+ \to \mathbb{R}$ is said to be a k-vanishing kernel on $\mathbb{R}^d$ if, for every integer $s \in \{1, 2, \dots, k\}$, the kernel satisfies:*

$$\int_{\mathbb{R}^d} |K(\|x\|_2)|\,\|x\|_2^s\,\mathrm{d}x < \infty, \quad \textit{and} \quad \int_{\mathbb{R}^d} K(\|x\|_2)\,\|x\|_2^s\,\mathrm{d}x = 0.$$

Case of a manifold $\mathcal{M}$.
Now, let us discuss the framework for the scenario where the measure μ is supported on a compact manifold $\mathcal{M}$ of dimension $d \geq 3$ (without boundary), smoothly embedded into a Euclidean space $(\mathbb{R}^m, \|\cdot\|_2)$.
Since $\mathcal{M}$ is smoothly embedded in $\mathbb{R}^m$, it inherits a natural Riemannian metric induced by the ambient Euclidean structure. With this metric, $\mathcal{M}$ becomes a Riemannian submanifold. We denote by ρ the geodesic distance associated with this induced metric.
For any probability measure $\mu \in \mathcal{P}(\mathcal{M})$ with density p with respect to the volume measure on $\mathcal{M}$. Let $(X_1, X_2, ..., X_n)$ be a sample of n i.i.d random variables of μ. In this scenario, we investigate the convergence of the kernel estimator $\widehat{\mu}^{\mathcal{M}}_{n,h}$ as the sample size n tends to infinity and the smoothing bandwidth h decreases appropriately to zero:

$$\widehat{\mu}^{\mathcal{M}}_{n,h}(\mathrm{d}y) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h^d}K\left(\frac{\|X_i - y\|_2}{h}\right)\mathrm{d}y, \tag{1.59}$$

where $\|\cdot\|_2$ is the distance with respect to $\mathbb{R}^m$, $\mathrm{d}y$ on the right side represents the volume measure on $\mathcal{M}$, and the kernel function $K : \mathbb{R}_+ \to \mathbb{R}$ is also a measurable bounded function with support in $[0, 1]$ such that satisfies Eq (4.5).
Note that, unlike in the previous scenario (cf. Eq. (4.6)), the mass of $\widehat{\mu}^{\mathcal{M}}_{n,h}$ need not equal 1 in general. To address this issue, the author in [37, p. 7] proposed replacing the kernel K by its pointwise normalized version in the definition of the kernel estimator $\widehat{\mu}^{\mathcal{M}}_{n,h}$. This normalization, however, introduces additional approximation steps and complexity into their analysis. In our treatment, we observe that such normalization may lead to avoidable computational complications. Hence, we retain the original kernel K and instead construct our estimation of μ via $\widehat{\mu}^{\mathcal{M}}_{n,h}$ differently.

1.6.3.2 Previous Work & Contributions

Within the Euclidean framework, our results partially overlap with those of [118, 90] in the case of compactly supported measures under the 1-Wasserstein metric. For this case, compared to [118, 90], we broaden the existing minimax results to include all probability densities possessing first moments, without restrictions on the boundedness of their support. Besides, the estimators given in [118, 90] are wavelet-based, while our choice is kernel estimators which are generally believed to be more basic and less computationally expensive [62]. Besides, our estimator achieved the minimax convergence rate.

For the compact manifold setting, in comparison with [37], our results emphasize that when focusing on the practically relevant 1-Wasserstein distance [10], the minimax results presented therein remain valid without imposing additional conditions on uniform lower and upper bounds for probability densities. Additionally, we establish that our estimator achieves convergence almost surely at the same rate. To the best of our knowledge, this stronger mode of convergence has not been previously demonstrated.
An additional refinement we introduced compared to [37], though of minimal practical significance, is the relaxed regularity requirement on the kernel function K. This adjustment was made primarily to deepen our theoretical understanding of the problem. More specifically, many calculations in [37] rely on a Taylor expansion up to relatively high order of K, which is a natural approach within the context of manifold learning. However, we have always believed that there must be a deeper geometric rationale behind why this seemingly 'brutal' Taylor expansion is effective.

1.6.3.3 Main results

Our primary theoretical contribution in $\mathbb{R}^d$ setting is summarized by the following Theorem 4.1.2 and its Corollary 4.1.3:

Notation 1.6.10 (Modified Vinogradov notations)**.** *Throughout this chapter, for $A \geq 0$ and $B \geq 0$, we use occasionally $A \lesssim_a B$ as shorthand for the inequality $A \leq C_a B$ for some constant C_a depending only on a. The same goes for $A \gtrsim_a B$. [112, p.5]*

Theorem 1.6.11. *Let $k \geq 1$ be an integer, assume $d \geq 3$, and suppose the kernel K is a k-vanishing kernel on $\mathbb{R}^d$ as specified in Definition 1.6.9.*
Then, there is constant C such that for all integers $s \in \{1, 2, \dots, k-1\}$, any real number $q > d$, and $h \in (0, 1)$, the following bound holds:

$$\mathbb{E}\left(\mathcal{W}_1(\widehat{\mu}_{n,h}, \mu)\right) \leq C \left(((M_q(\mu))^{1/2} + 1) \frac{h^{1-d/2}}{\sqrt{n}} + \|p\|_{H_1^s(\mathbb{R}^d)} h^{s+1} \right),$$

where the q-th moment of μ is defined as

$$M_q(\mu) := \int_{\mathbb{R}^d} \|x\|_2^q \, \mu(\mathrm{d}x). \tag{1.60}$$

Moreover, the constant factor C can be chosen to depend only on the integers k, q and the uniform norm $\|K\|_\infty := \sup_x |K(x)|$ of K.

Corollary 1.6.12. *Assume that $d \geq 3$. If the density p of μ satisfies that $M_{d+1}(\mu) < \infty$ and $\|p\|_{H_1^s(\mathbb{R}^d)} < \infty$,*
Then there exist an explicitly defined kernel measure estimator $\tilde{\mu}_n$ and a constant C such that:

$$\mathbb{E}(\mathcal{W}_1(\tilde{\mu}_n, \mu)) \leq C \times n^{-\frac{1+s}{d+2s}}, \tag{1.61}$$

where the constant C only depends on $d, \|K\|_\infty, M_{d+1}(\mu)$, and $\|p\|_{H_1^s(\mathbb{R}^d)}$.

We intentionally omit the cases $d = 1$ and $d = 2$, as these dimensions are already fully covered by classical results regarding empirical measure approximations [46]. Besides, $\widehat{\mu}_{n,h}$ is possibly be a signed measure, but this will not affect the definition of $\mathcal{W}_1$ in (4.3).
Our primary theoretical contribution in manifold setting is summarized by the following theorem and corollary:

Theorem 1.6.13. *Let $k \geq 1$ be an integer, assume $d \geq 3$, and suppose the kernel K is a k-vanishing kernel on $\mathbb{R}^d$ as specified in Definition 1.6.9. On top of that, we assume K is Lipschitz on* $[0, 1]$.
Then, there is a constant C, such that for all integers $s \in \{1, 2, \ldots, k-1\}$ and n, the following bound holds:

$$\mathbb{E}\left(\mathcal{W}_1(\widehat{\mu}^{\mathcal{M}}_{n,h}, \widehat{\mu}^{\mathcal{M}}_{n,h}(\mathcal{M})\mu)\right) \leq C\left(\frac{h^{1-d/2}}{\sqrt{n}} + \|p\|_{H^s_1(\mathcal{M})} h^{s+1}\right),$$

where the Wasserstein distance is defined as in Eq (4.3).
Moreover, the constant factor C can be chosen to depend only on $\mathcal{M} \subset \mathbb{R}^m$, the integer k, the uniform norm $\|K\|_\infty := \sup_x |K(x)|$ of K, and the Lipschitz constant of $K\big|_{[0,1]}$.

Corollary 1.6.14. *Assume that $d \geq 3$. If the density p of μ satisfies that $\|p\|_{H^s_1(\mathbb{R}^d)} < \infty$, Then there exist an explicitly defined kernel measure estimator $\tilde{\mu}_n$ such that:*

$$\mathbb{E}(\mathcal{W}_1(\tilde{\mu}_n, \mu)) \lesssim_{d, \|p\|_{H^s_1(\mathbb{R}^d)}, s} n^{-\frac{1+s}{d+2s}}. \tag{1.62}$$

Moreover, almost surely,

$$\limsup_{n\to\infty} n^{\frac{1+s}{d+2s}} \mathcal{W}_1(\tilde{\mu}_n, \mu) = \limsup_{n\to\infty} n^{\frac{1+s}{d+2s}} \mathbb{E}(\mathcal{W}_1(\tilde{\mu}_n, \mu)) < \infty. \tag{1.63}$$

1.7 Future works

Possible future research directions include:

- **Central limit theorems:** An interest that is frequently discussed is to construct central limit theorems for the empirical transportation cost $\mathcal{T}_c(\mu_n, \nu)$ [32, 110], where

$$\mathcal{T}_c(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x, y) \mathrm{d}\pi(x, y), \tag{1.64}$$

  where $\Pi(\mu, \nu)$ is the set of all joint probabilities on $\mathcal{X} \times \mathcal{Y}$ with marginals μ, ν. In an ongoing work, we expect to extend the previously known results on central limits of empirical transportation to some compact metric spaces. The limit law is proven to be a supremum of a centered Gaussian process $(B_f)_{f\in\mathcal{C}_b(\mathcal{X})}$ indexed by continuous bounded functions on $\mathcal{X}$ whose covariance function is

$$\begin{aligned} \mathrm{Cov}_B : \mathcal{C}_b(\mathcal{X}) \times \mathcal{C}_b(\mathcal{X}) &\longrightarrow \mathbb{R}_+ \\ (f, g) &\longmapsto \mathrm{Cov}(f(X_1), g(X_1)). \end{aligned} \tag{1.65}$$

  We expect to prove that under suitable but general assumptions on μ or $\mathcal{X}$:

$$\sqrt{n}(\mathcal{T}_c(\mu_n, \nu) - \mathcal{T}_c(\mu, \nu)) \xrightarrow[n\to\infty]{(d)} \sup_{(\psi,\phi)\in\Phi_c(\mu,\nu)} (B_\psi), \tag{1.66}$$

  where $\Phi_c(\mu, \nu)$ is the set of optimal potentials of the transportation problems $\mathcal{T}_c(\mu, \nu)$.

- **Branching diffusions on manifolds:** In [38](cf. Chapter 3) and [121], it has been shown that the occupation measure of diffusion processes on manifolds exhibits intriguing convergence properties when evaluated in the Wasserstein distance. However, there is a scarcity of literature addressing the convergence behavior—either in terms of the Wasserstein distance or through entropic optimal transport—of branching diffusion processes.

We expect that in a first step, we can be able to establish a convergence rate for the occupation measure of a branching process that is comparable to those known for standard diffusion processes. That is to study the convergence of:

$$\mathcal{W}_2(\widehat{\mu}_T, \mu),$$

where $\widehat{\mu}_T$ is a suitable choice of occupation measure for a branching process X.

- **Entropic optimal transport:** Recently, the theoretical statistics community has exhibited a growing interest in entropic optimal transport, culminating in several fascinating findings [97, 39]. One notable result is that the fluctuations (normalized with a $n^{1/2}$ factor) of the entropic transportation cost between an empirical measure $\widehat{\mu}_n$ and a fixed measure ν converge in distribution to a Gaussian random variable[34]. This Gaussian behavior is in stark contrast with the Wasserstein distance, for which such convergence typically requires additional conditions on the support of the measures and the underlying spaces[32, 33].

  We anticipate that similar results can be obtained for diffusion processes. More precisely, let $(X_t)_{t\geq 0}$ be a diffusion process in $\mathbb{R}^d$. We aim to establish a central limit theorem for the entropic optimal transportation cost in the form

$$\sqrt{n}\Big(S(\mu_T, \nu) - \mathbb{E}\big(S(\mu_T, \nu)\big)\Big) \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0, \sigma^2),$$

  where

$$\mu_T = \frac{1}{T}\int_0^T \delta_{X_s}\,\mathrm{d}s$$

  denotes the occupation measure of the process, $S(\mu_T, \nu)$ represents the entropic transportation cost between μ_T and ν, and σ^2 is an appropriate variance parameter.

Chapter 2

Convergence of graph laplacians

Contents

Consider n points independently sampled from a density p of class $\mathcal{C}^2$ on a smooth compact d-dimensional sub-manifold $\mathcal{M}$ of $\mathbb{R}^m$, and consider the random walk visiting these points according to a transition kernel K. We study the almost sure uniform convergence of the generator of this process to the diffusive Laplace-Beltrami operator when n tends to infinity, from which we establish the convergence of the random walk to a diffusion process on the manifold. This work extends known results of the past 15 years. In particular, our result does not require the kernel K to be continuous, which covers the cases of walks exploring kNN and geometric graphs, and convergence rates are given. The distance between the random walk generator and the limiting operator is separated into several terms: a statistical term, related to the law of large numbers, is treated with concentration tools and an approximation term that we control with tools from differential geometry. The case of kNN Laplacians is detailed. The convergence of the stochastic processes having these operators as generators is also studied, by establishing additional tightness results of their distributions on the space of càdlàg functions.

2.1 Introduction

Let $\mathcal{M}$ be a compact smooth d-dimensional submanifold without boundary of $\mathbb{R}^m$, which we embed with the Riemannian structure induced by the ambient space $\mathbb{R}^m$. Denote by $\|\cdot\|_2$, $\rho(\cdot,\cdot)$ and $\mu(\mathrm{d}x)$ respectively the Euclidean distance of $\mathbb{R}^m$, the geodesic distance on $\mathcal{M}$ and the volume measure on $\mathcal{M}$. Let $(X_i, i \in \mathbb{N})$ be a sequence of i.i.d. points in $\mathcal{M}$ sampled from the distribution $p(x)\mu(\mathrm{d}x)$, where $p \in \mathcal{C}^2$ is a continuous function such that $p(x)\mu(\mathrm{d}x)$ defines a probability measure on $\mathcal{M}$.
In this article, we study the limit of the random operators $(\mathcal{A}_{h_n,n}, n \in \mathbb{N})$:

$$\mathcal{A}_{h_n,n}(f)(x) := \frac{1}{nh_n^{d+2}} \sum_{i=1}^{n} K\left(\frac{\|x - X_i\|_2}{h_n}\right)(f(X_i) - f(x)), \quad x \in \mathcal{M} \tag{2.1}$$

where $K : \mathbb{R}_+ \to \mathbb{R}_+$ is a function of bounded variation and $(h_n, n \in \mathbb{N})$ is a sequence of positive real numbers converging to 0.
Such operators can be viewed as the infinitesimal generator of continuous time random walks visiting the points $(X_i)_{i\in[\![1,n]\!]}$, where $[\![1,n]\!] = \{1,\dots n\}$. Such process jumps from its position x to the new position X_i at a rate $K(\|x - X_i\|_2/h_n)/(nh_n^{d+2})$ that depends on the distance between x and X_i. Notice that here, the Euclidean distance is used. When walking on the manifold $\mathcal{M}$, using the geodesic distance and considering the operator

$$\widetilde{\mathcal{A}}_{h_n,n}(f)(x) := \frac{1}{nh_n^{d+2}} \sum_{i=1}^{n} K\left(\frac{\rho(x, X_i)}{h_n}\right)(f(X_i) - f(x)), \quad x \in \mathcal{M}$$

could be also very natural. In fact, for smooth manifolds, the limits of the two operators $\mathcal{A}_{h_n,n}$ and $\widetilde{\mathcal{A}}_{h_n,n}$ are the same, as is indicated by [51, Prop. 2]. In view of applications to manifold learning, when $\mathcal{M}$ is unknown and when only the sample points X_i's are available, using the norm of the ambient space $\mathbb{R}^m$ can be justified.
The operator (2.1) can also be seen as a graph Laplacian for a weighted graph with vertices being data points and their convergence has been studied extensively in machine learning literature to approximate the Laplace-Beltrami operator of $\mathcal{M}$ (see e.g. [108, 55, 83, 15, 16, 113]). Nonetheless, most of these results are done for Gaussian kernel, i.e., $K(a) = e^{-a^2}$, or sufficiently smooth kernels. These assumptions are too strong to include the case of ε-geometric graphs or the ' true' k-nearest neighbor graphs (abbreviated as kNN), and that correspond to choices of indicators for the kernel K. In recent years, many works had been done to relax the regularity of K and gave birth to many interesting papers (e.g. [25, 115]), as discussed below.

In the sequel, under a mild assumption on K (weaker than continuity, see Assumption 3 below) and a condition on the rate of convergence of (h_n), we show that almost surely, the sequence of operators $(\mathcal{A}_{h_n,n})$ converges uniformly on $\mathcal{M}$ to the second order differential operator $\mathcal{A}$ on $\mathcal{M}$ defined as

$$\mathcal{A}(f) := c_0\bigg(\langle \nabla_{\mathcal{M}}(p), \nabla_{\mathcal{M}}(f)\rangle + \frac{1}{2}p\Delta_{\mathcal{M}}(f)\bigg), \tag{2.2}$$

for all function $f \in \mathcal{C}^3(\mathcal{M})$, where $\nabla_{\mathcal{M}}$ and $\Delta_{\mathcal{M}}$ are respectively the gradient operator and Laplace-Beltrami operator of $\mathcal{M}$ (introduced in Section 2.3) and

$$c_0 := \frac{1}{d}\int_{\mathbb{R}^d} K(\|v\|_2)\,\|v\|_2^2 \mathrm{d}v = \frac{1}{d}S_{d-1}\int_0^\infty K(a)a^{d+1}\mathrm{d}a, \tag{2.3}$$

where S_{d-1} denotes the volume of the unit sphere of $\mathbb{R}^d$. Moreover, a convergence rate is also deduced, as stated in our main Theorem below (Theorem 2.1.1) that we will present after having enounced the assumptions needed on the kernel K:

Assumption 3. *The kernel* $K : \mathbb{R}_+ \to \mathbb{R}_+$ *is a measurable function with* $K(\infty) = 0$ *and of bounded variation* H *such that:*

$$\int_0^\infty a^{d+3}\, dH(a) < \infty. \tag{2.4}$$

Recall that the total variation H of a kernel K is defined for each nonnegative number a as $H(a) = \sup \sum_{i=1}^n |K(a_i) - K(a_{i-1})|$, where the supremum ranges over all $n \in \mathbb{N}$ and all subdivisions $0 = a_0 < \cdots < a_n = a$ of $[0, a]$. Assumption 3 is the key to avoid making continuity hypotheses on the kernel K.

Theorem 2.1.1 (Main theorem)**.** *Suppose that the density of points* $p(x)$ *on the compact smooth manifold* $\mathcal{M}$ *is of class* $\mathcal{C}^2$*. Suppose that Assumptions 3 for the kernel* K *are satisfied and that:*

$$\lim_{n\to+\infty} h_n = 0, \qquad \textit{and} \qquad \lim_{n\to+\infty} \frac{\log h_n^{-1}}{n h_n^{d+2}} = 0. \tag{2.5}$$

Then, with probability 1*, for all* $f \in \mathcal{C}^3(\mathcal{M})$*,*

$$\sup_{x\in\mathcal{M}} |\mathcal{A}_{h_n,n}(f)(x) - \mathcal{A}(f)(x)| = O\left(\sqrt{\frac{\log h_n^{-1}}{n h_n^{d+2}}} + h_n\right). \tag{2.6}$$

Notice that the window h_n that optimizes the convergence rate in (2.6) is of order $n^{-1/(d+4)}$, up to log factors, resulting in a convergence rate in $n^{-1/(4+d)}$. This corresponds to the optimal convergence rate announced in [66].
An important point in the assumptions of Theorem 2.1.1 is that K is not necessarily continuous nor with mass equal to 1. This can allow to tackle the cases of geometric or kNN graphs for example.

This theorem extends the convergence given by Giné and Koltchinskii [55, Th 5.1]. They consider the kernel $K(a) = e^{-a^2}$ and control the convergence of the generators uniformly over a class of functions f of class $\mathcal{C}^3$, uniformly bounded and with uniformly bounded derivatives up to the third order. For such class of functions, the constants in the right hand side (RHS) of (2.6) can be made independent of f and we recover a similar uniform bound.
The condition (2.5) results from a classical bias-variance trade-off that appears in a similar way in the work of Giné and Koltchinskii [55]. Notice that the speed $\sqrt{\log h_n^{-1}/(n h_n^{d+2})}$ is also obtained by these authors under the additional assumption that $n h_n^{d+4}/\log h_n^{-1} \to 0$. We do not make this assumption here. When the additional assumption of Giné and Koltchinskii is satisfied, our rate and their rate coincide as: $h_n^2 = o\big(\log h_n^{-1}/(n h_n^{d+2})\big)$. Hein et al. [65, 66] extended the results of Giné and Koltchinskii to other kernels K, but requesting in particular that these kernels are twice continuously differentiable and with exponential decays (see e.g. [65, Assumption 2] or [66, Assumption 20]). Singer [108], considering Gaussian kernels, upper bounds the variance term in a different manner compared to Hein et al., improving their convergence rate when p is the uniform distribution.
To our knowledge there are a few works where the consistency of graph Laplacians is proved without continuity assumptions on the kernel K. Ting et al. [115] also worked under the bounded variation assumption on K. Additionally, they had to assume that K is compactly supported. In [25], Calder and Garcia-Trillos considered a non-increasing kernel with support on $[0, 1]$ and Lipschitz continuous on this interval. This choice allows them to consider $K(a) = \mathbf{1}_{[0,1]}(a)$. Calder and García Trillos established Gaussian concentration of $\mathcal{A}_{h_n,n}(f)(x)$ and showed that

the probability that $|\mathcal{A}_{h_n,n}(f)(x) - \mathcal{A}f(x)|$ exceeds some threshold δ is exponentially small, of order $\exp(-C\delta^2 n h_n^{d+2})$, when $n \to +\infty$. In this paper, thanks to the uniform convergence in Theorem 2.1.1, we obtain a similar result with additional uniformity on the test functions f:

Corollary 2.1.2. *Suppose that the density p on the smooth manifold $\mathcal{M}$ is of class $\mathcal{C}^2$, and that Assumptions 3 and* (2.5) *are satisfied. Then there exists a constant $C' > 0$ (see* (2.58)*), such that for all n and $\delta \in \left[h_n \vee \sqrt{\frac{\log h_n^{-1}}{n h_n^{d+2}}}, 1\right]$, we have:*

$$\mathbf{P}\left(\sup_{f\in\mathcal{F}}\sup_{x\in\mathcal{M}} |\mathcal{A}_{h_n,n}(f)(x) - \mathcal{A}f(x)| > C'\delta\right) \leqslant \exp(-n h_n^{d+2}\delta^2), \tag{2.7}$$

where $\mathcal{F}$ is the family of $\mathcal{C}^3(\mathcal{M})$ functions bounded by 1 and with derivatives up to the third order also bounded by 1.

The fact that the convergence in Theorem 2.1.1 is uniform has several other applications. For example, it can be a step to study the spectral convergence for the graph Laplacian using the Courant-Fisher minmax principle (see e.g. [25]). Interestingly, the uniform convergence of the Laplacians is also used to study Gaussian free fields on manifolds [29].

The result of Theorem 2.1.1 can be extended to the convergence of kNN Laplacians in the following way. Recall that for $n, k \in \mathbb{N}$ fixed, such that $k \leqslant n$, the kNN graph on the vertices $\{X_1, \dots X_n\}$ is a graph for which the vertices have out-degree k. Each vertex has outgoing edges to its k-nearest neighbor for the Euclidean distance (again, the geodesic distance could be considered).
For $x \in \mathcal{M}$, the distance between x and its k-nearest neighbor is defined as:

$$R_{n,k}(x) = \inf\Big\{r \geqslant 0,\ \sum_{i=1}^{n} \mathbf{1}_{\|x-X_i\|_2 \leqslant r} \geqslant k\Big\}. \tag{2.8}$$

The Laplacian of the kNN-graph is then, for $x \in \mathcal{M}$,

$$\mathcal{A}_n^{\mathrm{kNN}}(f)(x) := \frac{1}{n R_{n,k_n}^{d+2}(x)} \sum_{i=1}^{n} \mathbf{1}_{[0,1]}\left(\frac{\|X_i - x\|_2}{R_{n,k_n}(x)}\right)(f(X_i) - f(x)). \tag{2.9}$$

A major difficulty here is that the width of the moving window, $R_{n,k_n}(x)$ is random and depends on $x \in \mathcal{M}$, contrary to the previous h_n. The above expression corresponds to the choice of the kernel $K(a) = \mathbf{1}_{[0,1]}(a)$. The case of kNN has been much discussed in the literature but to our knowledge, there are few works where the consistency of kNN graph Laplacians have been fully and rigorously considered, because: 1) of the non-regularity of the kernel K and 2) of the fact that the kNN graph is not symmetric, more precisely, the vertex X_i is among the k-nearest neighbors of a vertex X_j does not imply that X_j is among the k-nearest neighbors of X_i. Ting et al. [115] discussed that if there is a kind of Taylor expansion with respect to x of the window $R_{n,k_n}(x)$, one might prove a pointwise convergence for kNN graph Laplacian, without convergence rate. In the present proof, we do not require such Taylor-like expansion. Let us mention also the work of Calder and García Trillos [25] where the spectral convergence is established. In other papers such as [28], (2.8) is considered for defining the window width h_n but the kernel K remains continuous.

We will prove the following limit theorem for the rescaled kNN Laplacian:

Theorem 2.1.3. *Under Assumption 3, if the density $p \in \mathcal{C}^2(\mathcal{M})$ is such that for all $x \in \mathcal{M}$,*

$$0 < p_{\min} \leqslant p(x) \leqslant p_{\max}, \tag{2.10}$$

and if

$$\lim_{n\to+\infty}\frac{k_n}{n}=0, \quad \textit{and} \quad \lim_{n\to+\infty}\frac{1}{n}\left(\frac{k_n}{n}\right)^{-1-2/d}\log\left(\frac{k_n}{n}\right)=0, \tag{2.11}$$

we have with probability 1,

$$\sup_{x\in\mathcal{M}}\left|\mathcal{A}_n^{\mathrm{kNN}}(f)(x)-\mathcal{A}(f)(x)\right|=O\left(\sqrt{\log\left(\frac{n}{k_n}\right)}\frac{1}{\sqrt{k_n}}\left(\frac{n}{k_n}\right)^{1/d}+\left(\frac{k_n}{n}\right)^{1/d}\right). \tag{2.12}$$

This theorem is proved in Section 2.6. Notice that the important point in the assumption (2.10) is the lower bound, since in our case of compact manifold, any continuous function p is bounded. The condition (2.11) and the rate of convergence in (2.12) come from that fact that the random distance $R_{n,k_n}(x)$ stays with large probability in an interval $[\kappa^{-1}h_n,\kappa h_n]$ for some $\kappa>1$ independent of x and n, and for a sequence h_n independent of x. This property is based on a result of Cheng and Wu [28]. The proof of Theorem 2.1.3 follows the main steps presented in the proof of Theorem 2.1.1 with some slight modifications.
Notice that the assumption (2.11) is satisfied for

$$k_n=Cn^{1-\alpha}, \qquad \text{with } \alpha\in\big(0,\frac{1}{d+2}\big),$$

for instance. Optimizing the upper bound in (2.12) by varying α in the above choice gives:

$$k_n=Cn^{\frac{4}{d+4}},$$

yielding again a convergence rate of $O\Big(\sqrt{\log(n)}\;n^{-1/(d+4)}\Big)$.

Finally, we make the link between the convergence of the generators and the convergence of the associated stochastic processes. As mentioned at the beginning of the article, the generator $\mathcal{A}_{h_n,n}$ can be seen as the infinitesimal generator of continuous time random walks $(X^{(n)})_{n\geqslant0}$ visiting the points $(X_i)_{i\in[\![1,n]\!]}$. Their trajectories are described by the following stochastic differential equation (SDE):

$$X_t^{(n)}=X_0^{(n)}+\int_0^t\int_{\mathbf{N}}\int_{\mathbb{R}_+}\mathbf{1}_{i\leqslant n}\mathbf{1}_{\theta\leqslant\frac{1}{nh_n^{d+2}}K\left(\frac{\|X_i-X_{s_-}^{(n)}\|_2}{h_n}\right)}(X_i-X_{s_-}^{(n)})\;Q(\mathrm{d}s,\mathrm{d}i,\mathrm{d}\theta) \tag{2.13}$$

with initial condition $X_0^{(n)}$ and where $Q(\mathrm{d}s,\mathrm{d}i,\mathrm{d}\theta)$ is a Poisson point measure on $\mathbb{R}_+\times\mathbf{N}\times\mathbb{R}_+$, independent of $X_0^{(n)}$, and of intensity $\mathrm{d}s\otimes\mathrm{n}(\mathrm{d}i)\otimes\mathrm{d}\theta$, with $\mathrm{d}s$ and $\mathrm{d}\theta$ Lebesgue measures on $\mathbb{R}_+$ and $\mathrm{n}(\mathrm{d}i)$ the counting measure on $\mathbf{N}$ (see e.g. [68] for an introduction on SDEs driven by Poisson point measures). Consider $T>0$ a fixed time. These random walks on $[0,T]$ are stochastic processes with paths in the space $\mathbb{D}([0,T],\mathcal{M})$ of càdlàg $\mathcal{M}$-valued processes, embedded with the Skorokhod topology (see [21, 70]), and converge to a diffusion on the manifold $\mathcal{M}$ with generator $\mathcal{A}$:

Theorem 2.1.4. *Let $T>0$ be fixed. Suppose that the density p on the smooth manifold $\mathcal{M}$ is of class $\mathcal{C}^2$ and that Assumptions 3 and* (2.5) *are satisfied. Assume additionally that the initial conditions $(X_0^{(n)})_{n\geqslant0}$ converge in distribution to a probability measure ν on $\mathcal{M}$. Then, the sequence of random walks $(X^{(n)})_{n\geqslant0}$ converges in distribution, and in the space $\mathbb{D}([0,T],\mathcal{M})$, to the diffusion X that is the unique solution of the martingale problem associated with the operator $\mathcal{A}$ with initial distribution ν.*

Similarly, we introduce the random walk associated to the kNN-generator as the solution of the following SDE:

$$X_t^{(n),\mathrm{kNN}} = X_0^{(n),\mathrm{kNN}} \\ + \int_0^t \int_{\mathbf{N}} \int_{\mathbb{R}_+} \mathbf{1}_{i\leqslant n} \mathbf{1}_{\theta \leqslant \frac{1}{nR_{n,k_n}^{d+2}\left(X_{s_-}^{(n),\mathrm{kNN}}\right)}} \mathbf{1}_{\|X_i - X_{s_-}^{(n),\mathrm{kNN}}\|_2 \leqslant R_{n,k_n}^{d+2}\left(X_{s_-}^{(n),\mathrm{kNN}}\right)} (X_i - X_{s_-}^{(n)})\, Q(\mathrm{d}s, \mathrm{d}i, \mathrm{d}\theta). \tag{2.14}$$

Theorem 2.1.5. *Let $T > 0$ be fixed. Suppose that the density p on the smooth manifold $\mathcal{M}$ is of class $\mathcal{C}^2$ and that Assumptions 3, (2.10) and (2.11) are satisfied. Assume additionally that the initial conditions $(X_0^{(n),\mathrm{kNN}})_{n\geqslant 0}$ converge in distribution to a probability measure ν on $\mathcal{M}$. Then, the sequence of random walks $(X^{(n),\mathrm{kNN}})_{n\geqslant 0}$ converges in distribution, and in the space $\mathbb{D}([0,T],\mathcal{M})$, to the diffusion X that is the unique solution of the martingale problem associated with the operator $\mathcal{A}$ with initial distribution ν.*

The rest of the paper is organized as follows. In Section 2.2, we give the scheme of the proof. The term $|\mathcal{A}_{h_n,n}(f)(x) - \mathcal{A}(f)(x)|$ is separated into a bias error, a variance error and a term corresponding to the convergence of the kernel operator to a diffusion operator. In Section 2.3, we provide some geometric backgrounds that will be useful for the study of the third term, which is treated in Section 2.4. The two first statistical terms are considered in Section 2.5, which will end the proof of Theorem 2.1.1. Corollary 2.1.2 is then proved at the end of this section. In Section 2.6, we treat the convergence of kNN Laplacians: after recalling a concentration result for $R_{n,k_n}(x)$, the proof amounts to considering a uniform convergence over a range of window widths. The functional limit theorems, showing the convergence of the random walks to diffusive limits are shown in Section 2.7.

Notation 2.1.6. *In this paper diam($\mathcal{M}$), $B_{\mathbb{R}^d}(0,r)$ and S_{d-1} denote respectively the diameter of $\mathcal{M}$, $\max_{z,y\in\mathcal{M}}(\|z-y\|_2)$, the ball of $\mathbb{R}^d$ centered at 0 with radius r and the volume of the $(d-1)$-unit sphere of $\mathbb{R}^d$.*

2.2 Outline of the proof for Theorem 2.1.1

First, we focus on the proof of Theorem 2.1.1. Recall that $\rho(\cdot,\cdot)$ denotes the geodesic distance on $\mathcal{M}$ and that $\mu(\mathrm{d}x)$ is the volume measure on $\mathcal{M}$. We define two new operators $\mathcal{A}_h$, $\tilde{\mathcal{A}}_h$ for each $h > 0, x \in \mathcal{M}, f \in \mathcal{C}^3(\mathcal{M})$:

$$\mathcal{A}_h(f)(x) := \frac{1}{h^{d+2}} \int_{\mathcal{M}} K\left(\frac{\|x-y\|_2}{h}\right)(f(y) - f(x))p(y)\mu(\mathrm{d}y) \tag{2.15}$$

$$\tilde{\mathcal{A}}_h(f)(x) := \frac{1}{h^{d+2}} \int_{\mathcal{M}} K\left(\frac{\rho(x,y)}{h}\right)(f(y) - f(x))p(y)\mu(\mathrm{d}y). \tag{2.16}$$

The difference between $\mathcal{A}_h$ and $\tilde{\mathcal{A}}_h$ relies in the use of the extrinsic Euclidean distance $\|\cdot\|_2$ for $\mathcal{A}_h$ and of the intrinsic geodesic distance $\rho(\cdot,\cdot)$ for $\tilde{\mathcal{A}}_h$. Recall here that these two metrics are comparable for close x and y:

Theorem 2.2.1 (Approximation inequality for Riemannian distance)**.** *[51, Prop. 2] There is a constant c such that for $x, y \in \mathcal{M}$, we have:*

$$\|x-y\|_2 \leq \rho(x,y) \leq \|x-y\|_2 + c\|x-y\|_2^3.$$

□

Let us sketch the proof of Theorem 2.1.1. By the classical triangular inequality,

$$\begin{aligned} |\mathcal{A}_{h_n,n}(f)(x) - \mathcal{A}(f)(x)| \leqslant & \left|\mathcal{A}(f)(x) - \widetilde{\mathcal{A}}_{h_n}(f)(x)\right| \\ & + \left|\widetilde{\mathcal{A}}_{h_n}(f)(x) - \mathcal{A}_{h_n}(f)(x)\right| \\ & + |\mathcal{A}_{h_n}(f)(x) - \mathcal{A}_{h_n,n}(f)(x)| \end{aligned} \tag{2.17}$$

The first term in the RHS of (2.17) corresponds to the convergence of kernel-based generator to a continuous diffusion generator on $\mathcal{M}$. The following proposition is proved in Section 2.4.2:

Proposition 2.2.2 (Convergence of averaging kernel operators)**.** *Under Assumption 3, and if p is of class $\mathcal{C}^2$. Then, for all $f \in \mathcal{C}^3(\mathcal{M})$, we have:*

$$\sup_{x\in\mathcal{M}} \left|\tilde{\mathcal{A}}_h(f)(x) - \mathcal{A}(f)(x)\right| = O(h).$$

This approximation is based on tools from differential geometry and exploits the assumed regularities on K and p. Similar results have been obtained, in particular by [55, Th. 3.1] but with continuous assumptions on K that exclude the kNN cases.

The second term in (2.17) corresponds to the approximation of the Euclidean distance by the geodesic distance and is dealt with the following proposition, proved in Section 2.4.3:

Proposition 2.2.3. *Under Assumption 3, and for a bounded measurable function p, we have, for all f Lipschitz continuous on $\mathcal{M}$:*

$$\sup_{x\in\mathcal{M}} \left|\mathcal{A}_h(f)(x) - \tilde{\mathcal{A}}_h(f)(x)\right| = O(h).$$

For the last term in the RHS of (2.17), note that:

$$\mathbb{E}\left[\mathcal{A}_{h_n,n}f(x)\right] = \mathcal{A}_{h_n}f(x),$$

because $(X_i, i \in \mathbb{N})$ are i.i.d. This term corresponds to a statistical error. The following proposition will be proved in Section 2.5 using Vapnik-Chervonenkis theory:

Proposition 2.2.4. *Under Assumption 3 and for a bounded measurable function p, we have, for all $f \in \mathcal{C}^3(\mathcal{M})$,*

$$\sup_{x\in\mathcal{M}} |\mathcal{A}_{h_n,n}f(x) - \mathcal{A}_{h_n}f(x)| = O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}} + h_n\right), \textit{a.s.}$$

It is worth noticing that there is an interplay between Euclidean and Riemannian distances. On the one hand, the Vapnik-Chervonenkis theory is extensively studied for Euclidean distances, not for Riemannian distance. On the other hand, approximations on manifolds naturally use local coordinate representations for which the Riemannian distance is well adapted.

2.3 Some geometric backgrounds

2.3.1 Riemannian manifold

Let us recall some facts from differential geometry that will be useful. We refer the reader to [27, 79] for a more rigorous introduction to Riemannian geometry. Let $\mathcal{M}$ be a smooth d-dimensional submanifold of $\mathbb{R}^m$.

At each point x of $\mathcal{M}$, there is a tangent vector space $T_x\mathcal{M}$ that contains all the tangent vectors of $\mathcal{M}$ at x. The tangent bundle of $\mathcal{M}$ is denoted by $T\mathcal{M} = \sqcup_{x\in\mathcal{M}} T_x\mathcal{M}$. For each $x \in \mathcal{M}$, the canonical scalar product $\langle \cdot, \cdot \rangle_{\mathbb{R}^m}$ of $\mathbb{R}^m$ induces a natural scalar product on $T_x\mathcal{M}$, denoted by $\mathbf{g}(x)$. The application $\mathbf{g}$, which associates each point x with a scalar product on $T_x\mathcal{M}$, is then called the Riemannian metric on $\mathcal{M}$ induced by the ambient space $\mathbb{R}^m$. For $\xi, \eta \in T_x\mathcal{M}$, we use the classical notation $\langle \xi, \eta \rangle_{\mathbf{g}}$ to denote the scalar product of ξ and η w.r.t to the scalar product $\mathbf{g}(x)$.

Consider a coordinate chart $\Phi = (x^1, \dots x^d) : U \to \mathbb{R}^d$ on a neighborhood U of x. Denoting by $\left\{ \frac{\partial}{\partial x^1}\big|_x, \frac{\partial}{\partial x^2}\big|_x, \dots, \frac{\partial}{\partial x^d}\big|_x \right\}$ the natural basis of $T_x\mathcal{M}$ associated with the coordinates $(x^1, \dots x^d)$. Then, the scalar product $\mathbf{g}(x)$ is associated to a matrix $(g_{ij})_{i,j\in[\![1,d]\!]}$ in the sense that in this coordinate chart, for ξ and $\eta \in T_x\mathcal{M}$,

$$\langle \xi, \eta \rangle_{\mathbf{g}} = \sum_{i,j=1}^{d} g_{ij}(x) \xi^i \eta^j, \tag{2.18}$$

where $(\xi^i), (\eta^j)$ are the coordinates of ξ and η in the above basis of $T_x\mathcal{M}$. Notice that, for each $i, j \in [\![1, d]\!]$

$$g_{ij}(x) := \left\langle \frac{\partial}{\partial x^i}\bigg|_x, \frac{\partial}{\partial x^j}\bigg|_x \right\rangle_{\mathbf{g}}, \tag{2.19}$$

and $g_{ij} : U \subset M \to \mathbb{R}$ is smooth. For a real function f on $\mathcal{M}$, we will denote $\widehat{f}$ its expression in the local chart: $\widehat{f} = f \circ \Phi^{-1}$. Recall that the derivative $\frac{\partial f}{\partial x^j}$ is defined as:

$$\frac{\partial f}{\partial x^j} := \frac{\partial \widehat{f}}{\partial x^j} \circ \Phi.$$

Also we denote

$$\widehat{g}_{ij} := g_{ij} \circ \Phi^{-1}, \tag{2.20}$$

which will be called the coordinate representation of the Riemannian metric in the local chart Φ.

Charts (from $U \subset \mathcal{M} \to \mathbb{R}^d$) induce local parameterizations of the manifold (from $\mathbb{R}^d \to U \subset \mathcal{M}$). Among all possible local coordinate systems of a neighborhood of x in $\mathcal{M}$, there are *normal coordinate charts* (see [79, p. 131-132] or the remark below for a definition). We denote by $\mathcal{E}_x$ the Riemannian normal parameterization at x, i.e., $\mathcal{E}_x^{-1}$ is the corresponding normal coordinate chart.

Remark 2.3.1 (Construction of $\mathcal{E}_x^{-1}$)**.** *For the sake of completeness, we briefly recall the construction of [79]. Let U be an open subset of $\mathcal{M}$. There exists a local orthonormal frame $(E_i)_{i\in[\![1,d]\!]}$ over U, see [79, Prop. 2.8, p. 14]. The tangent bundle TU can be identified with $U \times \mathbb{R}^d$ thanks to the smooth map:*

$$F \;:\; \begin{array}{ccc} U \times \mathbb{R}^d & \to & TU \\ (x, (v_1, \dots v_d)) & \mapsto & v = \sum_{i=1}^d v_i E_i|_x. \end{array} \tag{2.21}$$

So for each $x \in U$, $F(x, \cdot)$ is an isometry between $\mathbb{R}^d$ and $T_x\mathcal{M}$.
Recall that by [79, Prop 5.19, p. 128], the exponential map $\exp(\cdot)$ of $\mathcal{M}$ can be defined on a non-empty open subset W of $T\mathcal{M}$ such that $\forall x \in \mathcal{M}$, $\overrightarrow{0}_x \in W$, where $\overrightarrow{0}_x$ is the zero element of $T_x\mathcal{M}$. Then, the map $\exp \circ F : (x, v) \mapsto \mathcal{E}_x(v) := \exp \circ F(x, (v_1, \dots, v_d))$ is well-defined on $F^{-1}(W \cap TU)$ and $\mathcal{E}_x^{-1}$ is a Riemannian normal coordinate chart at $x \in U$, smooth with respect to x.

Let us state some properties of the normal coordinate charts.

Theorem 2.3.2 (Derivatives of Riemannian metrics in normal coordinate charts)**.** *[79, Prop. 5.24] For $x \in \mathcal{M}$, let $\mathcal{E}_x^{-1} : U \subset \mathcal{M} \to \mathbb{R}^d$ be a normal coordinate chart at a point x such that $\mathcal{E}_x^{-1}(x) = 0$ and let $(\widehat{g}_{ij}; 1 \leq i, j \leq d)$ be the coordinate representation of the Riemannian metric of $\mathcal{M}$ in the local chart $\mathcal{E}_x^{-1}$. Then for all i, j,*

$$\widehat{g}_{ij}(0) = \delta_{ij}, \quad \widehat{g}'_{ij}(0) = 0, \tag{2.22}$$

where δ_{ij} is the Kronecker delta. Additionally, for all $y \in U$,

$$\rho(x, y) = \|\mathcal{E}_x^{-1}(y)\|_2. \tag{2.23}$$

Notation 2.3.3. *For any function $f : \mathbb{R}^d \to \mathbb{R}^k$, we denote by $f' : \mathbb{R}^d \to \mathbb{R}^k$ the linear map that represents the first order derivative of f. Similarly, we denote respectively by $f'' : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^k$ and $f''' : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^k$ the bi-linear map and the tri-linear map that represent the second order derivative and the third order derivative of f. Thus, the Taylor's expansion of f up to third order can be written as*

$$f(x + v) = f(x) + f'(x)(v) + \frac{1}{2} f''(x)(v, v) + \frac{1}{6} f'''(x + \varepsilon v)(v, v, v),$$

for some $\varepsilon \in (0, 1)$.

For the normal parameterizations $\mathcal{E}_x$, we now state some uniform controls that are keys for our computations in the sequel.

Theorem 2.3.4 (Existence of a "good" family of parameterizations.)**.** *There exist constants $c_1, c_2 > 0$ and a family $(\mathcal{E}_x, x \in \mathcal{M})$ of smooth local parameterizations of $\mathcal{M}$ which have the same domain $B_{\mathbb{R}^d}(0, c_1)$ such that for all $x \in \mathcal{M}$,*

i. $\mathcal{E}_x^{-1}$ is a normal coordinate chart of $\mathcal{M}$ and $\mathcal{E}_x(0) = x$.

ii. For $v \in B_{\mathbb{R}^d}(0, c_1)$, we denote by $(\widehat{g}^x_{ij}(v); 1 \leq i, j \leq d)$ the coordinate representation of the Riemannian metric $\mathbf{g}(\mathcal{E}_x(v))$ of $\mathcal{M}$ in the local parameterization $\mathcal{E}_x$. Then for all $v \in B_{\mathbb{R}^d}(0, c_1)$:

$$\left| \sqrt{\det \widehat{g}^x_{ij}(v)} - 1 \right| \leq c_2 \|v\|_2^2. \tag{2.24}$$

iii. We have $\|\mathcal{E}_x(v) - x\|_2 \leq \|v\|_2$. In addition, for all $v \in B_{\mathbb{R}^d}(0, c_1)$,

$$\|\mathcal{E}_x(v) - x - \mathcal{E}'_x(0)(v)\|_2 \leq c_2 \|v\|_2^2, \tag{2.25}$$

and

$$\|\mathcal{E}_x(v) - x - \mathcal{E}'_x(0)(v) - \frac{1}{2}\mathcal{E}''_x(0)(v, v)\|_2 \leq c_2 \|v\|_2^3, \tag{2.26}$$

Proof for Theorem 2.3.4. Let U be an open domain of a local chart of $\mathcal{M}$. Following Remark 2.3.1 and noticing that there is always an orthonormal frame over U, we can define a family of normal parameterizations $(\mathcal{E}_x)_{x \in U}$.
First, we easily note that $\|\mathcal{E}_x(v) - x\|_2 \leq \|v\|_2$ thanks to Theorem 2.2.1 and (2.23).
Restricting U if necessary (by an open subset with compact closure in U), there exists constants $c_1, c_2 > 0$ such that $\exp \circ F$ is well-defined on $U \times B_{\mathbb{R}^d}(0, c_1)$ and that all $v \in B_{\mathbb{R}^d}(0, c_1)$, (2.25)-(2.26) hold by Taylor expansions of $\mathcal{E}_x$. Equation (2.24) is a consequence of the smoothness of $\mathcal{E}_x$ and Theorem 2.3.4.
Clearly, for each point $y \in \mathcal{M}$, we can find an open neighborhood U of y and positive constants c_1 and c_2 such as above. Hence, such open sets form an open covering of $\mathcal{M}$. Therefore, by the compactness of $\mathcal{M}$, there exists a finite covering of $\mathcal{M}$ by such open sets U and therefore, the constants c_1 and c_2 can be chosen uniformly for all $\mathcal{E}_x$. □

2.3.2 Gradient operator, Laplace-Beltrami operator

Given a Riemannian manifold $(\mathcal{M}, \mathbf{g})$, the gradient operator $\nabla_{\mathcal{M}}$ and the Laplace-Beltrami operator $\Delta_{\mathcal{M}}$ are, as suggested by their names, the generalizations for differential manifolds of the gradient $\nabla_{\mathbb{R}^m}$, the Laplacian $\Delta_{\mathbb{R}^m}$ in the Euclidean space $\mathbb{R}^m$.
For a function f of class $\mathcal{C}^1$ on $\mathcal{M}$, the gradient $\nabla_{\mathcal{M}} f$ is expressed in local coordinates as

$$\nabla_{\mathcal{M}} f(x) = \sum_{i,j=1}^{d} g^{ij}(x) \frac{\partial f}{\partial x^i}(x) \left.\frac{\partial}{\partial x^j}\right|_x, \tag{2.27}$$

where $\left(g^{ij}\right)_{1\leqslant i,j,\leqslant d}$ is the inverse matrix of $(g_{ij})_{1\leqslant i,j,\leqslant d}$. Since $\sum_{j=1}^d g^{ij} g_{jk} = \delta_{ik}$, we note that for f, h functions of class $\mathcal{C}^1$,

$$\langle \nabla_{\mathcal{M}}(f), \nabla_{\mathcal{M}}(h) \rangle_{\mathbf{g}} = \sum_{i,j=1}^{d} g^{ij} \frac{\partial f}{\partial x^i} \frac{\partial h}{\partial x^j}. \tag{2.28}$$

The Laplace-Beltrami operator is defined by (see [67, Section 3.1])

$$\Delta_{\mathcal{M}} f := \sum_{i,j=1}^{d} \frac{1}{\sqrt{\det(\mathbf{g})}} \frac{\partial}{\partial x^i} \left(\sqrt{\det(\mathbf{g})} g^{ij} \frac{\partial f}{\partial x^j} \right). \tag{2.29}$$

When using normal coordinates, the expressions of the Laplacian and the gradient of a smooth function f at a point x match their definitions in $\mathbb{R}^d$.

Proposition 2.3.5. *Suppose that $\Phi : U \subset \mathcal{M} \to \mathbb{R}^d$ is a normal coordinate chart at a point x in $\mathcal{M}$ such that $\Phi(x) = 0$, then:*

i. $\langle \nabla_{\mathcal{M}} f(x), \nabla_{\mathcal{M}} h(x) \rangle_{\mathbf{g}} = \langle \nabla_{\mathbb{R}^d} \hat{f}(0), \nabla_{\mathbb{R}^d} \hat{h}(0) \rangle$.

ii. $\Delta_{\mathcal{M}} f(x) = \Delta_{\mathbb{R}^d} \hat{f}(0)$,

Proof for Proposition 2.3.5. Recall that $g_{ij}(x) = \widehat{g}_{ij}(0)$. By Theorem 2.3.2, we know that $\hat{g}_{ij}(0) = \delta_{ij}$, thus, $\widehat{g}^{ij}(0) = \delta_{ij}$ and *i.* is a consequence of (2.28). For the equality *ii.*, we use (2.29). Since for the normal coordinates $\det \widehat{\mathbf{g}}(0) = 1$ and since the derivatives of $\widehat{g}_{ij}$ and $\widehat{g}^{ij}$ vanish at 0, we have the conclusion. □

2.4 Some kernel-based approximations of $\mathcal{A}$

The aim of this Section is to prove the estimates for the two error terms in the RHS of (2.17) and prove the Propositions 2.2.2 and 2.2.3. Both error terms are linked with the geometry of the problem and use the results presented in Section 2.3. The first one deals with the approximation of the Laplace-Beltramy operator by a kernel estimator (see Section 2.4.2), while the second one treats the differences between the use of the Euclidean norm of $\mathbb{R}^m$ and the use of the geodesic distance (see Section 2.4.3).

2.4.1 Weighted moment estimates

We begin with an auxiliary estimation. The result is related to kernel smoothing and can also be useful in density estimation on manifolds (see e.g. [18]).

Lemma 2.4.1. *Under Assumption 3, uniformly in $x \in \mathcal{M}$, when h converges to 0, we have:*

$$\frac{1}{h^{d+2}} \int_{\mathcal{M}} \mathbb{1}_{\rho(x,y)\geq c_1} K\left(\frac{\rho(x,y)}{h}\right) \mu(\mathrm{d}y) = o(h), \tag{2.30}$$

$$\frac{1}{h^{d+2}} \int_{\mathcal{M}} \mathbb{1}_{\rho(x,y)\geq c_1} K\left(\frac{\|x-y\|_2}{h}\right) \mu(\mathrm{d}y) = o(h), \tag{2.31}$$

and there is a generic constant c such that for all point $x \in \mathcal{M}$ and positive number $h > 0$, we have:

$$\frac{1}{h^{d+2}} \int_{\mathcal{M}} K\left(\frac{\rho(x,y)}{h}\right) \|x-y\|_2^3 \mu(\mathrm{d}y) \leq ch, \tag{2.32}$$

$$\frac{1}{h^{d+2}} \int_{\mathcal{M}} K\left(\frac{\rho(x,y)}{h}\right) \|x-y\|_2^2 \mu(\mathrm{d}y) \leq c, \tag{2.33}$$

$$\frac{1}{h^{d+2}} \int_{\mathcal{M}} K\left(\frac{\|x-y\|_2}{h}\right) \|x-y\|_2^3 \mu(\mathrm{d}y) \leq ch, \tag{2.34}$$

$$\frac{1}{h^{d+2}} \int_{\mathcal{M}} K\left(\frac{\|x-y\|_2}{h}\right) \|x-y\|_2^2 \mu(\mathrm{d}y) \leq c. \tag{2.35}$$

Proof of Lemma 2.4.1. Using Lemma 2.9.5, we have:

$$\begin{aligned}
\int_{\mathcal{M}} \mathbb{1}_{\rho(x,y)\geq c_1} K\left(\frac{\rho(x,y)}{h}\right) \mu(\mathrm{d}y) &\leq \mu(\mathcal{M}) \sup_{r\geq c_1} K\left(\frac{r}{h}\right) \\
&\leq \mu(\mathcal{M}) \left[H(\infty) - H\left(\frac{c_1}{h}\right)\right] \\
&= \mu(\mathcal{M}) \int_{(c_1/h,\infty)} \mathrm{d}H(a) \\
&\leq h^{d+3} \frac{\mu(\mathcal{M})}{c_1^{d+3}} \int_{(c_1/h,\infty)} a^{d+3} \mathrm{d}H(a). \qquad (2.36)
\end{aligned}$$

Thanks to the boundedness of $\int_0^\infty a^{d+3}\mathrm{d}H(a)$, we obtain (2.30). Then, as a consequence of (2.30), by the compactness of $\mathcal{M}$, we easily observe that uniformly in $x \in \mathcal{M}$, when h converges to 0,

$$\frac{1}{h^{d+2}} \int_{\mathcal{M}} \mathbb{1}_{\rho(x,y)\geq c_1} K\left(\frac{\rho(x,y)}{h}\right) \|x-y\|_2^3 \mu(\mathrm{d}y) = o(h).$$

So, to prove Inequality (2.32), it is left to prove that uniformly in x, when h converges to 0,

$$I := \frac{1}{h^{d+3}} \int_{\mathcal{M}} \mathbb{1}_{\rho(x,y)<c_1} K\left(\frac{\rho(x,y)}{h}\right) \|x-y\|_2^3 \mu(\mathrm{d}y) = O(1). \tag{2.37}$$

Recall that in Theorem 2.3.4, we showed that for each point $x \in \mathcal{M}$, there is a local smooth parameterization $\mathcal{E}_x$ of $\mathcal{M}$ that has many nice properties, especially $\rho(x,y) = \|\mathcal{E}_x^{-1}(y)\|_2$ for all y within an appropriate neighborhood of x by (2.23). Thus, the term I in the left hand side (LHS) of (2.37) can be re-written in its coordinate representation under the parameterization $\mathcal{E}_x$ by using the change of variables $v = \mathcal{E}_x^{-1}(y)$:

$$I = \frac{1}{h^{d+3}} \int_{B_{\mathbb{R}^d}(0,c_1)} K\left(\frac{\|v\|_2}{h}\right) \|x - \mathcal{E}_x(v)\|_2^3 \sqrt{\det \widehat{g}_{ij}^x}(v)\mathrm{d}v.$$

Then, using Theorem 2.2.1 and Theorem 2.3.4 (*ii* and *iii*),

$$\begin{aligned}
I &\leq \frac{c_2}{h^{d+3}} \int_{B_{\mathbb{R}^d}(0,c_1)} K\left(\frac{\|v\|_2}{h}\right) \|v\|_2^3(1 + c_2\|v\|_2^2)\mathrm{d}v \\
&\leq \frac{c_2}{h^{d+3}} \int_{B_{\mathbb{R}^d}(0,c_1)} K\left(\frac{\|v\|_2}{h}\right) \|v\|_2^3(1 + c_2 c_1^2)\mathrm{d}v \\
&\leq \frac{c_2}{h^{d+3}} \int_{\mathbb{R}^d} K\left(\frac{\|v\|_2}{h}\right) \|v\|_2^3(1 + c_2 c_1^2)\mathrm{d}v. \\
&= c_2(1 + c_2 c_1^2) \int_{\mathbb{R}^d} K(\|v\|_2)\|v\|_2^3 \mathrm{d}v, \qquad (2.38)
\end{aligned}$$

Using the spherical coordinate system when $d \geqslant 2$:

$$\begin{aligned} I \leqslant & c_2(1+c_2c_1^2)\left[\int_0^\infty K(a)a^3 \times a^{d-1}\mathrm{d}a\right] \times \\ & \times \left[\int_{[0,2\pi]\times[0,\pi]^{d-2}} \sin^{d-2}(\theta_1)\sin^{d-3}(\theta_2)\cdots\sin(\theta_{d-2})\mathrm{d}\theta\right] \\ = & c_2(1+c_2c_1^2)\ S_{d-1}\left[\int_0^\infty K(a)a^{d+2}\mathrm{d}a\right]. \end{aligned}$$

For $d=1$, we use that $\int_{\mathbb{R}^1} K(|v|)|v|^3\mathrm{d}v = 2\times\int_0^\infty K(a)a^{1+2}\mathrm{d}a$.
Hence, by Lemma 2.9.5 in the Appendix, and Fubini's theorem, we have:

$$\begin{aligned} I &\leq c_2(1+c_2c_1^2)S_{d-1}\int_0^\infty (H(\infty)-H(a))\,a^{d+2}\mathrm{d}a \\ &= c_2(1+c_2c_1^2)(d+3)^{-1}S_{d-1}\int_{[0,\infty]} b^{d+3}\mathrm{d}H(b) < \infty. \end{aligned} \tag{2.39}$$

Therefore, Inequality (2.32) is proved. The proof of Inequality (2.33) is similar.
For Inequalities (2.31), (2.34) and (2.35), we observe that they are indeed consequences of (2.30), (2.32) and (2.33). Consider for example (2.31), using again Lemma 2.9.5 and Theorem 2.2.1:

$$\begin{aligned} &\frac{1}{h^{d+2}}\int_{\mathcal{M}} \mathbf{1}_{\rho(x,y)\geqslant c_1} K\left(\frac{\|x-y\|_2}{h}\right)\mu(\mathrm{d}y) \\ \leq&\frac{1}{h^{d+2}}\int_{\mathcal{M}} \mathbf{1}_{\rho(x,y)\geqslant c_1}\left[H(\infty)-H\left(\frac{\rho(x,y)}{h(1+c_3\|x-y\|_2^2)}\right)\right]\mu(\mathrm{d}y) \\ \leq&\frac{1}{h^{d+2}}\int_{\mathcal{M}} \mathbf{1}_{\rho(x,y)\geqslant c_1}\left[H(\infty)-H\left(\frac{\rho(x,y)}{h(1+c_3\mathrm{diam}(\mathcal{M})^2)}\right)\right]\mu(\mathrm{d}y) \\ =&\frac{1}{h^{d+2}}\int_{\mathcal{M}} \mathbf{1}_{\rho(x,y)\geqslant c_1}\tilde{K}\left(\frac{\rho(x,y)}{h}\right)\mu(\mathrm{d}y), \end{aligned} \tag{2.40}$$

for $\tilde{K}(a) := H(\infty) - H\left(\frac{a}{1+c_3\mathrm{diam}(\mathcal{M})^2}\right)$ and where the second inequality uses that H is a non-decreasing function. So Inequality (2.31) corresponds to Inequality 2.30 where K is replaced with $\tilde{K}$. Clearly, the function $\tilde{K}$ is of bounded variation and satisfies Assumption 3, which conclude the proof for (2.31). The arguments are similar for (2.34) and (2.35). □

2.4.2 Proof of Proposition 2.2.2

In this section, we prove Proposition 2.2.2, dealing with the approximation of the Laplace Beltrami operator by a kernel operator.
In the course of the proof, some quantities involving gradients and Laplacian will appear repetitively. The next lemma deal will be useful to deal with these expressions and its proof is postponed to Appendix 2.9.3:

Lemma 2.4.2. *(Some auxiliary calculations) Suppose that $f, h : \mathbb{R}^m \to \mathbb{R}$, $k : \mathbb{R}^d \to \mathbb{R}^m$ are $\mathcal{C}^2$-continuous functions, that $k(0) = x$ and suppose that $G : \mathbb{R}_+ \to \mathbb{R}$ is a locally bounded measurable function. Then, for all $c > 0$:*

$$\begin{aligned} &\int_{B_{\mathbb{R}^d}(0,c)} G(\|v\|_2)\langle\nabla_{\mathbb{R}^m} f(x), k'(0)(v)\rangle\langle\nabla_{\mathbb{R}^m} h(x), k'(0)(v)\rangle\ \mathrm{d}v \\ &\qquad = \langle\nabla_{\mathbb{R}^d}(f\circ k)(0), \nabla_{\mathbb{R}^d}(h\circ k)(0)\rangle\left(\frac{1}{d}\int_{B_{\mathbb{R}^d}(0,c)} G(\|v\|_2)\|v\|_2^2\mathrm{d}v\right), \end{aligned} \tag{2.41}$$

and

$$\int_{B_{\mathbb{R}^d}(0,c)} G(\|v\|_2)\left[\left\langle \nabla_{\mathbb{R}^m} f(x), k'(0)(v) + \frac{1}{2}k''(0)(v,v)\right\rangle + \frac{1}{2}f''(x)(k'(0)(v),k'(0)(v))\right]\mathrm{d}v = \frac{1}{2}\Delta_{\mathbb{R}^d}(f\circ k)(0)\left(\frac{1}{d}\int_{B_{\mathbb{R}^d}(0,c)} G(\|v\|_2)\|v\|_2^2\mathrm{d}v\right). \quad (2.42)$$

Let us now prove Proposition 2.2.2. As $\mathcal{M}$ is compact, $\mathcal{M}$ is a properly embedded submanifold of $\mathbb{R}^m$ (see [78, p.98]). Hence, any function of class $\mathcal{C}^3$ on $\mathcal{M}$ can be extended to a function of class $\mathcal{C}^3$ on $\mathbb{R}^m$, see [78, Lemma 5.34]. So without loss of generality, assume f and p are respectively $\mathcal{C}^3$ and $\mathcal{C}^2$ functions on $\mathbb{R}^m$ with compact supports.
Recall that we want to study $\left|\widetilde{\mathcal{A}}_h(f) - \mathcal{A}(f)\right|$ where $\mathcal{A}$ and $\widetilde{\mathcal{A}}_h$ have been respectively defined in (2.2) and (2.16). So, introducing the constant $c_1 > 0$ of Lemma 2.4.1 and noticing that f and p are uniformly bounded on the compact $\mathcal{M}$, to prove Proposition 2.2.2, we only have to prove that uniformly in $x \in \mathcal{M}$,

$$\left|\mathcal{A}(f)(x) - \frac{1}{h^{d+2}}\int_{\mathcal{M}} \mathbb{1}_{\rho(y,x)<c_1} K\left(\frac{\rho(x,y)}{h}\right)(f(y)-f(x))p(y)\mu(\mathrm{d}y)\right| = O(h).$$

Besides, thanks to the compactness of $\mathcal{M}$ and to the regularity of f and p, Taylor's expansion implies that there is a constant c_4 such that for all $x, y \in \mathcal{M}$:

$$\Big|(f(y)-f(x))p(y) - \left(\langle \nabla_{\mathbb{R}^m} f(x), y-x\rangle + \frac{1}{2}f''(x)(y-x,y-x)\right)p(x) - \langle \nabla_{\mathbb{R}^m} f(x), y-x\rangle\langle \nabla_{\mathbb{R}^m} p(x), y-x\rangle\Big| \le c_4\|x-y\|_2^3. \quad (2.43)$$

Hence, by Inequality (2.32), it is sufficient to prove that uniformly in x,

$$\begin{aligned} J_1 :=& \frac{1}{h^{d+2}}\int_{\mathcal{M}} \mathbb{1}_{\rho(y,x)<c_1} K\left(\frac{\rho(x,y)}{h}\right)\langle \nabla_{\mathbb{R}^m} f(x), y-x\rangle\langle \nabla_{\mathbb{R}^m} p(x), y-x\rangle\mu(\mathrm{d}y) \\ =& c_0\langle \nabla_{\mathcal{M}}(f)(x), \nabla_{\mathcal{M}}(p)(x)\rangle_{\mathbf{g}} + O(h). \end{aligned} \quad (2.44)$$

and

$$\begin{aligned} J_2 :=& \frac{1}{h^{d+2}}\int_{\mathcal{M}} \mathbb{1}_{\rho(y,x)<c_1} K\left(\frac{\rho(x,y)}{h}\right)\left[\langle \nabla_{\mathbb{R}^m} f(x), y-x\rangle + \frac{1}{2}f''(x)(y-x,y-x)\right]\mu(\mathrm{d}y) \\ =& \frac{1}{2}c_0\Delta_{\mathcal{M}}(f)(x) + O(h). \end{aligned} \quad (2.45)$$

The proof is similar than the study of I given by (2.37) in the proof of Lemma 2.4.1. We re-write the considered integrals in coordinate representations. Using the change of variables $v = \mathcal{E}_x^{-1}(y)$, we have

$$J_1 = \frac{1}{h^{d+2}}\int_{B_{\mathbb{R}^d}(0,c_1)} K\left(\frac{\|v\|_2}{h}\right)\langle \nabla_{\mathbb{R}^m} f(x), \mathcal{E}_x(v) - x\rangle\langle \nabla_{\mathbb{R}^m} p(x), \mathcal{E}_x(v) - x\rangle\sqrt{\det \widehat{g}_{ij}^x(v)}\mathrm{d}v.$$

By properties *ii.* and *iii.* in Theorem 2.3.4 we have

$$\begin{aligned}
&\left| J_1 - \int_{B_{\mathbb{R}^d}(0,c_1)} K\left(\frac{\|v\|_2}{h}\right) \langle \nabla_{\mathbb{R}^m} f(x), \mathcal{E}_x(v) - x\rangle \langle \nabla_{\mathbb{R}^m} p(x), \mathcal{E}_x(v) - x\rangle \mathrm{d}v \right| \\
&\le \frac{c_2}{h^{d+2}} \|\nabla_{\mathbb{R}^m} f(x)\|_2 \|\nabla_{\mathbb{R}^m} p(x)\|_2 \int_{B_{\mathbb{R}^d}(0,c_1)} K\left(\frac{\|v\|_2}{h}\right) \|v\|_2^2 . \|\mathcal{E}_x(v) - x\|_2^2 \mathrm{d}v \\
&\le \frac{c_2^3}{h^{d+2}} \|\nabla_{\mathbb{R}^m} f(x)\|_2 \|\nabla_{\mathbb{R}^m} p(x)\|_2 \int_{B_{\mathbb{R}^d}(0,c_1)} K\left(\frac{\|v\|_2}{h}\right) \|v\|_2^4 \mathrm{d}v \\
&\le \frac{c_2^3 c_1}{h^{d+2}} \|\nabla_{\mathbb{R}^m} f(x)\|_2 \|\nabla_{\mathbb{R}^m} p(x)\|_2 \int_{B_{\mathbb{R}^d}(0,c_1)} K\left(\frac{\|v\|_2}{h}\right) \|v\|_2^3 \mathrm{d}v.
\end{aligned}$$

As in the proof of Lemma 2.4.1, we deduce that the latter is bounded by $O(h)$.
Besides, using again Property *iii.* in Theorem 2.3.4, we have that uniformly in x,

$$\begin{aligned}
&\left| \frac{1}{h^{d+2}} \int_{B_{\mathbb{R}^d}(0,c_1)} K\left(\frac{\|v\|_2}{h}\right) \langle \nabla_{\mathbb{R}^m} f(x), \mathcal{E}_x(v) - x\rangle \times \right. \\
&\qquad \left. \langle \nabla_{\mathbb{R}^m} p(x), \mathcal{E}_x(v) - x\rangle \mathrm{d}v - J_{11} \right| = O(h) \qquad (2.46)
\end{aligned}$$

with

$$J_{11} := \frac{1}{h^{d+2}} \int_{B_{\mathbb{R}^d}(0,c_1)} K\left(\frac{\|v\|_2}{h}\right) \langle \nabla_{\mathbb{R}^m} f(x), \mathcal{E}'_x(0)(v)\rangle \langle \nabla_{\mathbb{R}^m} p(x), \mathcal{E}'_x(0)(v)\rangle \mathrm{d}v.$$

Let us now compare J_{11} to the first term of the generator $\mathcal{A}$. Using Equation (2.41) of Lemma 2.4.2, with $G(||v||_2) = K\left(\frac{||v||_2}{h}\right)$, $k = \mathcal{E}_x$, and Proposition 2.3.5, we have:

$$\begin{aligned}
J_{11} &= \frac{1}{h^{d+2}} \left(\frac{1}{d} \int_{B_{\mathbb{R}^d}(0,c_1)} K\left(\frac{\|v\|_2}{h}\right) \|v\|_2^2 \mathrm{d}v \right) \langle \nabla_{\mathbb{R}^d}(f \circ \mathcal{E}_x)(0), \nabla_{\mathbb{R}^d}(p \circ \mathcal{E}_x)(0)\rangle \\
&= \left(\frac{1}{d} \int_{B_{\mathbb{R}^d}(0,c_1/h)} K(\|v\|_2) \|v\|_2^2 \mathrm{d}v \right) \langle \nabla_{\mathcal{M}} f(x), \nabla_{\mathcal{M}} p(x)\rangle_{\mathbf{g}} \\
&= \left(\frac{1}{d} \int_{\mathbb{R}^d} K(\|v\|_2) \|v\|_2^2 \mathrm{d}v \right) \langle \nabla_{\mathcal{M}} f(x), \nabla_{\mathcal{M}} p(x)\rangle_{\mathbf{g}} + o(h),
\end{aligned}$$

where the last estimation is uniform in $x \in \mathcal{M}$ and comes from the second estimation in (2.73) in Lemma 2.9.5. Thus, we have proved Equation (2.44) for J_1.
The proof for J_2, given by (2.45), is similar to what we have done for J_1. For identifying the Laplace-Beltrami operator in the last step of the proof, we use Equation (2.42) of Lemma 2.4.2 and the second point of Proposition 2.3.5. Therefore, we have proved Proposition 2.2.2. □

2.4.3 Proof of Proposition 2.2.3

Let us now prove Proposition 2.2.3. This proposition deals with the difference between the geodesic distance on $\mathcal{M}$ and the Euclidean norm of $\mathbb{R}^m$.

By Inequalities (2.30) and (2.31) of Lemma 2.4.1, we know that uniformly in x, when h converges to 0,

$$\int_{\mathcal{M}} \left(K\left(\frac{\|x-y\|_2}{h}\right) + K\left(\frac{\rho(x-y)}{h}\right) \right) \mathbb{1}_{\rho(x,y)\ge c_1} \mu(\mathrm{d}y) = o(h^{d+3}).$$

Thus, by regularity of f, boundedness of p and compactness of $\mathcal{M}$, uniformly in x, when h converges to 0,

$$\int_{\mathcal{M}} \left(K\left(\frac{\|x-y\|_2}{h}\right) + K\left(\frac{\rho(x-y)}{h}\right)\right)|f(x)-f(y)|p(y)\mathbb{1}_{\rho(x,y)\geq c_1}\mu(\mathrm{d}y) = o(h^{d+3}).$$

So, we only have to prove that uniformly in x,

$$\int_{\mathcal{M}} \left| K\left(\frac{\rho(x,y)}{h}\right) - K\left(\frac{\|x-y\|_2}{h}\right)\right| |f(x)-f(y)|p(y)\mathbb{1}_{\rho(x,y)<c_1}\mu(\mathrm{d}y) = O(h^{d+3}),$$

Or equivalently, using the change of variables $v = \mathcal{E}_x^{-1}(y)$ and $\rho(x,y) = \|\mathcal{E}_x^{-1}(y)\|_2$ by (2.23),

$$\int_{B_{\mathbb{R}^d(0,c_1)}} \left| K\left(\frac{\|v\|_2}{h}\right) - K\left(\frac{\|\mathcal{E}_x(v)-x\|_2}{h}\right)\right| |f\circ\mathcal{E}_x(v) - f(x)|\, p\circ\mathcal{E}_x(v)\sqrt{\det\widehat{g}_{ij}^x(v)}\mathrm{d}v$$
$$= O(h^{d+3}).$$

Besides, by regularity of f, boundedness of p and compactness of $\mathcal{M}$, there is a constant c such that $|f(x)-f(y)| \leq c\|x-y\|_2$. Moreover, by Property ii. of Theorem 2.3.4, the function $v \mapsto \det\widehat{g}_{ij}^x(v)$ is bounded on $B_{\mathbb{R}^d(0,c_1)}$. Hence, it is sufficient to show that uniformly in x,

$$I := \int_{B_{\mathbb{R}^d(0,c_1)}} \left| K\left(\frac{\|v\|_2}{h}\right) - K\left(\frac{\|\mathcal{E}_x(v)-x\|_2}{h}\right)\right| \|\mathcal{E}_x(v)-x\|_2\mathrm{d}v = O(h^{d+3}).$$

Recall that $\|\mathcal{E}_x(v)-x\|_2 \leqslant \|v\|_2$ (by Theorem refTheorem: Existence of a "good" family of normal coordinate systemsA). By Inequation (2.72) in Lemma 2.9.5, we have

$$\begin{aligned} I &\leqslant \int_{B_{\mathbb{R}^d(0,c_1)}} \left(\int_{\left(\frac{\|\mathcal{E}_x(v)-x\|_2}{h}, \frac{\|v\|_2}{h}\right]} \mathrm{d}H(a)\right) \|v\|_2\mathrm{d}v \\ &= \int_{B_{\mathbb{R}^d(0,c_1)}} \left(\int_{\mathbb{R}_+} \mathbb{1}_{\|\mathcal{E}_x(v)-x\|_2 < ah \leq \|v\|_2}\mathrm{d}H(a)\right) \|v\|_2\mathrm{d}v. \end{aligned}$$

Also by Theorem 2.2.1, there exists a constant c_3 such that $\forall x,y\in\mathcal{M}$, $\rho(x,y) \leqslant c_3\|x-y\|_2^3 + \|x-y\|$. The polynomial function $z\mapsto z + c_3z^3$ is an increasing bijective function and we denote by φ its inverse. Thus, for all $x,y\in\mathcal{M}$, $\varphi(\rho(x,y)) \leq \|x-y\|_2$.
Consequently, introducing $\varphi(\rho(x,\mathcal{E}_x(v))) = \varphi(\|v\|_2)$, we deduce

$$\begin{aligned} I &\leq \int_{B_{\mathbb{R}^d(0,c_1)}} \left(\int_{\mathbb{R}_+} \mathbb{1}_{\varphi(\|v\|_2) < ah \leq \|v\|_2}\mathrm{d}H(a)\right) \|v\|_2\mathrm{d}v \\ &= \int_{\mathbb{R}^+} \left(\int_{B_{\mathbb{R}^d(0,c_1)}} \|v\|_2.\mathbb{1}_{ah \leq \|v\|_2 < ah + c_3(ah)^3}\mathrm{d}v\right) \mathrm{d}H(a), \end{aligned}$$

by Fubini's Theorem. Finally, using the spherical coordinate system as in the proof of Lemma 2.4.1, we see that:

$$\begin{aligned} I \leqslant & S_{d-1}\int_{\mathbb{R}^+} \left(\int_0^{c_1} r^d \mathbb{1}_{ah\leq r < ah + c_3(ah)^3}\mathrm{d}r\right)\mathrm{d}H(a) \\ \leq & S_{d-1}\int_{\mathbb{R}^+} \left(\mathbb{1}_{ah\leq c_1} \times \int_{ah}^{ah+c_3(ah)^3} r^d\mathrm{d}r\right)\mathrm{d}H(a) \\ \leq & S_{d-1}\int_{\mathbb{R}^+} \left(\mathbb{1}_{ah\leq c_1} \times c_3(ah)^3\left[ah + c_3(ah)^3\right]^d\right)\mathrm{d}H(a) \\ \leq & S_{d-1}\int_{\mathbb{R}_+} c_3(ah)^{d+3}(1+c_3c_1^2)^d\mathrm{d}H(a) \\ = & S_{d-1}c_3(1+c_3c_1^2)^d h^{d+3}\int_{\mathbb{R}^+} a^{d+3}\mathrm{d}H(a). \end{aligned}$$

This ends the proof of Proposition 2.2.3. □

2.5 Approximations by random operators

In this section, we study the statistical error and prove Proposition 2.2.4.

Notation 2.5.1. *For a $\mathcal{C}^3$-function $f:\mathcal{M}\to\mathbb{R}^k$, we denote respectively by $\|f'\|_\infty, \|f''\|_\infty, \|f'''\|_\infty$ the standard norm of multi-linear maps, i.e.*

$$\|f''\|_\infty = \sup_{\substack{x\in\mathcal{M},(v,w)\in(\mathrm{Rm})^2\\ \|v\|_2\le 1,\|w\|_2\le 1}} \big|f''(x)(v,w)\big|.$$

Recall that for $\alpha\in[\![1,m]\!]$ and $x\in\mathbb{R}^m$, we denote by x^α the α-th coordinate of x.

Let us consider the following collection $\mathcal{F}$ of $\mathcal{C}^3$-functions

$$\mathcal{F} := \{f\in\mathcal{C}^3(\mathcal{M}) : \|f\|_\infty\le 1, \|f'\|_\infty\le 1, \|f''\|_\infty\le 1, \|f'''\|_\infty\le 1\}. \tag{2.47}$$

Let X be a random variable with distribution $p(x)\mu(\mathrm{d}x)$ on $\mathcal{M}$. We introduce the following sequence of random variables $(Z_n, n\in\mathbb{N})$:

$$\begin{aligned} Z_n &:= \sup_{f\in\mathcal{F}}\sup_{x\in\mathcal{M}}\Big|\mathcal{A}_{h_n,n}(f)(x) - \mathbb{E}[\mathcal{A}_{h_n,n}(f)(x)]\Big| \\ &= \frac{1}{nh_n^{d+2}}\sup_{f\in\mathcal{F}}\sup_{x\in\mathcal{M}}\Bigg|\sum_{i=1}^n\Bigg(K\left(\frac{\|X_i-x\|_2}{h_n}\right)(f(X_i)-f(x)) \\ &\qquad -\mathbb{E}\Big[K\Big(\frac{\|X-x\|_2}{h_n}\Big)(f(X)-f(x))\Big]\Bigg)\Bigg|. \end{aligned}$$

Recall that for all function f and point x, $\mathbb{E}[\mathcal{A}_{h_n,n}(f)(x)] = \mathcal{A}_{h_n}(f)(x)$. We want to prove that with probability 1,

$$Z_n = O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}} + h_n\right). \tag{2.48}$$

The general idea to prove this estimation is that instead of proving directly this convergence speed for (Z_n), we show that its expectation also has this speed of convergence, that is:

$$\limsup_{n\to\infty}\left[\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}} + h_n\right)^{-1}\mathbb{E}(Z_n)\right] < \infty, \tag{2.49}$$

then (2.48) will follow easily from Talagrand's inequality (see Corollary 2.9.1 in Appendix) and Borel-Cantelli's theorem, as explained in Section 2.5.4. The detailed plan for the proof of (2.48) is as follows:

Step I: Use Taylor's expansion to divide Z_n into many simpler terms each.

Step II: Use Vapnik-Chernonenkis theory and Theorem 2.5.3 to bound the expectation of each terms.

Step III: Use Talagrand's inequality to conclude.

After using Talagrand's inequality, we have a non-asymptotic estimation of

$$\mathbb{P}\left(\sup_{f\in\mathcal{F}}\sup_{x\in\mathcal{M}}\Big|\mathcal{A}_{h_n,n}(f)(x) - \mathbb{E}[\mathcal{A}_{h_n,n}(f)(x)]\Big| \ge \delta\right)$$

for some suitable constant δ and will be able to prove the Corollary 2.1.2 at the end of this section. This term is of interest of many papers [83, 65, 25].

2.5.1 About the Vapnik-Chernonenkis theory

Before starting the proof, we first recall here the main definitions and an important result of the Vapnik-Chernonenkis theory for the Borelian space $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ we will need. Other useful results are given in Appendix 2.9.1. For more details on the Vapnik-Chernonenkis theory, we refer the reader to [35, 56, 92]. In this section, we will recall upper-bounds that exist for

$$\sup_{f \in \mathcal{F}} \mathbb{E}\left[\left|\sum_{i=1}^{n} (f(X_i) - \mathbb{E}[f(X_i)])\right|\right]$$

when the functions f range over certain VC classes of functions that are defined below.

Let (T, d) be a pseudo-metric space. Let $\varepsilon > 0$ and $N \in \mathbb{N} \cup \{+\infty\}$. A set of points $\{x_1, \ldots, x_N\}$ in T is an ε-cover of T if for any $x \in T$, there exists $i \in [1, N]$ such that $d(x, x_i) \leq \varepsilon$. Then, the *$\varepsilon$-covering number* of T is defined as:

$$N(\varepsilon, T, d) := \inf\{N \in \mathbb{N} \cup \{+\infty\} : \text{there are } N \text{ points in } T \text{ such that they form an } \varepsilon\text{-cover of } (T, d)\}.$$

For a collection of real-valued measurable functions $\mathcal{F}$ on $\mathbb{R}^m$, a real measurable function F defined on $\mathbb{R}^m$ is called *envelope* of $\mathcal{F}$ if for any $x \in \mathbb{R}$,

$$\sup_{f \in \mathcal{F}} |f(x)| \leq F(x).$$

This allows us to define VC classes of functions (see Definition 3.6.10 in [56]). Recall that for a probability measure Q on the measurable space $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$, the $L^2(Q)$-distance given by

$$(f, g) \mapsto \left(\int |f(x) - g(x)|^2 Q(\mathrm{d}x)\right)^{1/2}$$

defines a pseudo-metric on the collection of all bounded real measurable functions on $\mathbb{R}^m$.

Definition 2.5.2 (VC class of functions,)**.** *A class of measurable functions $\mathcal{F}$ is of VC type with respect to a measurable envelope F of $\mathcal{F}$ if there exist finite constants A, v such that for all probability measures Q and $\varepsilon \in (0, 1)$*

$$N(\varepsilon \|F\|_{L^2}, \mathcal{F}, L^2(Q)) \leq (A/\varepsilon)^v.$$

We will denote:

$$N(\varepsilon, \mathcal{F}) := \sup_{Q} N(\varepsilon, \mathcal{F}, L^2(Q)).$$

We now present a version of the useful inequality (2.5) of Giné and Guillou in [54] that gives a bound for the expected concentration rate. For a class of function $\mathcal{F}$, let us define for any real valued function $\varphi : \mathcal{F} \to \mathbb{R}$,

$$\|\varphi(f)\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |\varphi(f)|.$$

Theorem 2.5.3. *(see [54, Proposition 2.1 and Inequality (2.5)]) Consider n i.i.d random variables $X_1, \ldots, X_n$ with values in $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$.*
Let $\mathcal{F}$ be a measurable uniformly bounded VC-type class of functions on $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$. We introduce two positive real number σ^2 and U, such that

$$\sigma^2 \geq \sup_{f \in \mathcal{F}} Var\big(f(X_1)\big), \quad U \geq \sup_{f \in \mathcal{F}} \|f\|_\infty \quad \text{and} \quad 0 < \sigma \leq 2U.$$

Then there exists a constant R depending only on the VC-parameters A, v of $\mathcal{F}$ and on U, such that:

$$\mathbb{E}\left[\left\|\sum_{i=1}^{n} (f(X_i) - \mathbb{E}[f(X_i)])\right\|_{\mathcal{F}}\right] \leq R\Big(\sqrt{n}\sigma\sqrt{|\log \sigma|} + |\log \sigma|\Big).$$

Notice that there exists also a formulation of the previous result in term of deviation probability (see e.g. [86, Theorem 3]), that would lead to results similar to the ones established in [25].

2.5.2 Step I: decomposition of Z_n

We first upper bound the quantity Z_n with a sum of simpler terms.

Lemma 2.5.4. *For any function $f \in \mathcal{F}$, there is a constant $c > 0$ such that for all $n \geqslant 1$,*

$$nh_n^{d+2} Z_n \leq \sum_{\alpha=1}^{m} Y_n^{\alpha} + \sum_{\alpha,\beta=1}^{m} Y_n^{\alpha,\beta} + Y_n^{(3)} + 2nch_n^{d+3}, \tag{2.50}$$

where

$$Y_n^{\alpha} := \sup_{x\in\mathcal{M}} \left| \sum_{i=1}^{n} \left[K\left(\frac{\|X_i - x\|_2}{h_n}\right)(X_i^{\alpha} - x_i^{\alpha}) - \mathbb{E}\left(K\left(\frac{\|X - x\|_2}{h_n}\right)(X^{\alpha} - x^{\alpha})\right)\right]\right|$$

$$Y_n^{\alpha,\beta} := \sup_{x\in\mathcal{M}} \left| \sum_{i=1}^{n} \left[K\left(\frac{\|X_i - x\|_2}{h_n}\right)(X_i^{\alpha} - x^{\alpha})(X_i^{\beta} - x^{\beta}) - \mathbb{E}\left(K\left(\frac{\|X - x\|_2}{h_n}\right)(X^{\alpha} - x^{\alpha})(X^{\beta} - x^{\beta})\right)\right]\right|$$

$$Y_n^{(3)} := \sup_{x\in\mathcal{M}} \left| \sum_{i=1}^{n} K\left(\frac{\|X_i - x\|_2}{h_n}\right)\|X_i - x\|_2^3 - \mathbb{E}\left[K\left(\frac{\|X - x\|_2}{h_n}\right)\|X - x\|_2^3\right]\right|.$$

Proof. Since for any $f \in \mathcal{F}$, the differentials up to third order have operator norms bounded by 1, then, by the Taylor's expansion theorem, for any $(x, y) \in (\mathbb{R}^m)^2$, we have

$$f(y) - f(x) = f'(x)(y - x) + \frac{1}{2} f''(x)(y - x, y - x) + \tau_f(y; x)$$

where τ_f is some function satisfying

$$\sup_{f\in\mathcal{F}} |\tau_f(y, x)| \leq \|f'''\|_\infty \|x - y\|_2^3 = \|x - y\|_2^3. \tag{2.51}$$

Thus, using the notation of the lemma, we deduce

$$nh_n^{d+2} Z_n \leq \sum_{\alpha=1}^{m} Y_n^{\alpha} + \sum_{\alpha=1,\beta=1}^{m} Y_n^{\alpha,\beta} + Y_n^{r},$$

with

$$Y_n^{r} := \sup_{\substack{f\in\mathcal{F}\\ x\in\mathcal{M}}} \left| \sum_{i=1}^{n} \left(K\left(\frac{\|X_i - x\|_2}{h_n}\right)\tau_f(X_i, x) - \mathbb{E}\left[K\left(\frac{\|X - x\|_2}{h_n}\right)\tau_f(X, x)\right]\right)\right|.$$

Using (2.51), we now control Y_n^r by $Y_n^{(3)}$, as follows

$$\begin{aligned} Y_n^r &\leq \sup_{x\in\mathcal{M}} \left| \sum_{i=1}^{n} K\left(\frac{\|X_i - x\|_2}{h_n}\right)\|X_i - x\|_2^3 \right| + n \sup_{x\in\mathcal{M}} \mathbb{E}\left[K\left(\frac{\|X - x\|_2}{h_n}\right)\|X - x\|_2^3\right] \\ &\leq Y_n^{(3)} + 2n \sup_{x\in\mathcal{M}} \mathbb{E}\left[K\left(\frac{\|X - x\|_2}{h_n}\right)\|X - x\|_2^3\right]. \end{aligned}$$

Since the function p is bounded on the compact $\mathcal{M}$, using Inequation (2.34) of Lemma 2.4.1, we deduce that $Y_n^r \leqslant Y_n^{(3)} + 2nch_n^{d+3}$, which conclude the proof. ☐

2.5.3 Step II: Application of the Vapnik-Chervonenkis theory

2.5.3.1 Control the first order terms $\mathbb{E}[Y_n^\alpha]$

Let $\alpha \in [\![1, m]\!]$ be fixed. Given the kernel K, to bound the first order term Y_n^α, we introduce three families of real functions on $\mathcal{M}$:

$$\mathcal{G} := \{\varphi_{h,y,z} : y, z \in \mathcal{M}, h > 0\}, \quad \mathcal{G}_1 := \{\psi_{h,y} : y \in \mathcal{M}, h > 0\}$$
$$\text{and } \mathcal{G}_2 := \{\zeta_y(x) : y \in \mathcal{M}\},$$

with

$$\begin{array}{llll} \varphi_{h,y,z} : & x & \longmapsto & K\left(\frac{\|x-y\|_2}{h}\right)(x^\alpha - z^\alpha) \\ \psi_{h,y} : & x & \longmapsto & K\left(\frac{\|x-y\|_2}{h}\right) \\ \zeta_y : & x & \longmapsto & x^\alpha - y^\alpha. \end{array}$$

Since K is of bounded variation, by [92, Lemma 22], $\mathcal{G}_1$ is VC-type w.r.t a constant envelope. Since $\mathcal{M}$ is a compact manifold, by Lemma 2.9.4, $\mathcal{G}_2$ is VC-type wrt to a constant envelope. Thus, using Lemma 2.9.3, we deduce that $\mathcal{G}$ is a VC-type class of functions because $\mathcal{G} = \mathcal{G}_1 \cdot \mathcal{G}_2$. So, by Definition 2.5.2, there exist real values $A \geq 6, v \geq 1$ depending only on the VC-characteristics of $\mathcal{G}_1$ and $\mathcal{G}_2$ such that, for all $\varepsilon \in (0,1)$,

$$N(\varepsilon, \mathcal{G}) \leq \left(\frac{A}{2\varepsilon}\right)^v.$$

Now, let us consider the following sequence of families of real functions on $\mathcal{M}$:

$$\mathcal{H}_n = \{\varphi_{n,y} : y \in \mathcal{M}\}, \text{ with } \varphi_{n,y} : x \longmapsto K\left(\frac{\|x-y\|_2}{h_n}\right)(x^\alpha - y^\alpha).$$

Proposition 2.5.5. *Let $(X_i)_{i\geqslant 1}$ be a sample of i.i.d. random variables with distribution $p(x)\mu(\mathrm{d}x)$ on the compact manifold $\mathcal{M}$ and X a random variable with the same distribution. We assume that p is bounded on $\mathcal{M}$.*
Then, if the kernel K satisfies Assumption 3 and the sequence $(h_n)_{n\geqslant 0}$ satisfies Assumption (2.5)*, we have*

$$\frac{1}{nh_n^{d+2}}\mathbb{E}\left[\left\|\sum_{i=1}^n (f(X_i) - \mathbb{E}[f(X)])\right\|_{\mathcal{H}_n}\right] = O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}}\right). \tag{2.52}$$

Proof. Since $\mathcal{H}_n \subset \mathcal{G}$, by Lemma 2.9.2, for all n, we have $N(\varepsilon, \mathcal{H}_n) \leq \left(\frac{A}{\varepsilon}\right)^v$. Hence, by theorem 2.5.3, there exists a constant R depending only on A, v and U such that:

$$\mathbb{E}\left[\left\|\sum_{i=1}^n f(X_i) - \mathbb{E}(f(X))\right\|_{\mathcal{H}_n}\right] \leq R\left(\sqrt{n}\sigma\sqrt{|\log\sigma|} + |\log\sigma|\right).$$

where U is a constant such that $U \geq \sup_{f\in\mathcal{H}_n}\|f\|_\infty$, and σ is a constant such that $4U^2 \geq \sigma^2 \geq \sup_{f\in\mathcal{H}_n}\mathbb{E}[f^2(X)]$.
Since $\mathcal{H}_n \subset \mathcal{G}$, we can choose U to be the constant envelope of $\mathcal{G}$ (thus, independent of n). Besides, we see that:

$$\sup_{f\in\mathcal{H}_n}\mathbb{E}\left[f^2(X)\right] \leq \|K\|_\infty \sup_{x\in\mathcal{M}}\int_{\mathcal{M}} K\left(\frac{\|x-y\|}{h_n}\right)(x^\alpha - y^\alpha)^2 p(y)\mu(dy).$$

By Inequation 2.33 of Lemma 2.4.1, we deduce that, there is $c > 0$ such that

$$\sup_{f\in\mathcal{H}_n}\mathbb{E}\left[f^2(X)\right] \leq \|K\|_\infty\|p\|_\infty\mu(\mathcal{M})ch_n^{d+2}, \tag{2.53}$$

which goes to 0 when $n \to +\infty$. Choose $\sigma^2 := \sigma_n^2 = \|K\|_\infty \|p\|_\infty \mu(\mathcal{M}) c h_n^{d+2}$. For n large enough, $\sigma_n \leq 2U$. Hence, using Assumption (2.5) on the sequence $(h_n)_{n\geqslant 1}$, we deduce

$$\frac{1}{nh_n^{d+2}}\mathbb{E}\left[\left\|\sum_{i=1}^n f(X_i) - \mathbb{E}[f(X)]\right\|_{\mathcal{H}_n}\right] = O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}} + \frac{\log h_n^{-1}}{nh_n^{d+2}}\right)$$
$$= O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}}\right).$$

This concludes the proof. □

The conclusion of the above proposition means that:

$$\frac{1}{nh_n^{d+2}}\mathbb{E}[Y_n^\alpha] = O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}}\right).$$

2.5.3.2 Control the second order terms $\mathbb{E}\left[Y_n^{\alpha,\beta}\right]$

The way to bound the second order term $Y_n^{\alpha,\beta}$, for $\alpha, \beta \in [\![1, m]\!]$, is similar to the previous step, but instead of considering $\mathcal{H}_n$, we consider the following VC-type family of functions:

$$\mathcal{I}_n := \left\{\xi_{n,y,z} : x \mapsto K\left(\frac{\|x-y\|}{h_n}\right)(x^\alpha - y^\alpha)(x^\beta - q^\beta) : y \in \mathcal{M}, z \in \mathcal{M}, \right\}.$$

We notice that, for any r.v. X,

$$\mathbb{E}\left[\sup_{g\in\mathcal{I}_n} \left|g^2(X)\right|\right] \leqslant \operatorname{diam}(\mathcal{M})^2 \mathbb{E}\left[\sup_{f\in\mathcal{H}_n} \left|f^2(X)\right|\right].$$

Using (2.53), we deduce $\sup_{g\in\mathcal{I}_n} \mathbb{E}[g^2(X)] = O(h_n^{d+2})$, and

$$\frac{1}{nh_n^{d+2}}\mathbb{E}\sup_{g\in\mathcal{I}_n}\left|\sum_{i=1}^n (g(X_i) - \mathbb{E}[g(X_i)])\right| = O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}}\right).$$

Therefore, we conclude that:

$$\frac{1}{nh_n^{d+2}}\mathbb{E}[Y_n^{\alpha,\beta}] = O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}}\right).$$

2.5.3.3 Control the third order terms $\mathbb{E}\left[Y_n^{(3)}\right]$

This step is essentially the same as the two previous steps, except that the considered family of functions is a little bit different, which is:

$$\mathcal{K}_n := \left\{x \mapsto K\left(\frac{\|x-y\|}{h_n}\right)\|x-y\|^3 : y \in \mathcal{M}\right\}.$$

With the same arguments as before, we obtain:

$$\frac{1}{nh_n^{d+2}}\mathbb{E}\left[Y_n^{(3)}\right] = O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}}\right).$$

Now, thanks to Step I, Step II and Lemma 2.5.4, we have shown that:

$$\mathbb{E}[Z_n] = O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}} + h_n\right). \tag{2.54}$$

2.5.4 Step III: Conclusion

Recall that the set of function $\mathcal{F}$ is defined by (2.47). Since p is bounded on $\mathcal{M}$, by (2.35) of Lemma 2.4.1, there exists $c > 0$ such that $\forall f \in \mathcal{F}, \forall x \in \mathcal{M}$

$$\mathbb{E}\left[K\left(\frac{\|X-x\|}{h_n}\right)^2 (f(X)-f(x))^2\right]$$
$$\leq \|K\|_\infty \mathbb{E}\left[K\left(\frac{\|X-x\|}{h_n}\right)\|X-x\|^2\right] \leq \|K\|_\infty c h_n^{d+2}.$$

In other words,

$$\sup_{f\in\mathcal{F}}\sup_{x\in\mathcal{M}} \mathbb{E}\left[K\left(\frac{\|X-x\|}{h_n}\right)^2 (f(X)-f(x))^2\right] \leq \|K\|_\infty c h_n^{d+2}.$$

Thus by choosing $\sigma := \sigma_n = \sqrt{\|K\|_\infty c h_n^{d+2}}$, and using Massart version of Talagrand inequality (c.f. Corollary 2.9.1) with the functions of the form $y \mapsto K\left(\frac{\|y-x\|_2}{h_n}\right)(f(y)-f(x))$, for all n sufficiently large and any positive number $t_n > 0$, with probability at least $1-e^{-t_n}$,

$$\sup_{f\in\mathcal{F}}\sup_{y\in\mathcal{M}} nh_n^{d+2}|\mathcal{A}_{h_n,n}(f)(x) - \mathbb{E}[\mathcal{A}_{h_n,n}(f)(x)]| \leq 9\left(nh_n^{d+2}\mathbb{E}[Z_n] + \sigma_n\sqrt{nt_n} + bt_n\right), \tag{2.55}$$

where, in this case, the constant envelope b is equal to

$$b := \|K\|_\infty \text{diam}\mathcal{M}.$$

Choose $t_n = 2\log n$, by Borel-Cantelli's lemma, with probability 1

$$\sup_{f\in\mathcal{F}}\sup_{x\in\mathcal{M}} \left|\mathcal{A}_{h_n,n}(f)(x) - \mathbb{E}[\mathcal{A}_{h_n,n}(f)(x)]\right| = O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}} + h_n + \sqrt{\frac{\log n}{nh_n^{d+2}}}\right).$$

Besides, under Assumption (2.5) on the sequence $(h_n)_{n\geqslant 1}$, $\lim_{n\to+\infty} nh_n^{d+2} = +\infty$, hence $\log h_n^{-1} = O(\log n)$. Thus with probability 1,

$$\sup_{f\in\mathcal{F}}\sup_{x\in\mathcal{M}} \left|\mathcal{A}_{h_n,n}(f)(x) - \mathbb{E}[\mathcal{A}_{h_n,n}(f)(x)]\right| = O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}} + h_n\right).$$

This ends the proof of Proposition 2.2.4. Hence, Theorem 2.1.1 is proved.

2.5.5 Proof of Corollary 2.1.2

Using the results of the above sections, we can now prove Corollary 2.1.2. First, we see that by the proofs of Propositions 2.2.2, 2.2.3 and (2.54), there is a constant $C > 0$ such that for all $h > 0, n \in \mathbb{N}$:

$$\sup_{f\in\mathcal{F}}\sup_{x\in\mathcal{M}} |\mathbb{E}[\mathcal{A}_{h,n}(f)(x)] - \mathcal{A}(f)(x)| \leq Ch, \tag{2.56}$$

and

$$\mathbb{E}[Z_n] \leq C\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}} + h_n\right). \tag{2.57}$$

Then by choosing $t_n := \delta^2 n h_n^{d+2}$ in (2.55) with $\delta \in [h_n \vee \sqrt{\frac{\log h_n^{-1}}{n h_n^{d+2}}}, 1]$, we know that with probability at least $1 - e^{-\delta^2 n h_n^{d+2}}$,

$$\sup_{f\in\mathcal{F}}\sup_{y\in\mathcal{M}} |\mathcal{A}_{h_n,n}(f)(x) - \mathbb{E}[\mathcal{A}_{h_n,n}(f)(x)]| \leq \frac{9\left(n h_n^{d+2}\mathbb{E}[Z_n] + \sigma_n\sqrt{n t_n} + b t_n\right)}{n h_n^{d+2}}.$$

Besides, by (2.57), we have:

$$\begin{aligned}\frac{n h_n^{d+2}\mathbb{E}[Z_n] + \sigma_n\sqrt{n t_n} + b t_n}{n h_n^{d+2}} \leq & C\left(\sqrt{\frac{\log h_n^{-1}}{n h_n^{d+2}}} + h_n\right) + \frac{\sigma_n\sqrt{n t_n} + b t_n}{n h_n^{d+2}} \\ = & C\left(\sqrt{\frac{\log h_n^{-1}}{n h_n^{d+2}}} + h_n\right) + \sqrt{\|K\|_\infty c}\,\delta + \|K\|_\infty(\dim\mathcal{M})\delta^2 \\ \leq & \left(2C + \sqrt{\|K\|_\infty c} + \|K\|_\infty(\dim\mathcal{M})\right)\delta.\end{aligned}$$

In addition, after (2.56), we have:

$$\begin{aligned}\sup_{f\in\mathcal{F}}\sup_{x\in\mathcal{M}} |\mathcal{A}_{h_n,n}(f)(x) - \mathcal{A}(f)(x)| \leq & \sup_{f\in\mathcal{F}}\sup_{x\in\mathcal{M}} |\mathcal{A}_{h_n,n}(f)(x) - \mathbb{E}[\mathcal{A}_{h_n,n}(f)(x)]| \\ & + \sup_{f\in\mathcal{F}}\sup_{x\in\mathcal{M}} |\mathbb{E}[\mathcal{A}_{h_n,n}(f)(x)] - \mathcal{A}(f)(x)| \\ \leq & \sup_{f\in\mathcal{F}}\sup_{x\in\mathcal{M}} |\mathcal{A}_{h_n,n}(f)(x) - \mathbb{E}[\mathcal{A}_{h_n,n}(f)(x)]| + C h_n \\ \leq & \sup_{f\in\mathcal{F}}\sup_{x\in\mathcal{M}} |\mathcal{A}_{h_n,n}(f)(x) - \mathbb{E}[\mathcal{A}_{h_n,n}(f)(x)]| + C\delta.\end{aligned}$$

Therefore, by letting

$$C' := 9[2C + \sqrt{\|K\|_\infty c} + \|K\|_\infty(\dim\mathcal{M})] + C, \tag{2.58}$$

where C is the constant appearing in (2.56) and (2.57), we have:

$$\begin{aligned}&\mathbf{P}\left(\sup_{f\in\mathcal{F}}\sup_{x\in\mathcal{M}} |\mathcal{A}_{h_n,n}(f)(x) - \mathcal{A}(f)(x)| > C'\delta\right) \\ &\leq \mathbf{P}\left(\sup_{f\in\mathcal{F}}\sup_{x\in\mathcal{M}} |\mathcal{A}_{h_n,n}(f)(x) - \mathbb{E}[\mathcal{A}_{h_n,n}(f)(x)]| > C'\delta - C\delta\right) \\ &= \mathbf{P}\left(\sup_{f\in\mathcal{F}}\sup_{x\in\mathcal{M}} |\mathcal{A}_{h_n,n}(f)(x) - \mathbb{E}[\mathcal{A}_{h_n,n}(f)(x)]| > 9\left(2C + \sqrt{\|K\|_\infty c} + \|K\|_\infty(\dim\mathcal{M})\right)\delta\right) \\ &\leq \exp\left(-\delta^2 n h_n^{d+2}\right).\end{aligned}$$

This proves Corollary 2.1.2.

2.6 Convergence of kNN Laplacians

We now consider the case of random walks exploring the kNN graph on $\mathcal{M}$ built on the vertices $\{X_i\}_{i\geqslant 1}$, as defined in the introduction.
Recall that for $n \in \mathbf{N}$, $k \in \{1, \dots n\}$ and $x \in \mathcal{M}$, the distance between x and its k-nearest neighbor is defined in (2.8) and that the Laplacian of the kNN-graph is given by, for $x \in \mathcal{M}$,

$$\mathcal{A}_n^{\mathrm{kNN}}(f)(x) := \frac{1}{n R_{n,k_n}^{d+2}(x)} \sum_{i=1}^{n} \mathbf{1}_{[0,1]}\left(\frac{\|X_i - x\|_2}{R_{n,k_n}(x)}\right)(f(X_i) - f(x)). \tag{2.59}$$

Notice here that the width of the moving window, $R_{n,k_n}(x)$, is random and depends on $x \in \mathcal{M}$, contrary to h_n in the previous generator $\mathcal{A}_{h_n,n}$ defined by (2.1).

To overcome this difficulty, we use the result of Cheng and Wu [28, Th. 2.3], with $h = \mathbf{1}_{[0,1]}$, that allows us to control the randomness and locality of the window:

Theorem 2.6.1 (Cheng-Wu, Th. 2.3). *Under Assumption 3, if the density p satisfies* (2.10) *and if*

$$\lim_{n\to+\infty} \frac{k_n}{n} = 0, \qquad \textit{and} \lim_{n\to+\infty} \frac{k_n}{\log(n)} = +\infty,$$

then, with probability higher than $1 - n^{-10}$,

$$\sup_{x\in\mathcal{M}} \left| \frac{R_{n,k_n}(x)}{V_d^{1/d} p^{-1/d}(x) \left(\frac{k_n}{n}\right)^{1/d}} - 1 \right| = O\left(\left(\frac{k_n}{n}\right)^{2/d} + \frac{3\sqrt{13}}{d}\sqrt{\frac{\log n}{k_n}} \right), \tag{2.60}$$

where V_d is the volume of unit $d-$ball.

As a corollary for Theorem 2.6.1, we deduce that the distance $R_{n,k_n}(x)$ is, uniformly in x and with large probability, of the order of h_n:

$$\mathbf{P}(\forall x \in \mathcal{M},\ R_{n,k_n}(x) \in [h_n(x) - \gamma_n, h_n(x) + \gamma_n]) \geqslant 1 - n^{-10}, \tag{2.61}$$

with

$$h_n(x) = V_d^{1/d} p^{-1/d}(x) \left(\frac{k_n}{n}\right)^{1/d}, \text{ and } \gamma_n = 2\left(\left(\frac{k_n}{n}\right)^{2/d} + \frac{3\sqrt{13}}{d}\sqrt{\frac{\log n}{k_n}} \right). \tag{2.62}$$

We will then derive the limit Theorem 2.1.3 for the rescaling of the kNN Laplacian using next result, proved right after.

Theorem 2.6.2. *Suppose that the density of points p on the compact smooth manifold $\mathcal{M}$ is of class $\mathcal{C}^2$. Suppose that Assumptions 3 for the kernel K are satisfied and that $(h_n, n \in \mathbb{N})$ satisfies* (2.5)*, i.e.*

$$\lim_{n\to+\infty} h_n = 0, \qquad \textit{and} \qquad \lim_{n\to+\infty} \frac{\log h_n^{-1}}{n h_n^{d+2}} = 0.$$

Then, for all real number $\kappa > 1$, with probability 1*, for all $f \in \mathcal{C}^3(\mathcal{M})$,*

$$\sup_{\kappa^{-1}h_n \le r \le \kappa h_n} \sup_{x\in\mathcal{M}} |\mathcal{A}_{r,n}(f)(x) - \mathcal{A}(f)(x)| = O\left(\sqrt{\frac{\log h_n^{-1}}{n h_n^{d+2}}} + h_n \right), \tag{2.63}$$

where $\mathcal{A}_{r,n}$ and $\mathcal{A}$ are respectively defined by (2.1) *(replacing h_n with r) and* (2.2)*.*

Proof of Theorem 2.1.3. Assume that Theorem 2.6.2 is proved. We know that the event $\{\forall x \in \mathcal{M},\ R_{n,k_n}(x) \in [h_n(x) - \gamma_n, h_n(x) + \gamma_n]\}$ is of probability $1 - n^{-10}$. Therefore, by Borel-Cantelli's theorem, with probability 1, there exists $N := N(\omega) \in \mathbb{N}$ such that:

$$\forall n \ge N : \forall x \in \mathcal{M},\ R_{n,k_n}(x) \in [h_n(x) - \gamma_n, h_n(x) + \gamma_n].$$

Thus with probability 1, for all $n \ge N(\omega)$, we have:

$$\left| \mathcal{A}_n^{\text{kNN}}(f)(x) - \mathcal{A}(f)(x) \right| \leqslant \sup_{r\in[a_n,b_n]} |\mathcal{A}_{r,n}(f)(x) - \mathcal{A}(f)(x)|$$

with:

$$a_n = V_d^{1/d} p_{\max}^{-1/d} \left(\frac{k_n}{n}\right)^{1/d} - \gamma_n \quad \text{and} \quad b_n = V_d^{1/d} p_{\min}^{-1/d} \left(\frac{k_n}{n}\right)^{1/d} + \gamma_n. \tag{2.64}$$

Notice that for n large enough, a_n will be positive. Using Theorem 2.6.2 with $h_n = b_n$ and $\kappa = (p_{\max}/p_{\min})^{1/d} + 1$, we see that $[a_n, b_n] \subset [\kappa^{-1}h_n, \kappa h_n]$. The result follows with the choice of number of neighbors k_n in (2.11) coming from (2.5) with our choice of h_n. The rate of convergence in (2.12) result from (2.6). □

Proof for Theorem 2.6.2. The proof for the above theorem is essentially the same as the proof we presented for Theorem 2.1.1 except some necessary modifications. Decomposing the error term as in (2.17), we have to treat with similar terms. The approximations involving the geometry and corresponding to Propositions 2.2.2 and 2.2.3 can be generalized directly to account for a supremum in the window width $r \in [\kappa^{-1}h_n, \kappa h_n]$. Let us consider the statistical term.

We recall that $\mathcal{F}$ is defined by (2.47). Following the previous computations of Section 2.5, we introduce the following sequence of random variables $(\tilde{Z}_n, n \in \mathbb{N})$, where here $K = \mathbf{1}_{[0,1]}$:

$$\begin{aligned}\tilde{Z}_n &:= \sup_{f\in\mathcal{F}} \sup_{\kappa^{-1}h_n \le r \le \kappa h_n} \sup_{x\in\mathcal{M}} \left| \mathcal{A}_{r,n}(f)(x) - \mathbb{E}[\mathcal{A}_{r,n}(f)(x)] \right| \\ &= \frac{1}{nh_n^{d+2}} \sup_{f\in\mathcal{F}} \sup_{\kappa^{-1}h_n \le r \le \kappa h_n} \sup_{x\in\mathcal{M}} \left| \sum_{i=1}^n \left(K\left(\frac{\|X_i - x\|_2}{r}\right)(f(X_i) - f(x)) \right.\right. \\ &\qquad \left.\left. - \mathbb{E}\left[K\left(\frac{\|X - x\|_2}{r}\right)(f(X) - f(x))\right]\right)\right|.\end{aligned}$$

Similar to what we did in Section 2.5.2, we can show that there is a constant c independent of n such that:

$$nh_n^{d+2}\tilde{Z}_n \le \sum_{\alpha=1}^m \tilde{Y}_n^\alpha + \sum_{\alpha,\beta=1}^m \tilde{Y}_n^{\alpha,\beta} + \tilde{Y}_n^{(3)} + 2nch_n^{d+3}, \tag{2.65}$$

where

$$\begin{aligned}\tilde{Y}_n^\alpha &:= \sup_{\kappa^{-1}h_n \le r \le \kappa h_n} \sup_{x\in\mathcal{M}} \left| \sum_{i=1}^n \left[K\left(\frac{\|X_i - x\|_2}{r}\right)(X_i^\alpha - x_i^\alpha)\mathbb{E}\left(K\left(\frac{\|X - x\|_2}{r}\right)(X^\alpha - x^\alpha)\right)\right]\right| \\ \tilde{Y}_n^{\alpha,\beta} &:= \sup_{\kappa^{-1}h_n \le r \le \kappa h_n} \sup_{x\in\mathcal{M}} \left| \sum_{i=1}^n \left[K\left(\frac{\|X_i - x\|_2}{r}\right)(X_i^\alpha - x^\alpha)(X_i^\beta - x^\beta) - \right.\right. \\ &\qquad \left.\left. \mathbb{E}\left(K\left(\frac{\|X - x\|_2}{r}\right)(X^\alpha - x^\alpha)(X^\beta - x^\beta)\right)\right]\right| \\ \tilde{Y}_n^{(3)} &:= \sup_{\kappa^{-1}h_n \le r \le \kappa h_n} \sup_{x\in\mathcal{M}} \left| \sum_{i=1}^n K\left(\frac{\|X_i - x\|_2}{r}\right)\|X_i - x\|_2^3 - \mathbb{E}\left[K\left(\frac{\|X - x\|_2}{r}\right)\|X - x\|_2^3\right]\right|.\end{aligned}$$

We now treat these terms by applying Vapnik-Chernonenkis theory. Let us start with the control the first order terms $\mathbb{E}[\tilde{Y}_n^\alpha]$:
In Section 2.5.3.1, we have already shown that the family

$$\mathcal{G} := \left\{ \varphi_{h,y,z} : x \longmapsto K\left(\frac{\|x - y\|_2}{h}\right)(x^\alpha - z^\alpha) : y, z \in \mathcal{M}, h > 0 \right\}$$

is a VC class of functions, and that there exist real values $A \ge 6, v \ge 1$ such that, for all $\varepsilon \in (0,1)$, $N(\varepsilon, \mathcal{G}) \le (A/2\varepsilon)^v$.

Now, on top of this, we consider the following sequence of families of real functions on $\mathcal{M}$:

$$\tilde{\mathcal{H}}_n = \left\{\varphi_{r,y} : y \in \mathcal{M}, \kappa^{-1}h_n \leq r \leq \kappa h_n\right\},$$

with $\varphi_{r,y} : x \longmapsto K\left(\frac{\|x-y\|_2}{r}\right)(x^\alpha - y^\alpha)$. Because each $\tilde{\mathcal{H}}_n$ is a subfamily of $\mathcal{G}$, it is still a VC class for which we can use the Talagrand inequality 2.5.3. The latter can deal with the additional supremum with respect to the window width. Similarly to what we did in the proof of Proposition 2.5.5, we obtain that:

$$\frac{1}{nh_n^{d+2}}\mathbb{E}\left[\left\|\sum_{i=1}^n (f(X_i) - \mathbb{E}[f(X)])\right\|_{\tilde{\mathcal{H}}_n}\right] = O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}}\right),$$

which means that as $n \to \infty$,

$$\frac{1}{nh_n^{d+2}}\mathbb{E}[\tilde{Y}_n^\alpha] = O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}}\right).$$

The control the second and third order terms are done as in Sections 2.5.3.2 and 2.5.3.1, using the same trick and the classes of functions

$$\tilde{\mathcal{I}}_n := \left\{x \mapsto K\left(\frac{\|x-y\|}{r}\right)(x^\alpha - y^\alpha)(x^\beta - q^\beta) : y \in \mathcal{M}, q \in \mathcal{M}, \kappa^{-1}h_n \leq r \leq \kappa h_n\right\}$$

and

$$\tilde{\mathcal{K}}_n := \left\{x \mapsto K\left(\frac{\|x-y\|}{r}\right)\|x-y\|^3 : y \in \mathcal{M}, \kappa^{-1}h_n \leq r \leq \kappa h_n\right\}.$$

This provides:

$$\frac{1}{nh_n^{d+2}}\mathbb{E}[\tilde{Y}_n^{\alpha,\beta}] = O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}}\right), \quad \text{and} \quad \frac{1}{nh_n^{d+2}}\mathbb{E}[\tilde{Y}_n^{(3)}] = O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}}\right). \tag{2.66}$$

Therefore, we can deduce the conclusion by using the same argument presented in Section 2.5.4. □

2.7 Tightness and convergence of the random walks

2.7.1 Proof of Theorem 2.1.4

Let K be a kernel that satisfies Assumption 3. Let $n \geqslant 1$ be fixed. Recall that the generator $\mathcal{A}_{h_n,n}$ can be related to a random walk $X^{(n)}$ on the sample points $\{X_1, \dots, X_n\}$, which is solution of the following stochastic differential equation:

$$X_t^{(n)} = X_0^{(n)} + \int_0^t \int_{\mathbf{N}} \int_{\mathbb{R}_+} \mathbf{1}_{i \leqslant n} \mathbf{1}_{\theta \leqslant \frac{1}{nh_n^{d+2}} K\left(\frac{\|X_i - X_{s_-}^{(n)}\|_2}{h_n}\right)} (X_i - X_{s_-}^{(n)})\ Q(\mathrm{d}s, \mathrm{d}i, \mathrm{d}\theta)$$

with initial condition $X_0^{(n)}$ and where $Q(\mathrm{d}s, \mathrm{d}i, \mathrm{d}\theta)$ is a Poisson point measure on $\mathbb{R}_+ \times \mathbf{N} \times \mathbb{R}_+$ independent of $X_0^{(n)}$, and of intensity $\mathrm{d}s \otimes \mathrm{n}(\mathrm{d}i) \otimes \mathrm{d}\theta$, with $\mathrm{d}s$ and $\mathrm{d}\theta$ Lebesgue measures on $\mathbb{R}_+$ and $\mathrm{n}(\mathrm{d}i)$ the counting measure on $\mathbf{N}$.

Remark 2.7.1. *The initial distribution of $X_0^{(n)}$ can have support on the sample points $\{X_1, \dots, X_n\}$, but not necessarily. It can be any distribution on the manifold $\mathcal{M}$. In any case, the random walk reaches the sample points $\{X_1, \dots X_n\}$ – and stays in this set – after the first jump.*

Proposition 2.7.2. *For a fixed $n \geqslant 1$, a random variable $X_0^{(n)}$ and a Poisson point measure $Q(ds, di, d\theta)$, there exists a unique strong solution of the stochastic differential equation* (2.13). *For any real-valued measurable bounded function f on $\mathcal{M}$, we have that*

$$M_t^{n,f} = f(X_t^{(n)}) - f(X_0^{(n)}) - \int_0^t \mathcal{A}_{h_n,n}(f)(X_s^{(n)})\ \mathrm{d}s \tag{2.67}$$

is a square-integrable martingale with predictable quadratic variation:

$$\langle M^{n,f}\rangle_t = \int_0^t \frac{1}{nh_n^{d+2}} \sum_{i=1}^n K\left(\frac{\|X_i - X_s^{(n)}\|_2}{h_n}\right)\big(f(X_i) - f(X_s^{(n)})\big)^2\ \mathrm{d}s. \tag{2.68}$$

Proof. The proof is straightforward as the process takes its values in the compact manifold $\mathcal{M}$. The jump rate remains therefore bounded and the path of the random walk $X^{(n)}$ can be constructed algorithmically for any time $t \in \mathbb{R}_+$. The second part of the proposition comes from stochastic calculus for jump processes (see [68, Th. 5.1, page 66]). □

Let $T > 0$ be a positive fixed time. The proof of Theorem 2.1.4 is now divided into several steps. First, we prove that the sequence of processes $(X^{(n)})_{n\geqslant 0}$ is tight in the path-space $\mathbb{D}[0,T],\mathcal{M})$. By Prohorov's theorem (e.g. [21]), the sequence is therefore sequentially relatively compact. The convergence of the generators $\mathcal{A}_{h_n,n}$ to $\mathcal{A}$ defined in (2.2) will yield that any limiting value is solution of the martingale problem associated with $\mathcal{A}$, which is well-posed.

Lemma 2.7.3. *Under the hypotheses of Theorem 2.1.4, the sequence $(X^n)_{n\geqslant 0}$ is tight in $\mathbb{D}[0,T],\mathcal{M})$.*

Proof. We check conditions $(T1)$ and $(T2)$ in Aldous criteria for tightness (see e.g. [71, 6]). Because $\mathcal{M}$ is compact, only $(T2)$ needs to be considered. Thanks to Markov inequality, it is sufficient to show that for any $\varepsilon > 0$, there exists $\delta > 0$ such that for any couple of stopping times $(S_n, T_n)_{n\geqslant 0}$ satisfying $0 \leqslant S_n \leqslant T_n \leqslant (S_n + \delta) \wedge T$, we have for all n sufficiently large that:

$$\mathbf{P}\Big(\rho(X_{S_n}^{(n)}, X_{T_n}^{(n)}) > \varepsilon\Big) \leqslant \varepsilon. \tag{2.69}$$

By Markov inequality and Theorem 2.2.1, it is sufficient to study $\mathbb{E}\left[\left\|X_{S_n}^{(n)} - X_{T_n}^{(n)}\right\|_2^2\right]$. We observe that,

$$\begin{aligned}
&\mathbb{E}\left[\left\|X_{S_n}^{(n)} - X_{T_n}^{(n)}\right\|_2^2\right] \leqslant \mathbb{E}\left[\int_{S_n}^{T_n} \sum_{i=1}^n \frac{1}{nh_n^{d+2}} K\left(\frac{\|X_i - X_s^{(n)}\|_2}{h_n}\right)\left\|X_i - X_s^{(n)}\right\|_2^2 \mathrm{d}s\right]\\
\leqslant&\delta\, \mathbb{E}\left[\sup_{x\in\mathcal{M}} \sum_{i=1}^n \frac{1}{nh_n^{d+2}} K\left(\frac{\|X_i - x\|_2}{h_n}\right)\|X_i - x\|_2^2\right]\\
\leqslant&\delta \sup_{x\in\mathcal{M}} \frac{1}{h_n^{d+2}} \int K\left(\frac{\|x-y\|_2}{h_n}\right)\|x-y\|_2^2 p(y)\mu(\mathrm{d}y)\\
&+\delta \frac{1}{h_n^{d+2}} \mathbb{E}\left[\sup_{x\in\mathcal{M}}\left|\frac{1}{n}\sum_{i=1}^n K\left(\frac{\|X_i - x\|_2}{h_n}\right)\|X_i - x\|_2^2 - \mathbb{E}\left[K\left(\frac{\|X_1 - x\|_2}{h_n}\right)\|X_1 - x\|_2^2\right]\right|\right].
\end{aligned}$$

Since p is bounded on the compact $\mathcal{M}$ by continuity, using (2.35) of Lemma 2.4.1, and using the estimate of $Y_n^{(\alpha,\beta)}$ in Section 2.5.3.2 based on the Vapnik-Chervonenkis theory, we deduce that there is a constant $C > 0$, which does not depend on n nor ε, such that

$$\mathbb{E}\left[\left\|X_{S_n}^{(n)} - X_{T_n}^{(n)}\right\|_2^2\right] \leqslant C\delta. \tag{2.70}$$

Choosing $\delta = \varepsilon^2/C$ yields (2.69). □

Consider now a limiting value $Y \in \mathbb{D}([0,T],\mathcal{M})$ of the tight sequence $(X^{(n)})_{n\geqslant 0}$. There exists a subsequence converging in distribution to Y. Using Skorokhod representation theorem (see [20, Th. 29.6 p.399]) we can assume that this convergence holds almost surely and with an abuse of notation, we keep the notation $(X^{(n)})_{n\geqslant 0}$ for the subsequence converging to Y.

Lemma 2.7.4. *Under the hypothesis of Theorem 2.1.4, any limiting value Z of the sequence $(X^{(n)})_{n\geqslant 0}$ is a solution to the martingale problem associated to the generator $\mathcal{A}$ and with initial measure ν.*

Proof. By assumption, Z_0 has the distribution ν, so it is sufficient to show that for all $0 \leq t_0 < t_1 < t_2 < \cdots < t_k \leqslant s < t$ and functions $g_1, g_2, g_2, \ldots, g_k, g \in \mathcal{C}(\mathcal{M}), f \in \mathcal{C}^3(\mathcal{M})$, we have:

$$\mathbb{E}\left[\left(\prod_{i=0}^{k} g_i(Y_{t_i})\right)\left(f(Y_t) - f(Y_s) - \int_s^t \mathcal{A}f(Y_u)\mathrm{d}u\right)\right] = 0.$$

Let us denote by Ψ the application:

$$\Psi \ : \ y \in \mathbb{D}([0,T],\mathcal{M}) \mapsto \left(\prod_{i=0}^{k} g_i\big(y_{t_i}\big)\right)\left(f\big(y_t\big) - f\big(y_s\big) - \int_s^t \mathcal{A}f(y_u)\mathrm{d}u\right).$$

We have:

$$\begin{aligned}\Big|\mathbb{E}\big[\Psi\big(X^{(n)}\big)\big]\Big| \leqslant& \left|\mathbb{E}\left[\left(\prod_{i=0}^{k} g_i\big(X^{(n)}_{t_i}\big)\right)\left(f\big(X^{(n)}_t\big) - f\big(X^{(n)}_s\big) - \int_s^t \mathcal{A}_{h_n,n}f\big(X^{(n)}_u\big)\mathrm{d}u\right)\right]\right| \\ &+\left|\mathbb{E}\left[\left(\prod_{i=0}^{k} g_i\big(X^{(n)}_{t_i}\big)\right)\int_s^t\left(\mathcal{A}_{h_n,n}f\big(X^{(n)}_u\big) - \mathcal{A}f\big(X^{(n)}_u\big)\right)\mathrm{d}u\right]\right| \\ =&O\left((t-s)\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}} + h_n\right),\end{aligned}$$

by Proposition 2.7.2 and Theorem 2.1.1. The application Ψ is not continuous on $\mathbb{D}([0,T],\mathcal{M})$ in general, but since the limiting process is continuous a.s., we have that $\Psi(X^{(n)})$ converges to $\Psi(Y)$ a.s. We can then conclude the proof as $\mathbb{E}[\Psi(Y)] = 0$ by using the dominated convergence theorem. □

Proof of Theorem 2.1.4. From Lemmas 2.7.3 and 2.7.4, the limiting processes are all solutions of the same martingale problem associated with $(\mathcal{A}, \nu)$. The well-posedness of the latter martingale problem is a consequence of Theorem 1.2.9 in [67]. Hence the limiting processes all have the same distribution and the sequence $(X^{(n)})_{n\geqslant 0}$ converges in distribution to the limit stated in Theorem 2.1.4. □

2.7.2 Convergence of the kNN random walk

We prove Theorem 2.1.5. For the sake of notation, the random walk $X^{(n),\mathrm{kNN}}$ is now denoted by $X^{(n)}$. Let us recall its SDE:

$$X^{(n)}_t = X^{(n)}_0 + \int_0^t \int_{\mathbf{N}}\int_{\mathbb{R}_+} \mathbf{1}_{i\leqslant n}\mathbf{1}_{\theta \leqslant \frac{1}{nR^{d+2}_{n,k_n}(X^{(n)}_{s_-})}}\mathbf{1}_{[0,1]}\left(\frac{\|X_i - X^{(n)}_{s_-}\|_2}{R_{n,k_n}\left(X^{(n)}_{s^-}\right)}\right)(X_i - X^{(n)}_{s_-})\ Q(\mathrm{d}s,\mathrm{d}i,\mathrm{d}\theta)$$

with

$$R_{n,k}(x) = \inf\Big\{r \geqslant 0,\ \sum_{i=1}^{n}\mathbf{1}_{\|x-X_i\|_2\leqslant r} \geqslant k\Big\}.$$

The related martingale is thus,

$$M_t^{n,f} = f(X_t^{(n)}) - f(X_0^{(n)}) - \int_0^t \mathcal{A}_n^{\mathrm{kNN}}(f)(X_s^{(n)})\,\mathrm{d}s$$

with predictable quadratic variation:

$$\langle M^{n,f}\rangle_t = \int_0^t \sum_{i=1}^n \mathbf{1}_{[0,1]}\left(\frac{\|X_i - X_s^{(n)}\|_2}{R_{n,k_n}\left(X_s^{(n)}\right)}\right)\big(f(X_i) - f(X_s^{(n)})\big)^2\,\mathrm{d}s.$$

As in the proof of Lemma 2.7.3, to obtain the tightness of the distribution, thanks to Aldous criterion, it is sufficient to show that for any $\varepsilon > 0$, there exists $\delta > 0$ such that for any couple of stopping times $(S_n, T_n)_{n\geqslant 0}$ satisfying $0 \leqslant S_n \leqslant T_n \leqslant (S_n + \delta)\wedge T$, we have for all n sufficiently large that:

$$\mathbf{P}\Big(\rho\big(X_{S_n}^{(n)}, X_{T_n}^{(n)}\big) > \varepsilon\Big) \leqslant \varepsilon.$$

We observe that, using the quantities $h_n(x)$ and γ_n defined by (2.62) in Section 2.6,

$$\begin{aligned}\mathbf{P}\Big(\rho\big(X_{S_n}^{(n)}, X_{T_n}^{(n)}\big) > \varepsilon\Big) \leqslant & \mathbf{P}\Big(\rho\big(X_{S_n}^{(n)}, X_{T_n}^{(n)}\big) > \varepsilon,\ \sup_{x\in\mathcal{M}} |R_{n,k_n}(x) - h_n(x)| \leqslant \gamma_n\Big)\\ &+ \mathbf{P}\Big(\sup_{x\in\mathcal{M}} |R_{n,k_n}(x) - h_n(x)| > \gamma_n\Big)\\ \leqslant & \frac{1}{\varepsilon^2}\mathbb{E}\left(\left\|X_{S_n}^{(n)} - X_{T_n}^{(n)}\right\|_2^2 \mathbf{1}_{\sup_{x\in\mathcal{M}}|R_{n,k_n}(x) - h_n(x)|\leqslant\gamma_n}\right) + n^{-10}.\end{aligned}$$

Introducing (a_n, b_n) given by (2.64), we have $[h_n(x) - \gamma_n,\ h_n(x) + \gamma_n] \subset [a_n, b_n]$ for all $x \in \mathcal{M}$. Thus, on the event $\{\sup_{x\in\mathcal{M}} |R_{n,k_n}(x) - h_n(x)| \leqslant \gamma_n\}$, we have:

$$\begin{aligned}&\left\|X_{T_n}^{(n)} - X_{S_n}^{(n)}\right\|_2\\ &\quad\leqslant \int_{S_n}^{T_n}\int_{\mathbf{N}}\int_{\mathbb{R}_+} \mathbf{1}_{i\leqslant n}\mathbf{1}_{\theta\leqslant\frac{1}{nR_{n,k_n}^{d+2}(X_{s_-}^{(n)})}}\mathbf{1}_{[0,1]}\left(\frac{\|X_i - X_{s_-}^{(n)}\|_2}{R_{n,k_n}\left(X_{s^-}^{(n)}\right)}\right)\left\|X_i - X_{s_-}^{(n)}\right\|_2 Q(\mathrm{d}s, \mathrm{d}i, \mathrm{d}\theta)\\ &\quad\leqslant \int_{S_n}^{T_n}\int_{\mathbf{N}}\int_{\mathbb{R}_+} \mathbf{1}_{i\leqslant n}\mathbf{1}_{\theta\leqslant\frac{1}{na_n^{d+2}}}\mathbf{1}_{\|X_i - X_{s_-}^{(n)}\|_2\leqslant b_n}\left\|X_i - X_{s_-}^{(n)}\right\|_2 Q(\mathrm{d}s, \mathrm{d}i, \mathrm{d}\theta)\end{aligned}$$

We deduce that

$$\begin{aligned}&\mathbb{E}\left[\left\|X_{S_n}^{(n)} - X_{T_n}^{(n)}\right\|_2^2 \mathbf{1}_{\sup_{x\in\mathcal{M}}|R_{n,k_n}(x) - h_n(x)|\leqslant\gamma_n}\right]\\ &\quad\leqslant \mathbb{E}\left[\int_{S_n}^{T_n}\sum_{i=1}^n \frac{1}{na_n^{d+2}}\mathbf{1}_{\|X_i - X_{s_-}^{(n)}\|_2\leqslant b_n}\left\|X_i - X_s^{(n)}\right\|_2^2 \mathrm{d}s\right]\\ &\quad\leqslant \frac{\delta}{na_n^{d+2}}\mathbb{E}\left[\sup_{x\in\mathcal{M}}\sum_{i=1}^n \mathbf{1}_{\|X_i - x\|_2\leqslant b_n}\|X_i - x\|_2^2\right]\\ &\quad\leqslant \frac{\delta}{a_n^{d+2}}\int_{\mathcal{M}} \mathbf{1}_{\|x-y\|_2\leqslant b_n}\|x - y\|_2^2\, p(y)\mu(\mathrm{d}y)\\ &\qquad+ \frac{\delta}{a_n^{d+2}}\mathbb{E}\left[\sup_{x\in\mathcal{M}}\left|\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\|X_i - x\|_2\leqslant b_n}\|X_i - x\|_2^2 - \mathbb{E}\Big[\mathbf{1}_{\|X_1 - x\|_2\leqslant b_n}\|X_1 - x\|_2^2\Big]\right|\right]\end{aligned}$$

As in the proof of Lemma 2.7.3, we use for the first term in the right hand side that p is bounded on the compact $\mathcal{M}$ and (2.35) of Lemma 2.4.1. For the second term we use the estimate of $\tilde{Y}_n^{(\alpha,\beta)}$ in Section 2.6 based on the Vapnik-Chervonenkis theory, and the fact that $[a_n, b_n] \subset [\kappa^{-1}h_n, \kappa h_n]$. We deduce that there is a constant $C > 0$, which does not depend on n nor ε, such that

$$\mathbb{E}\left[\left\|X_{S_n}^{(n)} - X_{T_n}^{(n)}\right\|_2^2\right] \leqslant C\delta. \tag{2.71}$$

This shows that the sequence of random walks is tight. The identification of the limiting martingale problem follows the proof of Theorem 2.1.4 using Theorem 2.1.3.

2.8 Acknowledgements

The authors thank Frédéric Rochon for useful discussions.

The research of H.G. and D.T.N. is part of the Mathematics for Public Health program funded by the Natural Sciences and Engineering Research Council of Canada (NSERC). H.G. is also supported by NSERC discovery grant (RGPIN-2020-07239). D.T.N. and V.C.T. are supported by Labex Bézout (ANR-10-LABX-58), GdR GeoSto 3477.

2.9 Appendices

2.9.1 Some concentration inequalities

2.9.1.1 Talagrand's concentration inequality

As a corollary of Talagrand's inequality presented in Massart [86, Theorem 3], where for simplicity we choose $\varepsilon = 8$, we have the following deviation inequality:

Corollary 2.9.1 (Simplified version of Massart's inequality)**.** *Consider n independant random variables $\xi_1, \dots, \xi_n$ with values in some measurable space $(\mathbb{X}, \mathfrak{X})$. Let $\mathcal{F}$ be some countable family of real-valued measurable functions on $(\mathbb{X}, \mathfrak{X})$ such that for some positive real number b, $\|f\|_\infty \leq b$ for every $f \in \mathcal{F}$.*

$$Z := \sup_{f\in\mathcal{F}} \left|\sum_{i=1}^{n} \Big(f(\xi_i) - \mathbb{E}\left[f(\xi_i)\right]\Big)\right|.$$

then with $\sigma^2 = \sup_{f\in\mathcal{F}} Var(f(\xi_1))$, and for any positive real number x,

$$\mathbb{P}\big(Z \geq 9(\mathbb{E}[Z] + \sigma\sqrt{nx} + bx)\big) \leq e^{-x}.$$

2.9.1.2 Covering numbers and complexity of a class of functions

If $S \subset T$ is a subspace of T, it is not true in general that $N(\varepsilon, S, d) \leq N(\varepsilon, T, d)$ because of the constraints that the cencers x_i should belong to S. However, we can bound the covering number of S by T's as follows

Lemma 2.9.2. *If $S \subset T$ is a subspace of the metric space (T, d), then for any positive number ε*

$$N(2\varepsilon, S, d) \leq N(\varepsilon, T, d).$$

Proof. Let $\{x_1, ..., x_N\}$ be a ε-cover of T and for any $i \in [\![1, N]\!]$, let us define $K_i := \{x \in T : d(x, x_i) \leq \varepsilon\}$. Of course, K_i may not intersect S, hence, without loss of generality, assume that for a natural number $0 < m \leq N$ we have that $K_i \cap S \neq \emptyset$ if and only if $i \leq m$. Let y_i be any point in $K_i \cap S$ for $i \in [\![1, m]\!]$. Since $\{x_1, ..., x_N\}$ is a ε cover of T, for any $y \in S$, there exists a $i \leq m$ such that $y \in K_i \cap S$. Hence, $d(y, y_i) \leq 2\varepsilon$. Consequently, $y_1, ..., y_m$ be a 2ε-cover of (S, d). $\square$

Let us consider the Borel space $(\mathbb{R}^m, \mathcal{B}(R^m))$. If $\mathcal{F}, \mathcal{G}$ are two collections of measurable functions on $\mathbb{X}$, we are interested in the "complexity" of $\mathcal{F} \cdot \mathcal{G} = \{fg | f \in \mathcal{F}, g \in \mathcal{G}\}$.

Lemma 2.9.3 (Bound on ε-covering numbers)**.** *Let $\mathcal{F}, \mathcal{G}$ be two bounded collections of measurable functions, i.e, there are two constants c_1, c_2 such that*

$$\|f\|_\infty \leq c_1 \text{ and } \|g\|_\infty \leq c_2 \text{ for all } f \in \mathcal{F}\ ,\ g \in \mathcal{G}.$$

then for any probability measure Q,

$$N(2\varepsilon c_1 c_2, \mathcal{F} \cdot \mathcal{G}, L^2(Q)) \leq N(\varepsilon c_1, \mathcal{F}, L^2(Q)) N(\varepsilon c_2, \mathcal{G}, L^2(Q)).$$

Proof. If $f_1, f_2, ..., f_n$ is a εc_1-cover of $(\mathcal{F}, L^2(Q))$ and $g_1, g_2, ..., g_m$ is a εc_2-cover of $(\mathcal{G}, L^2(Q))$, then for any $(f, g) \in \mathcal{F} \times \mathcal{G}$, we have:

$$|f(x)g(x) - f_i(x)g_j(x)| \leq |f(x) - f_i(x)|c_2 + c_1|g(x) - g_i(x)|.$$

which implies that $\{f_i g_j : 1 \leq i \leq n \text{ and } 1 \leq j \leq m\}$ is a $2\varepsilon c_1 c_2$ -cover of $\mathcal{F} \cdot \mathcal{G}$. □

The following lemma is just a simplied version result of the theory of VC Hull class of functions (Section 3.6.3 in [56]).

Lemma 2.9.4. *If f is a bounded measurable function on the measurable space $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ and $D = [a, b] \subset \mathbb{R}$ is a compact interval, then*

$$\mathcal{F} := \{f + d : d \in D\},$$

is VC type with respect to a constant envelope.

Proof. Let $N = [\frac{b-a}{\varepsilon}]$, $f_i = f + i\varepsilon$ for all $i \in [\![1, N]\!]$. So, by the definition of $\mathcal{F}$, for all $g \in \mathcal{F}$, there is an $i \in [\![1, N]\!]$ such that $|g(x) - f_i(x)| < \varepsilon$ for all $x \in \mathbb{R}^m$. Thus, for all probability measure Q on $\mathbb{R}^m$, we have: $\|g - f_i\|_{L^2(Q)} \leq \varepsilon$, which makes $\mathcal{H} := \{f_i : i \in [\![1, N]\!]\}$ be a ε-cover of $L^2(Q)$. Hence,

$$N(\varepsilon, \mathcal{F}, L^2(Q)) \leq N \leq \frac{(b-a)}{\varepsilon}.$$

So $\mathcal{F}$ is a VC-type class of functions with $A = b-a$, $v = 1$,$F = \max(1, \|f\|_\infty + |a|, \|f\|_\infty + |b|)$. □

2.9.2 Some estimates using the total variation

Lemma 2.9.5. *If $K : [0, +\infty) \to \mathbb{R}$ is a bounded variation function with $H(a)$ its total variation on the interval $[0, a]$, for all $a, b \in [0, \infty]$, with $a \leqslant b$,*

$$|K(b) - K(a)| \leq H(b) - H(a). \tag{2.72}$$

Besides, if K satisfies Assumption 3, then, when b goes to infinity,

$$K(b)b^{d+3} = o(1) \qquad \text{and} \int_b^\infty K(a)a^{d+1}\mathrm{d}a = o(1/b). \tag{2.73}$$

Proof of Lemma 2.9.5. Inequality (2.72) comes directly from the definition of total variation. We note that:

$$b^{d+3}(H(\infty) - H(b)) \leq \int_b^\infty a^{d+3}\mathrm{d}H(a).$$

Then, by Assumption 3,

$$\lim_{b \to +\infty} b^{d+3}(H(\infty) - H(b)) = 0.$$

Then, as:

$$K(b)b^{d+3} \leq b^{d+3}(H(\infty) - H(b)),$$

we have proven the first estimation in (2.73).
For the second estimation, we see that:

$$\begin{aligned}(d+2)\int_b^{\infty} bK(a)a^{d+1}\mathrm{d}a &\leq (d+2)\int_b^{\infty} b(H(\infty)-H(a))a^{d+1}\mathrm{d}a \\ &= -\,b^{d+3}(H(\infty)-H(b)) + b\int_b^{\infty} a^{d+2}\mathrm{d}H(a) \\ &\leq -\,b^{d+3}(H(\infty)-H(b)) + \int_b^{\infty} a^{d+3}\mathrm{d}H(a).\end{aligned}$$

Therefore, we have the conclusion. □

2.9.3 Proof of Lemma 2.4.2

Thanks to the symmetry of the Euclidean norm $\|\cdot\|_2$, we observe that for any $i, j \in [\![1, d]\!]$,

$$\int_{B_{\mathbb{R}^d}(0,c)} G(\|v\|_2)v^i v^j \mathrm{d}v = \begin{cases} 0 & \text{if } i \neq j, \\ \frac{1}{d}\int_{B_{\mathbb{R}^d}(0,c)} G(\|v\|_2)\|v\|_2^2\mathrm{d}v & \text{if } i = j. \end{cases}$$

Thus, LHS of (2.41) is equal to:

$$\begin{aligned} &= \left[\frac{1}{d}\int_{B_{\mathbb{R}^d}(0,c)} G(\|v\|_2)\|v\|_2^2\mathrm{d}v\right]\left[\sum_{i=1}^{d}\left\langle \nabla_{\mathbb{R}^m} f(x), \frac{\partial k}{\partial x^i}(0)\right\rangle\left\langle \nabla_{\mathbb{R}^m} h(x), \frac{\partial k}{\partial x^i}(0)\right\rangle\right] \\ &= \left[\frac{1}{d}\int_{B_{\mathbb{R}^d}(0,c)} G(\|v\|_2)\|v\|_2^2\mathrm{d}v\right]\left[\sum_{i=1}^{d}\frac{\partial(f\circ k)}{\partial x^i}(0)\,\frac{\partial(h\circ k)}{\partial x^i}(0)\right] \\ &= \left[\frac{1}{d}\int_{B_{\mathbb{R}^d}(0,c)} G(\|v\|_2)\|v\|_2^2\mathrm{d}v\right]\big\langle \nabla_{\mathbb{R}^d}(f\circ k)(0), \nabla_{\mathbb{R}^d}(h\circ k)(0)\big\rangle. \end{aligned}$$

Hence, we have (2.41).
For (2.42), for all i, thanks again to the symmetry of the Euclidean norm $\|\cdot\|_2$, we have

$$\int_{B_{\mathbb{R}^d}(0,c)} G(\|v\|_2)v^i\mathrm{d}v = 0,$$

Thus, LHS of (2.42) is equal to

$$\left[\left\langle \nabla_{\mathbb{R}^m} f(x), \frac{1}{2}\sum_{i=1}^{d}\frac{\partial^2 k}{\partial x^i\partial x^i}(0)\right\rangle + \frac{1}{2}\sum_{i=1}^{d} f''(x)\left(\frac{\partial k}{\partial x^i}(0), \frac{\partial k}{\partial x^i}(0)\right)\right]\frac{1}{d}\int_{B_{\mathbb{R}^d}(0,c)} G(\|v\|_2)\|v\|_2^2\mathrm{d}v.$$

Besides, since $k(0) = x$,

$$\begin{aligned} &\left\langle \nabla_{\mathbb{R}^m} f(x), \sum_{i=1}^{d}\frac{\partial^2 k}{\partial x^i\partial x^i}(0)\right\rangle + \sum_{i=1}^{d} f''(x)\left(\frac{\partial k}{\partial x^i}(0), \frac{\partial k}{\partial x^i}(0)\right) \\ =& \sum_{i=1}^{d}\left[\sum_{j=1}^{m}\frac{\partial f}{\partial x^j}(x)\frac{\partial^2 k^j}{\partial x^i\partial x^i}(0) + \sum_{j,l=1}^{m}\frac{\partial^2 f}{\partial x^j\partial x^l}(x)\frac{\partial k^j}{\partial x^i}(0)\frac{\partial k^l}{\partial x^i}(0)\right] \\ =& \sum_{i=1}^{d}\left[\sum_{j=1}^{m}\frac{\partial}{\partial x^i}\left(\frac{\partial f}{\partial x^j}\circ k \times \frac{\partial k^j}{\partial x^i}\right)\Big|_0\right] \\ =& \sum_{i=1}^{d}\frac{\partial^2(f\circ k)}{\partial x^i\partial x^i}(0) = \Delta_{\mathbb{R}^d}(f\circ k)(0). \end{aligned}$$

This ends the proof of Lemma 2.4.2.

Chapter 3

Convergence in Wasserstein distance of occupation measures with convolution

Contents

This Chapter is a collaboration with Vincent Divol, Hélène Guérin and Viet Chi Tran, and has been submitted for publication. We would like to thank Frédéric Rochon for his advice on the differential geometry literature, which helped us to tackle some concepts more efficiently.

From the observation of a diffusion path $(X_t)_{t\in[0,T]}$ on a compact connected d-dimensional manifold $\mathcal{M}$ without boundary, we consider the problem of estimating the stationary measure μ of the process. Wang and Zhu (2023) showed that for the Wasserstein metric $\mathcal{W}_2$ and for $d \geq 5$, the convergence rate of $T^{-1/(d-2)}$ is attained by the occupation measure of the path $(X_t)_{t\in[0,T]}$ when $(X_t)_{t\in[0,T]}$ is a Langevin diffusion. We extend their result in several directions. First, we show that the rate of convergence holds for a large class of diffusion paths, whose generators are uniformly elliptic. Second, the regularity of the density p of the stationary measure μ with respect to the volume measure of $\mathcal{M}$ can be leveraged to obtain faster estimators: when p belongs to a Sobolev space of order $\ell > 0$, smoothing the occupation measure by convolution with a kernel yields an estimator whose rate of convergence is of order $T^{-(\ell+1)/(2\ell+d-2)}$. We further show that this rate is the minimax rate of estimation for this problem.

3.1 Introduction

The manifold hypothesis has become ubiquitous in the modern machine learning landscape, where it is commonly used to explain the efficiency of nonparametric methods in high-dimensional statistical models [23]. This paradigm has motivated statisticians to study inference problems under manifold constraints [91, 52, 3, 4, 36, 99]. Given n i.i.d. samples from a distribution μ supported by a d-dimensional manifold compact $\mathcal{M}$, the task of estimating either μ or geometric quantities related to $\mathcal{M}$ naturally arises. The picture is now well-understood. For example, minimax rates are known for the estimation of $\mathcal{M}$ with respect to the Hausdorff distance [53], for tangent space estimation [117, 4], and for curvature estimation [2, 4, 17, 1]. The estimation of the measure μ has been tackled in a pointwise manner [19], with respect to the Wasserstein distance [37, 109], or with respect to more general adversarial losses [111]. Once again, minimax rates of estimation are known and are typically achieved by kernel-like estimators. Yet, the literature is far less abundant when we leave the i.i.d. setting.
However, a framework in which the data is generated through an exploration process is also natural. This setting is especially relevant in scenarios where the manifold is seen as the continuum limit of a large graph, the latter being explored by a random walk (think for instance of the famous PageRank algorithm [94]). At the limit, this random walk converges to a continuous time diffusion exploring the manifold. Formally, we will consider that we have access to a sample path $(X_t)_{t\in[0,T]}$ on $[0,T]$ of a diffusion process on a submanifold $\mathcal{M}\subseteq\mathbb{R}^m$, that is generated by a uniformly elliptic $\mathcal{C}^2$-differential operator $\mathcal{A}$, essentially self-adjoint with respect to some invariant measure μ. Our goal is to propose reconstruction methods for the measure μ based on the observation of the sample path $(X_t)_{t\in[0,T]}$.
The general framework we work in encapsulates in particular operators of the form $\mathcal{A}_{pq}$, given for any test function f of class $\mathcal{C}^2$ on $\mathcal{M}$ by

$$\mathcal{A}_{pq}(f) := q\Delta f + \langle q\nabla \ln(pq), \nabla f\rangle, \tag{3.1}$$

where $p,\ q\in\mathcal{C}^2$ are two positive functions, with p being the density of the measure μ with respect to the volume measure $\mathrm{d}x$ on $\mathcal{M}$ and ∇ and Δ denote respectively the gradient and the Laplace-Beltrami operator on $\mathcal{M}$ (see e.g. [120]). When we take $q=\frac{p}{2}$, we recover the generator

$$\frac{p}{2}\Delta f + \langle\nabla p, \nabla f\rangle$$

studied in [25, 55, 61]. When $q=1$, we recover a Langevin diffusion, whose generator $\mathcal{L}$ is defined for any test function f of class $\mathcal{C}^2$ on $\mathcal{M}$ by

$$\mathcal{L}(f) :=\Delta f + \langle\nabla \ln p, \nabla f\rangle = \Delta f + \left\langle \frac{\nabla p}{p}, \nabla f\right\rangle. \tag{3.2}$$

Processes with this kind of generators can be obtained as the limits of random walks (without and with renormalization) visiting points sampled independently with identical distribution (i.i.d.) $\mu(\mathrm{d}x)=p(x)\mathrm{d}x$ on $\mathcal{M}$, see e.g. [25, 55, 61].
In $\mathbb{R}^m$, the question of estimating the invariant measure of a diffusion has been treated abundantly, see [30, 42, 95, 102] (notice that the problem could also be studied with the different point of view of non-parametric estimation for diffusion processes, see e.g. [31]). For manifold-valued data, the problem of reconstructing the stationary measure μ from a sample path was first addressed by Wang and Zhu [121] for the generator (3.2). They consider the occupation measure μ_T of the process, defined for every bounded measurable test function f by

$$\int_{\mathcal{M}} f(x)\mu_T(\mathrm{d}x) = \frac{1}{T}\int_0^T f(X_s)\ \mathrm{d}s. \tag{3.3}$$

Let $\mathcal{P}(\mathcal{M})$ be the space of probability measures on the compact connected d-dimensional Riemannian manifold $\mathcal{M}$. We introduce the 2-Wasserstein distance on $\mathcal{P}(\mathcal{M})$, defined by

$$\mathcal{W}_2(\mu_1, \mu_2) := \inf_{\pi \in \mathcal{C}(\mu_1,\mu_2)} \left(\int_{\mathcal{M}\times\mathcal{M}} \rho(x,y)^2 \pi(\mathrm{d}x, \mathrm{d}y) \right)^{1/2},$$

where ρ is the geodesic distance on $\mathcal{M}$ and $\mathcal{C}(\mu_1, \mu_2)$ is the set of measures on $\mathcal{M} \times \mathcal{M}$ with first marginal μ_1 and second marginal μ_2. Wang and Zhu [121, Theorem 1.2] showed that for the process with generator $\mathcal{L}$,

$$\mathbb{E}_x\left[\mathcal{W}_2^2(\mu_T, \mu)\right] \lesssim \begin{cases} T^{-1} & \text{when } d \leq 3 \\ T^{-1}\ln(1+T) & \text{when } d = 4 \\ T^{-\frac{2}{d-2}} & \text{when } d \geq 5, \end{cases} \tag{3.4}$$

where $\mathbb{E}_x$ stands for the expectation taken from the diffusion process starting at $x \in \mathcal{M}$. As noticed by Divol [37], in the context of i.i.d. random variables $X_1, \dots X_n$ sampled from μ on $\mathcal{M}$, the rate of convergence can be increased by smoothing the empirical measure. Our purpose here is to extend this result beyond the i.i.d. setting, by studying the convergence properties of an estimator $\widehat{\mu}_{T,h}$ of μ, obtained by smoothing the occupation measure μ_T with a kernel K of bandwidth $h > 0$. When $d \geq 5$ and for an appropriate choice of h, we obtain the rate of convergence

$$\mathbb{E}_x\left[\mathcal{W}_2^2(\widehat{\mu}_{T,h}, \mu)\right] \lesssim T^{-\frac{2\ell+2}{2\ell+d-2}}, \tag{3.5}$$

where μ has a density of regularity $\ell \geq 2$. The above rate does not only hold for the Langevin diffusion with generator $\mathcal{L}$, but for all diffusion paths $(X_t)_{t\in[0,T]}$ whose generator $\mathcal{A}$ is a uniformly elliptic $\mathcal{C}^2$-differential operator, essentially self-adjoint with respect to μ. Furthermore, we will show that these rates cannot be improved by providing minimax rates of convergence for this problem.

In Section 3.2, we define the estimator $\widehat{\mu}_{T,h}$ and enounce precisely our main result. In Section 3.3, we review some useful notions of Riemannian geometry. In Section 3.4, we start the proof by treating the stationary case, i.e., when the initial measure of the SDE is μ, and then give a generalization for a general initial measure.

Notation. Throughout the paper, we fix a smooth compact d-dimensional connected submanifold $\mathcal{M}$ of $\mathbb{R}^m$, without boundary, and embedded with the Riemannian structure induced by the ambient space $\mathbb{R}^m$. The volume measure on $\mathcal{M}$ is denoted by $\mathrm{d}x$. Without loss of generalization, we assume that $\mathrm{vol}(\mathcal{M}) = 1$. Unless stated otherwise, quantities $c_0, c_1, \dots$ are constants that are only allowed to depend on the manifold $\mathcal{M}$. We write c_a for a constant depending on an additional parameter a. The geodesic distance on $\mathcal{M}$ is denoted by ρ, and $\mathcal{B}(x,r)$ is the geodesic open ball centered at $x \in \mathcal{M}$ of radius $r \geq 0$. We also let $\mathcal{B}_{\mathbb{R}^d}(u,r)$ be the open ball centered at $u \in \mathbb{R}^d$ of radius r.

For μ a probability measure, we will denote by $L^2(\mu)$ the space of real-valued measurable functions f on $\mathcal{M}$ such that $\int |f|^2 \mathrm{d}\mu < \infty$. More generally, for $p \geq 1$, we let $L^p(\mu)$ denote the space of L^p functions with respect to μ, with $\|\cdot\|_{L^p(\mu)}$ the corresponding norm. For $k \geq 0$, we denote by $\mathcal{C}^k(\mathcal{M})$ the space of k-times continuously differentiable real-valued functions defined on $\mathcal{M}$, endowed with the norm

$$\|f\|_{\mathcal{C}^k(\mathcal{M})} := \sup_{0\leq i\leq k} \sup_{x\in\mathcal{M}} \|\nabla^i f(x)\|, \tag{3.6}$$

where $\nabla^i f(x)$ is the ith iterated covariant derivative of f at x. By abuse of notation, we use the same notation for the uniform norm $\|.\|_\infty$ on $\mathbb{R}_+$, on $\mathcal{M}$, and on $\mathbb{R}^d$.

3.2 Main results

We consider the framework described at the beginning of the paper. Let $\mu(\mathrm{d}x) = p(x)\mathrm{d}x$ be a probability measure on $\mathcal{M}$ with a positive density p of class $\mathcal{C}^2$, and $(X_t)_{t\geq 0}$ be a diffusion on $\mathcal{M}$ with generator $\mathcal{A}$.

Assumption 4. *$\mathcal{A} : \mathcal{C}^\infty(\mathcal{M}) \to \mathcal{C}^2(\mathcal{M})$ is a uniformly elliptic $\mathcal{C}^2$-differential operator of second order on $\mathcal{M}$, symmetric with respect to the measure μ.*

This diffusion admits μ as stationary measure. In the sequel, this generator is extended, as is usually done, to $L^2(\mu)$ functions. We introduce a concept closely associated with second-order differential operators, known as the "carré du champ" $\Gamma(f, g)$ for the operator $\mathcal{A}$:

$$\Gamma(f,g) = \frac{1}{2}\left(\mathcal{A}(fg) - f\mathcal{A}(g) - g\mathcal{A}(f)\right) \tag{3.7}$$

Given that $\mathcal{A}$ is symmetric with respect to μ, for any smooth functions f and g, it follows that:

$$\int_{\mathcal{M}} \Gamma(f,g)\mathrm{d}\mu = -\int_{\mathcal{M}} f\mathcal{A}(g)\mathrm{d}\mu = -\int_{\mathcal{M}} \mathcal{A}(f)g\mathrm{d}\mu. \tag{3.8}$$

Since $\mathcal{M}$ is compact, from the regularity of p, there exist $p_{\min}$, $p_{\max} > 0$ such that,

$$\forall x \in \mathcal{M}, \quad p_{\min} \leq p(x) \leq p_{\max}. \tag{3.9}$$

Furthermore, the uniform ellipticity, the continuity of $\mathcal{A}$ and the compactness of $\mathcal{M}$ imply that there exist positive constants $\kappa_{\min}$, $\kappa_{\max}$ such that for all functions f,

$$\kappa_{\min}|\nabla f|^2 \leq \Gamma(f,f) \leq \kappa_{\max}|\nabla f|^2. \tag{3.10}$$

We denote by $\mathbb{E}_{\mu_0}$ and $\mathbf{P}_{\mu_0}$ the expectation and probability taken for the diffusion process with initial distribution μ_0. When μ_0 is a Dirac measure δ_x, with $x \in \mathcal{M}$, we will simply write $\mathbb{E}_x$ and $\mathbf{P}_x$.

Let $T > 0$ be a time horizon, and assume that we observe the diffusion $(X_t)_{t\geq 0}$ on the time window $[0,T]$. We consider the occupation measure μ_T on $[0,T]$ of the diffusion $(X_t)_{t\geq 0}$ as the positive measure defined by (3.3). As explained in the introduction, the occupation measure can be seen as a first naive estimator of the measure μ, that will be improved upon by convolution with a kernel K_h.

Let $K : \mathbb{R}_+ \to \mathbb{R}$ be a (signed) Lipschitz-continuous function supported in $[0,1]$, with $\int_{\mathbb{R}^d} K(\|u\|)\mathrm{d}u = 1$. We define for $h > 0$ and $(x,y) \in \mathcal{M}^2$,

$$K_h(x,y) := \frac{1}{\eta_h(x)} K\left(\frac{\|x-y\|}{h}\right), \tag{3.11}$$

with $\eta_h(x) = \int_{\mathcal{M}} K\left(\frac{\|x-y\|}{h}\right)\mathrm{d}y$. We will show later (in Lemma 3.4.2) that $\eta_h > 0$ for h small enough, ensuring that the kernel K_h is well-defined. We consider the following estimator of the density p of μ obtained by convolution of the occupation measure μ_T with K_h. For $x \in \mathcal{M}$,

$$\begin{aligned} p_{T,h}(x) &:= \int_{\mathcal{M}} K_h(z,x)\mu_T(\mathrm{d}z) = \frac{1}{T}\int_0^T K_h(X_s,x)\mathrm{d}s \\ &= \frac{1}{T}\int_0^T \frac{1}{\eta_h(X_s)} K\left(\frac{\|X_s - x\|}{h}\right)\mathrm{d}s. \end{aligned} \tag{3.12}$$

Thanks to the definition of η_h, we notice that

$$\int_{\mathcal{M}} p_{T,h}(x)\mathrm{d}x = 1.$$

However, the function $p_{T,h}$ is not necessarily a density, as it may not be nonnegative everywhere (recall that K is a signed function). Still, we will show in this article that the function $p_{T,h}$ approximates the density p and is nonnegative, and is therefore a density, with high probability. Let x_0 be an arbitrary fixed point of $\mathcal{M}$. We introduce two random measures $\mu_{T,h}$, and $\widehat{\mu}_{T,h}$ on $\mathcal{M}$, defined by

$$\mu_{T,h}(\mathrm{d}x) = p_{T,h}(x)\mathrm{d}x, \text{ and } \widehat{\mu}_{T,h} = \begin{cases} \mu_{T,h} & \text{if } \mu_{T,h} \text{ is a nonnegative measure,} \\ \delta_{x_0} & \text{otherwise.} \end{cases} \tag{3.13}$$

The measure $\widehat{\mu}_{T,h}$ is introduced for purely technical purposes. Indeed, the measure $\mu_{T,h}$ may not be a probability measure (with exponentially small probability), so that the risk $\mathcal{W}_2(\mu_{T,h}, \mu)$ may not even be defined, whereas $\mathcal{W}_2(\widehat{\mu}_{T,h}, \mu)$ is always defined.

Remark 3.2.1. *It would have been arguably more convenient to work with a kernel based on the geodesic distance ρ, i.e. a kernel $\widetilde{K}_h$ defined by $\widetilde{K}_h(x,y) := \frac{1}{\widetilde{\eta}_h(x)} K\Big(\frac{\rho(x,y)}{h}\Big)$, with $\widetilde{\eta}_h(x) = \int_{\mathcal{M}} K\Big(\frac{\rho(x,y)}{h}\Big)\mathrm{d}y$. For instance, we can easily prove that $\big(h^{-d}\widetilde{\eta}_h\big)_{h>0}$ converges to 1 uniformly on $\mathcal{M}$, with a speed of convergence of order h^2. Such a property also holds for η_h, but only for sufficiently smooth kernels K satisfying some moments assumptions (see Definition 3.7.1). However, for statistical purposes, the use of the Euclidean distance in (3.11) seems natural in the context where the manifold $\mathcal{M}$ (and hence the geodesic distance ρ) is unknown. Yet, the study of the convolution with the kernel $\widetilde{K}_h$ is of own interest and is treated in [?]. Moreover, remark that since the manifold $\mathcal{M}$ is assumed to be compact, the geodesic distance $\rho(\cdot,\cdot)$ on $\mathcal{M}$ and the Euclidean distance $\|\cdot\|$ of $\mathbb{R}^m$ are known to be equivalent, see e.g. [51, Proposition 2].*

We are now in position to state our main result, which gives the rate of convergence of $\widehat{\mu}_{T,h}$ to μ when the diffusion path $(X_t)_{t\in[0,T]}$ exploring the manifold has generator $\mathcal{A}$. More precisely, our purpose is to upper-bound

$$\sup_{x\in\mathcal{M}} \mathbb{E}_x\big[\mathcal{W}_2^2(\widehat{\mu}_{T,h}, \mu)\big]. \tag{3.14}$$

For any initial measure μ_0, we can write $\mathbb{E}_{\mu_0}$ as a mixture $\mathbb{E}_{\mu_0}[\cdot] = \int \mathbb{E}_x[\cdot]\mu_0(\mathrm{d}x)$. Hence, a bound on the uniform risk defined in (3.14) automatically implies a bound on the risk for *any* initial measure. As often, such a bound is obtained by decomposing the loss into a bias term $\mathcal{W}_2^2(\mu_h, \mu)$ and a variance term $\mathbb{E}_x\big[\mathcal{W}_2^2(\widehat{\mu}_{T,h}, \mu_h)\big]$, where

$$\begin{aligned} \mu_h(\mathrm{d}x) = p_h(x)\mathrm{d}x \quad \text{with} \quad p_h(x) &:= \int_{\mathcal{M}} K_h(z,x)p(z)\mathrm{d}z \\ &= \int_{\mathcal{M}} \frac{1}{\eta_h(z)} K\Big(\frac{\|z-x\|}{h}\Big)p(z)\mathrm{d}z. \end{aligned} \tag{3.15}$$

is the intensity measure of the random measure $\mu_{T,h}$.
The control of the variance term $\mathbb{E}_x\big[\mathcal{W}_2^2(\widehat{\mu}_{T,h}, \mu_h)\big]$ relies on fine spectral properties of the generator of the diffusion. The proof of the following result is detailed in Section 3.4 for a diffusion starting from its invariant measure μ and in Section 3.5 for a diffusion starting from a general initial distribution. The bound on the variance in the following theorem depends on the ultracontractivity constant $u_{\mathcal{A}}$ of the generator $\mathcal{A}$, defined in Section 3.5.

Theorem 3.2.2 (Estimation from a diffusion with generator $\mathcal{A}$). *Let $d \geq 1$ and p be a positive $\mathcal{C}^2$ density function with associated measure μ. Let $(X_t)_{t\geq 0}$ be a diffusion with generator $\mathcal{A}$ satisfying Assumption 4. Let $T \geq 2$ and let $0 < h \leq h_0$ for some constant h_0 depending on $\mathcal{M}$ and K. Assume that either K is nonnegative or that $d \geq 4$ and that $Th^d \geq c\ln(T)$ (in which case, h_0 additionally depends on $p_{\min}$ and on the $\mathcal{C}^1$-norm of p). Then,*

$$\sup_{x\in\mathcal{M}} \mathbb{E}_x\big[\mathcal{W}_2^2(\widehat{\mu}_{T,h}, \mu_h)\big] \leq c_0 \frac{u_{\mathcal{A}} p_{\max}^2}{p_{\min}^2} \|K\|_\infty^2 \begin{cases} \frac{h^{4-d}}{T} & \text{if } d \geq 5 \\ \frac{\ln(1/h)}{T} & \text{if } d = 4 \\ \frac{1}{T} & \text{if } d \leq 3, \end{cases} \tag{3.16}$$

where c_0 depends on $\mathcal{M}$, and c depends on $\mathcal{M}$, K, $p_{\min}$, $p_{\max}$ and $\kappa_{\min}$.

The second term in the risk decomposition is the bias term $\mathcal{W}_2^2(\mu_h, \mu)$, which was already studied by Divol in [37]. Let $\ell \geq 0$. We introduce the Sobolev space $H^\ell(\mathcal{M})$ as the completion of the set of smooth functions on $\mathcal{M}$ with respect to the norm:

$$\|f\|_{H^\ell(\mathcal{M})} := \max_{0\leq i\leq \ell} \left(\int_{\mathcal{M}} \|\nabla^i f(x)\|^2 \, \mathrm{d}x \right)^{1/2}.$$

As $\mathcal{M}$ is compact, we note that for any $f \in \mathcal{C}^\ell(\mathcal{M})$, $\|f\|_{H^\ell(\mathcal{M})} \leqslant \|f\|_{\mathcal{C}^\ell(\mathcal{M})}$ and $\mathcal{C}^\ell(\mathcal{M})$ is a subset of $H^\ell(\mathcal{M})$.

Under additional technical conditions on the kernel K (recalled in Section 3.7), Divol showed that if $p \in H^\ell(\mathcal{M})$ for some $\ell \geq 0$, then

$$\mathcal{W}_2^2(\mu_h, \mu) \leq c_1 \frac{\|p\|^2_{H^\ell(\mathcal{M})}}{p^2_{\min}} h^{2\ell+2}, \tag{3.17}$$

where c_1 depends only on $\mathcal{M}$ and K, see Proposition 3.7.2. As a corollary of Theorem 3.2.2 and (3.17), we obtain a tight control on the risk of the estimator $\widehat{\mu}_{T,h}$.

Corollary 3.2.3. *Let $d \geq 5$ and p be a positive $\mathcal{C}^2$ density function with associated measure μ. Further assume that p has a controlled Sobolev norm $\|p\|_{H^\ell(\mathcal{M})}$ for some $\ell \geq 2$. Assume that K is a kernel of order larger than ℓ (in the sense of Definition 3.7.1). Let $(X_t)_{t\geq 0}$ be a diffusion with generator $\mathcal{A}$ satisfying Assumption 4. Let $T \geq 2$ and let $0 < h \leq h_0$ and assume that $Th^d \geq c\ln(T)$, where h_0, c are the constants from Theorem 3.2.2. Then,*

$$\sup_{x\in\mathcal{M}} \mathbb{E}_x\left[\mathcal{W}_2^2(\widehat{\mu}_{T,h}, \mu)\right] \leq 2\ c_1 \frac{\|p\|^2_{H^\ell(\mathcal{M})}}{p^2_{\min}} h^{2\ell+2} + 2\ c_0 \frac{u_{\mathcal{A}} p^2_{\max}}{p^2_{\min}} \|K\|^2_\infty \frac{h^{4-d}}{T} \tag{3.18}$$

where c_0 is the constant from Theorem 3.2.2 and c_1 is the constant in (3.17). *In particular, for h of order $T^{-1/(2\ell+d-2)}$, it holds that*

$$\sup_{x\in\mathcal{M}} \mathbb{E}_x\left[\mathcal{W}_2^2(\widehat{\mu}_{T,h}, \mu)\right] \lesssim T^{-\frac{2\ell+2}{2\ell+d-2}}. \tag{3.19}$$

The results of Theorem 3.2.2 and Corollary 3.2.3 are non-asymptotic results. It is a remarkable fact that the constants in the previous corollary only depend on the generator $\mathcal{A}$ through the uniform ellipticity constant $\kappa_{\min}$ and the ultracontractivity constant $u_{\mathcal{A}}$. From a statistical perspective, this implies that the knowledge of the exact SDE satisfied by the sample path $(X_t)_{t\in[0,T]}$ is not needed to estimate the invariant measure μ. Only an a priori estimate on the uniform ellipticity constant $\kappa_{\min}$ of the generator $\mathcal{A}$ of the sample path and its ultracontractivity constant $u_{\mathcal{A}}$ have to be known. For instance, the same reconstruction method will apply for estimating a sample path with either of the generators $\mathcal{L}$ or $\mathcal{A}_{pq}$ for $q = p/2$ mentioned in the introduction.

Comparing with the results of Wang and Zhu in [121] for the operator $\mathcal{L}$, we note that for $d \geq 5$, the rate $T^{-\frac{2\ell+2}{2\ell+d-2}}$ is faster than the rate of $T^{-\frac{2}{d-2}}$ that they obtained for the occupation measure μ_T, see (3.4). Actually, our results allow us to recover their rate of convergence, for *any* generator $\mathcal{A}$ satisfying Assumption 4.

Corollary 3.2.4. *Let $d \geq 1$ and p be a positive density function of class $\mathcal{C}^2$. Let $(X_t)_{t\geq 0}$ be a diffusion with generator $\mathcal{A}$ satisfying Assumption 4. Then, for all $T \geq 2$,*

$$\sup_{x\in\mathcal{M}} \mathbb{E}_x\left[\mathcal{W}_2^2(\mu_T, \mu)\right] \leq c_0 \left(1 + \frac{u_{\mathcal{A}} p^2_{\max}}{p^2_{\min}}\right) \begin{cases} T^{-1} & \text{when } d \leq 3 \\ T^{-1}\ln(1+T) & \text{when } d = 4 \\ T^{-\frac{2}{d-2}} & \text{when } d \geq 5, \end{cases} \tag{3.20}$$

for some constant c_0 depending on $\mathcal{M}$.

Proof. Let K be a nonnegative Lipschitz-continuous kernel supported in $[0,1]$ with $\int_{\mathbb{R}^d} K(\|u\|)\mathrm{d}u = 1$. As K is positive, $\mu_{T,h}$ is always a probability measure, so that $\widehat{\mu}_{T,h} = \mu_{T,h}$. For any probability measure ν, we define its convolution ν_h by K_h as in (3.15). We prove in Lemma 3.8.3 that for any measure ν and $h > 0$ small enough, $\mathcal{W}_2^2(\nu_h, \nu) \leq c\|K\|_\infty h^2$ for some constant c. Hence, both $\mathcal{W}_2^2(\mu_T, \mu_{T,h})$ and $\mathcal{W}_2^2(\mu, \mu_h)$ are of order h^2. Then,

$$\begin{aligned}\mathbb{E}_x\left[\mathcal{W}_2^2(\mu_T, \mu)\right] &\leq 4\mathbb{E}_x\left[\mathcal{W}_2^2(\mu_T, \mu_{T,h})\right] + 4\mathbb{E}_x\left[\mathcal{W}_2^2(\mu_{T,h}, \mu_h)\right] + 4\mathcal{W}_2^2(\mu_h, \mu) \\ &\leq 8c\|K\|_\infty h^2 + 4\mathbb{E}_x\left[\mathcal{W}_2^2(\mu_{T,h}, \mu_h)\right].\end{aligned}$$

We pick $h = T^{-1/(d-2)}$ for $d \geq 5$ and $h = T^{-1/2}$ for $d \leq 4$ and apply Theorem 3.2.2 to conclude. □

Finally, we address the optimality of our statistical procedure using minimax theory. Consider a class $\mathcal{P}_T$ of probability distributions of diffusion processes $(X_t)_{t\in[0,T]}$. For $\mathbf{P}_T \in \mathcal{P}_T$, we let $\mu(\mathbf{P}_T)$ be the invariant measure of the diffusion process $(X_t)_{t\geq 0}$ whose restriction to $[0,T]$ has distribution $\mathbf{P}_T$. The associated minimax rate of convergence is defined as

$$\mathcal{R}(\mathcal{P}_T) = \inf_{\hat{\mu}} \sup_{\mathbf{P}_T \in \mathcal{P}_T} \mathbb{E}_{\mathbf{P}_T}[\mathcal{W}_2(\hat{\mu}, \mu(\mathbf{P}_T))], \tag{3.21}$$

where the infimum is taken over all estimators $\hat{\mu}$, i.e. measurable functions of the observation $(X_t)_{t\in[0,T]}$, and where the expectation is over processes $(X_t)_{t\in[0,T]}$ distributed as $\mathbf{P}_T$. To put it another way, the minimax rate is the best risk an estimator can attain for the problem of estimating the invariant measure of some diffusion process $(X_t)_{t\in[0,T]}$ whose distribution lies in $\mathcal{P}_T$.

For parameters $\ell \geq 2$, $\kappa_{\min}$, $p_{\min}$, $p_{\max}$, $u_{\max}$, $L > 0$, we consider the statistical model $\mathcal{P}_{T,\ell} = \mathcal{P}_{T,\ell}(\kappa_{\min}, p_{\min}, p_{\max}, u_{\max}, L)$ consisting of all the distribution of diffusion processes $(X_t)_{t\in[0,T]}$ on $\mathcal{M}$ with arbitrary initial distribution, generator $\mathcal{A}$ satisfying Assumption 4 with constant $\kappa_{\min}$ and ultracontractivity constant $u_{\mathcal{A}}$ smaller than $u_{\max}$, and whose invariant measure μ has a $\mathcal{C}^2$ density p on $\mathcal{M}$ with $\|p\|_{H^\ell(\mathcal{M})} \leq L$ and $\|p\|_{\mathcal{C}^1(\mathcal{M})} \leq L$, satisfying $p_{\min} \leq p \leq p_{\max}$. Remark that the kernel density estimator $\widehat{\mu}_{T,h}$ attains (for $d \geq 5$) the rate of convergence $T^{-\frac{\ell+1}{2\ell+d-2}}$ uniformly on the class $\mathcal{P}_{T,\ell}$. Hence, the minimax rate satisfies $\mathcal{R}(\mathcal{P}_{T,\ell}) \lesssim T^{-\frac{\ell+1}{2\ell+d-2}}$. The next proposition, proved in Section 3.6, states that this rate cannot be improved.

Proposition 3.2.5. *Let $\ell \geq 2$ be an integer. Then, for $\kappa_{\min}$, $p_{\min}$ small enough and $p_{\max}$, $u_{\max}$, L large enough,*

$$\mathcal{R}(\mathcal{P}_{T,\ell}) \gtrsim \begin{cases} T^{-1/2} & \text{if } d \leq 4 \\ T^{-\frac{\ell+1}{2\ell+d-2}} & \text{if } d \geq 5. \end{cases} \tag{3.22}$$

For $d \geq 5$, these rates match with the rates obtained by our kernel-based estimator $\widehat{\mu}_{T,h}$, whereas for $d \leq 4$, the empirical estimator μ_T attains the minimax rate (up to logarithmic factors), according to the results of Wang and Zhu [121] for the Langevin generator $\mathcal{L}$, or according to Corollary 3.2.4 for a general generator $\mathcal{A}$.

Remark 3.2.6. *The estimator $\widehat{\mu}_{T,h}$ has a density with respect to the volume measure* dx *on $\mathcal{M}$. As a consequence, computing $\widehat{\mu}_{T,h}$ requires the knowledge of the manifold $\mathcal{M}$, prohibiting the use of this method in the situation where the manifold $\mathcal{M}$ is unknown. However, we expect that similar methods to the ones developed in [37] will allow us to create an estimator $\widehat{\mathrm{vol}_{\mathcal{M}}}$ of the volume measure. Such an estimator of the volume measure is based on a patch-based reconstruction of the underlying manifold $\mathcal{M}$ developed by Aamari and Levrard [4]. Guarantees for this method are only known in the case of i.i.d. samples, although we believe that the results of [4] could be adapted to the setting of this paper, where a diffusion path is observed. Such an estimator of the volume measure could then be used through a plug-in method to design a new estimator $\widetilde{\mu}_{T,h}$ that would not require the knowledge of the manifold $\mathcal{M}$.*

3.3 Preliminaries

We detail in this section some tools, notation, and general results on elliptic operators on compact manifolds. Recall that $\mathcal{M}$ is a smooth compact d-dimensional connected Riemannian manifold without boundary.

3.3.1 The Laplace-Beltrami operator

Recall that by Green's theorem, the Laplace-Beltrami operator is symmetric, with for all $f, g \in \mathcal{C}^2(\mathcal{M})$,

$$\int_{\mathcal{M}} (\Delta f) g \mathrm{d}x = -\int_{\mathcal{M}} \langle \nabla f, \nabla g \rangle \, \mathrm{d}x = \int_{\mathcal{M}} f(\Delta g) \mathrm{d}x. \tag{3.23}$$

The operator Δ defines an essentially self-adjoint operator over $L^2(\mathrm{d}x)$ with a discrete spectrum (see e.g. [26, Chapter I, Section 3]). We denote by

$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \cdots$$

the eigenvalues of $(-\Delta)$, and by $(\phi_i)_{i\geq 0}$ the respective eigenfunctions of Δ (note that they are of class $\mathcal{C}^\infty$ on $\mathcal{M}$). The first eigenfunction ϕ_0 is constant and the family $(\phi_i)_{i\geq 0}$ forms a Hilbert basis of $L^2(\mathrm{d}x)$: for any $f \in L^2(\mathrm{d}x)$,

$$f = \sum_{i=0}^{+\infty} \beta_i \phi_i, \qquad \text{with} \qquad \beta_i = \int_{\mathcal{M}} f \phi_i \mathrm{d}x.$$

The first nonzero eigenvalue λ_1 is of particular importance and is called the spectral gap. The inverse operator Δ^{-1} is defined on $L_0^2(\mathrm{d}x) := \left\{ f \in L^2(\mathrm{d}x) : \int_{\mathcal{M}} f \mathrm{d}x = 0 \right\}$ by

$$\Delta^{-1}(f) := -\sum_{i=1}^{\infty} \frac{\beta_i}{\lambda_i} \phi_i. \tag{3.24}$$

We introduce the operator $(-\Delta)^{-1/2}$ defined for $f \in L_0^2(\mathrm{d}x)$ by

$$(-\Delta)^{-1/2}(f) := \sum_{i=1}^{\infty} \frac{\beta_i}{\sqrt{\lambda_i}} \phi_i.$$

3.3.2 Green function of the Laplace Beltrami operator

The Green function G of Δ is a linear bounded operator $G : L^2(\mathrm{d}x) \to L^2(\mathrm{d}x)$ which is, in some sense, an inverse of Δ in $L^1(\mathrm{d}x)$.

Proposition 3.3.1 (See Appendix A in [7], or Theorem 4.13 in [11]). *We define* $\mathrm{diag}(\mathcal{M}) = \{(x, x) : x \in \mathcal{M}\}$. *There exists a unique continuous function* $G \in \mathcal{C}^\infty(\mathcal{M} \times \mathcal{M} \setminus \mathrm{diag}(\mathcal{M}))$, *which has the following properties*

(i) $\forall x \in \mathcal{M}$, $G(x, \cdot) \in L^1(\mathrm{d}x)$ *with* $\int_{\mathcal{M}} G(x, \cdot) \mathrm{d}x = 0$;

(ii) G *is symmetric* : $\forall (x, y) \in \mathcal{M}^2 \setminus \mathrm{diag}(\mathcal{M})$, $G(x, y) = G(y, x)$;

(iii) $\forall f \in \mathcal{C}^2(\mathcal{M}), \forall x \in \mathcal{M}$, *we have*

$$\int_{\mathcal{M}} G(x, y) \Delta f(y) \mathrm{d}y = f(x) - \int_{\mathcal{M}} f(y) \mathrm{d}y;$$

(iv) there exists a constant $\kappa > 0$ such that $\forall (x,y) \in \mathcal{M}^2 \setminus \mathrm{diag}(\mathcal{M})$,

$$|G(x,y)| \leq \kappa \begin{cases} 1 & \text{when } d = 1, \\ 1 + |\ln \rho(x,y)| & \text{when } d = 2, \text{ and} \\ \rho(x,y)^{2-d} & \text{when } d \geq 3. \end{cases}$$

The function G is called the Green function of the operator Δ and it naturally defines an operator from $L^2(\mathrm{d}x)$ to $L^2(\mathrm{d}x)$, which is also denoted by G, as follows: $\forall f \in L^2(\mathrm{d}x)$, $\forall x \in \mathcal{M}$,

$$(Gf)(x) := \int_{\mathcal{M}} G(x,y) f(y) \mathrm{d}y. \tag{3.25}$$

From Proposition 3.3.1 (*iii*) and (3.25), we remark that for $f \in L^2_0(\mathrm{d}x)$ we have

$$Gf = \Delta^{-1}(f).$$

The Green function will play a central role to control the variance of $\widehat{\mu}_{T,h}$, in particular through the following lemma, which controls the behavior of Gf for a function f localized on a small ball.

Lemma 3.3.2. *Let $h > 0$ sufficiently small. For any $(x,z) \in \mathcal{M}^2$, any continuous function f with support in the ball $\mathcal{B}(x,h)$, there exist constants $\kappa_1, \kappa_2 > 0$ depending only on $\mathcal{M}$ such that:*

- *when $d = 1$, $|(Gf)(z)| \leq \kappa_1 \|f\|_\infty h$,*
- *when $d = 2$, $|(Gf)(z)| \leq \kappa_1 \|f\|_\infty h^2 (1 + \ln|h|)$,*
- *when $d \geq 3$,*

$$|(Gf)(z)| \leq \kappa_1 \|f\|_\infty h^2 \quad \text{when } \rho(x,z) \leq 2h,$$
$$|(Gf)(z)| \leq \kappa_2 \|f\|_\infty h^d \rho(x,z)^{2-d} \quad \text{when } \rho(x,z) > 2h.$$

The proof, given in Appendix 3.8.3, uses the Riemannian normal parametrization of the manifold $\mathcal{M}$.

3.3.3 The elliptic operator $\mathcal{A}$

Consider an elliptic operator $\mathcal{A}$ on $\mathcal{M}$ satisfying Assumption 4. As mentioned previously, using the carré du champ Γ defined by (3.7), $\mathcal{A}$ satisfies a Green's formula (3.8). Furthermore, as a uniformly elliptic operator of second order on a compact manifold without boundary, symmetric with respect to the measure μ, the operator $\mathcal{A}$ is essentially self-adjoint with respect to $L^2(\mu)$, and its spectrum is discrete (see e.g. [44, Chapter 6] or Theorem 1.5.39 of Chapter 1). Consequently, there exist functions $(\psi_i)_{i\in\mathbb{N}}$ and a sequence of real numbers $0 = \gamma_0 < \gamma_1 \leq \gamma_2 \leq \cdots$ such that

(i) for each i, ψ_i is an eigenfunction of $-\mathcal{A}$ associated to the eigenvalue γ_i (counted with multiplicity);

(ii) $(\psi_i)_{i\in\mathbb{N}}$ is an orthonormal basis of $L^2(\mu)$.

Let $L^2_0(\mu) := \left\{ f \in L^2(\mu) : \int_{\mathcal{M}} f \mathrm{d}\mu = 0 \right\}$. As for the Laplace-Beltrami operator, we introduce the operators $\mathcal{A}^{-1}$ and $(-\mathcal{A})^{-1/2}$ defined for $f \in L^2_0(\mu)$ by

$$\mathcal{A}^{-1}(f) := -\sum_{i=1}^{\infty} \frac{\alpha_i}{\gamma_i} \psi_i \quad \text{and} \quad (-\mathcal{A})^{-1/2}(f) := \sum_{i=1}^{\infty} \frac{\alpha_i}{\sqrt{\gamma_i}} \psi_i, \tag{3.26}$$

with $\alpha_i = \int_{\mathcal{M}} f\psi_i \mathrm{d}\mu$. At last, we denote by $(P_t)_{t\geq 0}$ the semigroup of a diffusion process $(X_t)_{t\geq 0}$ with generator $\mathcal{A}$. Then, for $f \in L^2(\mu)$, $f = \sum_{i\geq 0} \alpha_i \psi_i$ with $\alpha_i = \int_{\mathcal{M}} f\psi_i \mathrm{d}\mu$, we have

$$\forall x \in \mathcal{M},\ P_t f(x) = \mathbb{E}_x[f(X_t)] = \sum_{i=0}^{\infty} \mathrm{e}^{-\gamma_i t} \alpha_i \psi_i(x).$$

This result is a consequence of the Dynkin formula [76], which implies that for any function f in the domain of $\mathcal{A}$

$$\forall x \in \mathcal{M},\ P_t f(x) = \mathbb{E}_x[f(X_t)] = f(x) + \int_0^t P_s \mathcal{A} f(x) \mathrm{d}s.$$

We now give a Poincaré inequality for the general operator $\mathcal{A}$.

Proposition 3.3.3. *(Poincaré's inequality for $\mathcal{A}$) For any $f \in \mathcal{C}^2(\mathcal{M})$ such that $\int_{\mathcal{M}} f \mathrm{d}\mu = 0$,*

$$\int_{\mathcal{M}} -f\mathcal{A}f \mathrm{d}\mu \geq \frac{p_{\min}\kappa_{\min}}{p_{\max}} \lambda_1 \int_{\mathcal{M}} f^2 \mathrm{d}\mu,$$

where λ_1 is the spectral gap of Δ, and $\kappa_{\min}$ is defined in (3.10).

Proof. By the symmetry of $\mathcal{A}$, Equation (3.10) and the Poincaré's inequality of Δ, we have:

$$\begin{aligned}\int_{\mathcal{M}} -f\mathcal{A}f \mathrm{d}\mu &= \int_{\mathcal{M}} \Gamma(f,f)\mathrm{d}\mu \geq p_{\min}\kappa_{\min} \int_{\mathcal{M}} |\nabla f|^2 \mathrm{d}x \geq p_{\min}\kappa_{\min}\lambda_1 \int_{\mathcal{M}} (f-\overline{f})^2 \mathrm{d}x \\ &\geq \frac{p_{\min}\kappa_{\min}}{p_{\max}} \lambda_1 \int_{\mathcal{M}} (f-\overline{f})^2 \mathrm{d}\mu \geq \frac{p_{\min}\kappa_{\min}}{p_{\max}} \lambda_1 \int_{\mathcal{M}} f^2 \mathrm{d}\mu,\end{aligned}$$

where $\overline{f} := \int_{\mathcal{M}} f \mathrm{d}x$. We develop the square and use that that $\int_{\mathcal{M}} f \mathrm{d}\mu = 0$ to obtain the conclusion. □

Remark 3.3.4. *The elliptic operators $\mathcal{A}_{pq}$ and $\mathcal{L}$ defined in Section 3.1 are essentially Laplace operators. Indeed, the operator $\mathcal{L}$ is a weighted Laplacian with weight $\mu(\mathrm{d}x) = p(x)\mathrm{d}x$ as defined in [59, Section 3.6], while the operator $\mathcal{A}_{pq}$ is a weighted Laplacian on a weighted manifold, i.e. a manifold $\mathcal{M}$ with a new Riemannian metric depending on the weight q, see Section 3.6 and Exercise 3.11 in [59].*

3.4 Variance term for the stationary process

Before proving Theorem 3.2.2, we state a simpler version of the theorem, where the initial measure δ_x for the sample path (X_t) is replaced by the invariant measure μ. For such a choice, it holds that for all $y \in \mathcal{M}$,

$$\mathbb{E}_\mu[p_{T,h}(y)] = \int_{\mathcal{M}} K_h(z,y)p(z)\mathrm{d}z = p_h(y). \tag{3.27}$$

We will then explain in Section 3.5 how we can extend the result to any initial distribution μ_0 using the ultracontractivity of the semi-group $(P_t)_{t\geq 0}$.

Proposition 3.4.1. *Let $d \geq 1$ and p be a positive $\mathcal{C}^2$ density function with associated measure μ. Let $(X_t)_{t\geq 0}$ be a diffusion with generator $\mathcal{A}$ satisfying Assumption 4. Let $0 < h \leq h_0$ for some constant h_0 depending on $\mathcal{M}$ and K. Assume that either K is nonnegative or that $d \geq 4$ and that $Th^{d-2} \geq c\ln(T)$ (in which case, h_0 additionally depends on $p_{\min}$ and on the $\mathcal{C}^1$-norm of p). Then,*

$$\mathbb{E}_\mu\left[\mathcal{W}_2^2(\widehat{\mu}_{T,h}, \mu_h)\right] \leq c_0 \frac{p_{\max}^2}{p_{\min}^2} \|K\|_\infty^2 \begin{cases} \frac{h^{4-d}}{T} & \text{if } d \geq 5 \\ \frac{\ln(1/h)}{T} & \text{if } d = 4 \\ \frac{1}{T} & \text{if } d \leq 3, \end{cases} \tag{3.28}$$

where c_0 depends on $\mathcal{M}$, and c depends on $\mathcal{M}$, K, $p_{\min}$, $p_{\max}$ and $\kappa_{\min}$.

The main difference between Proposition 3.4.1 and Theorem 3.2.2 is the choice of the initial measure, which is of the form δ_x for Theorem 3.2.2 and is equal to μ in Proposition 3.4.1. As for any distribution μ_0 and random variable U, $\mathbb{E}_{\mu_0}[U] = \int \mathbb{E}_x[U]\mathrm{d}\mu_0(x)$, Theorem 3.2.2 is stronger, and implies that the convergence holds for *any* initial measure μ_0.
The remainder of this section is dedicated to proving Proposition 3.4.1. We first state some useful properties on K_h. The following lemma is stated in [37, Lemma 10] under stronger regularity hypotheses on the kernel K. We can relax these assumptions and a proof is given in Appendix 3.8.1. The first point of the lemma guarantees that K_h is well defined on $\mathcal{M}^2$ for h small enough.

Lemma 3.4.2. *Assume that K is a continuous function on $\mathbb{R}$ with support in $[0,1]$, such that $\int_{\mathbb{R}^d} K(\|z\|)\mathrm{d}z = 1$. Let $h>0$, and consider K_h defined by (3.11) with the renormalizing factor η_h. Then,*

(i) $h^{-d}\eta_h$ converges to 1 uniformly on $\mathcal{M}$;

(ii) $\forall x \in \mathcal{M}, \displaystyle\int_{\mathcal{M}} K_h(x,y)\mathrm{d}y = 1$;

(ii) there exists $h_c > 0$, depending on $\mathcal{M}$ and K, such that for all $h < h_c$, K_h is bounded on $\mathcal{M}\times\mathcal{M}$ with $\|K_h\|_\infty \le 2\|K\|_\infty h^{-d}$;

(iv) if furthermore K is Lipschitz continuous, then for all $h < h_c$, and all $x \in \mathcal{M}$, $y \mapsto K_h(x,y)$ is Lipschitz continuous with constant $2\mathrm{Lip}(K)h^{-d-1}$, where $\mathrm{Lip}(K)$ is the Lipschitz constant of K.

We also require the following elementary convergence result, proved in Appendix 3.8.2.

Lemma 3.4.3. *Let p_h be defined by (3.15) for $h>0$. Under the assumptions of Lemma 3.4.2, when h goes to 0, $(p_h)_{h>0}$ converges to p uniformly on $\mathcal{M}$. Moreover, there exists h_c depending on $\mathcal{M}$, K, $p_{\min}$ and the $\mathcal{C}^1$-norm of p such that for all $0 < h \le h_c$, $\inf_{y\in\mathcal{M}} p_h(y) \ge \frac{p_{\min}}{2}$ and $\sup_{y\in\mathcal{M}} p_h(y) \le 2p_{\max}$.*

Hence, for h small enough, μ_h is indeed a probability measure. Recall that the function K is a signed kernel, so that $\mu_{T,h}$ is a priori a signed measure. We introduce the event $E_{T,h}$ defined by

$$E_{T,h} = \{p_{T,h} \ge 0\}. \tag{3.29}$$

On this event, we have $\widehat{\mu}_{T,h} = \mu_{T,h}$ and we notice that for $0 < h < h_c$,

$$\mathbb{E}_\mu\left[\mathcal{W}_2^2(\widehat{\mu}_{T,h}, \mu_h)\right] \le \mathbb{E}_\mu\left[\mathcal{W}_2^2(\mu_{T,h}, \mu_h)\mathbf{1}_{E_{T,h}}\right] + \mathrm{diam}(\mathcal{M})^2\mathbf{P}_\mu(E_{T,h}^c), \tag{3.30}$$

where $\mathrm{diam}(\mathcal{M})$ is the diameter of $\mathcal{M}$. Of course, when K is nonnegative, $p_{T,h}$ is also nonnegative, so that in that case, the event $E_{T,h}$ is satisfied for all $h>0$.
To prove Proposition 3.4.1, we will need several intermediate results related to the spectral decompositions of the operator Δ, $\mathcal{A}$, and their inverses. We first recall a useful result given by Peyre [96, Corollary 2.3] (see also [106, Section 5.5.2] on the negative Sobolev norm), which links the Wasserstein distance to the inverse Laplace operator. Let us remind that the inverse operator Δ^{-1} is defined on $L_0^2(\mathrm{d}x) = \left\{f \in L^2(\mathrm{d}x) : \int_{\mathcal{M}} f\mathrm{d}x = 0\right\}$ (see Section 3.3.2).

Lemma 3.4.4. *Let $f_1, f_2 \in L^2(\mathrm{d}x)$ be two probability density functions with respect to the volume measure $\mathrm{d}x$, with f_1 lower bounded by some positive constant $f_{\min} > 0$. Then, we have*

$$\mathcal{W}_2^2(f_1\mathrm{d}x, f_2\mathrm{d}x) \le \frac{4}{f_{\min}}\int_{\mathcal{M}} \left|(-\Delta)^{-1/2}(f_1 - f_2)\right|^2\mathrm{d}x. \tag{3.31}$$

Besides, ellipticity yields the following relation between the general operator $\mathcal{A}$ and the Laplace-Beltrami operator Δ.

Lemma 3.4.5. *For any function $f \in L_0^2(\mu)$, we have*

$$\int_{\mathcal{M}} \left|(-\mathcal{A})^{-1/2} f\right|^2 \mathrm{d}\mu \leq \frac{1}{p_{\min}\kappa_{\min}} \int_{\mathcal{M}} \left|(-\Delta)^{-1/2}(fp)\right|^2 \mathrm{d}x.$$

Proof. From Equation (3.8),

$$\int_{\mathcal{M}} \Gamma(\mathcal{A}^{-1}f, \mathcal{A}^{-1}f)\mathrm{d}\mu = \int_{\mathcal{M}} f(-\mathcal{A}^{-1})f\mathrm{d}\mu = \int_{\mathcal{M}} \left|(-\mathcal{A})^{-1/2}f\right|^2 \mathrm{d}\mu.$$

Besides, since $\mathcal{A}^{-1}f \in L^2(\mu)$ and since $\mathcal{C}^1(\mathcal{M})$ is dense in this space,

$$\begin{aligned}
&\sqrt{\int_{\mathcal{M}} \Gamma(\mathcal{A}^{-1}f, \mathcal{A}^{-1}f)\mathrm{d}\mu} \\
&\leq \sup\left\{\int_{\mathcal{M}} \Gamma(-\mathcal{A}^{-1}f, g)\mathrm{d}\mu : g \in \mathcal{C}^1(\mathcal{M}) \text{ such that } \int_{\mathcal{M}} \Gamma(g,g)\mathrm{d}\mu \leq 1\right\} \\
&= \sup\left\{\int_{\mathcal{M}} fg\mathrm{d}\mu : g \in \mathcal{C}^1(\mathcal{M}) \text{ such that } \int_{\mathcal{M}} \Gamma(g,g)\mathrm{d}\mu \leq 1\right\}.
\end{aligned}$$

Then, because the simple fact that the supremum of a given set is always bigger than the supremum on any subset, and using (3.10), we have

$$\begin{aligned}
&\sqrt{\int_{\mathcal{M}} \Gamma(\mathcal{A}^{-1}f, \mathcal{A}^{-1}f)\mathrm{d}\mu} \\
&\leq \sup\left\{\int_{\mathcal{M}} (fp)g\mathrm{d}x : g \in \mathcal{C}^1(\mathcal{M}) \text{ such that } \int_{\mathcal{M}} |\nabla g|^2 p_{\min}\kappa_{\min}\mathrm{d}x \leq 1\right\} \\
&= (p_{\min}\kappa_{\min})^{-1/2} \sup\left\{\int_{\mathcal{M}} (fp)g\mathrm{d}x : g \in \mathcal{C}^1(\mathcal{M}) \text{ such that } \int_{\mathcal{M}} |\nabla g|^2 \mathrm{d}x \leq 1\right\}.
\end{aligned}$$

Besides, using Green's theorem, we have that for all $g \in \mathcal{C}^1(\mathcal{M})$

$$\int_{\mathcal{M}} (fp)g\mathrm{d}x = \int_{\mathcal{M}} g\Delta(\Delta^{-1})(fp)\mathrm{d}x = -\int_{\mathcal{M}} \langle \nabla(\Delta^{-1})(fp), \nabla g\rangle \mathrm{d}x.$$

Hence, by Hölder's inequality, we conclude that:

$$\sqrt{\int_{\mathcal{M}} \Gamma((\mathcal{A}^{-1})f, (\mathcal{A}^{-1})f)\mathrm{d}\mu} \leq (p_{\min}\kappa_{\min})^{-1/2} \sqrt{\int_{\mathcal{M}} \left|(-\Delta)^{-1/2}(fp)\right|^2 \mathrm{d}x}.$$

□

The following standard result is crucial: it bounds the variance of the random variable $\mu_T(f)$ for some function f in terms of the generator $\mathcal{A}$.

Lemma 3.4.6. *Let $(X_t)_{t\geq 0}$ be a diffusion with generator $\mathcal{A}$, starting from its invariant measure μ. We have for any $f \in L_0^2(\mu)$,*

$$\mathbb{E}_\mu\left[\left(\frac{1}{T}\int_0^T f(X_s)\mathrm{d}s\right)^2\right] \leq \frac{2}{T}\int_{\mathcal{M}} \left|(-\mathcal{A})^{-1/2}f\right|^2 \mathrm{d}\mu.$$

Proof. Recall that $(\gamma_i)_{i\geq 0}$ are the eigenvalues of $-\mathcal{A}$, with $0 = \gamma_0 < \gamma_1 \leq \gamma_2 \leq \cdots$, and $(\psi_i)_{i\geq 0}$ their respective eigenfunctions. We also remind the reader that $(P_t)_{t\geq 0}$ denotes the semigroup of a process $(X_t)_{t\geq 0}$ with generator $\mathcal{A}$ (see Section 3.3.3).

Since $f \in L_0^2(\mu)$, we write $f = \sum_{i=1}^{\infty} \alpha_i \psi_i$ with $\alpha_i = \int_{\mathcal{M}} f \psi_i \mathrm{d}\mu$, and $\alpha_0 = 0$. By the Markov property, denoting by $(\mathcal{F}_t)_{t\geq 0}$ the natural filtration of the process $(X_t)_{t\geq 0}$, we have

$$\begin{aligned}\mathbb{E}_\mu\left[\left(\frac{1}{T}\int_0^T f(X_s)\mathrm{d}s\right)^2\right] &= \frac{2}{T^2}\mathbb{E}_\mu\left[\int_0^T \int_s^T f(X_t)f(X_s)\mathrm{d}t\mathrm{d}s\right] \\ &= \frac{2}{T^2}\int_0^T \int_s^T \mathbb{E}_\mu[\mathbb{E}[f(X_t)|\mathcal{F}_s]f(X_s)]\mathrm{d}t\mathrm{d}s \\ &= \frac{2}{T^2}\int_0^T \int_s^T \mathbb{E}_\mu[P_{t-s}f(X_s)f(X_s)]\mathrm{d}t\mathrm{d}s.\end{aligned}$$

By assumption, the distribution of X_t is μ for any $t \geq 0$, and computing the expectation using the link between the semigroup $(P_t)_{t\geq 0}$ and the generator $\mathcal{A}$ of the process, we then obtain

$$\begin{aligned}\mathbb{E}_\mu\left[\left(\frac{1}{T}\int_0^T f(X_s)\mathrm{d}s\right)^2\right] &= \frac{2}{T^2}\sum_{i=1}^{\infty}\alpha_i^2 \int_0^T \int_s^T e^{-\gamma_i(t-s)}\mathrm{d}t\mathrm{d}s \\ &\leq \frac{2}{T}\sum_{i=1}^{\infty}\frac{\alpha_i^2}{\gamma_i}.\end{aligned}$$

The lemma is then proved by definition of $(-\mathcal{A})^{-1/2}$ given by Equation (3.26). □

3.4.1 Estimation of the probability $\mathbf{P}_\mu(E_{T,h}^c)$

Recall the definition of the event $E_{T,h} = \{p_{T,h} \geq 0\}$. In view of (3.30), we need to bound the probability $\mathbf{P}_\mu(E_{T,h}^c)$. If the kernel K is nonnegative, then $p_{T,h} \geq 0$ and $\mathbf{P}_\mu(E_{T,h}^c) = 0$. Otherwise, when K is signed, we will use that

$$\mathbf{P}_\mu(E_{T,h}^c) \leq \mathbf{P}_\mu\left(\inf_{y\in\mathcal{M}} p_{T,h}(y) < p_{\min}/8\right), \tag{3.32}$$

and provide a bound for the right-hand side of the inequality.
From Lemma 3.4.3, and in what follows, we choose $0 < h \leq h_c$ so that $p_{\min}/2 \leq p_h \leq 2p_{\max}$.

Proposition 3.4.7. *Assume that $d \geq 4$. There exist c_0 depending on $\mathcal{M}$, K, $p_{\min}$, $p_{\max}$ and $\kappa_{\min}$, and h_0 depending on $\mathcal{M}$, K, $p_{\min}$ and the $\mathcal{C}^1$-norm of p such that for any $T > 0$ and $y \in \mathcal{M}$, when $0 < h \leq h_0$, we have*

$$\mathbf{P}_\mu(p_{T,h}(y) < p_{\min}/4) \leq \exp\left(-c_0 T h^{d-2}\right).$$

Proof. Let $y \in \mathcal{M}$ be fixed. We first start with a pointwise concentration bound for $p_{T,h}(y)$ around its expectation $p_h(y)$. We apply the Bernstein's bound obtained in [49, Theorem 3.5] to the function $g(x) := -K_h(x,y) + p_h(y)$, with $\Phi(u) = \Psi(u) = u^2/2$ (using the notation of [49]). Note that $\int_{\mathcal{M}} g \mathrm{d}\mu = 0$. Write $a_+ = \max(a, 0)$ for $a \in \mathbb{R}$. Let $M = \|g_+\|_{L^2(\mu)}$ and

$$\sigma^2 = \lim_{T\to+\infty}\frac{1}{T}\mathrm{Var}_\mu\left(\int_0^T g(X_s)\mathrm{d}s\right).$$

Then, using Poincaré's inequality for $\mathcal{A}$ given in Proposition 3.3.3, it holds that

$$\mathbf{P}_\mu(p_{T,h}(y) - p_h(y) < -p_{\min}/4) \leq \exp\left(-\frac{Tp_{\min}^2}{32(\sigma^2 + Mp_{\max}/(4\kappa_{\min}\lambda_1))}\right). \tag{3.33}$$

We first bound M. As $\int_{\mathcal{M}} g \mathrm{d}\mu = 0$, and by assumption on the kernel K, we deduce

$$\begin{aligned} M^2 &\leq \int (K_h(x,y) - p_h(y))^2 \mu(\mathrm{d}x) = \int K_h(x,y)^2 \mu(\mathrm{d}x) - p_h(y)^2 \leqslant \int K_h(x,y)^2 \mu(\mathrm{d}x) \\ &\leq p_{\max} \|K_h\|_\infty^2 \int_{\|x-y\| \leq h} \mathrm{d}x \leq 4 p_{\max} \|K\|_\infty^2 h^{-2d} c_1 h^d, \end{aligned}$$

where we use the fact the geodesic distance is equivalent to the Euclidean distance (see [51, Proposition 2]), Lemma 3.8.2, and Lemma 3.4.2-(*iii*). Hence, $M \leq c_2 \|K\|_\infty \sqrt{p_{\max}} h^{-d/2}$ for h small enough.
We then bound σ^2: according to Lemma 3.4.5 and Lemma 3.4.6, and introducing the Green operator G defined in (3.25),

$$\begin{aligned} \sigma^2 &\leq \frac{2}{p_{\min}\kappa_{\min}} \int_{\mathcal{M}} |(-\Delta)^{-1/2}(gp)|^2 \mathrm{d}x \\ &= \frac{2}{p_{\min}\kappa_{\min}} \int_{\mathcal{M}} gp\, G(gp) \mathrm{d}x \\ &\leq \frac{2}{p_{\min}\kappa_{\min}} \left(p_{\max} \|g\|_\infty \int_{\|x-y\| \leq h} |G(gp)(x)| \mathrm{d}x + 4 p_{\max}^3 \int_{\|x-y\| \geq h} |G(p)(x)| \mathrm{d}x \right), \end{aligned}$$

because $g(x) = p_h(y)$ on $\{\|x-y\| \geq h\}$ and $p_h \leq 2p_{\max}$. It remains to bound the right-hand side. First, according to Lemma 3.4.2 and Lemma 3.4.3, $\|g\|_\infty \leq 2p_{\max} + 2\|K\|_\infty h^{-d} \leq c_3 h^{-d}$ for $h \leq 1$. Second, $|G(gp)(x)| \leq |G(K_h(\cdot, y)p)(x)| + p_h(y)|G(p)(x)|$, with

$$|G(p)(x)| \leq p_{\max} \int_{\mathcal{M}} |G(x,z)| \mathrm{d}z \leq p_{\max} c_4$$

according to Proposition 3.3.1-(*iv*) and Lemma 3.8.2. Then, using again the equivalence between the geodesic and the Euclidean distances, as the function $K_h(\cdot, y)p$ is supported on a geodesic ball $\mathcal{B}(y, c_5 h)$, according to Lemma 3.3.2 and Lemma 3.4.2-(*iii*), for all $x \in \mathcal{B}(y, c_5 h)$,

$$|G(K_h(\cdot, y)p)(x)| \leq \kappa_1 \|K_h(\cdot, y)p\|_\infty h^2 \leq 2\kappa_1 p_{\max} \|K\|_\infty h^{2-d}.$$

In total, as $\int_{\|x-y\| \leq h} \mathrm{d}x \leq c_6 h^d$ according to Lemma 3.8.2 and by assumption $\mathrm{vol}(\mathcal{M}) = 1$, it holds that for $h \leq 1$,

$$\begin{aligned} \sigma^2 &\leq \frac{2p_{\max}^2}{p_{\min}\kappa_{\min}} \left(2c_3 c_5 \left(\kappa_1 \|K\|_\infty h^{2-d} + p_{\max} c_4 \right) + 4 p_{\max}^2 c_4 \right) \\ &\leq c_8 h^{2-d}. \end{aligned}$$

As $d \geq 4$, M is smaller than $c_2 \|K\|_\infty \sqrt{p_{\max}} h^{2-d}$. As $p_h(y) \geq p_{\min}/2$ for $h \leq h_c$, we have $\mathbf{P}_\mu(p_{T,h}(y) < p_{\min}/4) \leq \mathbf{P}_\mu(p_{T,h}(y) - p_h(y) < -p_{\min}/4)$ for such a value of h. We therefore obtain the desired result. □

We then conclude with a standard union bound argument by using a covering of $\mathcal{M}$.

Proposition 3.4.8. *Assume that $d \geq 4$. There exists h_0 depending on $\mathcal{M}$, K, $p_{\min}$ and the $\mathcal{C}^1$-norm of p such that for any $T > 0$ and $0 < h \leq h_0$, we have*

$$\mathbf{P}_\mu \left(\inf_{y \in \mathcal{M}} p_{T,h}(y) < p_{\min}/8 \right) \leq c_1 h^{-d(d+1)} \exp\left(-c_0 T h^{d-2} \right)$$

for some constant c_1 depending only on $p_{\min}$, K and $\mathcal{M}$, where c_0 is the constant of Proposition 3.4.7.

Proof. To have a uniform estimation of $p_{T,h}$, we will make use of the Lipschitz continuity of K and the covering number $N_\delta(\mathcal{M})$ of $\mathcal{M}$, for $\delta > 0$, i.e. the smallest number N such that there exists a subset E of N distinct points of $\mathcal{M}$ such that $\max_{y\in\mathcal{M}} \min_{x\in E} \rho(x, y) \le \delta$.
By Lemma 3.4.2, the function $p_{T,h}$ is Lipschitz continuous with constant $2\text{Lip}(K)h^{-d-1}$ as an average of Lipschitz continuous functions. Let $\delta > 0$. We consider the covering number $N_\delta(\mathcal{M})$ of $\mathcal{M}$. By [60, 2.2A], there are constants $c_\mathcal{M}$ and $\delta_\mathcal{M}$ depending only on $\mathcal{M}$ such that for all $0 < \delta \le \delta_\mathcal{M}$,

$$N_\delta(\mathcal{M}) \le c_\mathcal{M}\delta^{-d}. \tag{3.34}$$

Consequently, for $h \le h_0$ (where h_0 is the constant of Proposition 3.4.7), if $y_1, \dots, y_{N_\delta}$ is a minimal δ-covering of $\mathcal{M}$,

$$\begin{aligned}\mathbf{P}_\mu\Big(\inf_{y\in\mathcal{M}} p_{T,h}(y) < p_{\min}/8\Big) &\le \sum_{i=1}^{N_\delta(\mathcal{M})} \mathbf{P}_\mu(\exists y \in \mathcal{B}(y_i,\delta) : p_{T,h}(y) < p_{\min}/8)\\ &\le \sum_{i=1}^{N_\delta(\mathcal{M})} \mathbf{P}_\mu\Big(p_{T,h}(y_i) < p_{\min}/8 + 2\text{Lip}(K)h^{-d-1}\delta\Big).\end{aligned}$$

Choose $\delta = \frac{p_{\min}h^{d+1}}{16\text{Lip}(K)}$, which is smaller than $\delta_\mathcal{M}$ as long as h is small enough with respect to $\mathcal{M}$ and K, as $p_{\min} \le 1$. By Proposition 3.4.7 and Equation (3.34), we easily deduce

$$\mathbf{P}_\mu(\inf_{y\in\mathcal{M}} p_{T,h}(y) < p_{\min}/8) \le c_{K,\mathcal{M}} p_{\min}^{-d} h^{-d(d+1)} \exp\Big(-c_0 T h^{d-2}\Big),$$

and the result is proved. □

3.4.2 Proof of Proposition 3.4.1

We first give a last useful result related to the diffusion $(X_t)_{t\ge0}$ with generator $\mathcal{A}$ in relation with the operators.

Notation. Given a space E, for any function $f : E \times E \to \mathbb{R}$ and any operator $\mathcal{J} : D \subset \mathbb{R}^E \to \mathbb{R}^E$, we define $\mathcal{J}_1 f$ when the operator is applied to the first variable of f and $\mathcal{J}_2 f$ when it is applied to the second variable of f:

$$\mathcal{J}_1 f(x, y) := (\mathcal{J} f(., y))(x) \quad \text{and} \quad \mathcal{J}_2 f(x, y) := (\mathcal{J} f(x, .))(y).$$

Proposition 3.4.9. *Let $R \in L^2(\mu \otimes \mathrm{d}y)$ be a function such that for all $(x, y) \in \mathcal{M}^2$, $R(x,\cdot) \in L^2_0(\mathrm{d}y)$ and $R(\cdot, y) \in L^2_0(\mu)$. Then, when the initial distribution of the diffusion $(X_t)_{t\ge0}$ is its invariant measure μ, we have*

$$\mathbb{E}_\mu\left[\int_\mathcal{M} \left|(-\Delta)^{-1/2}\Big(\frac{1}{T}\int_0^T R(X_s,\cdot)\mathrm{d}s\Big)\right|^2 \mathrm{d}y\right] \le \frac{2}{T}\iint_{\mathcal{M}^2} \left|(-\mathcal{A})_1^{-1/2}(-\Delta)_2^{-1/2} R\right|^2 \mathrm{d}\mu\mathrm{d}y.$$

Proof. Denote by $\widehat{R}$ the function $(-\Delta)_2^{-1/2}R$. We observe that $\widehat{R} \in L^2(\mu\otimes \mathrm{d}y)$ and for any y, $\widehat{R}(\cdot, y) \in L^2_0(\mu)$. Hence, by applying Lemma 3.4.6, we have the following sequence of equalities:

$$\begin{aligned}&\mathbb{E}_\mu\left[\int_\mathcal{M} \left|(-\Delta)^{-1/2}\Big(\frac{1}{T}\int_0^T R(X_s,\cdot)\mathrm{d}s\Big)\right|^2 \mathrm{d}y\right] = \mathbb{E}_\mu\left[\int_\mathcal{M} \left|\frac{1}{T}\int_0^T (-\Delta)^{-1/2}R(X_s,\cdot)\mathrm{d}s\right|^2 \mathrm{d}y\right]\\ &= \int_\mathcal{M} \mathbb{E}_\mu\left[\left|\frac{1}{T}\int_0^T \widehat{R}(X_s, y)\mathrm{d}s\right|^2\right]\mathrm{d}y \le \frac{2}{T}\int_\mathcal{M}\Big(\int_\mathcal{M}\left|(-\mathcal{A})_1^{-1/2}\widehat{R}\right|^2\mathrm{d}\mu\Big)\mathrm{d}y.\end{aligned}$$

Therefore, the proposition is proved. □

Using the decomposition (3.30), we can now prove Proposition 3.4.1. We first use the estimate of the probability of $E^c_{T,h}$ given in Proposition 3.4.8, and then give an explicit estimate of the variance term on the event $E_{T,h}$ for a diffusion $(X_t)_{t\geq 0}$ with generator $\mathcal{A}$ starting from its invariant measure μ.

Proof of Proposition 3.4.1.
According to (3.32) and Proposition 3.4.8, the probability of the event $E^c_{T,h}$ is negligible. Indeed, when $Th^{d-2} > \frac{d^2+4}{c_0}\ln(T)$ and $d \geq 4$, for T large enough the second term in the decomposition (3.30) is smaller than $\frac{p_{\max}^2}{p_{\min}^2}\|K\|_\infty^2 h^{4-d}T^{-1}$. Furthermore, it is equal to 0 when K is nonnegative. It remains to bound the first term.
On the event $E_{T,h}$, both μ_h and $\mu_{T,h}$ are probability measures with respective density functions $p_{T,h}$ and p_h. Furthermore, for $h < h_c$, $p_h \geq p_{\min}/2$. Hence, by Lemma 3.4.4, we have:

$$\mathcal{W}_2^2(p_{T,h}, p_h) \leq \frac{8}{p_{\min}} \int_{\mathcal{M}} \left|(-\Delta)^{-1/2}(p_{T,h} - p_h)\right|^2 \mathrm{d}x. \tag{3.35}$$

Now, consider the function $R_h(x,y) = K_h(x,y) - \int_{\mathcal{M}} K_h(z,y)\mathrm{d}\mu(z)$. First, we observe that

$$p_{T,h}(y) - p_h(y) = \frac{1}{T}\int_0^T R_h(X_s, y)\mathrm{d}s.$$

As R_h is continuous on a compact manifold $\mathcal{M}$, we have $R_h \in L^2(\mu \otimes \mathrm{d}y)$. Thus, for each $x, y \in \mathcal{M}$, $R_h(x,\cdot) \in L_0^2(\mathrm{d}y)$ and $R_h(\cdot, y) \in L_0^2(\mu)$. Therefore, due to Proposition 3.4.9,

$$\mathbb{E}_\mu\left[\int_{\mathcal{M}} \left|(-\Delta)^{-1/2}(p_{T,h} - p_h)\right|^2 \mathrm{d}y\right] \leq \frac{2}{T}\iint_{\mathcal{M}^2} \left|(-\mathcal{A})_1^{-1/2}(-\Delta)_2^{-1/2} R_h\right|^2 \mathrm{d}\mu\mathrm{d}y. \tag{3.36}$$

Besides, by Lemma 3.4.5, we have:

$$\iint_{\mathcal{M}^2} \left|(-\mathcal{A})_1^{-1/2}(-\Delta)_2^{-1/2} R_h\right|^2 \mathrm{d}\mu\mathrm{d}y \leq \frac{1}{p_{\min}\kappa_{\min}} \iint_{\mathcal{M}^2} \left|(-\Delta)_1^{-1/2}(M_p)_1(-\Delta)_2^{-1/2} R_h\right|^2 \mathrm{d}x\mathrm{d}y, \tag{3.37}$$

where $M_p : L_0^2(\mu) \to L_0^2(\mathrm{d}x)$ is the bounded multiplication operator $f \mapsto pf$.
Therefore, after Inequalities (3.35), (3.36), (3.37), we have:

$$\mathbb{E}_\mu\left[\mathcal{W}_2(\mu_{T,h}, \mu_h)^2 \mathbf{1}_{E_{T,h}}\right] \leq \frac{8}{p_{\min}^2\kappa_{\min}T} \iint_{\mathcal{M}^2} \left|(-\Delta)_1^{-1/2}(M_p)_1(-\Delta)_2^{-1/2} R_h\right|^2 \mathrm{d}x\mathrm{d}y.$$

Note that M_p and $\Delta^{-1/2}$ are bounded operators, which implies $(M_p)_1$ and $(\Delta^{-1/2})_2$ are commutative. In other words, $(M_p)_1(-\Delta)_2^{-1/2} R_h = (-\Delta)_2^{-1/2}(M_p)_1 R_h$, which means

$$\mathbb{E}_\mu\left[\mathcal{W}_2^2(\mu_{T,h}, \mu_h)\mathbf{1}_{E_{T,h}}\right] \leq \frac{8}{p_{\min}^2\kappa_{\min}T} \iint_{\mathcal{M}^2} \left|(-\Delta)_1^{-1/2}(-\Delta)_2^{-1/2} S_h\right|^2 \mathrm{d}x\mathrm{d}y,$$

where $S_h(x,y) = p(x)R_h(x,y) \in L^2(\mathrm{d}x \otimes \mathrm{d}y)$.
Recall that $(\lambda_i)_{i\geq 0}$ and $(\phi_i)_{i\geq 0}$ are respectively the eigenvalues and the eigenfunctions of $(-\Delta)$. As p is upper bounded, $\forall x \in \mathcal{M}$ $S_h(x,.) \in L_0^2(\mathrm{d}y)$, $\forall y \in \mathcal{M}$ $S_h(.,y) \in L_0^2(\mathrm{d}x)$ and $S_h \in L_0^2(\mathrm{d}x \otimes \mathrm{d}y)$. Thus, there are $(\alpha_{i,j}(h))_{i,j\geq 0}$ such that S_h has the following decomposition (with $\alpha_{i,0} = \alpha_{0,j} = 0$): for all $x, y \in \mathcal{M}$

$$S_h(x,y) = \sum_{i,j\geq 1} \alpha_{ij}(h)\phi_i(x)\phi_j(y),$$

with $\alpha_{ij}(h) = \iint_{\mathcal{M}^2} S_h(x,y)\phi_i(x)\phi_j(y)\mathrm{d}x\mathrm{d}y$. Consequently, for all $x, y \in \mathcal{M}$

$$(-\Delta)_1^{-1/2}(-\Delta)_2^{-1/2} S_h(x,y) = \sum_{i,j\geq 1} \frac{\alpha_{ij}(h)}{\sqrt{\lambda_i\lambda_j}}\phi_i(x)\phi_j(y),$$
$$(-\Delta)_1^{-1} S_h(x,y) = \sum_{i,j\geq 1} \frac{\alpha_{ij}(h)}{\lambda_i}\phi_i(x)\phi_j(y),$$
$$(-\Delta)_2^{-1} S_h(x,y) = \sum_{i,j\geq 1} \frac{\alpha_{ij}(h)}{\lambda_j}\phi_i(x)\phi_j(y).$$

Therefore,

$$\iint_{\mathcal{M}^2} \left|(-\Delta)_1^{-1/2}(-\Delta)_2^{-1/2} S_h\right|^2 \mathrm{d}x\mathrm{d}y = \sum_{i,j\geq 1} \frac{\alpha_{ij}^2(h)}{\lambda_i\lambda_j}$$
$$= \iint_{\mathcal{M}^2} (\Delta_1^{-1}S_h)(\Delta_2^{-1}S_h)\mathrm{d}x\mathrm{d}y = \iint_{\mathcal{M}^2} (G_1S_h)(x,y)(G_2S_h)(x,y)\mathrm{d}x\mathrm{d}y,$$

where G is the Green function of Δ introduced in Section 3.3.2.

Let $d \geq 5$. Using the fact that K has compact support and that the geodesic and Euclidean distances are equivalent, the functions $x \mapsto K_h(x,y)$ and $y \mapsto K_h(x,y)$ are supported on a geodesic ball of radius $c_1 h$ for some $c_1 > 0$ depending only on $\mathcal{M}$. Hence, Lemma 3.3.2 implies that for h sufficiently small and for all $x \in \mathcal{M}$,

$$\begin{aligned}
&\int_{\mathcal{M}} |G_1S_h(x,y)G_2S_h(x,y)|\mathrm{d}y \\
&\leq c_2\|S_h\|_\infty^2 \left(h^4 \int_{\mathcal{B}(x,2c_1h)} \mathrm{d}y + h^{2d}\int_{\mathcal{M}\setminus\mathcal{B}(x,2c_1h)} \rho(x,y)^{4-2d}\mathrm{d}y\right) \\
&\leq 4c_2 p_{\max}^2 \|K_h\|_\infty^2 \left(h^4 \int_{\mathcal{B}(x,2h)} \mathrm{d}y + h^{2d}\int_{\mathcal{M}\setminus\mathcal{B}(x,2h)} \rho(x,y)^{4-2d}\mathrm{d}y\right) \\
&\leq c_3 p_{\max}^2 \|K\|_\infty^2 h^{-2d}\left(h^{4+d} + h^{2d}h^{4-d}\right) \\
&\leq 2c_3 p_{\max}^2 \|K\|_\infty^2 h^{-2d} \times h^{4+d},
\end{aligned}$$

where we also use Lemma 3.8.2 and Lemma 3.4.2. This proves the proposition when $d \geq 5$. The computations for $d \leq 4$ are similar, and left to the reader. □

3.5 Transition to a general initial measure

In the previous section, we have obtained a control of the variance term $\mathcal{W}_2^2(\widehat{\mu}_{T,h}, \mu_h)$ when the stochastic process $(X_t)_{t\geq 0}$ starts from its invariant measure μ. In this section, we explain how to extend the result to an initial measure of type Dirac measure δ_x (which will imply the result for any initial measure). The main idea is to use the ultracontractivity of the diffusion $(X_t)_{t\geq 0}$. Let us first introduce this notion.

Lemma 3.5.1. *[120, Theorem 3.5.5.] The semigroup (P_t) associated to the operator $\mathcal{A}$ is ultracontractive. In other words, for each $t > 0$, there is a minimal positive value $c_t > 0$, such that for any bounded measurable function f, we have*

$$\|P_t f\|_\infty \leq c_t \|f\|_{L^1(\mu)}. \tag{3.38}$$

In Section 3.8.5, an explicit form of the ultracontractivity term c_t is given. We denote by $u_{\mathcal{A}} = c_1$ the ultracontractivity constant at time $t = 1$.

Remark 3.5.2. *Wang only considers operators of the form $\Delta + \nabla p$ in [120]. However, as explained in Remark 3.3.4 (see Section 3.8.5 for more details), for any $\mathcal{C}^2$ second order elliptic differential operator $\mathcal{A}$ on a smooth manifold $\mathcal{M}$ such that $\mathcal{A}$ is symmetric with respect to the measure $\mu(\mathrm{d}x) = p(x)\mathrm{d}x$, there is always a $\mathcal{C}^2$-Riemannian metric $\tilde{\mathbf{g}}$ on $\mathcal{M}$ such that $\mathcal{A} = \tilde{\Delta} + \tilde{\nabla} p$, where $\tilde{\Delta}$ and $\tilde{\nabla}$ are respectively the Laplacian and the gradient operator of $(\mathcal{M}, \tilde{\mathbf{g}})$. Hence, [120, Theorem 3.5.5.] can readily be applied.*

Remark 3.5.3. *In [120], Wang actually defines the ultra-contractivity of a semigroup $(P_t)_{t\geq 0}$ in a slightly different way: for any $t > 0$, there should exist $c_t > 0$ such that for any measurable bounded function f, $\|P_t f\|_\infty \leq c_t \|f\|_{L^2(\mu)}$. However, Wang's definition implies (3.38) in our setting with $\mathcal{M}$ compact. Assume that $(P_t)_{t\geq 0}$ is ultra-contractive in Wang's sense, then for $t > 0$ and for any measurable bounded function f,*

$$\|P_t f\|_\infty \leq c_{t/2} \big\|P_{t/2} f\big\|_{L^2(\mu)} \leq \left(c_{t/2}\right)^2 \|f\|_{L^1(\mu)}$$

because $\|T\|_{2\to\infty} = \|T\|_{1\to 2}$ for a symmetric operator T.

We now use the ultracontractivity constant $u_{\mathcal{A}}$ to control the distance $\mathbb{E}_x\left[\mathcal{W}_2^2(\widehat{\mu}_{T,h}, \mu)\right]$.

Proposition 3.5.4. *Let $x \in \mathcal{M}$. There exist constants c, h_0 depending only on $\mathcal{M}$, and on $\mathcal{M}$ and K respectively such that for any $x \in \mathcal{M}$ and any $T > 1$, and any $h \leq h_0$, we have that:*

$$\begin{aligned}\mathbb{E}_x\left[\mathcal{W}_2^2(\widehat{\mu}_{T,h}, \mu)\right] \leq c\frac{\|K\|_\infty}{T} + 2\mathrm{diam}(\mathcal{M})^2 \mathbf{P}_x(E_{T,h}^c) + 2u_{\mathcal{A}}\mathrm{diam}(\mathcal{M})^2 \mathbf{P}_\mu(E_{T,h}^c)\\ + 2u_{\mathcal{A}}\mathbb{E}_\mu\big[\mathcal{W}_2^2(\widehat{\mu}_{T,h}, \mu)\big].\end{aligned}$$

Proof. Consider functions $F(x) = \mathbb{E}_x\big[\mathcal{W}_2^2(\widehat{\mu}_{T,h}, \mu)\big]$ and $H(x) = \mathbf{P}_x(E_{T,h}^c)$. Note that the shifted process $(\widetilde{X}_t)_{t\geq 0} = (X_{t+1})_{t\geq 0}$ is still a diffusion process with generator $\mathcal{A}$. Therefore, if we consider the following shifted quantities

$$\begin{aligned}\widetilde{p}_{T,h}(y) &= \frac{1}{T}\int_0^T K_h(X_{s+1}, y)\mathrm{d}s,\\ \widetilde{\mu}_{T,h} &= \begin{cases}\widetilde{p}_{T,h}\mathrm{d}x & \text{if } \widetilde{p}_{T,h}\mathrm{d}x \text{ is positive measure}\\ \delta_{x_0} & \text{otherwise,}\end{cases}\\ \widetilde{E}_{T,h} &= \{\widetilde{p}_{T,h} \geq 0\},\end{aligned}$$

we have

$$\begin{aligned}\mathbb{E}_x\left[\mathcal{W}_2^2(\widetilde{\mu}_{T,h}, \mu)\right] &= \mathbb{E}_x\left[\mathbb{E}_{X_1}\big[\mathcal{W}_2^2(\widehat{\mu}_{T,h}, \mu)\big]\right] = \mathbb{E}_x[F(X_1)] = P_1 F(x),\\ \mathbf{P}_x(\widetilde{E}_{T,h}^c) &= \mathbb{E}_x\big[\mathbf{P}_x(E_{T,h}^c)\big] = \mathbb{E}_x[H(X_1)] = P_1 H(x).\end{aligned}$$

Thus, by ultracontractivity of the process $(X_t)_{t\geq 0}$, at time $t = 1$,

$$\mathbb{E}_x\left[\mathcal{W}_2^2(\widetilde{\mu}_{T,h}, \mu)\right] \leq \|P_1 F\|_\infty \leq u_{\mathcal{A}} \int_{\mathcal{M}} F(y)\mu(\mathrm{d}y) = u_{\mathcal{A}}\mathbb{E}_\mu\big[\mathcal{W}_2^2(\widehat{\mu}_{T,h}, \mu)\big], \tag{3.39}$$

$$\mathbf{P}_x(\widetilde{E}_{T,h}^c) \leq u_{\mathcal{A}}\mathbf{P}_\mu(E_{T,h}^c), \tag{3.40}$$

where $u_{\mathcal{A}}$ has been defined after Lemma 3.5.1.

Recall that the Wasserstein distance is controlled by the total variation distance [119, Theorem 6.15]. Besides, on the event $E_{T,h} \cap \widetilde{E}_{T,h}$, $p_{T,h}\mathrm{d}x$ and $\widetilde{p}_{T,h}\mathrm{d}x$ are both positive measures. Hence, on $E_{T,h} \cap \widetilde{E}_{T,h}$, by Lemma 3.4.2, when h is sufficiently small depending on $\mathcal{M}$ and K,

$$\begin{aligned}
&\mathcal{W}_2^2(\widehat{\mu}_{T,h}, \widetilde{\mu}_{T,h}) \leq 2\mathrm{diam}(\mathcal{M})^2 \int_{\mathcal{M}} |p_{T,h}(y) - \widetilde{p}_{T,h}(y)|\mathrm{d}y \\
&\leq 2\mathrm{diam}(\mathcal{M})^2 \frac{1}{T} \int_{\mathcal{M}} \Big(\int_0^1 |K_h(X_s,y)|\mathrm{d}s + \int_T^{T+1} |K_h(X_s,y)|\mathrm{d}s \Big)\mathrm{d}y \\
&= 2\mathrm{diam}(\mathcal{M})^2 \frac{1}{T} \Big(\int_0^1 \Big(\int_{\mathcal{M}} |K_h(X_s,y)|\mathrm{d}y \Big)\mathrm{d}s + \int_T^{T+1} \Big(\int_{\mathcal{M}} |K_h(X_s,y)|\mathrm{d}y \Big)\mathrm{d}s \Big) \\
&\leq 2\mathrm{diam}(\mathcal{M})^2 \frac{1}{T} 2\|K\|_\infty h^{-d} \Big(\int_0^1 \int_{\mathcal{M}} \mathbf{1}_{\|X_s - y\| \leq h}\mathrm{d}y + \int_T^{T+1} \int_{\mathcal{M}} \mathbf{1}_{\|X_s - y\| \leq h}\mathrm{d}y\mathrm{d}s \Big) \\
&\leq c_0 \frac{\|K\|_\infty}{T},
\end{aligned}$$

where c_0 is a constant depending on $\mathcal{M}$ and we used Lemma 3.8.2 for the last inequality and the equivalence between the Euclidean and geodesic distances.
Thus, by (3.40), for h sufficiently small,

$$\begin{aligned}
\mathbb{E}_x\big[W_2^2(\widehat{\mu}_{T,h}, \widetilde{\mu}_{T,h})\big] &\leq c_0 \frac{\|K\|_\infty}{T} + \mathrm{diam}(\mathcal{M})^2 \mathbf{P}_x\Big(E_{T,h}^c \cup \widetilde{E}_{T,h}^c\Big) \\
&\leq c_0 \frac{\|K\|_\infty}{T} + \mathrm{diam}(\mathcal{M})^2 \mathbf{P}_x(E_{T,h}^c) + u_{\mathcal{A}} \mathrm{diam}(\mathcal{M})^2 \mathbf{P}_\mu(E_{T,h}^c). \qquad (3.41)
\end{aligned}$$

Consequently, by triangular inequality, (3.39), and (3.41), we deduce that there exists a constant c depending only on $\mathcal{M}$ such that for h sufficiently small,

$$\begin{aligned}
&\frac{1}{2}\mathbb{E}_x\big[\mathcal{W}_2^2(\widehat{\mu}_{T,h}, \mu)\big] \leq \mathbb{E}_x\big[\mathcal{W}_2^2(\widehat{\mu}_{T,h}, \widetilde{\mu}_{T,h})\big] + \mathbb{E}_x\big[\mathcal{W}_2^2(\widetilde{\mu}_{T,h}, \mu)\big] \\
&\leq c_0 \frac{\|K\|_\infty}{T} + \mathrm{diam}(\mathcal{M})^2 \mathbf{P}_x(E_{T,h}^c) + u_{\mathcal{A}} \mathrm{diam}(\mathcal{M})^2 \mathbf{P}_\mu(E_{T,h}^c) + u_{\mathcal{A}} \mathbb{E}_\mu\big[\mathcal{W}_2^2(\widehat{\mu}_{T,h}, \mu)\big],
\end{aligned}$$

which is the desired conclusion. □

Hence, using Proposition 3.4.1 and Proposition 3.4.8, it only remains to bound $\mathbf{P}_x(E_{T,h}^c)$. This probability is 0 if K is nonnegative. Otherwise we assume that $d \geq 4$ and that $Th^d \geq c_0$ for a constant c_0 to fix.

Lemma 3.5.5. *Assume that $d \geq 4$. Let h_0, c_0 and c_1 be the constants of Proposition 3.4.8. Then, for any $T \geq 2$, if $0 < h \leq h_0$ and $Th^d > 32\|K\|_\infty / p_{\min}$, we have*

$$\mathbf{P}_x\big(E_{T,h}^c\big) \leq u_{\mathcal{A}} c_1 h^{-d(d+1)} \exp\Big(-\frac{c_0}{2} T h^{d-2}\Big).$$

Proof. We have the following observation for all $y \in \mathcal{M}$ and $T \geq 2$:

$$\begin{aligned}
p_{T,h}(y) =& \frac{1}{T} \int_0^1 K_h(X_s, y)\mathrm{d}s + \frac{T-1}{T} \times \frac{1}{T-1} \int_0^{T-1} K_h(X_{s+1}, y)\mathrm{d}s \\
\geq& -\frac{1}{T} \int_0^1 \|K_h\|_\infty \mathrm{d}s + \frac{1}{2(T-1)} \int_0^{T-1} K_h(X_{s+1}, y)\mathrm{d}s.
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathbf{P}_x\Big(\inf_{y\in\mathcal{M}} p_{T,h}(y) < 0\Big) &\leq \mathbf{P}_x\Big(\frac{1}{2} \inf_{y\in\mathcal{M}} \Big(\frac{1}{T-1} \int_0^{T-1} K_h(X_{s+1}, y)\mathrm{d}s\Big) < \frac{\|K_h\|_\infty}{T}\Big) \\
&= \mathbb{E}_x\Big(\mathbf{P}_{X_1}\Big(\inf_{y\in\mathcal{M}} p_{(T-1),h}(y) < \frac{2\|K_h\|_\infty}{T}\Big)\Big) \\
&\leq u_{\mathcal{A}} \mathbf{P}_\mu\Big(\inf_{y\in\mathcal{M}} p_{(T-1),h}(y) < \frac{2\|K_h\|_\infty}{T}\Big),
\end{aligned}$$

where we used ultracontractivity at the last step. By Lemma 3.4.2, we have $\|K_h\|_\infty \leq 2\|K\|_\infty h^{-d}$ for h small enough. Hence, if $Th^d > 32\|K\|_\infty/p_{\min}$, then $\frac{2\|K_h\|_\infty}{T} < p_{\min}/8$, and we obtain thanks to Proposition 3.4.8 that

$$\begin{aligned}\mathbf{P}_x\big(E^c_{T,h}\big) &\leq u_{\mathcal{A}}\mathbf{P}_\mu\bigg(\inf_{y\in\mathcal{M}} p_{(T-1),h}(y) < p_{\min}/8\bigg)\\ &\leq u_{\mathcal{A}}c_1 h^{-d(d+1)}\exp(-c_0(T-1)h^{d-2}).\end{aligned}$$

As $T-1 \geq T/2$, for $T \geq 2$, the result holds. □

We put together all the estimations obtained so far to conclude. Consider $d \geq 4$ and $Th^d \geq c_0'$, where the constant c_0' is chosen so that $Th^d > 32\|K\|_\infty/p_{\min}$ and Th^{d-2} is large enough with respect to $\ln(T)$, so that the upper bound on $\mathbf{P}_x\Big(E^c_{T,h}\Big)$ given in Lemma 3.5.5 is negligible. The proof is similar when $d < 4$ and K is nonnegative. By Proposition 3.5.4, Proposition 3.4.1, Lemma 3.5.5 and Proposition 3.4.8, for any $x \in \mathcal{M}$, for h sufficiently small, and $T \geq 2$, there are constants $c, \tilde{c} > 0$ such that

$$\begin{aligned}&\mathbb{E}_x\left[\mathcal{W}_2^2(\widehat{\mu}_{T,h},\mu)\right]\\ &\leq c\frac{\|K\|_\infty}{T} + 2\mathrm{diam}(\mathcal{M})^2\mathbf{P}_x(E^c_{T,h}) + 2u_{\mathcal{A}}\mathrm{diam}(\mathcal{M})^2\mathbf{P}_\mu(E^c_{T,h}) + cu_{\mathcal{A}}\frac{p^2_{\max}}{p^2_{\min}}\|K\|^2_\infty\frac{h^{4-d}}{T}\\ &\leq \tilde{c}\,u_{\mathcal{A}}\frac{p^2_{\max}}{p^2_{\min}}\|K\|^2_\infty\frac{h^{4-d}}{T}.\end{aligned}$$

3.6 Minimax lower bound

The proof of the minimax lower bound (Proposition 3.2.5) relies crucially on the computation of the Kullback-Leibler divergence between the law of two diffusion processes (X_t) and (X_t') on $\mathcal{M}$. Recall that for two probability measures $\mathbf{P}$ and $\mathbf{Q}$, the Kullback-Leibler divergence $\mathrm{KL}(\mathbf{P}\|\mathbf{Q})$ is defined as

$$\mathrm{KL}(\mathbf{P}\|\mathbf{Q}) = \int \ln\left(\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{Q}}\right)\mathrm{d}\mathbf{P} \tag{3.42}$$

whenever $\mathbf{P}$ is absolutely continuous with respect to $\mathbf{Q}$.

Recall that we have denoted, in Section 3.2, by $\mathcal{P}_{T,\ell} = \mathcal{P}_{T,\ell}(\kappa_{\min}, p_{\min}, p_{\max}, u_{\max}, L)$ the class of all the laws of diffusion paths $(X_t)_{t\in[0,T]}$ on $\mathcal{M}$ with arbitrary initial distribution, generator $\mathcal{A}$ satisfying Assumption 4 with constant $\kappa_{\min}$ and having an ultracontractivity constant smaller than $u_{\max}$, and invariant measure μ having a $\mathcal{C}^2$ positive p density on $\mathcal{M}$ with a Sobolev norm $\|p\|_{H^\ell(\mathcal{M})} \leq L$ and a $\mathcal{C}^1$-norm smaller than L, satisfying $p_{\min} \leq p \leq p_{\max}$.

Our minimax lower bound follows from an application of Assouad's lemma, see [116, Theorem 2.12].

Lemma 3.6.1 (Assouad's lemma)**.** *Consider a statistical model $\mathcal{P}$. Let $J > 0$ be an integer and consider a subfamily $\{\mathbf{P}_\tau\}_\tau \subset \mathcal{P}$ indexed by $\tau \in \{\pm1\}^J$. Define the Hamming distance $d_H(\tau,\tau') = \sum_{j=1}^J \mathbf{1}_{\{\tau_j\neq\tau_j'\}}$ on the hypercube $\{\pm1\}^J$. Assume that there exists A such that for every $\tau,\tau' \in \{\pm1\}^J$, $\mathcal{W}_2(\mu(\mathbf{P}_\tau),\mu(\mathbf{P}_{\tau'})) \geq Ad_H(\tau,\tau')$ and that whenever $d_H(\tau,\tau') = 1$,* $\mathrm{KL}(\mathbf{P}_\tau\|\mathbf{P}_{\tau'}) \leq 1/2$. *Then, the minimax risk over $\mathcal{P}$ defined in* (3.21) *satisfies*

$$\mathcal{R}(\mathcal{P}) \geq \frac{AJ}{4}. \tag{3.43}$$

To apply Assouad's lemma, we build an appropriate family $\{\mu_\tau\}$ of probability measures on $\mathcal{M}$ indexed by the hypercube $\{\pm1\}^J$ by perturbating the uniform measure by small bumps. We let p_τ be the density of μ_τ and $\mathbf{P}_\tau$ be the law of a diffusion path $(X_t)_{t\in[0,T]}$ with initial distribution μ_τ and generator $\mathcal{A}_{p_\tau q}$ (defined in (3.1)) with $q \equiv 1$. In that case, the associated carré du champ is equal to $\Gamma(f,f) = |\nabla f|^2$, so that we can take $\kappa_{\min} = 1$.

Lemma 3.6.2 (Existence of bumps). *Let $\ell \geq 0$ be an integer. There exist constants κ, $\varepsilon_0 > 0$ depending only on ℓ and $\mathcal{M}$ such that for all $x \in \mathcal{M}$ and all $0 \leq \varepsilon \leq \varepsilon_0$, there exists a smooth function $\phi_{x,\varepsilon} : \mathcal{M} \to \mathbb{R}$ supported on $\mathcal{B}(x,\varepsilon)$, with $\int_{\mathcal{M}} \phi_x \mathrm{d}y = 0$, $\int_{\mathcal{M}} \phi_x^2 \mathrm{d}y = \varepsilon^d$ and for all integers $0 \leq i \leq \ell$,*

$$\|\phi_{x,\varepsilon}\|_{\mathcal{C}^i(\mathcal{M})} \leq \kappa\varepsilon^{-i}. \tag{3.44}$$

Proof. Fix $x \in \mathcal{M}$. Consider a chart Ψ around x. For ε small enough, the chart $\Psi : \mathcal{B}_{\mathbb{R}^d}(0,\varepsilon/2) \to \mathcal{M}$ is a diffeomorphism whose image is contained in $\mathcal{B}(x,\varepsilon)$. Let $\phi_0 : \mathbb{R}^d \to [-1,1]$ be a smooth function of integral 0, with support included in $\mathcal{B}_{\mathbb{R}^d}(0,1/2)$, equal to 1 on $\mathcal{B}_{\mathbb{R}^d}(0,1/6)$. Let $J\Psi : \mathcal{B}_{\mathbb{R}^d}(0,\varepsilon/2) \to \mathbb{R}$ be the Jacobian of Ψ. We define for $u \in \mathcal{B}_{\mathbb{R}^d}(0,\varepsilon/2)$ and $y = \Psi(u)$

$$\phi_{x,\varepsilon}(y) = C_{x,\varepsilon}\frac{\phi_0(u/\varepsilon)}{J\Psi(u)},$$

and $\phi_{x,\varepsilon}$ is 0 outside $\Psi(\mathcal{B}_{\mathbb{R}^d}(0,\varepsilon/2))$. By construction, $\int_{\mathcal{M}} \phi_{x,\varepsilon}\mathrm{d}y = C_{x,\varepsilon}\int_{\mathcal{B}_{\mathbb{R}^d}(0,\varepsilon/2)} \phi_0(u/\varepsilon)\mathrm{d}u = 0$. We choose $C_{x,\varepsilon}$ so that $\int_{\mathcal{M}} \phi_{x,\varepsilon}^2 \mathrm{d}y = \varepsilon^d$. It remains to bound the $\mathcal{C}^i$-norm of $\phi_{x,\varepsilon}$ for $0 \leq i \leq \ell$. For ε small enough (uniformly over x as $\mathcal{M}$ is compact), the chart Ψ can be chosen so that the $\mathcal{C}^\ell$-norm of the function $J\Psi : \mathcal{B}_{\mathbb{R}^d}(0,\varepsilon/2) \to \mathbb{R}$ is uniformly bounded, with $J\Psi(u) \geq 1/2$ for all $u \in \mathcal{B}_{\mathbb{R}^d}(0,\varepsilon/2)$. In particular, as ϕ_0 is equal to 1 on $\mathcal{B}_{\mathbb{R}^d}(0,1/6)$, this implies that the constant $C_{x,\varepsilon}$ is larger than c_0 for some constant c_0 depending only on $\mathcal{M}$. The smoothness of ϕ_0 and of $J\Psi$ then imply the desired controls on the $\mathcal{C}^i$-norm of $\phi_{x,\varepsilon}$ (applying Leibniz formula for the derivative of a product). □

Fix $0 < \varepsilon \leq \varepsilon_0$ and consider a set of points $x_1, \dots, x_J \in \mathcal{M}$ that are all at least 2ε apart. Note that by a simple covering argument, the number J can be chosen to be of order ε^{-d}. For $j = 1, \dots, J$, write $\phi_j = \phi_{x_j,\varepsilon}$. Assume without loss of generality that the manifold $\mathcal{M}$ has unit volume. For $\tau \in \{\pm 1\}^J$, consider the probability measure

$$p_\tau = 1 + \frac{v}{2\kappa}\sum_{j=1}^J \tau_j\phi_j.$$

for some $0 \leq v \leq \varepsilon^\ell$. Note that according to Lemma 3.6.2, p_τ satisfies $1/2 \leq p_\tau \leq 3/2$ and integrates to 1. According to Lemma 3.6.2, the $\mathcal{C}^\ell$-norm of p_τ is smaller than $1/2$ (and therefore so is its Sobolev norm).

Using Remark 3.8.6, we see that for a choice of $p_{\min} \leq 1/2$, $p_{\max} \geq 3/2$, $\kappa_{\min} \leq 1$, $L \geq 1/2$, all the associated measures μ_τ are in the statistical model $\mathcal{P}_{T,\ell} = \mathcal{P}_{T,\ell}(\kappa_{\min}, p_{\min}, p_{\max}, u_{\max}, L)$ for some $u_{\max}$ large enough, as long as $\ell \geq 2$.

Let $\tau, \tau' \in \{\pm 1\}^J$. Consider the function $f = \sum_{j=1}^J \mathbf{1}_{\{\tau_j \neq \tau_j'\}}\tau_j\phi_j$. The function $f : \mathcal{M} \to \mathbb{R}$ is Lipschitz continuous with Lipschitz constant $\kappa\varepsilon^{-1}$. As the 2-Wasserstein distance $\mathcal{W}_2$ is larger than the 1-Wasserstein distance $\mathcal{W}_1$, it holds by duality [119, Particular Case 5.16] that

$$\begin{aligned}\mathcal{W}_2(\mu_\tau, \mu_{\tau'}) \geq \mathcal{W}_1(\mu_\tau, \mu_{\tau'}) &\geq \frac{1}{\kappa\varepsilon^{-1}}\left(\int_{\mathcal{M}} f p_\tau \mathrm{d}y - \int_{\mathcal{M}} f p_{\tau'}\mathrm{d}y\right) \\ &= \frac{v}{2\kappa^2\varepsilon^{-1}}\sum_{j=1}^J \mathbf{1}_{\{\tau_j \neq \tau_{j'}\}} 2\tau_j \int_{\mathcal{M}} \phi_j f \mathrm{d}y.\end{aligned}$$

By construction of f and as $\int_{\mathcal{M}} \phi_j^2 \mathrm{d}y = \varepsilon^d$, this quantity is equal to $\frac{v\varepsilon^{d+1}}{\kappa^2} d_H(\tau,\tau')$, that is the first condition in Assouad's lemma holds for $A = v\varepsilon^{d+1}/\kappa^2$.

The Kullback-Leibler divergence between $\mathbf{P}_\tau$ and $\mathbf{P}_{\tau'}$ can be controlled using Girsanov theorem. Indeed, Girsanov theorem provides an explicit formula for the Kullback-Leibler divergence between $\mathbf{P}_\tau$ and $\mathbf{P}_{\tau'}$ (see Section 3.10):

$$\begin{aligned}\mathrm{KL}(\mathbf{P}_\tau||\mathbf{P}_{\tau'}) &= \frac{T}{4}\int_{\mathcal{M}} \|\nabla \ln p_\tau - \nabla \ln p_{\tau'}\|^2 p_\tau^2 \mathrm{d}y = \frac{T}{4}\int_{\mathcal{M}} \|\nabla p_\tau - \frac{p_\tau}{p_{\tau'}}\nabla p_{\tau'}\|^2 \mathrm{d}y \\ &\leq \frac{T}{2}\int_{\mathcal{M}} \|\nabla p_\tau - \nabla p_{\tau'}\|^2 \mathrm{d}y + \frac{T}{2}\int_{\mathcal{M}} \frac{\|\nabla p_{\tau'}\|^2}{p_{\tau'}^2}(p_\tau - p_{\tau'})^2 \mathrm{d}y.\end{aligned}$$

Consider $\tau, \tau' \in \{\pm 1\}^J$ with $d_H(\tau, \tau') = 1$, and assume without loss of generality that the two vectors differ only by their first entry. Using the available bound on the $\mathcal{C}^1$-norm of p_τ and the fact that $p_\tau \geq 1/2$, we obtain that

$$\begin{aligned}\mathrm{KL}(\mathbf{P}_\tau||\mathbf{P}_{\tau'}) &\leq \frac{T}{2}\int_{\mathcal{M}} \|\nabla p_\tau - \nabla p_{\tau'}\|^2 \mathrm{d}y + \frac{v^2\varepsilon^{-2}}{2} T \int_{\mathcal{M}} (p_\tau - p_{\tau'})^2 \mathrm{d}y \\ &\leq \frac{T}{2}\frac{v^2}{\kappa^2}\int_{\mathcal{B}(x_1,\varepsilon)} \|\nabla \phi_1\|^2 \mathrm{d}y + \frac{v^4\varepsilon^{-2}}{2\kappa^2} T \int_{\mathcal{M}} \phi_1^2 \mathrm{d}y \\ &\leq \frac{T}{2} v^2 \varepsilon^{-2} \int_{\mathcal{B}(x_1,\varepsilon)} \mathrm{d}y + \frac{v^4}{2\kappa^2}\varepsilon^{-2+d} T \lesssim T v^2 \varepsilon^{d-2},\end{aligned}$$

where we use at the last line that the volume of a ball of radius ε is of order ε^d for ε small enough (see Lemma 3.8.2), and $v \in [0, \varepsilon^\ell]$.
When $d \geq 5$, choose $\varepsilon = cT^{-1/(2\ell+d-2)}$ and $v = \varepsilon^\ell$. For c small enough, $\mathrm{KL}(\mathbf{P}_\tau \| \mathbf{P}_{\tau'}) \leq 1/2$. By Assouad's lemma, and recalling that we can pick J of order ε^{-d}, we obtain that

$$\mathcal{R}(\mathcal{P}_{T,\ell}) \geq \frac{AJ}{4} = \frac{\varepsilon^{\ell+d+1}J}{4\kappa^2} \gtrsim \varepsilon^{\ell+1} \gtrsim T^{-\frac{\ell+1}{2\ell+d-2}}, \tag{3.45}$$

concluding the proof of Proposition 3.2.5 in this case. For $d \leq 4$, we let $\varepsilon = 1$ and $v = cT^{-1/2}$ for c small enough, so that $\mathrm{KL}(\mathbf{P}_\tau||\mathbf{P}_{\tau'}) \leq 1/2$ and J is of order 1. Then, Assouad's lemma gives

$$\mathcal{R}(\mathcal{P}_{T,\ell}) \geq \frac{AJ}{4} = \frac{v\varepsilon^{d+1}J}{4\kappa^2} \gtrsim T^{-1/2}. \tag{3.46}$$

This concludes the proof. □

3.7 Control of the bias term

In kernel density estimation, variance terms can be controlled with minimal assumptions on the kernel K (say, boundedness), whereas choosing a kernel having specific properties is required to control bias terms [116]. The situation is not different for the estimation of μ: we are able to control the variance term $\mathbb{E}_x\left[\mathcal{W}_2^2(\widehat{\mu}_{T,h}, \mu_h)\right]$ with few assumptions on K (see Theorem 3.2.2). However, controlling the bias requires the kernel K to be of sufficiently high order in the following sense.

Definition 3.7.1 (Order of a kernel). *For a multi-index $\alpha = (\alpha^1, ..., \alpha^d) \in \mathbb{Z}_+^d$, we denote $|\alpha| := \alpha^1 + \cdots + \alpha^d$, $z^\alpha = \prod_{j=1}^d z_j^{\alpha_j}$ and $\partial^\alpha \mathbf{K}$ the partial derivative of $\mathbf{K}$ in the direction α. A function $\mathbf{K} : \mathbb{R}^d \to \mathbb{R}$ is called kernel of order r if the function is of class $\mathcal{C}^r$, and $\mathbf{K}$ satisfies*

$$\int_{\mathbb{R}^d} \partial^\alpha \mathbf{K}(z) z^{\tilde{\alpha}} \mathrm{d}z = 0,$$

for any multi-index $\alpha, \tilde{\alpha}$ such that $|\alpha| < r$ and $|\tilde{\alpha}| < r + |\alpha|$, with $|\alpha| > 0$ when $\tilde{\alpha} = 0$.

When K is of order larger than $\ell + 1$, we obtain a tight control of the bias following [37].

Proposition 3.7.2 (Bias term). *Let K be a kernel of order larger than $\ell+1$, and let p be a density of class $\mathcal{C}^2$ with a finite Sobolev norm $\|p\|_{H^\ell(\mathcal{M})}$. Then, for h small enough,*

$$\mathcal{W}_2^2(\mu_h, \mu) \leq \frac{c\|p\|^2_{H^\ell(\mathcal{M})}}{p^2_{\min}} h^{2\ell+2}, \tag{3.47}$$

where c depends on $\mathcal{M}$ and K.

Proof. As μ has a lower bounded density, it holds according to Lemma 3.4.4 that

$$\mathcal{W}_2^2(\mu_h, \mu) \leq 4p_{\min}^{-1} \int_{\mathcal{M}} \left|(-\Delta)^{-1/2}(p - p_h)\right|^2 \mathrm{d}x,$$

whereas it is proved in [37, Proposition 9] that

$$\int_{\mathcal{M}} \left|(-\Delta)^{-1/2}(p - p_h)\right|^2 \mathrm{d}x \leq c\|p\|^2_{H^\ell(\mathcal{M})} h^{2\ell+2}$$

when K is a kernel of order larger than $\ell+1$, where c depends on $\mathcal{M}$ and K. □

3.8 Some technical proofs

3.8.1 Proof of Lemma 3.4.2

(i) The proof of the uniform convergence of $(h^{-d}\eta_h(.))_{h>0}$ is given in [37, Lemma 10] when the kernel K has some regularity. We detail below another proof assuming only that K is continuous with compact support in $[0,1]$.

Recall that since $\mathcal{M}$ is compact, the Euclidean norm and the geodesic distance are equivalent: there is a constant $c_{\mathcal{M}} > 1$ such that $\forall (x,y) \in \mathcal{M}^2$

$$\|x - y\| \leq \rho(x,y) \leq c_{\mathcal{M}}\|x - y\|.$$

Consider a family $\mathcal{E}_x : \mathcal{B}_{\mathbb{R}^d}(0, c_x) \to \mathcal{M}$ of Riemannian normal parametrizations at $x \in \mathcal{M}$ (see [61, Section 3.1] or [79, Proposition 5.24]). Thanks to [61, Theorem 3.4] (for more details see [79, Proposition 10.37]), there exists a constant $c_1 > 0$, such that all the parametrizations $(\mathcal{E}_x)_{x\in\mathcal{M}}$ have the same domain $\mathcal{B}_{\mathbb{R}^d}(0, c_1)$, and for any $h < \frac{c_1}{c_{\mathcal{M}}}$ the Jacobian determinant J_x of the change of variables $v = \mathcal{E}_x^{-1}(y)$ is uniformly bounded in x.

Consequently,

$$\begin{aligned} h^{-d}\eta_h(x) &= h^{-d} \int_{\mathcal{M}} K\left(\frac{\|x - y\|}{h}\right) \mathbf{1}_{\|x-y\|\leq h} \mathrm{d}y \\ &= h^{-d} \int_{\mathcal{B}_{\mathbb{R}^d}(0,c_1)} K\left(\frac{\|x - \mathcal{E}_x(v)\|}{h}\right) \mathbf{1}_{\|x-\mathcal{E}_x(v)\|\leq h} J_x(v) \mathrm{d}v \\ &= \int_{\mathcal{B}_{\mathbb{R}^d}\left(0,\frac{c_1}{h}\right)} K\left(\left\|\frac{x - \mathcal{E}_x(hv)}{h}\right\|\right) \mathbf{1}_{\left\|\frac{x-\mathcal{E}_x(hv)}{h}\right\|\leq 1} J_x(hv) \mathrm{d}v. \end{aligned}$$

We know that $\rho(x,y) = \left\|\mathcal{E}_x^{-1}(y)\right\|$ (see [61, Theorem 3.2]). Then by equivalence of the distances, we have $\|x - \mathcal{E}_x(hv)\| \geq \frac{\rho(x,\mathcal{E}_x(hv))}{c_{\mathcal{M}}} = h\frac{\|v\|}{c_{\mathcal{M}}}$. Consequently,

$$\left\{\left\|\frac{x - \mathcal{E}_x(hv)}{h}\right\| \leq 1\right\} = \left\{\left\|\frac{x - \mathcal{E}_x(hv)}{h}\right\| \leq 1\right\} \cap \{\|v\| \leq c_{\mathcal{M}}\}.$$

Beside, by [61, Theorem 3.4], there exists $c_2 > 0$, independent of x, such that $|J_x(v) - 1| \leq c_2\|v\|^2$ and $\|x - \mathcal{E}_x(v) - v\| \leq c_2\|v\|^2$ (since $\mathcal{E}'_x(0) = Id$).

We deduce that

$$\left|K\left(\left\|\frac{x - \mathcal{E}_x(hv)}{h}\right\|\right)\right| \mathbf{1}_{\left\|\frac{x-\mathcal{E}_x(hv)}{h}\right\| \leq 1} J_x(hv) \mathbf{1}_{\mathcal{B}_{\mathbb{R}^d}(0,\frac{c_1}{h})}(v) \leq \|K\|_\infty (1 + c_2 c_1^2) \mathbf{1}_{\|v\| \leq c_\mathcal{M}},$$

where the upper-bound is an integrable function on $\mathbb{R}^d$. We also remark that $(J_x(hv))_{h>0}$ converges to 1 and $\left(\left\|\frac{x-\mathcal{E}_x(hv)}{h}\right\|\right)_{h>0}$ converges to $\|v\|$ when $h \to 0$, both uniformly on $\mathcal{M}$. As K is continuous on $\mathbb{R}^d$ with compact support, the function is thus uniformly continuous on $\mathbb{R}^d$. We then deduce that $\left(K\left(\left\|\frac{x-\mathcal{E}_x(hv)}{h}\right\|\right)\right)_{h>0}$ converges to $K(\|v\|)$ when $h \to 0$, uniformly on $\mathcal{M}$.

By the dominated convergence theorem, we deduce that, when $h \to 0$, $\left(h^{-d}\eta_h\right)_{h>0}$ converges to $\int_{\mathbb{R}^d} K(\|v\|)\mathrm{d}v = 1$, uniformly on $\mathcal{M}$.

(ii) By uniform convergence, $\eta_h > 0$ when h is small enough, so that K_h is well-defined. The result then follows from a straightforward computation.

(iii) We note that the support of K_h is included in $\left\{(x, y) \in \mathcal{M}^2 : \|x - y\| \leq h\right\}$, and

$$|K_h(x, y)| \leq \frac{1}{|h^{-d}\eta_h(x)|}\|K\|_\infty h^{-d}.$$

There is uniform convergence on $\mathcal{M}$ of $(h^{-d}\eta_h(.))_{h>0}$ to 1 when $h \to 0$. Then there exists $h_c > 0$ such that $\forall h < h_c$, $\forall (x, y) \in \mathcal{M}^2$, we have $h^{-d}\eta_h(x) \geq \frac{1}{2}$ and

$$|K_h(x, y)| \leq 2\|K\|_\infty h^{-d}.$$

(iv) Let $x \in \mathcal{M}$ and $y, y' \in \mathcal{M}$. Let L be the Lipschitz constant of K. Then, by the triangle inequality when $h < h_c$, so that $h^{-d}\eta_h(x) \geq \frac{1}{2}$, we have

$$\begin{aligned}|K_h(x, y) - K_h(x, y')| &\leq \frac{1}{|h^{-d}\eta_h(x)|} h^{-d} |K\left(\frac{\|x - y\|}{h}\right) - K\left(\frac{\|x - y'\|}{h}\right)| \\ &\leq 2Lh^{-d-1}\|y - y'\| \leq 2Lh^{-d-1}\rho(y - y').\end{aligned}$$

Remark 3.8.1. *As mentioned in Remark 3.2.1, it is simpler to work with a kernel $\widetilde{K}_h$ based on the geodesic distance ρ, $\widetilde{K}_h(x, y) := \frac{1}{\widetilde{\eta}_h(x)} K\left(\frac{\rho(x,y)}{h}\right)$, with $\widetilde{\eta}_h(x) = \int_\mathcal{M} K\left(\frac{\rho(x,y)}{h}\right)\mathrm{d}y$. In that case, we can easily prove, without regularity assumptions, that when K is an integrable function with support in $[0, 1]$ and $\int_{\mathbb{R}^d} K(\|v\|)\mathrm{d}v = 1$, there is a constant $\kappa > 0$ such that $\left\|h^d\widetilde{\eta}_h - 1\right\|_\infty \leq \kappa h^2$. Actually, using the Riemannian normal parametrization $\mathcal{E}_x$ at $x \in \mathcal{M}$ and $\rho(x, y) = \left\|\mathcal{E}_x^{-1}(y)\right\|$, as in the previous proof of Lemma 3.4.2, we obtain for $x \in \mathcal{M}$*

$$\begin{aligned}\left|h^{-d}\widetilde{\eta}_h(x) - 1\right| &= \left|h^{-d}\int_\mathcal{M} K\left(\frac{\rho(x, y)}{h}\right)\mathbf{1}_{\rho(x,y)\leq h}\mathrm{d}y - 1\right| \\ &= h^{-d}\left|\int_{\mathcal{B}_{\mathbb{R}^d}(0,c_1)} K\left(\frac{\|v\|}{h}\right)\mathbf{1}_{\|v\|\leq h} J_x(v)\mathrm{d}v - \int_{\mathcal{B}_{\mathbb{R}^d}(0,h)} K\left(\frac{\|v\|}{h}\right)\mathrm{d}v\right| \\ &\leq h^{-d}\int_{\mathcal{B}_{\mathbb{R}^d}(0,h)} \left|K\left(\frac{\|v\|}{h}\right)\right| |J_x(v) - 1|\mathrm{d}v \\ &\leq c_2 h^2 \int_{\mathcal{B}_{\mathbb{R}^d}(0,1)} |K(\|v\|)|\|v\|^2\mathrm{d}v\end{aligned}$$

since $|J_x(v) - 1| \leq c_2\|v\|^2$.

3.8.2 Proof of Lemma 3.4.3

Using the same proof as in Appendix 3.8.1, we first remark that uniformly in $x \in \mathcal{M}$ the term $\left(h^{-d}\int_{\mathcal{M}} \mathbf{1}_{\|x-y\|\leqslant h}\mathrm{d}y\right)_{h>0}$ converges to $\mathrm{vol}(B_{\mathbb{R}^d}(0,1))$. Besides, by the triangular inequality,

$$\begin{aligned}
&|p_h(x) - p(x)| = \left|\int_{\mathcal{M}} \frac{1}{\eta_h(y)} K\left(\frac{\|x-y\|}{h}\right) p(y)\mathrm{d}y - p(x)\right| \\
&\leq \left|\int_{\mathcal{M}} \left(\frac{1}{\eta_h(y)} - h^{-d}\right) K\left(\frac{\|x-y\|}{h}\right) p(y)\mathrm{d}y\right| + \left|\int_{\mathcal{M}} h^{-d} K\left(\frac{\|x-y\|}{h}\right)(p(y)-p(x))\mathrm{d}y\right| + \\
&\quad + \left|\int_{\mathcal{M}} h^{-d} K\left(\frac{\|x-y\|}{h}\right) p(x)\mathrm{d}y - p(x)\right| \\
&\leq \left\|\frac{1}{h^{-d}\eta_h} - 1\right\|_{\infty} \|K\|_{\infty}\|p\|_{\infty} \int_{\mathcal{M}} h^{-d} 1_{\|x-y\|\leq h}\mathrm{d}y + \|K\|_{\infty}\mathrm{Lip}(p)\int_{\mathcal{M}} h^{-d}1_{\|x-y\|\leq h}\ \|x-y\|\ \mathrm{d}y + \\
&\quad + |h^{-d}\eta_h(x) - 1||p(x)|,
\end{aligned}$$

where $\mathrm{Lip}(p)$ is the Lipschitz constant of p with respect to the Euclidean distance, which is bounded up to a constant by the $\mathcal{C}^1$-norm of p (recall that the Euclidean and geodesic distances are equivalent). Hence, by the remark at the beginning of this proof, there is a constant $C > 0$ such that for all $x \in \mathcal{M}$,

$$|p_h(x) - p(x)| \leqslant C\|K\|_{\infty}\ \|p\|_{\mathcal{C}^1}\left(\left\|\frac{1}{h^{-d}\eta_h} - 1\right\|_{\infty} + h + \left\|h^{-d}\eta_h - 1\right\|_{\infty}\right), \tag{3.48}$$

where the constant C depends only on the embedding $\mathcal{M} \subset \mathbb{R}^m$ and where the parenthesis in the right term converges to zero by Lemma 3.4.2(i). Therefore, we deduce the uniform convergence of $(p_h)_{h>0}$ to p on $\mathcal{M}$, and the second result of the lemma is straightforward a consequence of Equation (3.48). We have the desired conclusion.

3.8.3 Proof of Lemma 3.3.2

We will make use of the following estimates.

Lemma 3.8.2. *Let $x \in \mathcal{M}$. For all $h > 0$ small enough, it holds that for $\alpha \geq d$*

$$\int_{\rho(x,y)\geq h} \rho(x,y)^{-\alpha}\mathrm{d}y \leq C_\alpha \begin{cases} h^{d-\alpha} & \text{if } \alpha > d \\ \ln(1/h) & \text{if } \alpha = d \end{cases} \tag{3.49}$$

and that for $\alpha < d$

$$\int_{\rho(x,y)\leq h} \rho(x,y)^{-\alpha}\mathrm{d}y \leq C_\alpha h^{d-\alpha}. \tag{3.50}$$

Before giving the proof, let us note that since the geodesic distance and the Euclidean norm are equivalent on the compact manifold $\mathcal{M}$ (see [51, Proposition 2]), we have a similar result with the Euclidean norm instead of the geodesic distance in Lemma 3.8.2.

Proof. Using the change of variables $v = \mathcal{E}_x^{-1}(y)$ as in the proof of Lemma 3.4.2 (see Section 3.8.1), we have for $h_0 > 0$ a constant larger than h

$$\int_{\rho(x,y)\geq h} \rho(x,y)^{-\alpha}\mathrm{d}y \leq \int_{h_0>\rho(x,y)\geq h} \rho(x,y)^{-\alpha}\mathrm{d}y + ch_0^{-\alpha}.$$

We pick h_0 small enough so that, by a change of variable

$$\int_{h_0>\rho(x,y)\geq h} \rho(x,y)^{-\alpha}\mathrm{d}y = \int_{h_0>\|v\|\geq h} \|v\|^{-\alpha} J_x(v)\mathrm{d}v,$$

where $J_x(v)$ is a Jacobian, that is bounded by 2 for all x and v when h_0 is chosen small enough. We then conclude by computing the integral. The proof of the second statement is similar. ☐

We only prove Lemma 3.3.2 the case $d \geq 3$, the cases $d = 1$ and $d = 2$ being treated with minimal modifications. By Proposition 3.3.1 (*iv*) on the Green function G, for $d \geq 3$, there exists a constant $\kappa > 0$ such that $\forall (x, y) \in \mathcal{M}^2 \setminus \operatorname{diag}(\mathcal{M})$,

$$|G(x, y)| \leq \kappa \rho(x, y)^{2-d}.$$

Hence,

$$|(Gf)(z)| \leq \int_{\mathcal{M}} |G(z, y)||f(y)|\mathrm{d}y \leq \|f\|_\infty \kappa \int_{\mathcal{M}} \rho(z, y)^{2-d} \mathbf{1}_{\rho(x,y)\leq h} \mathrm{d}y.$$

When $\rho(x, z) > 2h$, we have $\rho(x, y) \leq \rho(x, z)/2$ and

$$\rho(z, y) \geq \rho(x, z) - \rho(x, y) \geq \frac{1}{2}\rho(x, z).$$

Then, as $2 - d < 0$, we obtain

$$|(Gf)(z)| \leq 2^{d-2} \|f\|_\infty \kappa \int_{\mathcal{M}} \mathbf{1}_{\rho(x,y)\leq h} \mathrm{d}y \ \rho(x, z)^{2-d}.$$

According to Lemma 3.8.2,

$$|(Gf)(z)| \leq C\|f\|_\infty h^d \rho(x, z)^{2-d}.$$

We now turn to estimates when $\rho(x, z) < 2h$. Notice that

$$|(Gf)(z)| \leq \|f\|_\infty \kappa \int_{\mathcal{M}} \rho(z, y)^{2-d} \mathbf{1}_{\rho(z,y)\leq 3h} \mathrm{d}y.$$

The conclusion then also follows from Lemma 3.8.2.

3.8.4 Wasserstein distance between a measure and its convolution

Lemma 3.8.3. *Let ν be a probability measure supported on $\mathcal{M}$. Let K be a nonnegative kernel, with $\int_{\mathbb{R}^d} K(\|u\|)\mathrm{d}u = 1$. For $h > 0$, let ν_h be the measure with density $q_h(x) = \int K_h(z, x)\mathrm{d}\nu(z)$, where K_h is defined in* (3.11)*. Then, there exists h_0 depending only on $\mathcal{M}$ such that for $h \leq h_0$*

$$\mathcal{W}_2^2(\nu, \nu_h) \leq c_1 \|K\|_\infty h^2 \tag{3.51}$$

for some constant c_1 depending on $\mathcal{M}$.

Proof. By the convexity of the Wasserstein distance, it holds that

$$\mathcal{W}_2^2(\nu, \nu_h) \leq \int \mathcal{W}_2^2(\delta_x, (\delta_x)_h)\mathrm{d}\nu(x) \tag{3.52}$$

Let c_0 be such that $\rho(x, y) \leq c_0\|x - y\|$ for all $(x, y) \in \mathcal{M}^2$ (recall that the geodesic distance and the Euclidean distance are equivalent). We apply Lemma 3.4.2 for h small enough to obtain

$$\begin{aligned} \mathcal{W}_2^2(\delta_x, (\delta_x)_h) &= \int K_h(x, z)\rho(x, z)^2 \mathrm{d}z \\ &\leq 2\|K\|_\infty h^{-d} \int_{\rho(x,z)\leq c_0 h} \rho(x, z)^2 \mathrm{d}z \leq c_1\|K\|_\infty h^2, \end{aligned}$$

where we use Lemma 3.8.2 at the last line. ∎

3.8.5 The ultracontractivity term

We give in this section an explicit control of the ultracontractivity constant introduced at the beginning of Section 3.5. To this aim we need to introduce the iterated carré du champs.

Definition 3.8.4. *Given a differential operator $\mathcal{A}$, its iterated carré du champs Γ_2 is defined as:*

$$\Gamma_2(f,f) := \frac{1}{2}[\mathcal{A}\Gamma(f,f) - 2\Gamma(\mathcal{A}f,f)],$$

where Γ is the carré du champs of $\mathcal{A}$.

Lemma 3.8.5. *Let $p \in \mathcal{C}^2(\mathcal{M})$ be a positive density on $\mathcal{M}$ with respect to the volume measure $\mathrm{d}x$, and $\mathcal{A}$ a $\mathcal{C}^2$-elliptic second-order differential operator, which is symmetric with respect to the probability measure $\mu = p\mathrm{d}x$. Then, there is a constant κ, which can be negative, such that:*

$$\Gamma_2(f,f) \geq \kappa\Gamma(f,f), \tag{3.53}$$

In particular, the associated semigroup $(P_t)_{t\geqslant 0}$ satisfies, for all $t > 0$,

$$\|P_t f\|_\infty \leq \exp\left[\frac{\kappa \mathrm{diam}(\mathcal{M})^2}{2(e^{2\kappa t}-1)}\right] \times \|f\|_{L^2(\mu)}. \tag{3.54}$$

Proof. Because $\mathcal{A}$ is a $\mathcal{C}^2$-elliptic operator of second-order, there is a $\mathcal{C}^2$-metric $\tilde{\mathbf{g}}$ on $\mathcal{M}$ such that $\Gamma(f,f) = \left\langle \tilde\nabla f, \tilde\nabla f \right\rangle_{\tilde{\mathbf{g}}}$, where $\tilde\nabla$ is the gradient of the new metric (see [67, eq. 1.3.3]). Hence, due to the symmetry of $\mathcal{A}$ with respect to μ, $\mathcal{A} = \tilde\Delta + \tilde\nabla \ln(\tilde p)$, where $\tilde\Delta$ is the Laplacian of the new metric and $\tilde p = \frac{\mathrm{d}\mu}{\mathrm{dvol}_{\tilde{\mathbf{g}}}}$, where $\mathrm{vol}_{\tilde{\mathbf{g}}}$ is the volume measure in the new metric.
Consequently, by [12, eq. C.5.3] we have $\Gamma_2(f,f) = |\nabla\nabla f|^2 + \left(\mathrm{Ricc}_{\tilde{\mathbf{g}}} - \tilde\nabla\tilde\nabla \ln(\tilde p)\right)\left(\tilde\nabla f, \tilde\nabla f\right)$, where $\mathrm{Ricc}_{\tilde{\mathbf{g}}}$ denotes the Ricci tensor. Therefore, as in [12, eq C.6.3], (3.53) is equivalent to

$$\mathrm{Ricc}_{\tilde{\mathbf{g}}} - \tilde\nabla\tilde\nabla \ln(\tilde p) \geq \kappa. \tag{3.55}$$

Hence, from the compactness of $\mathcal{M}$ and the $\mathcal{C}^2$ continuity of both $\mathcal{A}$ and p, we have the desired conclusion for the first part.
For the second part, we have from the implication (1) $\to$ (3) of Theorem 2.3.3 in [120] with $p = 2$, that for any x,

$$|P_t f(x)|^2 = \int_{\mathcal{M}} |P_t f(x)|^2 \mu(\mathrm{d}y) \leq \int_{\mathcal{M}} P_t|f|^2(y) \exp\left[\frac{\kappa \mathrm{diam}(\mathcal{M})^2}{e^{2\kappa t}-1}\right]\mu(\mathrm{d}y) =$$
$$\exp\left[\frac{\kappa \mathrm{diam}(\mathcal{M})^2}{e^{2\kappa t}-1}\right] \int_{\mathcal{M}} P_t|f|^2(y)\mu(\mathrm{d}y) = \exp\left[\frac{\kappa \mathrm{diam}(\mathcal{M})^2}{e^{2\kappa t}-1}\right] \|f\|^2_{L^2(\mu)},$$

since μ is the invariant measure of the underlying process. Therefore, we have the second inequality, which is the ultracontractivity of $\mathcal{A}$. □

Remark 3.8.6. *We note that for the operator $\mathcal{A}_{pq}$ defined by (3.1) with $q \equiv 1$, there is no need to change the metric to obtain the result. Consequently, the constant κ appearing in Lemma 3.8.5 depends only on $\mathcal{M}$, an upper bound on $\|p\|_{\mathcal{C}^2(\mathcal{M})}$ and $p_{\min}$. This enables the choice of a uniform ultracontractivity constant $u_{\max}$ for the minimax lower bound in Section 3.6.*

3.9 SDEs for the diffusions with generator $\mathcal{A}_{pq}$

In this section, we give the SDEs satisfied by the diffusion processes of generator (3.1) and (3.2). Recall that the Stratonovich integral (see [98, page 82]) is defined as follows:

Definition 3.9.1 (Stratonovich integral). *Let X, Y be two continuous real-valued semimartingales. The Stratonovich integral of Y with respect to X, denoted by $\int_0^t Y_s \circ \mathrm{d}X_s$, is defined by*

$$\int_0^t Y_s \circ \mathrm{d}X_s := \int_0^t Y_s \mathrm{d}X_s + \frac{1}{2}\langle Y, X\rangle_t,$$

where the first term is the Itô integral of Y with respect to X and $\langle .,. \rangle$ is the bracket process (also known as the quadratic covariation process).

With the Stratonovich integral, the classical Itô formula can then be written is the following way (see [98, Theorems 20-21, pages 277-278]), for a continuous d-dimensional semimartingale X and a function $f : \mathbb{R}^d \to \mathbb{R}$ of class $\mathcal{C}^2$: $f(X)$ is a semimartingale and

$$f(X_t) - f(X_0) = \sum_{i=1}^{d} \int_0^t \frac{\partial f}{\partial x^i}(X_s) \circ \mathrm{d}X_s^i. \tag{3.56}$$

Because of this chain rule, the Stratonovich integral is better fitted to differential calculus on manifolds than the usual Itô integral.

We recall that a vector field V on a manifold $\mathcal{M}$ is a family $\{V(x)\}_{x\in\mathcal{M}}$ such that $\forall x \in \mathcal{M}$, $V(x) \in T_x\mathcal{M}$ (see for e.g. [77, Chapter 4]). In local coordinates $(x^1, x^2, ..., x^d)$, a smooth vector field V can be represented as

$$V(x) = \sum_{i=1}^{d} V^i(x) \left.\frac{\partial}{\partial x^i}\right|_x,$$

where $V^1, \ldots, V^d$ are real smooth functions on the domain of the local coordinate system, and where $\left\{\frac{\partial}{\partial x^i}\right\}_{1\le i\le d}$ denotes a basis of $T_x\mathcal{M}$.

Proposition 3.9.2 (Theorem 1.2.9 in [67]). *Let $l \geq 1$. Consider the Stratonovich SDE*

$$\mathrm{d}X_t = \sum_{\alpha=1}^{l} V_\alpha(X_t) \circ \mathrm{d}B_t^\alpha + V_0(X_t)\mathrm{d}t \tag{3.57}$$

where $(V_\alpha)_{0\le\alpha\le l}$ are $\mathcal{C}^2$ vector fields and $B = (B^\alpha)_{1\le\alpha\le l}$ is the standard l-dimensional Brownian motion. Then, there exists a unique strong solution to (3.57) (up to explosion time) whose infinitesimal generator is

$$\mathcal{A}f(x) = \frac{1}{2}\sum_{\alpha=1}^{l} \left(V_\alpha^2 f\right)(x) + (V_0 f)(x),$$

where $\left(V_\alpha^2 f\right)(x) := (V_\alpha(V_\alpha f))(x)$, and whose carré du champ operator is given by $\Gamma(f, g) = \frac{1}{2}\sum_{\alpha=1}^{l} V_\alpha(f)V_\alpha(g)$.

Let $\{e_\alpha\}_{1\le\alpha\le m}$ be an orthonormal basis on $\mathbb{R}^m$. For each $x \in \mathcal{M}$, we consider $P_\alpha(x)$ the orthogonal projection of e_α to $T_x\mathcal{M}$. Let us note that P_α is a vector field on $\mathcal{M}$. In a local coordinate system $(x^1, x^2, ..., x^d)$,

$$P_\alpha(x) = \sum_{i=1}^{d} P_\alpha^i(x) \left.\frac{\partial}{\partial x^i}\right|_x.$$

Then the Laplace-Beltrami operator satisfies (see [67, Theorem 3.1.4])

$$\Delta = \sum_{\alpha=1}^{m} P_\alpha^2. \tag{3.58}$$

Remark that for two real-valued functions of class $\mathcal{C}^2$ on $\mathcal{M}$, we have:

$$\langle \nabla f, \nabla h\rangle = \sum_{\alpha=1}^{m} (P_\alpha f)(P_\alpha h). \tag{3.59}$$

Then, as an application of Proposition 3.9.2, there exists a unique strong solution starting at $x \in \mathcal{M}$ to the following SDE

$$\mathrm{d}X_t = \sum_{\alpha=1}^{m} \sqrt{2q(X_t)} P_\alpha(X_t) \circ \mathrm{d}B_t^\alpha + \sum_{\alpha=1}^{m} \left(\frac{1}{2}(P_\alpha q)(X_t) + q(P_\alpha(\ln p))(X_t) \right) (P_\alpha f)(X_t) \tag{3.60}$$

for $(B^\alpha)_{1\le \alpha \le m}$ independent euclidean 1-dimensional Brownian motions, and whose infinitesimal generator is

$$\mathcal{A}_{pq} f = q\Delta f + \langle q \nabla \ln(pq), \nabla f\rangle. \tag{3.61}$$

In the particular case where $q \equiv 1$, we deduce that the unique solution to the SDE

$$\mathrm{d}X_t = \sqrt{2} \sum_{\alpha=1}^{m} P_\alpha(X_t) \circ \mathrm{d}B_t^\alpha + \sum_{\alpha=1}^{m} P_\alpha(\ln p)(X_t) P_\alpha(X_t)\mathrm{d}t, \tag{3.62}$$

has the infinitesimal generator

$$\mathcal{L} f = \Delta f + \langle \nabla \ln p, \nabla f\rangle.$$

Proof of (3.61). To compute the infinitesimal generator of (3.60), we apply Proposition 3.9.2 with $V_\alpha = \sqrt{2q} P_\alpha$ and $V_0 = \sum_{\alpha=1}^{m} \left(\frac{1}{2}(P_\alpha q) + q(P_\alpha(\ln p))\right) P_\alpha$. From this proposition, we know that the generator of X is, for a test function f of class $\mathcal{C}^2$,

$$\mathcal{A}_{pq} f(x) = \sum_{\alpha=1}^{m} \frac{1}{2} (V_\alpha^2 f)(x) + (V_0 f)(x).$$

We have:

$$\left(V_\alpha^2 f\right)(x) = 2\sqrt{q} P_\alpha(\sqrt{q}(P_\alpha f))(x) = 2q(x)\left(P_\alpha^2 f\right)(x) + (P_\alpha q)(x)(P_\alpha f)(x)$$

Thus,

$$\mathcal{A}_{pq} f(x) = q(x)\Delta f(x) + \sum_{\alpha=1}^{m} (P_\alpha q)(x)(P_\alpha f)(x) + q \sum_{\alpha=1}^{m} (P_\alpha(\ln p)) P_\alpha.$$

We conclude thanks to (3.59). □

At last, since $\mathcal{A}_{pq}$ is self-adjoint with respect to μ, we deduce the following proposition.

Proposition 3.9.3. *The measure* $\mu = p(x)\mathrm{d}x$ *is invariant for* $\mathcal{A}_{pq}$.

3.10 Kullback-Leibler divergence of path space measures on manifolds

In this section, the operator $\mathcal{L}$ will be denoted by $\mathcal{L}_p$ to highlight its dependence with respect to the density p of the measure μ. Let us denote by $\mathbf{P}_{(p,T)}$ the probability measure on $\mathcal{C}([0,T], \mathcal{M})$, given by the distribution of the diffusion with generator $\mathcal{L}_p$, defined for $f \in \mathcal{C}^2(\mathcal{M})$ by:

$$\mathcal{L}_p f = \Delta f + \langle \nabla f, \nabla \ln p\rangle,$$

and starting from its invariant measure $\mu = p\mathrm{d}x$.

We will denote by $\mathbb{E}_{(p,T)}$ the expectation in the distribution $\mathbf{P}_{(p,T)}$ and we define the Kullback-Leibler divergence as:

$$\mathrm{KL}(\mathbf{P}_{(p,T)} || \mathbf{P}_{(q,T)}) = \mathbb{E}_{(p,T)} \left[\ln \frac{\mathrm{d}\mathbf{P}_{(p,T)}}{\mathrm{d}\mathbf{P}_{(q,T)}} \right] \tag{3.63}$$

where $\mathrm{d}\mathbf{P}_{(p,T)}/\mathrm{d}\mathbf{P}_{(q,T)}$ stands for the density of $\mathbf{P}_{(p,T)}$ with respect to the measure $\mathbf{P}_{(q,T)}$.

Theorem 3.10.1. *For any two $\mathcal{C}^1$ strictly positive probability densities p and q on $\mathcal{M}$,*

$$\mathrm{KL}(\mathbf{P}_{(p,T)}||\mathbf{P}_{(q,T)}) = \frac{T}{4}\int_{\mathcal{M}} \|\nabla \ln p - \nabla \ln q\|^2 p^2 \mathrm{d}x. \tag{3.64}$$

The proof of Theorem 3.10.1 relies crucially on Girsanov's theorem.

Proposition 3.10.2 (Girsanov's theorem for embedded manifolds)**.** *Consider two continuous tangent vector fields Z_1, Z_2 that are tangent to $\mathcal{M}$. Suppose that $(X_t)_{t\geq 0}$ satisfies the Stratonovich SDE:*

$$\mathrm{d}X_t = Z_2(X_t)\mathrm{d}t + \sqrt{2}\sum_{\alpha=1}^{m} P_\alpha(X_t)\circ \mathrm{d}B_t^\alpha, \qquad X_0 = x_0 \in \mathcal{M},$$

where $B = (B^\alpha)_{1\leq\alpha\leq m}$ is a m-dimensional standard Brownian motion under $\mathbb{P}$, and $(P_\alpha)_{1\leq\alpha\leq m}$ is the tangent projection of the standard basis of $\mathbb{R}^m$ on $\mathcal{M}$ (see Section 3.9).
Then, defining for any $T > 0$,

$$\mathcal{E}_T = \exp\left(-\int_0^T \frac{1}{\sqrt{2}}(Z_2 - Z_1)(X_t)\mathrm{d}B_t + \frac{1}{4}\int_0^T \|Z_1(X_t) - Z_2(X_t)\|_2^2 \mathrm{d}t\right), \tag{3.65}$$

it follows that:

$$\mathbb{E}[\mathcal{E}_T] = 1$$

and under $\mathrm{d}\mathbb{Q} = \mathcal{E}_T\mathrm{d}\mathbb{P}$, $(X_t)_{t\geq 0}$ is a solution to the SDE:

$$\mathrm{d}X_t = Z_1(X_t)\mathrm{d}t + \sqrt{2}\sum_{\alpha=1}^{m} P_\alpha(X_t)\circ \mathrm{d}\widetilde{B}_t^\alpha, \qquad X_0 = x_0 \in \mathcal{M},$$

with $\widetilde{B}$ being a Brownian motion under $\mathbb{Q}$.

Proof for Proposition 3.10.2. Let $T > 0$. The first part follows from the fact that the Itô integral $\int_0^T (Z_2 - Z_1)(X_t)\mathrm{d}B_t$ is a local martingale with bounded quadratic variation (since the Z_i's are continuous and $\mathcal{M}$ is compact, hence the Z_is are bounded on $\mathcal{M}$). For the second part, by Girsanov's theorem in $\mathbb{R}^m$, the stochastic process $(\widetilde{B}_t)_{0\leq t\leq T}$ defined by:

$$\widetilde{B}_t = B_t + \int_0^T u(X_t)\mathrm{d}t,$$

with $u(x) = \frac{1}{\sqrt{2}}(Z_2 - Z_1)(x)$, is a standard Brownian motion under $\mathbb{Q}$. Besides, the Stratonovich SDE for X_t can be rewritten as:

$$\begin{aligned}
\mathrm{d}X_t &= Z_2(X_t)\mathrm{d}t + \sqrt{2}\sum_{\alpha=1}^{m} P_\alpha(X_t)\circ \mathrm{d}B_t^\alpha \\
&= Z_2(X_t)\mathrm{d}t + \sqrt{2}\sum_{\alpha=1}^{m} P_\alpha(X_t)\circ \left(\mathrm{d}\widetilde{B}_t^\alpha - u^\alpha(X_t)\mathrm{d}t\right) \\
&= \sqrt{2}\sum_{\alpha=1}^{m} P_\alpha(X_t)\circ \mathrm{d}\widetilde{B}_t^\alpha + (Z_2(X_t) - \sum_{\alpha=1}^{m}\sqrt{2}u^\alpha(X_t)P_\alpha(X_t))\mathrm{d}t \\
&= Z_1(X_t)\ \mathrm{d}t + \sqrt{2}\sum_{\alpha=1}^{m} P_\alpha(X_t)\circ \mathrm{d}\widetilde{B}_t^\alpha,
\end{aligned}$$

hence, we imply the desired conclusion. □

Let us now prove Theorem 3.10.1.

Proof of Theorem 3.10.1. Let $(X_t)_{t\geq 0}$ be a solution to the SDE:

$$\mathrm{d}X_t = \nabla \ln p(X_t)\mathrm{d}t + \sqrt{2}\sum_{\alpha=1}^{m} P_\alpha \circ \mathrm{d}B_t^\alpha,$$

with X_0 uniform on $\mathcal{M}$ and with B being a m-dimensional standard Brownian motion under some probability space $\mathbb{P}$. The infinitesimal generator of $(X_t)_{t\geq 0}$ is $\mathcal{L}_p$ by (3.62). By Proposition 3.10.2, $(X_t)_{t\geq 0}$ is also a solution to the SDE:

$$\mathrm{d}X_t = \nabla \ln q(X_t)\mathrm{d}t + \sqrt{2}\sum_{\alpha=1}^{m} P_\alpha \circ \mathrm{d}\widetilde{B}_t^\alpha,$$

with X_0 uniform on $\mathcal{M}$, and where $\widetilde{B}$ is a m-dimensional standard Brownian motion under the probability measure $\mathbb{Q}$ defined by $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \mathcal{E}_T$ with

$$\mathcal{E}_T = \exp\left(-M_T + \frac{1}{2}\langle M\rangle_T\right), \quad M_T = \int_0^T \frac{1}{\sqrt{2}}\left(\nabla \ln p - \nabla \ln q\right)(X_t)\mathrm{d}B_t.$$

Therefore, by the definition of Kullback-Leibler divergence,

$$\mathrm{KL}(\mathbf{P}_{(p,T)}||\mathbf{P}_{(q,T)}) = \mathbb{E}\left[\ln\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right] = \mathbb{E}\left[-M_T + \frac{1}{2}\langle M\rangle_T\right] = \frac{1}{2}\mathbb{E}[\langle M\rangle_T].$$

We have the desired conclusion by computing $\mathbb{E}[\langle M\rangle_T]$, because the distribution of X_t is $\mu = p\mathrm{d}x$ at each time $t \geq 0$. $\square$

Chapter 4

1-Wasserstein minimax estimation for general smoothed probability densities

Contents

This chapter is a research note I prepared during my work on minimax estimators with respect to the Wasserstein distance metric for Chapter 3.
In this analysis, we investigate the problem of estimating a probability measure from i.i.d. samples under the 1-Wasserstein metric. Specifically, we establish minimax convergence rates for measure estimation in the Wasserstein metric. Our work broadens the existing framework by removing the strict positivity condition and relaxing the restrictive boundedness requirements on probability densities. As a result, a wider class of distributions is encompassed, including Gaussian, lognormal, and Student's t-distributions. Furthermore, by employing kernel smoothing estimators, we derive efficient minimax estimators that are expected to be more intuitive and basic than previous wavelet-based methods. Our findings demonstrate that, for densities with finite moments and bounded Sobolev norms, kernel-based smoothing attains optimal minimax rates in various general settings. Moreover, an almost sure convergence speed result is also established.

4.1 Introduction

Optimal transportation theory, originally developed in the 18th century by Gaspard Monge [88] and later extended by Leonid Kantorovich [74] in the 20th century, has become a pivotal tool in modern mathematical and applied sciences. This theory addresses the problem of transforming one probability distribution into another in the most efficient way, measured by a cost function [119]. More precisely, given two Polish spaces $(\mathcal{X}, \rho_{\mathcal{X}})$ and $(\mathcal{Y}, \rho_{\mathcal{Y}})$ and two probability Borel measures $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$, where $\mathcal{P}(\mathcal{X})$ and $\mathcal{P}(\mathcal{Y})$ are respectively the spaces of all Borel probability measures on $\mathcal{X}$ and $\mathcal{Y}$. The transportation cost between μ and ν with respect to a transportation cost function $c : \mathcal{X} \times \mathcal{Y} \to [0, \infty)$ is defined via the transportation cost with respect to the intrinsic metric as:

$$\mathcal{T}_c(\mu,\nu) := \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\mathrm{d}\pi(x,y), \tag{4.1}$$

where $\Pi(\mu,\nu)$ is the set of all joint probabilities on $\mathcal{X}\times\mathcal{Y}$ with marginals μ,ν.
A particular case of interest of optimal transportation cost is when two Polish spaces are identical, i.e., $\mathcal{X}=\mathcal{Y}=(\mathcal{M},\rho)$ for some Polish space $(\mathcal{M},\rho)$. In this framework, for any real positive number $q\geq 1$, the q-th Wasserstein distance between two probability measures μ and ν is defined as:

$$\mathcal{W}_q(\mu,\nu) := \left(\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{M}\times\mathcal{M}} \rho(x,y)^q\mathrm{d}\pi(x,y)\right)^{1/q}. \tag{4.2}$$

It is worth noticing that specifically for $q=1$, 1-Wasserstein distance between two measures can also be represented as result of a maximization problem over the space of Lipschitz functions on $\mathcal{M}$ [45, Proposition 2.6.6]:

$$\mathcal{W}_1(\mu,\nu) := \sup_{f:\mathcal{M}\to\mathbb{R}\text{ is 1-Lipschitz}} \left(\int_{\mathcal{M}} f\mathrm{d}\mu - \int_{\mathcal{M}} f\mathrm{d}\nu\right). \tag{4.3}$$

In recent years, Wasserstein distances between probability measures have found their applications in various domains, particularly in machine learning and statistics. In machine learning, optimal transport provides a robust framework for comparing probability distributions, which is fundamental in numerous tasks such as generative modeling, domain adaptation, and clustering. For example, the Wasserstein distance, derived from optimal transport theory, has proven to be a powerful metric to measure the similarity between distributions, offering advantages over traditional metrics such as the Kullback-Leibler divergence or total variation distance [10, 47]. Its ability to capture the geometric structure of the data makes it particularly effective in high-dimensional spaces, where data often lie on lower-dimensional manifolds [22]. The significance of optimal transport in statistics is equally profound. It offers novel methods for nonparametric estimation and hypothesis testing [100, 90, 50]. Using the principles of optimal transport, statisticians can tackle a wider range of problems and achieve more interpretable results in areas such as empirical distribution approximation and estimation of dependency structures [37].
In this chapter, we revisit the problem of approximating probability measures with smooth density under the Wasserstein metric, a topic extensively studied in recent literature [118, 90, 37]. Specifically, given a sample consisting of n independent and identically distributed (i.i.d.) random variables drawn from an unknown probability measure μ, our objective is to construct from this sample an estimator $\tilde{\mu}_n$ for μ that attains optimal asymptotic convergence rates with respect to the Wasserstein metric $\mathcal{W}_1(\tilde{\mu}_n,\mu)$ when the sample size n goes to infinity.
Recognizing that convergence rates of empirical measures in Wasserstein distance crucially depend on the dimensional characteristics of the underlying space [46, 90, 37], we divide our analysis into two distinct scenarios. The first scenario addresses the case where the measure μ is absolutely continuous with respect to the Lebesgue measure on the Euclidean space $\mathbb{R}^d$. The second scenario considers the setting in which μ is supported on a low-dimensional space $\mathcal{M}$, which we assume to be a compact d-dimensional manifold without boundary, smoothly embedded in a high-dimensional Euclidean space $\mathbb{R}^m$ $(m>d)$.
Besides, in statistics, the regularity control on density functions is usually expressed in terms of Besov norms [57]. Nevertheless, Besov norms are interpolations of Sobolev norms $H^s_q(\mathcal{M})$ [80, p.152,153]:

$$\|f\|_{H^s_q(\mathbb{R}^d)} = \left(\int_{\mathbb{R}^d} \max_{1\leq i\leq s} \|\nabla^i p(x)\|^q_{\mathrm{op}}\,\mathrm{d}x\right)^{1/q} \quad \text{with } s\in\mathbb{N} \text{ and } f\in\mathcal{C}^\infty(\mathbb{R}^d),$$

$$\|f\|_{H^s_q(\mathcal{M})} = \left(\int_{\mathcal{M}} \max_{1\leq i\leq s} \|\nabla^i p(x)\|^q_{\mathrm{op}}\,\mathrm{d}x\right)^{1/q} \quad \text{with } s\in\mathbb{N} \text{ and } f\in\mathcal{C}^\infty(\mathcal{M}),$$

where in $\mathbb{R}^d$ setting, $\nabla^i p(x) : (\mathbb{R}^d)^i \to \mathbb{R}$ denotes the standard i^{th}-order derivative of p at x, and in manifold setting, $\nabla^i p(x) : (T_x\mathcal{M})^i \to \mathbb{R}$ denotes the i^{th}-order covariant derivatives of p at x (cf. [64, p.6, 21] or Section 1.4.6).
For the sake of simplicity, in this chapter, Sobolev norms are the only measure of regularity for density functions we will use.

Let us begin by examining the Euclidean scenarios $(\mathbb{R}^d, \|\cdot\|_2)$.
Let μ be a probability measure on $\mathbb{R}^d$ with density p with respect to the Lebesgue measure of $\mathbb{R}^d$ and $X_1, X_2, ..., X_n$ be a sample of n i.i.d. random elements sampled from μ. In this setting, we analyze the asymptotic behavior of the kernel measure estimator $\mu_{n,h}$ as the sample size n tends to infinity and the smoothing bandwidth h decreases at an appropriate speed to zero. This kernel estimator is explicitly defined by:

$$\widehat{\mu}_{n,h}(\mathrm{d}y) = \frac{1}{n}\sum_{i=1}^{n} h^{-d} K\left(\frac{\|X_i - y\|_2}{h}\right)\mathrm{d}y, \tag{4.4}$$

where h is the smoothing parameter, and the kernel function $K : \mathbb{R}_+ \to \mathbb{R}$ is bounded, measurable, and supported on $[0, 1]$, satisfying the normalization condition:

$$\int_{\mathbb{R}^d} K(\|x\|_2)\,\mathrm{d}x = 1. \tag{4.5}$$

This method of estimator construction is called kernel smoothing [63, Chapter 6]. Note that, the normalization condition Eq (4.5) implies that

$$\widehat{\mu}_{n,h}(\mathbb{R}^d) = 1, \tag{4.6}$$

regardless of the choice of n and h.
Our primary theoretical contribution in this setting is summarized by the following Theorem 4.1.2 and its Corollary 4.1.3:

Notation 4.1.1 (Modified Vinogradov notations)**.** *Throughout this chapter, for $A \geq 0$ and $B \geq 0$, we use occasionally $A \lesssim_a B$ as shorthand for the inequality $A \leq C_a B$ for some constant C_a depending only on a. The same goes for $A \gtrsim_a B$. [112, p.5]*

Theorem 4.1.2. *Let $k \geq 1$ be an integer, assume $d \geq 3$, and suppose the kernel K is a k-vanishing kernel on $\mathbb{R}^d$ as specified in Definition 4.1.4.*
Then, there is constant C such that for all integers $s \in \{1, 2, \ldots, k-1\}$, any real number $q > d$, and $h \in (0, 1)$, the following bound holds:

$$\mathbb{E}\left(\mathcal{W}_1(\widehat{\mu}_{n,h}, \mu)\right) \leq C\left(\left((M_q(\mu))^{1/2} + 1\right)\frac{h^{1-d/2}}{\sqrt{n}} + \|p\|_{H_1^s(\mathbb{R}^d)} h^{s+1}\right),$$

where the q-th moment of μ is defined as

$$M_q(\mu) := \int_{\mathbb{R}^d} \|x\|_2^q\,\mu(\mathrm{d}x). \tag{4.7}$$

Moreover, the constant factor C can be chosen to depend only on the integers k, q and the uniform norm $\|K\|_\infty := \sup_x |K(x)|$ of K.

Corollary 4.1.3. *Assume that $d \geq 3$. If the density p of μ satisfies that $M_{d+1}(\mu) < \infty$ and $\|p\|_{H_1^s(\mathbb{R}^d)} < \infty$, there exist an explicitly defined kernel measure estimator $\tilde{\mu}_n$ and a constant C such that:*

$$\mathbb{E}(\mathcal{W}_1(\tilde{\mu}_n, \mu)) \leq C \times n^{-\frac{1+s}{d+2s}}, \tag{4.8}$$

where the constant C only depends on $d, \|K\|_\infty, M_{d+1}(\mu)$, and $\|p\|_{H_1^s(\mathbb{R}^d)}$.

We intentionally omit the cases $d = 1$ and $d = 2$, as these dimensions are already fully covered by classical results regarding empirical measure approximations [46]. Besides, $\widehat{\mu}_{n,h}$ is possibly be a signed measure, but this will not affect the definition of $\mathcal{W}_1$ in (4.3).
Finally, we formalize the definition of "k-vanishing kernel" used previously:

Definition 4.1.4 (k-vanishing kernel)**.** *Let k be a positive integer. A kernel function $K : \mathbb{R}_+ \to \mathbb{R}$ is said to be a k-vanishing kernel on $\mathbb{R}^d$ if, for every integer $s \in \{1, 2, \ldots, k\}$, the kernel satisfies:*

$$\int_{\mathbb{R}^d} |K(\|x\|_2)| \, \|x\|_2^s \, \mathrm{d}x < \infty, \quad \textit{and} \quad \int_{\mathbb{R}^d} K(\|x\|_2) \, \|x\|_2^s \, \mathrm{d}x = 0.$$

Within this context, our results partially overlap with those of [118, 90] in the case of compactly supported measures under the 1-Wasserstein metric. For this case, compared to [118, 90], we broaden the existing minimax results to include all probability densities possessing first moments, without restrictions on the boundedness of their support. Besides, the estimators given in [118, 90] are wavelet-based, while our choice is kernel estimators which are generally believed to be more basic [62].
Besides, our estimator is minimax since:

Theorem 4.1.5. *[90, Theorem 3] For any $d \geq 2, s \geq 0$ and constant $C > 0$, we have:*

$$\inf_{\tilde{\mu}} \sup_{\substack{p \text{ is supported in } [0,1]^d \\ \|p\|_{H_1^s(\mathbb{R}^d)} \leq C}} \mathbb{E}(\mathcal{W}_1(\tilde{\mu}, \mu)) \gtrsim_{C,d,s} n^{-\frac{1+s}{d+2s}},$$

where the infimum is taken over the space of all possible measure estimators $\tilde{\mu}$ constructed from a sample of n observations.

We now consider the scenario where the measure μ is supported on a compact manifold $\mathcal{M}$ of dimension $d \geq 3$ (without boundary), smoothly embedded into a Euclidean space $(\mathbb{R}^m, \| \cdot \|_2)$.
Since $\mathcal{M}$ is smoothly embedded in $\mathbb{R}^m$, it inherits a natural Riemannian metric induced by the ambient Euclidean structure. With this metric, $\mathcal{M}$ becomes a Riemannian submanifold. We denote by ρ the geodesic distance associated with this induced metric.
For any probability measure $\mu \in \mathcal{P}(\mathcal{M})$ with density p with respect to the volume measure on $\mathcal{M}$. Let $(X_1, X_2, ..., X_n)$ be a sample of n i.i.d random variables of μ. In this scenario, we investigate the convergence of the kernel estimator $\widehat{\mu}_{n,h}^{\mathcal{M}}$ as the sample size n tends to infinity and the smoothing bandwidth h decreases appropriately to zero:

$$\widehat{\mu}_{n,h}^{\mathcal{M}}(\mathrm{d}y) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h^d} K\left(\frac{\|X_i - y\|_2}{h} \right) \mathrm{d}y, \tag{4.9}$$

where $\| \cdot \|_2$ is the distance with respect to $\mathbb{R}^m$, $\mathrm{d}y$ on the right side represents the volume measure on $\mathcal{M}$, and the kernel function $K : \mathbb{R}_+ \to \mathbb{R}$ is also a measurable bounded function with support in $[0, 1]$ such that satisfies Eq (4.5).
Note that, unlike in the previous scenario (cf. Eq. (4.6)), the mass of $\widehat{\mu}_{n,h}^{\mathcal{M}}$ need not equal 1 in general. To address this issue, the author in [37, p. 7] proposed replacing the kernel K by its pointwise normalized version in the definition of the kernel estimator $\widehat{\mu}_{n,h}^{\mathcal{M}}$. This normalization, however, introduces additional approximation steps and complexity into their analysis. In our treatment, we observe that such normalization may lead to avoidable computational complications. Hence, we retain the original kernel K and instead construct our estimation of μ via $\widehat{\mu}_{n,h}^{\mathcal{M}}$ differently.
Our primary theoretical contribution in this setting is summarized by the following theorem and corollary:

Theorem 4.1.6. *Let $k \geq 1$ be an integer, assume $d \geq 3$, and suppose the kernel K is a k-vanishing kernel on $\mathbb{R}^d$ as specified in Definition 4.1.4. On top of that, we assume K is Lipschitz on $[0,1]$.*
Then, there is a constant C, such that for all integers $s \in \{1,2,\ldots,k-1\}$, $h \in (0,1]$ and n, the following bound holds:

$$\mathbb{E}\left(\mathcal{W}_1(\widehat{\mu}_{n,h}^{\mathcal{M}}, \widehat{\mu}_{n,h}^{\mathcal{M}}(\mathcal{M})\mu)\right) \leq C\left(\frac{h^{1-d/2}}{\sqrt{n}} + \|p\|_{H_1^s(\mathcal{M})} h^{s+1}\right),$$

where the Wasserstein distance is defined as in Eq (4.3).
Moreover, the constant factor C can be chosen to depend only on $\mathcal{M} \subset \mathbb{R}^m$, the integer k, the uniform norm $\|K\|_\infty := \sup_x |K(x)|$ of K, and the Lipschitz constant of $K\big|_{[0,1]}$.

Note that the random mass $\widehat{\mu}_{n,h}^{\mathcal{M}}(\mathcal{M})$ converges to 1 almost surely when h converges to 0, regardless of n (cf. Lemma 4.3.16).

Corollary 4.1.7. *Assume that $d \geq 3$. If the density p of μ satisfies that $\|p\|_{H_1^s(\mathbb{R}^d)} < \infty$, Then there exist an explicitly defined kernel measure estimator $\tilde{\mu}_n$ such that:*

$$\mathbb{E}(\mathcal{W}_1(\tilde{\mu}_n, \mu)) \lesssim_{d, \|p\|_{H_1^s(\mathbb{R}^d)}, s} n^{-\frac{1+s}{d+2s}}, \tag{4.10}$$

Moreover, almost surely,

$$\limsup_{n\to\infty} n^{\frac{1+s}{d+2s}} \mathcal{W}_1(\tilde{\mu}_n, \mu) = \limsup_{n\to\infty} n^{\frac{1+s}{d+2s}} \mathbb{E}(\mathcal{W}_1(\tilde{\mu}_n, \mu)) < \infty. \tag{4.11}$$

In comparison with [37], we emphasize that when focusing on the practically relevant 1-Wasserstein distance [10], the minimax results presented therein remain valid without imposing additional conditions on uniform lower and upper bounds for probability densities. Additionally, we establish that our estimator achieves convergence almost surely at the same rate. To the best of our knowledge, this stronger mode of convergence has not been previously demonstrated.
An additional refinement we introduced compared to [37], though of minimal practical significance, is the relaxed regularity requirement on the kernel function K. This adjustment was made primarily to deepen our theoretical understanding of the problem. More specifically, many calculations in [37] rely on a Taylor expansion up to relatively high order of K, which is a natural approach within the context of manifold learning. However, we have always believed that there must be a deeper geometric rationale behind why this seemingly 'brutal' Taylor expansion is effective.
The remainder of this chapter is organized into two main sections. In Section 4.2, we address the first setting—the minimax convergence rate for measure estimation in $\mathbb{R}^d$. The proofs of Theorem 4.1.2 and Corollary 4.1.3 will be given in this section. Similarly, in Section 4.3, we examine the second setting—the minimax convergence rate for measure estimation on manifolds. The proofs of Theorem 4.1.6 and Corollary 4.1.7 will be given in this section.

4.2 Minimax measure estimation in $\mathbb{R}^d$

The goal of section is to give the proofs for Theorem 4.1.2 and Corollary 4.1.3. For the demonstration of Theorem 4.1.2, our primary idea is to control the approximation error $\mathcal{W}_1(\widehat{\mu}_{n,h}, \mu)$ by treating each term in the majoration:

$$\mathcal{W}_1(\widehat{\mu}_{n,h}, \mu) \leq \underbrace{\mathcal{W}_1(\widehat{\mu}_{n,h}, \widehat{\mu}_h)}_{\text{stochastic error term}} + \underbrace{\mathcal{W}_1(\widehat{\mu}_h, \mu)}_{\text{bias term}}, \tag{4.12}$$

where:

$$\widehat{\mu}_h(\mathrm{d}y) := \left(\int_{\mathbb{R}^d} p(x) h^{-d} K\left(\frac{\|x-y\|_2}{h}\right) \mathrm{d}x\right) \mathrm{d}y. \tag{4.13}$$

Remark 4.2.1. *Informally,* $\widehat{\mu}_h = \mathbb{E}(\widehat{\mu}_{n,h})$.

To avoid repetition, in the rest of this section, we fix an integer $k \geq 1$ and assume that:

Assumption 5. *K is k-vanishing, the dimension d of $\mathcal{M}$ is at least 3, and s is a positive integer smaller than $k-1$.*

Notation 4.2.2 (k-fold duplication)**.** *Given an element e of set E, in the rest of this chapter, we denote by $e^{\times k}$ the element $\underbrace{(e, ..., e)}_{k \text{ times}}$ of the cartesian product E^k.*

Let us begin with giving an upper bound for the bias term in Eq (4.12).

4.2.1 Bias term estimation

Firstly, we prove that:

Theorem 4.2.3. *There is a constant C depending only on k and $\|K\|_\infty$ such that for all compactly supported smooth function p, 1-Lipschitz function f, and $s \in [0, k-1]$ we have:*

$$\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \frac{1}{h^d} K\left(\frac{\|x-y\|_2}{h}\right) f(x)(p(x)-p(y))\mathrm{d}x\mathrm{d}y \leq Ch^{s+1} \times \|p\|_{H_1^s}. \tag{4.14}$$

Proof. For the sake of simplicity, we let $p^{(s)}$ denote $\nabla^s p$.
We consider first the case when s is even.
By Taylor's expansion up to the order s, we see that:

$$\begin{aligned}\int_{y\in\mathbb{R}^d} K\left(\frac{\|x-y\|}{h}\right) f(x)(p(y)-p(x))\mathrm{d}y &= \sum_{j=1}^{s}\int_{y\in\mathbb{R}^d} \frac{1}{j!}K\left(\frac{\|x-y\|}{h}\right) f(x)p^{(j)}(x)\left[(y-x)^{\times j}\right]\mathrm{d}y \\ + \int_{\lambda=0}^{1}\left(\int_{y\in\mathbb{R}^d}\frac{(1-\lambda)^{s-1}}{(s-1)!}K\left(\frac{\|x-y\|}{h}\right)\right.&\left. f(x)\left(p^{(s)}(x+\lambda(y-x)) - p^{(s)}(x)\right)\left[(y-x)^{\times s}\right]\mathrm{d}y\right)\mathrm{d}\lambda.\end{aligned} \tag{4.15}$$

Because K is k vanishing for $\mathbb{R}^d$, for every $1 \leq j \leq s \leq k-1$,

$$\int_{y\in\mathbb{R}^d} \frac{1}{j!}K\left(\frac{\|x-y\|_2}{h}\right) f(x)p^{(j)}(x)\left[(y-x)^{\times j}\right]\mathrm{d}y = 0,$$

where $(y-x)^{\times j} \in (\mathbb{R}^d)^j$ denotes the k-fold duplication of $y-x$(cf. Notation 4.2.2).
Besides, by the change of variables: $y \to \frac{z-(1-\lambda)x}{\lambda}$, we observe that:

$$\begin{aligned}&\int_{y\in\mathbb{R}^d} K\left(\frac{\|x-y\|}{h}\right) f(x)\left(p^{(s)}(x+\lambda(y-x)) - p^{(s)}(x)\right)\left[(y-x)^{\times s}\right]\mathrm{d}y \\ &= \int_{z\in\mathbb{R}^d} K\left(\frac{\|x-z\|}{\lambda h}\right) f(x)\left(p^{(s)}(z) - p^{(s)}(x)\right)\left[(z-x)^{\times s}\right]\frac{1}{\lambda^d}\mathrm{d}z.\end{aligned} \tag{4.16}$$

Let

$$A(\lambda) := \int_{x\in\mathbb{R}^d}\int_{y\in\mathbb{R}^d} K\left(\frac{\|x-z\|}{\lambda h}\right) f(x)\left(p^{(s)}(z) - p^{(s)}(x)\right)\left[(z-x)^{\times s}\right]\frac{1}{\lambda^d}\mathrm{d}z\mathrm{d}x.$$

Because s is even, by interchanging the role of z and x, we see that:

$$\begin{aligned}A(\lambda) &= -\int_{x\in\mathbb{R}^d}\int_{y\in\mathbb{R}^d} K\left(\frac{\|x-z\|}{\lambda h}\right) f(z)\left(p^{(s)}(z) - p^{(s)}(x)\right)\left[(z-x)^{\times s}\right]\frac{1}{\lambda^d}\mathrm{d}z\mathrm{d}x \\ &= \frac{1}{2}\int_{x\in\mathbb{R}^d}\int_{y\in\mathbb{R}^d} K\left(\frac{\|x-z\|}{\lambda h}\right) (f(x)-f(z))\left(p^{(s)}(z) - p^{(s)}(x)\right)\left[(z-x)^{\times s}\right]\frac{1}{\lambda^d}\mathrm{d}z\mathrm{d}x.\end{aligned}$$

On top of that, K is supported in $[0,1]$ and f is 1-Lipschitz. Therefore,

$$\begin{aligned}|A(\lambda)| &\leq \frac{1}{2}\|K\|_\infty \int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{\mathbf{1}_{\|x-z\|_2\leq\lambda h}}{\lambda^d}\|x-z\|^{s+1}\Big(\|p^{(s)}(x)\|_{op}+\|p^{(s)}(z)\|_{op}\Big)\mathrm{d}x\mathrm{d}z\\ &= \|K\|_\infty\Big(\int_{\mathbb{R}^d}\|p^{(s)}(x)\|_{op}\mathrm{d}x\Big)\Big(\int_{z\in\mathbb{R}^d}\frac{1_{|z|\leq\lambda h}}{\lambda^d}|z|^{s+1}\mathrm{d}z\Big)\\ &= h^{s+1}\times\|K\|_\infty\Big(\int_{\mathbb{R}^d}\|p^{(s)}(x)\|_{op}\mathrm{d}x\Big)\Big(\lambda^{s+1}\int_{z\in\mathbb{R}^d}1_{|z|\leq 1}|z|^{s+1}\mathrm{d}z\Big).\end{aligned}$$

Therefore, we have the desired conclusion for s even.
Now, we consider the case when s is odd, which means that $s-1, s+1$ are both even. Therefore, after we have shown previously, there is a constant C such that for all f 1-Lipschitz, we have:

$$\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\frac{1}{h^d}K\Big(\frac{\|x-y\|_2}{h}\Big)f(x)(p(x)-p(y))\mathrm{d}x\mathrm{d}y \leq Ch^s\times\|p\|_{H_1^{s-1}}, \tag{4.17}$$

and

$$\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\frac{1}{h^d}K\Big(\frac{\|x-y\|_2}{h}\Big)f(x)(p(x)-p(y))\mathrm{d}x\mathrm{d}y \leq Ch^{s+2}\times\|p\|_{H_1^{s+1}}. \tag{4.18}$$

Fix a 1-Lipschitz function f, we consider the linear mapping:

$$\begin{aligned}T:\mathcal{C}_c^\infty(\mathbb{R}^d) &\longrightarrow \mathbb{R}\\ p &\longmapsto \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\frac{1}{h^d}K\Big(\frac{\|x-y\|_2}{h}\Big)f(x)(p(x)-p(y)).\end{aligned}$$

We have that for all $p\in\mathcal{C}_c^\infty(\mathbb{R}^d)$,

$$|Tp|\leq Ch^{s+2}\times\|p\|_{H_1^{s+1}(\mathbb{R}^d)}\ ,\ |Tp|\leq Ch^s\times\|p\|_{H_1^{s-1}(\mathbb{R}^d)}.$$

Then, after Theorem 4.1.2 in [80], for all p,

$$|Tp|\leq Ch^{s+1}\|p\|_{[H^{s-1}(\mathbb{R}^d),H^{s+11}(\mathbb{R}^d)]_{1/2}},$$

where $[H^{s-1}(\mathbb{R}^d),H^{s+11}(\mathbb{R}^d)]_{1/2}$ the interpolation space of $H^{s-1}(\mathbb{R}^d)$ with $H^{s+11}(\mathbb{R}^d)$ with exponent $1/2$.
Besides, after [80, Theorem 6.4.5], $H_1^s(\mathbb{R}^d)\simeq[H^{s-1}(\mathbb{R}^d),H^{s+11}(\mathbb{R}^d)]_{1/2}$, which implies that there is a constant α such that for all $p\in\mathcal{C}^\infty$:

$$\|p\|_{[H^{s-1}(\mathbb{R}^d),H^{s+11}(\mathbb{R}^d)]_{1/2}}\leq\alpha\|p\|_{H_1^s(\mathbb{R}^d)}.$$

Therefore, we have the desired conclusion for s odd, which finishes the proof. □

Remark 4.2.4. *For the case where s is odd, an alternative approach to the above proof involves establishing an upper bound for*

$$\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\mathbb{1}_{\|x-y\|\leq h}\|p^{s-1}(x)-p^{s-1}(y)\|_{op}\|x-y\|_2^{s-1}.$$

However, we think that carrying out this approach rigorously is rather tedious. Instead, we opted for the interpolation argument presented above, which provides a more straightforward and efficient justification.

Now, thanks to the above theorem, we have the following estimation for our bias term.

Corollary 4.2.5. *Let k be an integer $k\geq 1$ and assume that $d\geq 3$. Then, there is a constant C depending only on $\|K\|_\infty$ and k such that for all integer $s\in[0,k-1]$,*

$$\mathcal{W}_1(\widehat{\mu}_h,\mu)\leq C\times h^{s+1}\times\|p\|_{H_1^s}. \tag{4.19}$$

4.2.2 Stochastic error term estimation

In this section, we treat the stochastic error in approximation error $\mathcal{W}_1(\widehat{\mu}_{h,n}, \widehat{\mu}_h)$.
We begin with recalling a few results on Green functions in $\mathbb{R}^d$ which we will use intensively in this section.

4.2.2.1 Premilinaries on Green functions in $\mathbb{R}^d$

Theorem 4.2.6. *[11] The Green's function $G(x, y)$ for $\mathbb{R}^d$ with the Laplace operator Δ is a fundamental solution to the Poisson equation:*

$$\Delta G(x, y) = \delta(x - y) ,$$

where δ is Dirac at 0. The explicit form of the Green's function depends on the dimension d:

$$\begin{aligned} G(x, y) &= -\frac{1}{2}|x - y| && d = 1, \\ G(x, y) &= -\frac{1}{2\pi}\log \|x - y\|_2 && d = 2, \\ G(x, y) &= \frac{1}{(d-2)\omega_d} \cdot \frac{1}{\|x - y\|_2^{d-2}} && d \geq 3, \end{aligned}$$

where ω_d is the surface area of the unit sphere in $\mathbb{R}^d$, given by $\omega_d = \frac{2\pi^{d/2}}{\Gamma(d/2)}$.

Then, for any function $f \in L^1 \cap L^\infty$, we denote by $Gf(x)$ the integral:

$$Gf(x) := \int_{\mathbb{R}^d} G(x, y)f(y)\mathrm{d}y. \tag{4.20}$$

The following lemma will be useful to our estimation of functions of the form ∇Gf.

Lemma 4.2.7. *There is a constant C' depending on d such that for all z,*

$$\left|\int_{\mathbb{R}^d} \frac{h^{-d}}{\omega_d}\frac{(x-y)}{\|x-y\|_2^d}K\left(\frac{\|z-x\|_2}{h}\right)\mathrm{d}x\right| \leq C' \times \|K\|_\infty \times \left(h^{1-d}\mathbf{1}_{\|y-z\|_2 \leq 2h} + \|y-z\|_2^{1-d}\mathbf{1}_{\|y-z\|_2 > 2h}\right). \tag{4.21}$$

Proof. If $\|y - z\|_2 \leq 2h$, we observe that there are two possible situations for x: Either $\|z - x\|_2 \leq h$ or $\|z - x\|_2 > h$. For the first situation, $\|y - x\|_2 \leq 3h$. For the second situation, $K\left(\frac{\|z-x\|_2}{h}\right) = 0$.
Therefore, if $\|y - z\|_2 \leq 2h$, we have:

$$\begin{aligned} \left|\int_{\mathbb{R}^d} \frac{h^{-d}}{\omega_d}\frac{(x-y)}{\|x-y\|_2^d}K\left(\frac{\|z-x\|_2}{h}\right)\mathrm{d}x\right| &\leq h^{-d}\frac{1}{\omega_d} \times \|K\|_\infty \times \int_{B_{\mathbb{R}^d}(y,3h)} \frac{1}{\|x-y\|_2^{d-1}}\mathrm{d}x \\ &= (\omega_d)^{-1}\|K\|_\infty h^{1-d}\int_{B_{\mathbb{R}^d}(0,3)} \frac{1}{\|x\|_2^{d-1}}\mathrm{d}x \\ &= (\omega_d)^{-1}\|K\|_\infty h^{1-d}\int_{\mathbb{S}^{d-1}}\int_0^3 \frac{1}{r^{d-1}}r^{d-1}\mathrm{d}r\mathrm{d}\theta \\ &= 3\|K\|_\infty h^{1-d}. \end{aligned}$$

If $\|y - z\|_2 > 2h$, again, there are two possible situations for x: Either $\|z - x\|_2 \leq h$ or $\|z - x\|_2 > h$. For the first situation, $\|x - y\|_2 \leq \|z - y\|_2 - \|x - z\|_2 \geq \frac{1}{2}\|y - z\|_2$. For the second situation, $K\left(\frac{\|z-x\|_2}{h}\right) = 0$.
Therefore, we have the desired conclusion for $\|y - z\|_2 > 2h$. Thus, we have finished the proof. □

4.2.2.2 Upper bounds for Wasserstein distance using Green's function

The following proposition gives the first upper bounder for 1-Wasserstein distance using Green operator.

Proposition 4.2.8. *For any two signed measures μ_1 and μ_2 on $\mathbb{R}^d$ with bounded densities p_1 and p_2 such that $\mu_1(\mathbb{R}^d) = \mu_2(\mathbb{R}^d)$. If μ_1 and μ_2 are supported in a bounded domain $B \subset \mathbb{R}^d$ with $\mathcal{C}^1$ boundary, we have:*

$$\mathcal{W}_1(\mu_1, \mu_2) \leq \int_B |\nabla G(p_1 - p_2)| \mathrm{d}x + \mathrm{diam}(B) \int_{\partial B} |\nabla G(p_1 - p_2)| \mathrm{d}x,$$

where $\mathrm{diam}(B) := \sup_{x,y \in B} \|x - y\|_2$ *is the diameter of B and ∂B is the boundary of B.*

Remark 4.2.9. *Note that this inequality becomes an equality when $B = [0, 1]$.*

Proof of Proposition 4.2.8. For any f continuously differentiable and any point $o \in B$, by Green-Ostrogradsky's theorem:

$$\begin{aligned}\int_B f(p_1 - p_2)\mathrm{d}x &= \int_B (f - f(o))(p_1 - p_2)\mathrm{d}x \\ = -\int_B \langle \nabla f, \nabla G(p_1 - p_2)\rangle \mathrm{d}x &+ \int_{\partial B} (f - f(o))\langle \nabla G(p_1 - p_2), n\rangle \mathrm{d}x,\end{aligned}$$

where n is the outward pointing vector of ∂B. The desired conclusion follows directly. □

While the above proposition is sharp, to simply our calculations, we will use the following proposition:

Proposition 4.2.10. *For any two signed measures μ_1 and μ_2 on $\mathbb{R}^d$ with bounded densities p_1 and p_2 such that $\mu_1(\mathbb{R}^d) = \mu_2(\mathbb{R}^d)$. If μ_1 and μ_2 are supported in a bounded domain $B \subset \mathbb{R}^d$ with $\mathcal{C}^1$ boundary, we have:*

$$\mathcal{W}_1(\mu_1, \mu_2) \leq \int_{B^*} |\nabla G(p_1 - p_2)| \mathrm{d}x,$$

where

$$B^* := B + B_{\mathbb{R}^d}(0, 2\mathrm{diam}B) = \{x + z : x \in B, |z| \leq 2\mathrm{diam}B\}. \tag{4.22}$$

Proof. Fix a point $o \in B$. For any f 1-Lipschitz on B such that $f(o) = \mathrm{diam}(B)$, we denote by f^* a function on B^* defined as follows:

$$f^*(y) = (\max_{x \in B}(f(x) - \|x - y\|_2))_+ \quad \forall y \in B^*.$$

Because f is 1-Lipschitz, $\max_{x \in B}(f(x) - \|x - y\|_2) = f(y)$ for all $y \in B$. Indeeds, for all x and y in B, we have that:

$$f(y) = f(x) + f(y) - f(x) \geq f(x) - \|x - y\|_2.$$

Thus,

$$f^* \equiv f \text{ on } B.$$

Also, $f(o) = \mathrm{diam}(B)$, $f(x)$ is positive on B. Moreover, for any $y \in \partial B^*$, $\min_{x \in B} \|y - x\|_2 = 2\mathrm{diam}(B)$. Therefore, $\max_{x \in B}(f(x) - \|x - y\|_2) \leq 0$ for all $y \in \partial B^*$. In other words,

$$f^* = 0 \text{ on } \partial B.$$

Clearly, by construction, f^* is 1-Lipschizt on B^*.
Thus, by Green–Ostrogradsky's theorem, we have that for all f 1-Lipschitz on B such that $f(o) = \mathrm{diam}(B)$, we have:

$$\begin{aligned}
&\int_B f(p_1 - p_2)\mathrm{d}x = \int_{B^*} f^*(p_1 - p_2)\mathrm{d}x \\
&= -\int_{B^*} \langle \nabla f^*, \nabla G(p_1 - p_2)\rangle \mathrm{d}x + \int_{\partial B^*} f^* \langle \nabla G(p_1 - p_2), n\rangle \mathrm{d}x, \\
&= -\int_{B^*} \langle \nabla f^*, \nabla G(p_1 - p_2)\rangle \mathrm{d}x \leq \int_{B^*} |\nabla G(p_1 - p_2)|\mathrm{d}x.
\end{aligned}$$

Therefore, the conclusion follows. □

4.2.2.3 An upper bound for the stochastic term

Now, let us present the main result of this section which is:

Theorem 4.2.11. *There is a constant C depending only on d, q $(q > d)$ such that for all $h \leq 1$,*

$$\mathbb{E}(\mathcal{W}_1(\widehat{\mu}_{n,h}, \widehat{\mu}_h)) \leq_{d,q} \left(\sqrt{M_q(\mu)} + 1\right) \times \frac{1}{\sqrt{n}} \times h^{1-d/2}.$$

To prove this, we first treat a weakened version of the above theorem as a Lemma:

Lemma 4.2.12. *Suppose that the underlying probability measure μ is supported in a bounded open set $B \subset \mathbb{R}^d$ with $\mathcal{C}^1$ boundary. Then for all $q > 1$, we have:*

$$\mathbb{E}(\mathcal{W}_1(\widehat{\mu}_{n,h}, \widehat{\mu}_h)) \lesssim_d C'' \times \|K\|_\infty \times \sqrt{|B^*|} \times \frac{1}{\sqrt{n}} \times h^{1-d/2},$$

where $B^* := B + B_{\mathbb{R}^d}(0, 2\mathrm{diam}(B))$. (cf. (4.22))

Proof of Lemma 4.2.12. For each x, we define :

$$p_{x,h}(y) := \frac{1}{h^d} K\left(\frac{\|x-y\|_2}{h}\right), \; p_h(y) := \left(\int_{\mathbb{R}^d} K\left(\frac{\|x-y\|_2}{h}\right) p(x)\mathrm{d}x\right)\mathrm{d}y. \tag{4.23}$$

Recall that by definition of $\widehat{\mu}_{n,h}$, we have:

$$n\widehat{\mu}_{n,h}(\mathrm{d}y) = \frac{1}{n}\left(\sum_{i=1}^n p_{X_i,h}(y)\right)\mathrm{d}y,$$

Thus, after Proposition 4.2.10 and Hölder's inequality, we have:

$$\begin{aligned}
\mathbb{E}(\mathcal{W}_1(\widehat{\mu}_{n,h}, \widehat{\mu}_h)) &\leq \mathbb{E}\left(\int_{B^*} \left|\nabla G\left(\frac{1}{n}(p_{X_1,h} + \cdots + p_{X_n,h}) - p_h\right)\right|(y)\mathrm{d}y\right) \\
&\leq \sqrt{|B^*|}\mathbb{E}\left(\sqrt{\int_{B^*} \left|\nabla G\left(\frac{1}{n}(p_{X_1,h} + \cdots + p_{X_n,h}) - p_h\right)\right|^2(y)\mathrm{d}y}\right) \\
&\leq \sqrt{|B^*|}\sqrt{\int_{B^*} \frac{1}{n}\mathbb{E}\Big(|\nabla G(p_{X_1,h})|^2(y)\Big)\mathrm{d}y}.
\end{aligned}$$

Besides, after Lemma 4.2.7, we have that for each y,

$$\begin{aligned}
&\int_{B^*} \mathbb{E}\Big(|\nabla G(p_{X_1,h})|^2(y)\Big)\mathrm{d}y \\
&\lesssim_d \int_{B^*} \times \|K\|_\infty \times \Big(h^{2-2d}\mathbb{P}(|X_1 - y| \leq 2h) + \mathbb{E}(|X_1 - y|^{2-2d} \times \mathbf{1}_{|X_1-y|\geq 2h})\Big)\mathrm{d}y \\
&\lesssim_d \int_{\mathbb{R}^d} (C')^2 \times \|K\|_\infty \times \Big(h^{2-2d}\mathbb{P}(|X_1 - y| \leq 2h) + \mathbb{E}(|X_1 - y|^{2-2d} \times \mathbf{1}_{|X_1-y|\geq 2h})\Big)\mathrm{d}y.
\end{aligned}$$

Then, by using Fubini to calculate first the integral with respect to y, we obtain that:

$$\int_{B^*} \mathbb{E}\Big(|\nabla G(p_{X_1,h})|^2(y)\Big)\mathrm{d}y$$
$$\lesssim_d \|K\|_\infty \times \Big(h^{2-2d} \times \frac{\omega_d}{d} \times (2h)^d + \omega_d \times (2h)^{2-d}\Big) \lesssim_d \|K\|_\infty \times h^{2-d}.$$

Thus, we have the desired conclusion. □

We can now proceed to prove Theorem 4.2.11. This proof is inspired by [46].

Proof of Theorem 4.2.11. As long as we can prove that C does not depend on the choice of p, we can assume without loss of generality that p is smooth for simplicity.
We fix for now the window size h and the number n- the size of random sample- and decompose firstly $\mathbb{R}^d$ into

$$\mathbb{R}^d = \bigsqcup_{j=0}^{\infty} B_j, \tag{4.24}$$

where $B_0 = B_{\mathbb{R}^d}(0,1)$- the ball centered at 0 with radius 1 of $\mathbb{R}^d$, and

$$B_j = B_{\mathbb{R}^d}(0,2^j)\backslash B_{\mathbb{R}^d}(0,2^{j-1}), \tag{4.25}$$

for all $j \geq 1$.(see Figure 4.1).

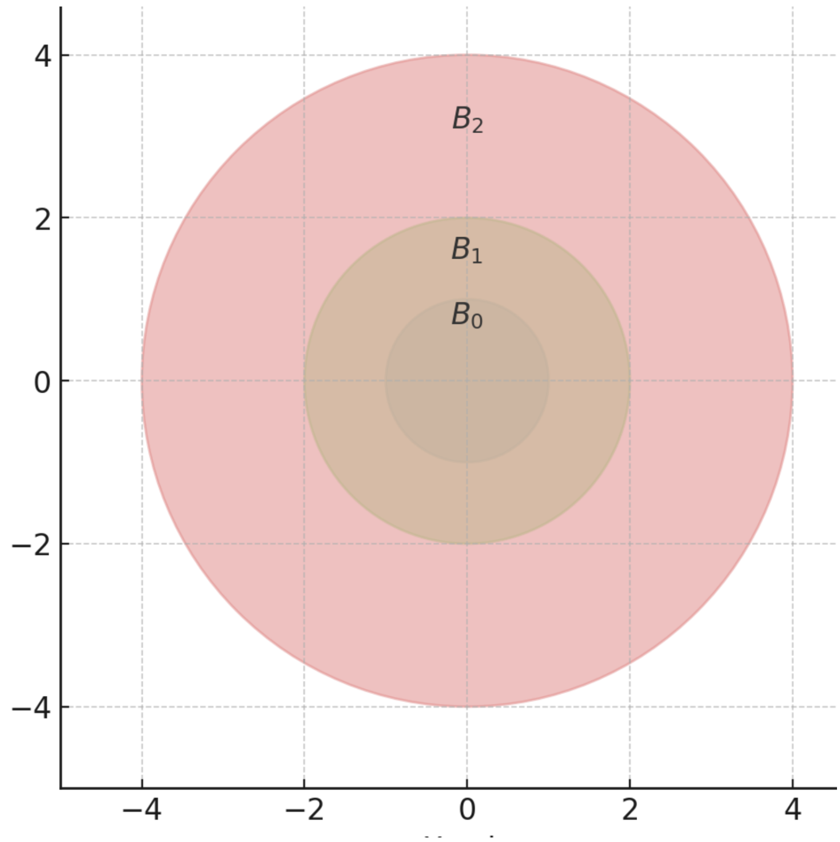


Figure 4.1: 2D Partition

For each $j \geq 0$, let N_j and $X_1^j, ..., X_{N_j}^j$ denote respectively the number of random points (X_i) lying in the region B_j and these random points. Note that when N_j are fixed, in other words, under an event $\{N_0 = n_0, N_1 = n_1, ...\}$, these random points X_i^j are independant and for each j, $(X_1^j, ..., X_{N_j}^1)$ are i.i.d. with distribution μ^j (see Definition (4.26)).

Then, for each j, we denote by $\mu^j, \widehat{\mu}_h^j$ the measures:

$$\mu^j := \mu(\cdot|B_j) = \frac{\mu(\mathrm{d}x)\mathbf{1}_{B_j}}{\mu(B_j)}. \tag{4.26}$$

$$\mu_{x,h}(\mathrm{d}y) := \frac{1}{h^d} K\left(\frac{\|x-y\|_2}{h}\right)\mathrm{d}y. \tag{4.27}$$

$$\widehat{\mu}_h^j := \mathbb{E}(\mu_{X_1^j,h}). \tag{4.28}$$

Without loss of generality, for the sake of well-definedness, we assume that $\mu(B_j) > 0$ for all j. Recall that $n\widehat{\mu}_{n,h} = \mu_{X_1,h} + \mu_{X_2,h} + \cdots + \mu_{X_n,h}$. Thus,

$$n\widehat{\mu}_{n,h} = \sum_{j\geq 0}\left(\sum_{i=1}^{N_j} \mu_{X_i^j,h}\right). \tag{4.29}$$

Hence, due to the triangle inequality and the convexity of Wasserstein distance, we have:

$$\begin{aligned}\mathcal{W}_1(\widehat{\mu}_{n,h}, \widehat{\mu_h}) &\leq \mathcal{W}_1\left(\widehat{\mu}_{n,h}, \frac{\sum_{j\geq 0} N_j \hat{\mu}_h^j}{n}\right) + \mathcal{W}_1\left(\frac{\sum_{j\geq 0} N_j \hat{\mu}_h^j}{n}, \widehat{\mu}_h\right)\\ &\leq \underbrace{\sum_{j\geq 0} \frac{N_j}{n}\mathcal{W}_1\left(\frac{\mu_{X_1^j,h} + \mu_{X_2^j,h} + \ldots + \mu_{X_{N_j}^j,h}}{N_j}, \widehat{\mu}_h^j\right)}_{=A} + \underbrace{\mathcal{W}_1\left(\frac{\sum_{j\geq 0} N_j \hat{\mu}_h^j}{n}, \widehat{\mu}_h\right)}_{=B}.\end{aligned} \tag{4.30}$$

We will treat each term A and B separately.

Firstly, for A, after Lemma 4.2.12 and Hölder's inequality, we know that:

$$\begin{aligned}\mathbb{E}(A) &\lesssim_d \|K\|_\infty \times \mathbb{E}\left(\sum_{j\geq 0}\sqrt{\frac{N_j}{n}} \times \sqrt{|B_j^*|}\right) \times \frac{1}{\sqrt{n}} \times h^{1-d/2}\\ &\leq \|K\|_\infty \times \left(\sum_{j\geq 0}\sqrt{\mu(B_j)} \times \sqrt{|B_j^*|}\right) \times \frac{1}{\sqrt{n}} \times h^{1-d/2}\\ &\leq \|K\|_\infty \times \sqrt{\sum_{j\geq 0} 2^{jq}\mu(B_j)} \times \sqrt{2^{-jq}|B_j^*|} \times \frac{1}{\sqrt{n}} \times h^{1-d/2}.\end{aligned}$$

And by our choice of (B_j) and $q > d$,

$$\sum_{j\geq 0} 2^{jq}\mu(B_j) \leq 1 + 2^q M_q(\mu),$$

$$\sum_{j\geq 0} 2^{-jq}|B_j^*| \leq \sum_{j\geq 0} 2^{-jq}(3\times 2^q)^d = \frac{3^d}{1-2^{d-q}}.$$

Hence, we have bounded correctly A. Now for B, we observe that:

$$\widehat{\mu}_h = \mathbb{E}(\widehat{\mu}_{n,h}) = \sum_{j\geq 0}\mathbb{E}\left(\frac{N_j}{n} \times \widehat{\mu}_h^j\right) = \sum_{j\geq 0}\mu(B_j)\widehat{\mu}_h^j.$$

Besides, for all 1-Lipschitz function f and measure ν such that $\nu(\mathbb{R}^d) = 0$, we have:

$$\int f\nu = \int (f(x) - f(0))\nu(\mathrm{d}x) \leq \int \|x\|_2|\nu|(\mathrm{d}x).$$

Therefore,

$$\begin{aligned}\mathbb{E}(B) &\leq \mathbb{E}\left(\int_{\mathbb{R}^d}\|x\|_2\left(\sum_{j\geq 0}\left|\frac{N_j}{n}-\mu(B_j)\right||\widehat{\mu}_h^j|(\mathrm{d}x)\right)\right)\\ &= \sum_{j\geq 0}\left(\mathbb{E}\left|\frac{N_j}{n}-\mu(B_j)\right|\right)\left(\int_{\mathbb{R}^d}\|x\|_2|\mu_h^j|(\mathrm{d}x)\right) \leq \sum_{j\geq 0}\frac{1}{\sqrt{n}}\sqrt{\mu(B_j)}\int_{\mathbb{R}^d}\|x\|_2|\mu_h^j|(\mathrm{d}x).\end{aligned}$$

On top of that,

$$\begin{aligned}\int_{\mathbb{R}^d}\|x\|_2|\mu_h^j|(\mathrm{d}x) &\leq \int_{\mathbb{R}^d}\|x\|_2\left|\frac{1}{\mu(B_j)}\int_{\mathbb{R}^d}h^{-d}K\left(\frac{\|x-y\|_2}{h}\right)p(y)\mathbb{1}_{y\in B_j}\mathrm{d}y\right|\mathrm{d}x\\ &=\frac{1}{\mu(B_j)}\int_{B_j}\left(h^{-d}\int_{\mathbb{R}^d}\|x+y\|_2\left|K\left(\frac{\|x\|_2}{h}\right)\right|\mathrm{d}x\right)p(y)\mathrm{d}y\\ &\lesssim_d \frac{1}{\mu(B_j)}\times\|K\|_\infty\times\int_{B_j}(h+\|y\|_2)p(y)\mathrm{d}y\\ &\lesssim_d \frac{1}{\mu(B_j)}\times\|K\|_\infty\times\int_{B_j}(h+2^j)p(y)\mathrm{d}y = \|K\|_\infty\times(h+2^j).\end{aligned}$$

Hence,

$$\begin{aligned}\mathbb{E}(B) &\lesssim_d \frac{1}{\sqrt{n}}\sum_{j\geq 0}\sqrt{\mu(B_j)(h^2+2^{2j})} \lesssim_d \frac{1}{\sqrt{n}}\sqrt{\left(\sum_{j\geq 0}2^{-j}\right)\left(\sum_{j\geq 0}2^j h^2\mu(B_j)+2^{3j}\mu(B_j)\right)}\\ &\lesssim_d \frac{1}{\sqrt{n}}\sqrt{1+h^2M_1(\mu)+M_3(\mu)},\end{aligned}$$

where $M_1(\mu)$ and $M_3(\mu)$ are respectively the first order moment and the third order moment of μ (cf. Eq. (4.7)). Therefore, we imply the desired conclusion. □

Proof of Theorem 4.1.2. This main theorem follows directly from Theorem 4.2.11 and Corollary 4.2.5. □

Proof of Corollary 4.1.7. This Corollary follows directly from Theorem 4.1.2 by choosing $h := h_n = n^{-\frac{1}{s+d/2}}$. □

4.3 Minimax measure estimation on manifolds

The aim of this section is to establish Theorem 4.1.6 and Corollary 4.1.7.
Since the framework of this section is independent of that in Section 4.2, and for the sake of simplicity (although by abuse of notation), we shall refer to $\widehat{\mu}_{n,h}^{\mathcal{M}}$ simply as $\widehat{\mu}_{n,h}$ for the remainder of this chapter.
Let $\mathcal{F}$ denote the space of all 1 Lipschitz functions $f : \mathcal{M} \to \mathbb{R}$ on $\mathcal{M}$ such that $\int_{\mathcal{M}} f(x)\mathrm{d}x = 0$.
For any Borel signed measure ν on $\mathcal{M}$, we define $\|\nu\|_{\mathcal{F}}$ to be:

$$\|\nu\|_{\mathcal{F}} := \sup_{f\in\mathcal{F}}|\nu(f)|. \tag{4.31}$$

Thus, $\mathcal{W}_1(\widehat{\mu}_{n,h}, \widehat{\mu}_{n,h}(\mathcal{M})\mu) = \|\widehat{\mu}_{n,h} - \widehat{\mu}_{n,h}(\mathcal{M})\mu\|_{\mathcal{F}}$.
Therefore, by triangle inequality, we have:

$$\begin{aligned}\mathcal{W}_1(\widehat{\mu}_{n,h}, \widehat{\mu}_{n,h}(\mathcal{M})\mu) &\leq \|\widehat{\mu}_{n,h}-\widehat{\mu}_h\|_{\mathcal{F}} + \|\widehat{\mu}_h(\mathcal{M})\mu - \widehat{\mu}_{n,h}(\mathcal{M})\mu\|_{\mathcal{F}} + \|\widehat{\mu}_h - \widehat{\mu}_h(\mathcal{M})\mu\|_{\mathcal{F}}\\ &\leq \underbrace{\|\widehat{\mu}_{n,h}-\widehat{\mu}_h\|_{\mathcal{F}}}_{\text{stochastic error}} + |\widehat{\mu}_h(\mathcal{M}) - \widehat{\mu}_{n,h}(\mathcal{M})|\|\mu\|_{\mathcal{F}} + \underbrace{\|\widehat{\mu}_h - \widehat{\mu}_h(\mathcal{M})\mu\|_{\mathcal{F}}}_{\text{bias}},\end{aligned} \tag{4.32}$$

where:

$$\widehat{\mu}_h(\mathrm{d}y) := \left(\int_{\mathcal{M}} h^{-d} K\left(\frac{\|x-y\|_2}{h}\right) p(x)\mathrm{d}x\right)\mathrm{d}y.$$

Note that, by definition of $\|\cdot\|_{\mathcal{F}}$, $|\widehat{\mu}_h(\mathcal{M}) - \widehat{\mu}_{n,h}(\mathcal{M})| \leq \|\widehat{\mu}_{n,h} - \widehat{\mu}_h\|_{\mathcal{F}}$. Therefore, to give a convergence rate for $\mathcal{W}_1(\widehat{\mu}_{n,h}, \widehat{\mu}_{n,h}(\mathcal{M})\mu)$, it is sufficient to control the stochastic error term and the bias term in the right side of the above inequality, as an analogue to what we did in the previous scenario.
This section is then organized as follows:

- In Section 4.3.1, we introduce a technical geometric result—Theorem 4.3.3. Although its proof is somewhat lengthy, this theorem provides us various insights into many manifold-related calculations, as discussed in Section 4.1.
- In Section 4.3.2, we present a preliminary analysis of the term $\mathcal{W}_1(\widehat{\mu}_{n,h}, \widehat{\mu}_{n,h}(\mathcal{M})\mu)$ and provide an estimate for the "bias" component of this term.
- In Section 4.3.3.1, we derive an estimate for the "stochastic error" component of the same term.
- Finally, in Section 4.3.4, we prove Theorem 4.1.6 and Corollary 4.1.7.

Let us begin with the technical result.

4.3.1 Morse lemma and its extension for manifold estimations

Recall the standard setting of a submanifold $\mathcal{M}$ embedded in a Euclidean space $\mathbb{R}^m$. There are two natural ways to measure distances between points on $\mathcal{M}$. One is the *geodesic distance*, which is coherent with the intrinsic geometry of the manifold, and the other is the *Euclidean distance* provided by the ambient space $\mathbb{R}^m$. While the geodesic distance is better aligned to the underlying structure of $\mathcal{M}$, the Euclidean distance is often preferred in practice due to its practically computational advantage.
However, this practical choice is not without drawbacks. Since the Euclidean distance is not intrinsic to the manifold $\mathcal{M}$, it can introduce certain difficulties in proving the robustness of estimators built on it. This discrepancy between the computational framework and the geometric nature of the data may lead various additional computational complexity.
In this section, we present an attempt to mitigate this geometric inconsistency, by proposing a family of coordinate systems on $\mathcal{M}$ whose construction is based on the Euclidean distance of $\mathbb{R}^m$. This family of coordinate systems is expected to be an analog, which translates better the Euclidean distance in local computations, to the standard normal coordinates systems on $\mathcal{M}$.
More precisely, we show that:

Theorem 4.3.1 (Morse coordinate system)**.** *Let $\mathcal{M} \subset \mathbb{R}^m$ be a compact Riemannian submanifold without border of $\mathbb{R}^m$.*
Then, there is a radius $r > 0$ such that for each point $o \in \mathcal{M}$, there are an open subset U_o of $\mathcal{M}$ and a local coordinate system $x_o = (x_o^1, ..., x_o^d) : U_o \to B_{\mathbb{R}^d}(0, r)$ centered at o such that:

i. $\|p - o\|_2^2 = \left(x_o^1(p)\right)^2 + \left(x_o^2(p)\right)^2 + ... + \left(x_o^d(p)\right)^2$ *for every point $p \in U_o$.*

Moreover,

ii. This family of local coordinate systems is smooth up to isometry. In other words, for all o, there is a isometry $I_o : \mathbb{R}^d \to T_o\mathcal{M}$ such that: $o \mapsto I_o \circ x_o$ is smooth.

Remark 4.3.2. *Recall that a local coordinate system (or chart) is a homeomorphism (often a diffeomorphism) from an open subset of a manifold to an open subset of $\mathbb{R}^d$, where d is the dimension of the manifold (cf. [78, p.4]). For the definition of the tangent space $T\mathcal{M}$ and concepts in differential geometry, we refer to [78] as our primary source.*

In this chapter, we refer these maps as *Morse coordinate systems.* The term *Morse* in the name of this tool is intended to highlight its inspiration from Morse theory[82], representing a modest extension of Morse's lemma.
The first property of this family is precisely what we sought: in these local coordinate systems, the Euclidean distances in the ambient space are largely preserved and can be translated into equivalent Euclidean distances within the local coordinates. While the second property is important when one seeks to give uniform bounds for mathematical expressions using these local coordinate systems.
Similar to how the existence of normal coordinate systems directly follows from the existence and properties of the exponential map, the existence of these Morse coordinate systems is a direct consequence of the following analog of the exponential map. This analog is carefully designed for our end which is to work with Euclidean distances of ambient spaces.
In application, we have the following result:

Theorem 4.3.3. *Let $\mathcal{M} \subset \mathbb{R}^m$ be a Riemannian submanifold withour border of $\mathbb{R}^m$. There exists a smooth mapping $\mathfrak{m} : \mathscr{M} \to \mathcal{M}$ satisfying the following properties:*

i. $\mathscr{M} := \sqcup_{o\in\mathcal{M}} \mathscr{M}_o \subset T\mathcal{M}$ and contains the image of the zero section of $T\mathcal{M}$, and each set $\mathscr{M}_o \subset T_o\mathcal{M}$ is star-shaped with respect to $0_{T_o\mathcal{M}}$, the zero vector of $T_o\mathcal{M}$.

ii. $\|o - \mathfrak{m}(o, v)\|_2 = \|v\|$ for all $o \in \mathcal{M}$, $v \in T_o\mathcal{M}$.

iii. For each $o \in \mathcal{M}$, $\mathfrak{m}_o : \mathscr{M}_o \to \mathcal{M}$ is a diffeomorphism onto its image.

To my humble knowledge, this theorem has not been proposed and proven.
Its proof is presented in Section 4.3.1.2. The proof of this theorem is based on the path method, following the approach proposed by Palais in [13].
For practical applications, the Morse coordinate map can be used to approximate various kernel-based integrals as follows:

Proposition 4.3.4. *Let $\mathcal{M}$ be a d-dimensional smooth compact submanifold without border of $\mathbb{R}^m$ and $K : \mathbb{R}_+ \to \mathbb{R}$ be a k-vanishing (cf. Definition4.1.4)bounded function for $\mathbb{R}^d$ with support in $[0, 1]$.*
Then, there are a constant h_0 such that for all real number $0 < h < h_0$, all smooth function ϕ on $\mathcal{M}$, we have that :

$$\sup_{x\in\mathcal{M}} \left| \int_{\mathcal{M}} \frac{1}{h^d} K\left(\frac{\|x-y\|_2}{h}\right)(\phi(y) - \phi(x))\mathrm{d}y \right| \lesssim_{\mathcal{M},k} \|K\|_\infty \times h^{k+1} \times \|\phi\|_{\mathcal{C}^{k+1}}, \tag{4.33}$$

where $\|x - y\|_2$ is the Euclidean distance between x and y in $\mathbb{R}^m$, $\|K\|_\infty := \sup |K|$, and

$$\|\phi\|_{\mathcal{C}^l} = \sup_x \sup_{0\le s\le k} \|\nabla^s \phi(x)\|.$$

Notation 4.3.5. *In Section of Morse charts, we write $A \lesssim B$(with $B > 0$) to say that there is a constant α depending only on $k, \|K\|_\infty, \mathcal{M} \subset \mathbb{R}^m$.*

Moreover, h_0 and the constant factor in the above inequality only depend on k and the embedding $\mathcal{M} \subset \mathbb{R}^m$. In other words, they only depend on k, $\mathcal{M}$, m, and the relation between $\mathcal{M}$ and $\mathbb{R}^m$.

As shown in the statement of the above proposition and its later proof, Morse coordinate map has some advantage in calculations. For example, we need not differentiate K up to power k to have the desired approximation, hence, improving the majoration.
Before going into tedious technical details in Section 4.3.1.1 and Section 4.3.1.2, let us first give the proof of Proposition 4.3.4.

Proof of Proposition 4.3.4. After Theorem 4.3.1, we consider the family of Morse coordinate system $(x_o : U_o \to B_{\mathbb{R}^d}(0, r_0))_{o\in\mathcal{M}}$ with a constant $r_0 > 0$.
Choose $h_0 = r_0$ and shrink h_0 if necessary so that $x_o(B_{\mathbb{R}^d}(0, h_0)) = \mathcal{M} \cap B_{\mathbb{R}^m}(o, h_0)$ for all o. (i.e, taking $h_0 = \min(r_0, \tau_{\mathcal{M}})$) , where $\tau_{\mathcal{M}}$ is the reach of $\mathcal{M}$.)
Then, for any $o \in \mathcal{M}$, there is a smooth function $a_o : B_{\mathbb{R}^d}(0, h_0) \to \mathbb{R}_+$, which represents the density of the volume form under the local coordinate x_o, such that for all $0 < h < h_0$:

$$\int_{\mathcal{M}} K\left(\frac{\|y-o\|_2}{h}\right)(\phi(y)-\phi(o))\mathrm{d}y = \int_{B_{\mathbb{R}^d}(0,h)} K\left(\frac{\|y\|_2}{h}\right)(\phi(\varphi^o(y)) - \phi(\varphi^o(0)))a_o(y)\mathrm{d}y,$$

where φ^o is the local parametrization $\mathcal{M}$ that is associated to x_o.
Besides, we have the following claim:

Claim 4.3.6. *For all smooth function $\phi : \mathbb{R}^d \to \mathbb{R}$, we have:*

$$\frac{1}{h^d}\int_{\mathbb{R}^d} K(\|z\|_2/h)(\phi(z)-\phi(0))\mathrm{d}z \lesssim_k \|K\|_\infty \times h^{k+1} \times \|\phi\|_{\mathcal{C}^{k+1}(\mathbb{R}^d)}.$$

Proof of Claim. This follows directly from the Taylor's expansion and the fact that K is supported in $[0, 1]$. □

Thus, we have concluded the proposition. □

4.3.1.1 Morse's lemma and some auxiliary lemmas

In this section, we recall the statement of Morse's lemma and some auxiliary lemmas.

Theorem 4.3.7 (Lemma of Morse)**.** *[82, p.12] Let c be a nondegenerate critical point of the function $f : V \to \mathbb{R}$, where V is an open subset of $\mathbb{R}^n$. There exist a neighborhood U of c and a diffeomorphism $\varphi : (U, c) \to (\mathbb{R}^n, 0)$ such that*

$$f \circ \varphi^{-1}(x_1, \ldots, x_n) = f(c) - \sum_{j=1}^{i} x_j^2 + \sum_{j=i+1}^{n} x_j^2.$$

In order to achieve our generalization in Theorem 4.3.3, we follow Palais's method as presented [13]. In this method's perspective, the desired smooth embedding comes from the flow of a time-dependent differential equations based on a family of vector fields $(\xi_t)_t$. Thus, to arrive at your end, we need several intermediate lemmas to construct the suitable vector fields. The following lemma is the very first step.

Lemma 4.3.8 (Existence of a smooth vector field on tangent space)**.** *Let $\mathcal{M}$ be a smooth manifold and $f, g : T\mathcal{M} \to \mathbb{R}$ be two smooth functions on the tangent space $T\mathcal{M}$ such that for each $o \in \mathcal{M}$, 0 is a critical point of $g\big|_{T_o\mathcal{M}}$ and also a non-degenerate critical point of $f\big|_{T_o\mathcal{M}}$. Then there is a smooth vector field $\xi : \mathscr{U} \to T\mathcal{M}$ on an open subset $\mathscr{U} = \bigsqcup_{o\in\mathcal{M}} \mathscr{U}_o$ of $T\mathcal{M}$ such that:*

i. For each $o \in \mathcal{M}$, $\xi\big|_{\mathscr{U}_o}$ takes values $T_o\mathcal{M}$.

ii. Each set $\mathscr{U}_o$ is a neighborhood of $0_{T_o\mathcal{M}}$.

iii. $\nabla_{\xi_o} f_o = g_o$ *with* $g_o := g\big|_{T_o\mathcal{M}}$ *and* $f_o := f\big|_{T_o\mathcal{M}}$.

To prove this lemma, we do need the following lemma:

Lemma 4.3.9. *Let V be a finite dimensional $\mathbb{R}$-vector space and $f, g : V \to \mathbb{R}$ be two smooth functions on V such that 0 is a critical point of g and also a non-degenerate critical point of f. Then, there is a way to associate each triple (f, g, V) with a function: $\xi^{f,g,V} : U^{f,V} \to V$ such that $U^{f,V}$ is a neighborhood of 0 and*

$$\nabla_{\xi^{f,g,V}(v)} f(v) = g(v) \qquad \forall v \in U^{f,V}. \tag{4.34}$$

Moreover, this choice of $\xi^{f,g,V}$ is invariant up to vector space isomorphism, that is, if there are a vector space isomorphism $T : \tilde{V} \to V$ and another triple $(\tilde{V}, \tilde{f}, \tilde{g})$ such that $\tilde{f} = f \circ T, \tilde{g} = g \circ T$ then:

$$T\Big(U^{f,V}\Big) = U^{\tilde{f},\tilde{V}} \text{ and } T^{-1} \circ \xi^{f,g,V} \circ T = \xi^{\tilde{f},\tilde{g},\tilde{V}}.$$

We begin with the proof of Lemma 4.3.9.

Proof of Lemma 4.3.9. Note that for any smooth function h on a vector space V, we have the following Taylor's expansion:

$$h(v) = h(0) + (\nabla_v h)(0) + \int_0^1 (1-s)\nabla^2 h_{sv}(v, v)\mathrm{d}s,$$

where $\nabla^2 h_x : V \times V \to \mathbb{R}$ is a bilinear map defined by: $\nabla^2 h_x(v, w) = (\nabla_v \nabla_w h)(x)$.
Therefore, due to the fact that 0 is a critical point of both f and g, we have:

$$g(v) = \left(\int_0^1 (1-s)\nabla^2 g_{sv}(v, v)\mathrm{d}s\right), \qquad \text{and}$$
$$(\nabla_w f)(v) = \int_0^1 \nabla^2 f_{sv}(v, w)\mathrm{d}s.$$

Besides, when $v = 0$, $\int_0^1 \nabla^2 f_{sv}\mathrm{d}s = \nabla^2 f_0$ which is a nondegenerate bilinear form on $V \times V$ due to the fact that 0 is a non-degenerate critical point of f.
Thus, if we choose:

$$U^{f,V} := \{v \in V : \int_0^1 \nabla^2 f_{sv}\mathrm{d}s \text{ is a nondegenerate bilinear form.}\} \tag{4.35}$$

Then, $U^{f,V}$ is an open neighborhood of 0 (because if we impose any inner product on V, then $U^{f,V} := \{v \in V : \det\left(\int_0^1 \nabla^2 f(sv)\mathrm{d}s\right) \neq 0\}$).
In consequence, due to our choice of $U^{f,V}$, for all $v \in U^{f,V}$, there is a unique $w_v \in V$ such that:

$$\left(\int_0^1 (1-s)\nabla^2 g_{sv}\mathrm{d}s\right)(v', v) = \left(\int_0^1 \nabla^2 f_{sv}\mathrm{d}s\right)(v', w_v) \quad \forall v' \in V. \tag{4.36}$$

Thus, there is a function: $\xi^{f,g,V} : U^{f,V} \to V$ such that $U^{f,V}$ is a neighborhood of 0 and

$$\nabla_{\xi^{f,g,V}(v)} f(v) = g(v) \qquad \forall v \in U^{f,V}. \tag{4.37}$$

To establish invariance up to isomorphism, it suffices to observe that the defining equation (4.36) of $\xi^{f,g,V}$ is invariant up to isomorphism. More precisely, it is due to the invariance of derivatives. For example, for all smooth function $h : V \to \mathbb{R}$, let $\tilde{h} = h \circ T$ then for all $x, v, w \in V$:

$$\nabla^2 h_x(v, w) = \nabla^2 \tilde{h}_{T^{-1}(x)}(T^{-1}v, T^{-1}w).$$

□

Now, we provide the proof of Lemma 4.3.8.

Proof of Lemma 4.3.8. For each $o \in \mathcal{M}$, by apply the Lemma 4.3.9 for each triple $(f_o, g_o, T_o\mathcal{M})$ where $f_o = f\big|_{T_o\mathcal{M}}$ and $g_o = g\big|_{T_o\mathcal{M}}$, we imply the existence of a vector field $\xi : \mathscr{U} \to T(T\mathcal{M})$ that satisfies *i*, *ii* and *iii*.
What is left is to prove that $\mathscr{U} \subset T\mathcal{M}$ is open and that ξ is smooth.
Besides, these properties are local. (That is, they are satisfied if and only if for all $o \in \mathcal{M}$, there is an open neighborhood U_o of o such that $\bigsqcup_{o \in U_o} \mathscr{U}_o$ is open and that ξ_{U_o} is smooth.)
Therefore, we can assume further that $\mathcal{M}$ is parallelizable (cf. [78, p. 179]) (Note that for each $o \in \mathcal{M}$, there is an open neighborhood of o that is parallelizable). More precisely, we assume that

$$T\mathcal{M} \cong \mathcal{M} \times \mathbb{R}^d,$$

where d is the dimension of $\mathcal{M}$.
Thus, based on the above isomorphism and thanks to the invarance up to isomorphism of our choice of ξ (due to Lemma 4.3.9), our problem is proven as soon as we show that:

Lemma 4.3.10. *Let $\mathcal{M}$ be a smooth manifold and $f, g : \mathcal{M} \times \mathbb{R}^d \to \mathbb{R}$ be two smooth functions such that for each $o \in \mathcal{M}$, 0 is a critical point of $g(o, \cdot)$ and also a non-degenerate critical point of $f(o, \cdot)$.*
Then there is a smooth vector field $\xi : \mathscr{U} \to \mathbb{R}^d$ on an open subset $\mathscr{U} = \bigsqcup_{o \in \mathcal{M}} \mathscr{U}_o$ of $\mathcal{M} \times \mathbb{R}^d$ such that:

i. *Each set $\mathscr{U}_o$ is a neighborhood of 0 in $\mathbb{R}^d$.*

ii. $\nabla_{\xi_o(v)} f_o(v) = g_o(v)$,

where $\xi_o := \xi(o, \cdot)$, $f_o := f(o, \cdot)$ and $g_o := g(o, \cdot)$.

But as we show right below, this claim is true. Therefore, we have the desired conclusion. ☐

Proof of Lemma 4.3.10. In order to prove this claim, recall that if we denote by $H(h)$ the Hessian matrix of a function $h : \mathbb{R}^d \to \mathbb{R}$, then by definition,

$$\nabla^2 h_x(v, v) = \langle v, H(h)_x v \rangle.$$

Therefore, after the defining Eq 4.36, we have for each $o \in \mathcal{M}$, and $v \in \mathbb{R}^d$,

$$\left(\int_0^1 (1 - s) H(g_o)_{sv} \mathrm{d}s \right) v = \left(\int_0^1 H(f_o)_{sv} \mathrm{d}s \right) \xi_o(v), \tag{4.38}$$

and

$$\mathscr{U} = \left\{ (o, v) | \det \left(\int_0^1 H(f_o)_{sv} \mathrm{d}s \right) \neq 0 \right\}. \tag{4.39}$$

Recall that $(o, v) \to \int_0^1 H(f_o)_{sv} \mathrm{d}s$ is smooth due to the smoothness of f and that with v being 0, $\int_0^1 H(f_o)_{sv} \mathrm{d} = H(f_o)_0$ which is non-degenerate by hypothesis. Thus, $\mathscr{U}$ is open.
On top of that,

$$\xi_o(v) = \left(\int_0^1 H(f_o)_{sv} \mathrm{d}s \right)^{-1} \left(\int_0^1 (1 - s) H(g_o)_{sv} \mathrm{d}s \right) v,$$

which is smooth on $(o, v) \in \mathcal{U}$ thanks to the smoothness of f and g. ☐

4.3.1.2 Proofs of Theorem 4.3.3 and Theorem 4.3.1

Proof of Theorem 4.3.3. Let $\mathfrak{e} : \mathscr{E} = \bigsqcup_{o\in\mathcal{M}} \mathscr{E}_o \subset T\mathcal{M} \to \mathcal{M}$ denote the exponential map of $\mathcal{M}$. Define the smooth functions $f : \mathscr{E} \to \mathbb{R}$ and $A : T\mathcal{M} \to \mathbb{R}$ by the equations:

$$f(v) = \|o - \mathfrak{e}(v)\|_2^2, \qquad \forall o \in \mathcal{M}, v \in \mathscr{E}_o,$$
$$A(v) = \|v\|_2^2, \qquad \forall v \in T\mathcal{M},$$

where in $\|\cdot\|_2$ referes to the Euclidean distance of $\mathbb{R}^m$, while $\|\cdot\|_2$ in $A(v)$ refers to the norm in $T\mathcal{M}$ with respect to Riemannian structure.
Let $\mathscr{W}$ be an open subset of $\mathscr{E}$ such that its closure $\overline{\mathscr{W}}$ is also contained in $\mathscr{E}$ and that $0_{T_o\mathcal{M}} \in \mathscr{W}$ for all $o \in \mathcal{M}$.
Let $F : T\mathcal{M} \to \mathbb{R}$ be a smooth extension of $f\big|_{\mathscr{W}}$.
Denote respectively by F_o, A_o the functions $F\big|_{T_o\mathcal{M}}$ and $A\big|_{T_o\mathcal{M}}$.
Clearly, for each $o \in \mathcal{M}$, 0 is a critical point of F_o and that 0 is a non-degenerate critical point of A.
Besides, after Proposition 2.2 in [55], the second order of Taylor's expansion of $F_o - A_o$ is null. Hence, the second order derivative of F_o must be equal to the one of A_o.
Thus, for all $t \in (-3, 3)$ and $o \in \mathcal{M}$, we define:

$$F_o^t := A_o + t(F_o - A_o)$$

has a non-degenerate critical point at 0.
Therefore, after Lemma 4.3.8, for there is an open set $\mathscr{U} \subset T\mathcal{M}$ such that $\mathscr{U} \ni 0_{T_o\mathcal{M}}$ for all $o \in \mathcal{M}$ and that there is a smooth family of smooth vector field $(\xi^t, t \in (-2, 2))$ on $\mathscr{U}$ such that:

$$\nabla_{\xi_o^t(v)} F_o^t = A_o(v) - F_o. \tag{4.40}$$

Consider the following flow φ of the time-dependent smooth vector field ξ_t.
By the existence of the flow, there is an open set $D = \sqcup_{o\in\mathcal{M}} D_o \subset (-2, 2) \times \mathscr{U}$ and a smooth flow $\varphi : D \to \mathscr{U}$ such that such that $0 \times \mathscr{U} \subset D$ and that for all $o \in \mathcal{M}$, $x \in D_o$,

$$\frac{\mathrm{d}\varphi_t}{\mathrm{d}t}(x) = \xi_o^t(\varphi_t(x)); \quad \varphi_0(x) = x. \tag{4.41}$$

Claim 4.3.11. *There is an open set $\mathscr{V} := \bigsqcup_{o\in\mathcal{M}} \mathscr{V}_o \subset \mathscr{U}$ such that for all $o \in \mathcal{M}$,*

i. $(-2, 2) \times \mathscr{V}_o \subset D_o$.

ii. $0_{T_o\mathcal{M}} \subset \mathscr{V}_o$ *for all* $o \in \mathcal{M}$.

iii. $\mathscr{V}_o$ *is star-shaped at* $0_{T_o\mathcal{M}}$.

Proof of Claim 4.3.11. First, for the sake of simplicity, let us consider a $\mathbb{R}^d$ version of Equation 4.41 as follows:

$$\frac{\mathrm{d}\phi^t}{\mathrm{d}t} = V^t(\phi^t), \quad \phi^0(v) = v, \tag{4.42}$$

for all $(t, v) \in (-2, 2) \times U_0$ for some open neighborhood U of 0 in $\mathbb{R}^d$ and time-dependent smooth vector field $V_t : U \to \mathbb{R}^d$ such that: $V^t(0) = 0$.
We observe that the above ODE has a unique smooth solution $(t, x) \mapsto \phi^t(x)$ on $(-2, 2) \times \tilde{U}$, where $\tilde{U}$ is defined by:

$$\tilde{U} := \{x \in U_o : \|V_t(sx)\| < \frac{1}{2} \text{ for all } -2 < t < 2, 0 < s < 1\}.$$

Therefore, for the Equation 4.41, we can conclude this lemma by choosing: $\mathscr{V} := \bigsqcup_{o\in\mathcal{M}} \mathscr{V}$ where:

$$\mathscr{V}_o := \{v \in D_o : sv \in D_o \text{ for all } 0 < s < 1 \text{ and } \sup_{-2<t<2, 0<s<1} \|\xi^t(sv)\| < \frac{1}{2}\}.$$

□

Let us continue the proof of Theorem 4.3.3.
Consider the the function $G^t = F^t(\varphi_t)$ on $\mathscr{V}$. We observe that

$$G^0(v) = F^0(\varphi_0(v)) = A(v) = \|v\|^2,$$

and that (thanks to Eq 4.41 and Eq 4.40):

$$\begin{aligned}\frac{\mathrm{d}G^t}{\mathrm{d}t} &= \frac{\mathrm{d}F^t(\varphi_t)}{\mathrm{d}t} = (\dot{F^t}) \circ (\varphi_t) + \nabla_{\dot{\varphi}_t} F^t \\ &= (F - A) \circ (\varphi) + \nabla_{\xi^t(\varphi_t)} F^t = (F - A) \circ (\varphi) + (\nabla_{\xi^t} F^t) \circ (\varphi_t) \\ &= 0.\end{aligned}$$

Thus, $G^1(v) = \|v\|^2$ for all $v \in \mathscr{V}$.
In other words, $\varphi_1 : \mathscr{V} \to T\mathcal{M}$ is a smooth map such that:

$$(F \circ \varphi_1)(v) = \|v\|^2.$$

That means, for all $o \in \mathcal{M}$ and $v \in \mathscr{V}_o \cap \mathscr{W}_o$, we have:

$$\|o - \mathfrak{e}(v)\|^2 = \|v\|^2. \tag{4.43}$$

Besides, for each o, Eq (4.41) defines a flow on $T_o\mathcal{M}$. Thus, φ_1^o is a diffeomorphism onto its image, where $\varphi_1^o = \varphi_1\big|_{T_o\mathcal{M}}$.
In summary, thus far, we have shown that φ_1 satisfies *ii* and *iii*.
In order to have *i*, we only need to shrink the domain of φ_1.
The smoothness of φ_1 comes from the fact that it is a solution of Eq 4.36.
Thus, we have the desired conclusion. □

Proof of Theorem 4.3.1. Theorem 4.3.1 is a direct consequence of Theorem 4.3.3 because of each open set U enough of the Riemannian manifold $\mathcal{M}$, we have the smooth isometry:

$$TU \cong U \times \mathbb{R}^d.$$

□

4.3.2 An estimation of bias term

To avoid repetition, in the rest of this section, we fix an integer $k \geq 1$ and assume that:

Assumption 6. *K is k-vanishing, the dimension d of $\mathcal{M}$ is at least 3, and s is a positive integer smaller than $k - 1$.*

$$\widehat{\mu}_h(\mathrm{d}y) := \left(\int_{\mathcal{M}} h^{-d} K\left(\frac{\|x - y\|_2}{h}\right) p(x)\mathrm{d}x\right)\mathrm{d}y.$$

The main result of this section is as follows:

Theorem 4.3.12. *There are a constant C depending only on $\mathcal{M} \subset \mathbb{R}^m$, k and $Lip(K) := \sup_{x,y\in\mathbb{R}:x\neq y} \frac{|K(x)-K(y)|}{|x-y|}$ such that for all $h \in (0,1,$*

$$\|\widehat{\mu}_h - \widehat{\mu}_h(\mathcal{M})\mu\|_{\mathcal{F}} \leq C \times h^{s+1} \times \|p\|_{H_1^s(\mathcal{M})},$$

where p is the density of μ.

In order to prove this theorem, we prove first its local version.

Proposition 4.3.13. *Let $\mathfrak{d} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ be any function such that*

- *$\mathfrak{d}(x, y) = \mathfrak{d}(y, x)$ for all x, y.*
- *$\mathfrak{d}(x, x) = 0$ for all x.*
- *There is a constant C so that for all x and $1 \leq j \leq k-1$, we have:*

$$\left|\frac{1}{j!}\int_{\mathbb{R}^d} h^{-d} K\left(\frac{\mathfrak{d}(x,y)}{h}\right)(x-y)^{\times j}\mathrm{d}y\right| \leq C \times h^{k+1}, \tag{4.44}$$

  where $(x-y)^{\times j}$ is defined as in Notation 4.2.2.
- *There is a constant C' such that $C'\mathfrak{d}(x, y) \geq \|x - y\|_2$ for all x, y.*
- *The third derivative of $\mathfrak{d}^2$ is uniformly bounded.*

Let $a : \mathbb{R}^d \to \mathbb{R}^d$ be another compactly supported smooth function.
Suppose that $Lip(K)$ is finite. Then there is a constant C'' depending only on d, k, $Lip(K)$, $\mathfrak{d}$, and a such that for all integer $s \in \{0, ..., k-1\}$, compactly supported smooth functions f and p, and real positive number h, we have:

$$\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} h^{-d}K\left(\frac{\mathfrak{d}(x,y)}{h}\right) f(x)(p(y)-p(x))a(y)\mathrm{d}x\mathrm{d}y \leq C'' \times (\|f\|_\infty + \|\nabla f\|_\infty) \times \|p\|_{H_1^s(\mathbb{R}^d)}.$$

Remark 4.3.14. *Note that because $\mathfrak{d}(x, x) = 0$ for all x, under the boundedness of the third derivative of $\mathfrak{d}^2$, there is a constant $\tilde{C}$ such that for all $x, y \in \mathbb{R}^d$ and $\lambda \in (0, 1)$, (cf. Lemma 4.3.15)*

$$\left|\mathfrak{d}\left(x, \frac{z-(1-\lambda)x}{\lambda}\right) - \frac{1}{\lambda}\mathfrak{d}(x,z)\right| \leq \frac{\tilde{C}}{\lambda^2}\|x-z\|_2^2.$$

This fact will be used in the proof of Proposition 4.3.13. .

Proof of Proposition 4.3.13. The proof of this proposition follows the same steps as of Theorem 4.2.3. For the sake of simplicity, we let $p^{(s)}$ denote $\nabla^s p$.

We consider first the case when s is even.
By Taylor's expansion up to the order s, we see that:

$$\begin{aligned}\int_{y\in\mathbb{R}^d} K\left(\frac{\mathfrak{d}(x,y)}{h}\right) f(x)(p(y)-p(x))a(y)\mathrm{d}y &= \sum_{j=1}^{s}\int_{y\in\mathbb{R}^d}\frac{1}{j!}K\left(\frac{\mathfrak{d}(x,y)}{h}\right) f(x)p^{(j)}(x)\left[(y-x)^{\times j}\right]a(y)\mathrm{d}y\\ &+ \int_{\lambda=0}^{1}\left(\int_{y\in\mathbb{R}^d}\frac{(1-\lambda)^{s-1}}{(s-1)!}K\left(\frac{\mathfrak{d}(x,y)}{h}\right)f(x)\left(p^{(s)}(x+\lambda(y-x))-p^{(s)}(x)\right)\left[(y-x)^{\times s}\right]a(y)\mathrm{d}y\right)\mathrm{d}\lambda.\end{aligned} \tag{4.45}$$

Because of the hypothesis 4.44, by using Taylor's expansion of the function a at x, we have that for every $1 \leq j \leq s \leq k-1$,

$$\left|\int_{y\in\mathbb{R}^d} \frac{1}{j!} K\left(\frac{\mathfrak{d}(x,y)}{h}\right) f(x) p^{(j)}(x)\left[(y-x)^{\times j}\right] a(y)\mathrm{d}y\right| \leq_{C,a} h^{k+1} \times |f(x)| \times \left\|p^{(j)}(x)\right\|_{op}.$$

Hence,

$$\left|\sum_{j=1}^{s}\int_{x\in\mathbb{R}^d}\int_{y\in\mathbb{R}^d} \frac{1}{j!} K\left(\frac{\mathfrak{d}(x,y)}{h}\right) f(x) p^{(j)}(x)\left[(y-x)^{\times j}\right] a(y)\mathrm{d}y\mathrm{d}x\right|$$
$$\lesssim_{C,a} k \times C \times h^{k+1} \times \|f\|_\infty \times \|p\|_{H_1^s(\mathbb{R}^d)}. \quad (4.46)$$

Then, by the change of variables: $y \to \frac{z-(1-\lambda)x}{\lambda}$, we observe that:

$$\begin{aligned}&\int_{y\in\mathbb{R}^d} K\left(\frac{\mathfrak{d}(x,y)}{h}\right) f(x)\Big(p^{(s)}(x+\lambda(y-x)) - p^{(s)}(x)\Big)\left[(y-x)^{\times s}\right] a(y)\mathrm{d}y\\ =&\int_{z\in\mathbb{R}^d} K\left(\frac{\mathfrak{d}\left(x, \frac{z-(1-\lambda)x}{\lambda}\right)}{h}\right) f(x)\Big(p^{(s)}(z) - p^{(s)}(x)\Big)\left[(z-x)^{\times s}\right] a\left(\frac{z-(1-\lambda)x}{\lambda}\right)\frac{1}{\lambda^d}\mathrm{d}z \quad (4.47)\end{aligned}$$

Now, let

$$\begin{aligned}A'(\lambda) &:= \int_{x\in\mathbb{R}^d}\int_{y\in\mathbb{R}^d} K\left(\frac{\mathfrak{d}\left(x, \frac{z-(1-\lambda)x}{\lambda}\right)}{h}\right) f(x)\Big(p^{(s)}(z) - p^{(s)}(x)\Big)\left[(z-x)^{\times s}\right] a\left(\frac{z-(1-\lambda)x}{\lambda}\right)\frac{1}{\lambda^d}\mathrm{d}z\mathrm{d}x.\\ A''(\lambda) &:= \int_{x\in\mathbb{R}^d}\int_{y\in\mathbb{R}^d} K\left(\frac{\mathfrak{d}(x,z)}{\lambda h}\right) f(x)\Big(p^{(s)}(z) - p^{(s)}(x)\Big)\left[(z-x)^{\times s}\right] a\left(\frac{z-(1-\lambda)x}{\lambda}\right)\frac{1}{\lambda^d}\mathrm{d}z\mathrm{d}x.\\ A(\lambda) &:= \int_{x\in\mathbb{R}^d}\int_{y\in\mathbb{R}^d} K\left(\frac{\mathfrak{d}(x,z)}{\lambda h}\right) f(x)\Big(p^{(s)}(z) - p^{(s)}(x)\Big)\left[(z-x)^{\times s}\right] a(x)\frac{1}{\lambda^d}\mathrm{d}z\mathrm{d}x.\end{aligned}$$

We will proceed to bound $|A'(\lambda) - A''(\lambda)|$, $|A''(\lambda) - A(\lambda)|$ and $A(\lambda)$.
Let us begin with $|A'(\lambda) - A''(\lambda)|$. By the Lipschitz continuity of K and $\mathfrak{d}$, and the boundedness condition on $\mathfrak{d}$, we have:

$$\begin{aligned}&\left|A'(\lambda) - A''(\lambda)\right|\\ \leq&\|a\|_\infty \times \|f\|_\infty \int_{x\in\mathbb{R}^d}\int_{y\in\mathbb{R}^d}\left|K\left(\frac{\mathfrak{d}(x,z)}{\lambda h}\right) - K\left(\frac{\mathfrak{d}\left(x, \frac{z-(1-\lambda)x}{\lambda}\right)}{h}\right)\right|\left(\left\|p^{(s)}(z)\right\|_{op} + \left\|p^{(s)}(x)\right\|_{op}\right)\\ &\qquad(\|x-z\|_2)^s\frac{1}{\lambda^d}\mathrm{d}z\mathrm{d}x\\ \leq&\|a\|_\infty \times \tilde{C} \times Lip(K) \times \|f\|_\infty \int_{x\in\mathbb{R}^d}\int_{y\in\mathbb{R}^d}\frac{\|x-z\|_2^2}{\lambda^2 h}\mathbb{1}_{\|x-z\|_2\leq C'\lambda h}\left(\left\|p^{(s)}(z)\right\|_{op} + \left\|p^{(s)}(x)\right\|_{op}\right)\times\\ &\qquad(\|x-z\|_2)^s\frac{1}{\lambda^d}\mathrm{d}z\mathrm{d}x\\ =&\|a\|_\infty \times 2\tilde{C} \times Lip(K) \times \|f\|_\infty \int_{x\in\mathbb{R}^d}\int_{y\in\mathbb{R}^d}\frac{\|x-z\|_2^2}{\lambda^2 h}\mathbb{1}_{\|x-z\|_2\leq C'\lambda h}\left\|p^{(s)}(z)\right\|_{op}(\|x-z\|_2)^s\frac{1}{\lambda^d}\mathrm{d}z\mathrm{d}x,\end{aligned}$$

where the indicator $\mathbb{1}_{\|x-z\|_2\leq C'\lambda h}$ in the last two lines comes from Remark 4.3.14 and the fact that K is supported in $[0,1]$.

Therefore,

$$\left|A'(\lambda) - A''(\lambda)\right| \lesssim_{d,C',a} \tilde{C} \times Lip(K) \times \|f\|_\infty \times h^{d+s+1} \times \lambda^s \times \|p\|_{H_1^s(\mathbb{R}^d)}. \quad (4.48)$$

Next, for $|A''(\lambda) - A(\lambda)|$, we have:

$$\begin{aligned}
&\left|A'(\lambda) - A''(\lambda)\right| \\
\leq &\|f\|_\infty \int_{x\in\mathbb{R}^d}\int_{y\in\mathbb{R}^d} \left|K\left(\frac{\mathfrak{d}(x,z)}{\lambda h}\right)\right| \left(\left\|p^{(s)}(z)\right\|_{op} + \left\|p^{(s)}(x)\right\|_{op}\right)(\|x-z\|_2)^s \frac{1}{\lambda^d}\left|a(x) - a\left(\frac{z-(1-\lambda)x}{\lambda}\right)\right| \mathrm{d}z\mathrm{d}x. \\
\leq &Lip(a) \times \|f\|_\infty \int_{x\in\mathbb{R}^d}\int_{y\in\mathbb{R}^d} \left|K\left(\frac{\mathfrak{d}(x,z)}{\lambda h}\right)\right| \left(\left\|p^{(s)}(z)\right\|_{op} + \left\|p^{(s)}(x)\right\|_{op}\right)(\|x-z\|_2)^s \frac{1}{\lambda^d}\frac{\|z-x\|_2}{\lambda}\mathrm{d}z\mathrm{d}x. \\
\leq &Lip(a) \times \|f\|_\infty \times \|K\|_\infty \int_{x\in\mathbb{R}^d}\int_{y\in\mathbb{R}^d} \mathbb{1}_{\|x-z\|\leq C'h}\left(\left\|p^{(s)}(z)\right\|_{op} + \left\|p^{(s)}(x)\right\|_{op}\right)(\|x-z\|_2)^s \frac{1}{\lambda^d}\frac{\|z-x\|_2}{\lambda}\mathrm{d}z\mathrm{d}x.
\end{aligned}$$

Therefore,

$$\left|A''(\lambda) - A(\lambda)\right| \lesssim_{d,C',a,\tilde{C}} \tilde{C} \times \|K\|_\infty \times \|f\|_\infty \times h^{d+s+1} \times \lambda^s \times \|p\|_{H_1^s(\mathbb{R}^d)}. \quad (4.49)$$

Now, we bound $A(\lambda)$.
Because s is even, by interchanging the role of z and x, we see that:

$$\begin{aligned}
A(\lambda) &= -\int_{x\in\mathbb{R}^d}\int_{y\in\mathbb{R}^d} K\left(\frac{\mathfrak{d}(x,z)}{\lambda h}\right) f(z)\left(p^{(s)}(z) - p^{(s)}(x)\right)\left[(z-x)^{\times s}\right] a(z)\frac{1}{\lambda^d}\mathrm{d}z\mathrm{d}x \\
&= \frac{1}{2}\int_{x\in\mathbb{R}^d}\int_{y\in\mathbb{R}^d} K\left(\frac{\mathfrak{d}(x,z)}{\lambda h}\right)(a(x)f(x) - a(z)f(z))\left(p^{(s)}(z) - p^{(s)}(x)\right)\left[(z-x)^{\times s}\right]\frac{1}{\lambda^d}\mathrm{d}z\mathrm{d}x
\end{aligned}$$

On top of that, we have that K is supported in $[0,1]$ and $C'\mathfrak{d}(x,y) \geq \|x-y\|_2$ for all x and y. Therefore,

$$\begin{aligned}
|A(\lambda)| &\leq \frac{1}{2}\|K\|_\infty \times Lip(af) \times \int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{\mathbf{1}_{\|x-y\|_2\leq\lambda C'h}}{\lambda^d}\|x-z\|^{s+1}\left(\|p^{(s)}(x)\|_{op} + \|p^{(s)}(z)\|_{op}\right)\mathrm{d}x\mathrm{d}z \\
&= \|K\|_\infty \times Lip(af) \times \left(\int_{\mathcal{M}}\|p^{(s)}(x)\|_{op}\mathrm{d}x\right)\left(\int_{z\in\mathbb{R}^d}\frac{1_{\|z\|_2\leq C'\lambda h}}{\lambda^d}|z|^{s+1}\mathrm{d}z\right) \\
&= h^{s+1} \times \|K\|_\infty \times Lip(af) \times \left(\int_{\mathbb{R}^d}\|p^{(s)}(x)\|_{op}\mathrm{d}x\right)\left(\lambda^{s+1}\int_{z\in\mathbb{R}^d} 1_{|z|\leq C'}|z|^{s+1}\mathrm{d}z\right). \quad (4.50)
\end{aligned}$$

Therefore, we have the desired conclusion for s even.
For s odd, we only have to follow the same interpolation argument presented in Proof of Theorem 4.2.3. Therefore, we have the desired conclusion for all s. □

Now, we are ready to give the proof of Theorem 4.3.12.

Proof of Theorem 4.3.12. Let $\tilde{\mu}_h$ denote:

$$\tilde{\mu}_h(\mathrm{d}y) = \eta_h(y)p(y)\mathrm{d}y,$$

where $\eta_h(y) = \int_{\mathcal{M}} h^{-d}K\left(\frac{\|x-y\|_2}{h}\right)\mathrm{d}x$.
Then after Lemma 4.3.16, we have that:

$$\|\tilde{\mu}_h - \widehat{\mu}_h(\mathcal{M})\mu\|_{\mathcal{F}} \leq \|\tilde{\mu}_h - \mu\|_{\mathcal{F}} + \|\mu - \widehat{\mu}_h(\mathcal{M})\mu\|_{\mathcal{F}} \lesssim_{\mathcal{M},\|K\|_\infty} h^k.$$

Hence, it is sufficient to give an upper bound for $\|\widehat{\mu}_h - \tilde{\mu}_h\|_{\mathcal{F}}$.
Recall that by definition,

$$\|\widehat{\mu}_h - \tilde{\mu}_h\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} \left(\int_{\mathcal{M}} \int_{\mathcal{M}} h^{-d} K\left(\frac{\|x-y\|_2}{h} \right) f(x)(p(y)-p(x))\mathrm{d}x\mathrm{d}y \right)$$

Consider a finite smooth partition of unity $(U_i, \tau_i : \mathcal{M} \to \mathbb{R})_{1\leq i\leq N}$ of $\mathcal{M}$, that is, $(U_i)_{1\leq i\leq N}$ is an open over of $\mathcal{M}$ and $\sum_{i=1}^N \tau_i(x) = 1$ for all x with τ_i being supported in U_i, such that there is a local chart $\varphi_i : U_i \to \mathbb{R}^d$ of $\mathcal{M}$. We will use extensively this family of local charts in the later proof.

Clearly, for each 1-Lipschitz function $f : \mathcal{M} \to \mathbb{R}$ and each i, $\tau_i f$ is also a Lipschitz function with Lipschitz coefficient

$$\|\tau_i\|_\infty + \|\nabla \tau_i\|_\infty \|\mathrm{diam}(\mathcal{M})\|_\infty,$$

where we have the used that fact that for all x,

$$|f(x)| = \left| f - \frac{1}{\mathrm{vol}(\mathcal{M})} \int_{\mathcal{M}} f(y)\mathrm{d}y \right| = \left| f - \frac{1}{\mathrm{vol}(\mathcal{M})} \int_{\mathcal{M}} (f(x)-f(y))\mathrm{d}y \right| \leq \sup_{x,y} \rho(x,y) = \mathrm{diam}(\mathcal{M}).$$

Define:

$$\mathcal{F}_i := \{ f : \mathcal{M} \to \ \mid Lip(f) \leq \|\tau_i\|_\infty + \|\nabla\tau_i\|_\infty \|\mathrm{diam}(\mathcal{M})\|_\infty \ \text{ and } \ \mathrm{supp}(f) \subset U_i \}.$$

Because of the decomposition $f = \sum_{i=1}^N f\tau_i$, we have that:

$$\|\widehat{\mu}_h - \tilde{\mu}_h\|_{\mathcal{F}} \leq \sum_{i=1}^N \sup_{f\in\mathcal{F}_i} \left(\int_{\mathcal{M}} \int_{\mathcal{M}} h^{-d} K\left(\frac{\|x-y\|_2}{h} \right) f(x)(p(y)-p(x))\mathrm{d}x\mathrm{d}y \right).$$

Thus, the initial problem reduces to bound individually each term

$$\sup_{f\in\mathcal{F}_i} \left(\int_{\mathcal{M}} \int_{\mathcal{M}} h^{-d} K\left(\frac{\|x-y\|_2}{h} \right) f(x)(p(y)-p(x))\mathrm{d}x\mathrm{d}y \right).$$

Without loss of generality, assume that y is sufficiently small so that for all i, for all $x \in \mathrm{supp}\,\tau_i$,:

$$\{ y \in \mathcal{M} : \|x-y\|_2 \leq h \} \subset U_i.$$

Then, fix an integer $i \in [1, N]$, by considering the local representation, we see that:

$$\begin{aligned}
&\sup_{f\in\mathcal{F}_i} \left(\int_{\mathcal{M}} \int_{\mathcal{M}} h^{-d} K\left(\frac{\|x-y\|_2}{h} \right) f(x)(p(y)-p(x))\mathrm{d}x\mathrm{d}y \right) \\
&= \sup_{f\in\mathcal{F}_i} \left(\int_{U_i} \int_{U_i} h^{-d} K\left(\frac{\|x-y\|_2}{h} \right) f(x)(p(y)-p(x))\mathrm{d}x\mathrm{d}y \right) \\
&= \sup_{g\in\mathcal{G}_i} \left(\int_{\varphi(U_i)} \int_{\varphi(U_i)} h^{-d} K\left(\frac{\mathfrak{d}_i(x,y)}{h} \right) g(x)(\tilde{p}(y)-\tilde{p}(x)) a(x)a(y)\mathrm{d}x\mathrm{d}y \right) \\
&= \sup_{g\in\tilde{\mathcal{G}}_i} \left(\int_{\varphi(U_i)} \int_{\varphi(U_i)} h^{-d} K\left(\frac{\mathfrak{d}_i(x,y)}{h} \right) g(x)(\tilde{p}(y)-\tilde{p}(x)) a(y)\mathrm{d}x\mathrm{d}y \right).
\end{aligned}$$

where $\tilde{p}$ is the local representation of p under the chart φ_i, $\mathfrak{d}$ is defined on $\varphi_i(U_i)$ such that :

$$\mathfrak{d}_i(\varphi_i(x), \varphi_i(y)) = \|x-y\|_2 \quad \forall x,y \in U_i, \tag{4.51}$$

and a is the local representation of the density of the volume measure of $\mathcal{M}$, $G_i := \{ f \circ \varphi_i^{-1} : \varphi(U_i) \to \mathbb{R} \mid f \in \mathcal{F}_i \}$, and $\tilde{\mathcal{G}}_i := \{ af : f \in \mathcal{G}_i \}$.

However, this is exactly the mathematical expression we have studied in Proposition 4.3.13. Therefore, in order to have the desired conclusion, it is sufficient to check that if function $\mathfrak{d}_i : \varphi(U_i) \times \varphi(U_i) \to \mathbb{R}_+$ is extensible to a function $\mathfrak{d} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ that satisfies all the hypothesis in Proposition 4.3.13.

Besides, the closure of $\varphi(U_i)$ is compact. Therefore, for a such $\mathfrak{d}$ exists, it is sufficient that $\mathfrak{d}_i$ verifies all the hypothesis in Proposition 4.3.13, which is true due to the fact that

- $(x, y) \mapsto \|x - y\|_2^2$ is a smooth function on $\mathcal{M} \times \mathcal{M}$.
- and that we have Proposition 4.3.4.

Therefore, we have the desired conclusion. □

Lemma 4.3.15. *Under the hypothesis of Proposition 4.3.13, we have that there is a constant $\tilde{C}$ such that for all $x, y \in \mathbb{R}^d$ and $\lambda \in (0,1)$, (cf. Lemma 4.3.15)*

$$\left|\mathfrak{d}\left(x, \frac{z-(1-\lambda)x}{\lambda}\right) - \frac{1}{\lambda}\mathfrak{d}(x,z)\right| \leq \frac{\tilde{C}}{\lambda^2}\|x-z\|_2^2.$$

Proof. It is sufficient to prove this estimation for $x = 0$, let $F(z) := \mathfrak{d}(0, z)$. We have that $F : \mathbb{R}^d \to \mathbb{R}$ is continuously differentiable, $F \geq 0$, and that $F(0) = 0$. Hence, $\nabla F(0) = 0$. Therefore, by using Taylor's expansion to the second order, we have:

$$\left|F(z/\lambda) - \frac{1}{\lambda^2}F(z)\right| = \frac{1}{2}\left|\frac{1}{\lambda^2}\int_0^1 \nabla^2 F(sz/\lambda)(z^{\times 2})\mathrm{d}s - \frac{1}{\lambda^2}\int_0^1 \nabla^2 F(sz)(z^{\times 2})\mathrm{d}s\right|$$
$$\leq \frac{\|z\|_2^3}{2\lambda^2}\sup_{x\in\mathbb{R}^d}\|\nabla^3 F(x)\|_{op}.$$

Besides, under the given hypothesis, $C'\sqrt{F(z)} \geq \|z\|_2$. Hence,

$$\left|\mathfrak{d}(0, z/\lambda) - \frac{1}{\lambda}\mathfrak{d}(0,z)\right| \leq \frac{\frac{\|z\|_2^3}{2\lambda^2}\sup_{x\in\mathbb{R}^d}\|\nabla^3 F(x)\|_{op}}{2C'\|z\|_2/\lambda} = \frac{\sup_{x\in\mathbb{R}^d}\|\nabla^3 F(x)\|_{op}}{4C'} \times \|z\|_2^2/\lambda,$$

Therefore, we have the desired conclusion. □

Lemma 4.3.16 (An estimation lemma)**.** *Let $\eta_h(y) = \int_{\mathcal{M}} h^{-d}K\left(\frac{\|x-y\|_2}{h}\right)\mathrm{d}x$, then there is a constant C depending only on $\|K\|_\infty$ and $\mathcal{M}$ such that for all $x \in \mathcal{M}$ and $h \in (0,1)$, we have:*

$$|\eta_h(x) - 1| \leq Ch^k.$$

Proof. Consider a family of Morse charts $(\varphi_x : U_x \to B_{\mathbb{R}^d}(0,r), x \in \mathcal{M})$. For any point $x \in \mathcal{M}$, let $(z^1, z^2, \ldots, z^d)$ represent the local coordinates in the chart φ_x. In this context, let (a_{ij}) denote the local representation of the Riemannian metric on $\mathcal{M}$.
Given our choice of local chart, for any $h < r$, it holds that

$$\int_{\mathcal{M}} \frac{1}{h^d}K\left(\frac{\|x-y\|_2}{h}\right)\mathrm{d}y = \int_{\mathbb{R}^d}\frac{1}{h^d}K\left(\frac{\|z\|_2}{h}\right)\sqrt{\det(a_{ij}(z))}\,dz,$$

Therefore, as in Proof of Proposition 4.3.4, it is sufficient to demonstrate that:

$$\det(a_{ij}(0)) = 1.$$

We first prove that:

$$\det(a_{ij}(0)) = \frac{1}{\det(\nabla^2(F)(x))}, \tag{4.52}$$

where $F(y) := \frac{1}{2}\|x-y\|_2^2$.

Recall that $\nabla^2(F)(x)$ is a bilinear form on the tangent space $T\mathcal{M}_x$ and $(\frac{\partial}{\partial z^1}\big|_x, \dots, \frac{\partial}{\partial z^d}\big|_x)$ forms a linear basis of $T\mathcal{M}_x$. Therefore, by definition of the determinant,

$$\det(\nabla^2(F)(x)) = \frac{\det\left(\nabla^2(F)(x)[\frac{\partial}{\partial z^i}, \frac{\partial}{\partial z^j}]\right)}{\det\left(\langle \frac{\partial}{\partial z^i}, \frac{\partial}{\partial z^j}\rangle\right)} = \frac{\det\left(\nabla^2(F)(x)[\frac{\partial}{\partial z^i}, \frac{\partial}{\partial z^j}]\right)}{\det(a_{ij}(0))}. \tag{4.53}$$

Furthermore, since x is a critical point of F, for any i and j, by definition of Hessian,

$$\nabla^2(F)(x)[\frac{\partial}{\partial z^i}, \frac{\partial}{\partial z^j}] = \partial_j \partial_i (F \circ \varphi_x^{-1}), \tag{4.54}$$

which simplifies to 1 if $i = j$ and 0 otherwise, given that $F \circ \varphi_x^{-1}(z) = \frac{1}{2}\|z\|_2^2$.
Now, we prove that

$$\det(\nabla^2(F)(x)) = 1$$

Indeed, we will reuse the result in Eq 4.53 but with another local chart different from φ_x. Consider a normal coordinate system $\varphi : U \to \mathbb{R}^d$ around a neighborhood U of x such that $\varphi(x) = 0$, and let (b_{ij}) denote the local representation of the Riemannian metric on $\mathcal{M}$ under this normal coordinate chart.
Then by properties of normal coordinates, we have that $b_{ij}(0) = 1$ for all $i = j$ and 0 otherwise. Therefore,

$$\det(b_{ij}(0)) = 1.$$

Besides,

$$\partial_j \partial_i (F \circ \varphi^{-1})\big|_0 = \left\langle \partial_i(\varphi^{-1})\big|_0, \partial_j(\varphi^{-1})\big|_0 \right\rangle_{\mathbb{R}^m},$$

where φ^{-1} is regarded as a function from $\varphi(U)$ to $\mathbb{R}^m$.
On top of that, because the Riemannian metric of $\mathcal{M}$ is induced by $\mathbb{R}^m$, we have:

$$\left\langle \partial_i(\varphi^{-1})\big|_0, \partial_j(\varphi^{-1})\big|_0 \right\rangle_{\mathbb{R}^m} = \left\langle \partial_i(\varphi^{-1})\big|_0, \partial_j(\varphi^{-1})\big|_0 \right\rangle_{\mathcal{M}},$$

where $\partial_i(\varphi^{-1})\big|_0, \partial_j(\varphi^{-1})\big|_0$ in right side are understood as tangent vector of $\mathcal{M}$.
But, by definition, $b_{ij}(0) = \left\langle \partial_i(\varphi^{-1})\big|_0, \partial_j(\varphi^{-1})\big|_0 \right\rangle_{\mathcal{M}}$.
Thus,

$$\det(\nabla^2(F)(x)) = 1.$$

Therefore, we have the desired conclusion. □

Remark 4.3.17. *Note that the most precise way to handle the induced Riemannian metric is through the language of immersions. In particular, for the above proof, the expression should technically be written as $F \circ \iota \circ \varphi_x^{-1}$, where ι is the immersion $\mathcal{M} \hookrightarrow \mathbb{R}^m$. However, throughout this chapter, we have implicitly identified $\mathcal{M}$ with $\iota(\mathcal{M})$ and thus have omitted explicit reference to ι. To maintain consistency, the proof above does not mention ι at all. Nonetheless, this choice of notation may cause confusion regarding how the induced metric works, and it is worth keeping in mind that the immersion ι underlies the identification of $\mathcal{M}$ with its image in $\mathbb{R}^m$.*

4.3.3 An estimation of stochastic term

The main result of this section is as follows:

Theorem 4.3.18. *For all n and $h \in (0, 1]$, we have:*

$$\mathbb{E}\left(\|\widehat{\mu}_{n,h} - \widehat{\mu}_h\|_{\mathcal{F}}\right) \lesssim_{\mathcal{M}} \|K\|_\infty \frac{1}{\sqrt{n}} h^{1-d/2}. \tag{4.55}$$

Before giving its proof, we begin with a preliminary on Green functions on manifolds.

4.3.3.1 A preliminary on Green functions on manifolds

On compact Riemannian manifolds, we have many estimation results analogues to what we have in $\mathbb{R}^d$. We begin first with the existence of Green functions.

Theorem 4.3.19. *[11, ch. 4] The Green's function $G(x,y) : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ for $\mathcal{M}$ with the Laplace-Beltrami operator Δ is a fundamental solution to the Poisson equation in distribution sense. That is, for all smooth function $f \in \mathcal{C}^\infty(\mathcal{M})$ such that $\int_{\mathcal{M}} f(x)\mathrm{d}x = 0$,*

$$\int_{\mathcal{M}} G(x,y)\Delta f(y)\mathrm{d}y = f(x).$$

Then if the dimension d of $\mathcal{M}$ is bigger than or equal to 3, this function G exists and is smooth on $(\mathcal{M} \times \mathcal{M})\backslash diag(\mathcal{M})$, where $diag(\mathcal{M}) := \{(x,y)|x \in \mathcal{M}\}$.
Moreover, there is a constant C depending only on $\mathcal{M}$ such that:

$$|G(x,y)| \leq C\rho(x,y)^{2-d} \ , \ \|\nabla_1 G(x,y)\|_2 \leq C\rho(x,y)^{1-d} \ , \ \textit{for all } x,y \in \mathcal{M}, \tag{4.56}$$

where $\|\cdot\|_2$ in the above inequality denotes the norm of tangent vectors and ∇_1 denotes the derivative on the first variable.
In particular, for all x, $y \mapsto \|\nabla_1 G(x,y)\|_2$ is $L^1(\mathcal{M}, \mathrm{d}x)$ and continuous on $\mathcal{M}\backslash\{x\}$.

Then, for any function $f \in L^\infty$, we denote by $Gf(x)$ the integral:

$$Gf(x) := \int_{\mathcal{M}} G(x,y)f(y)\mathrm{d}y. \tag{4.57}$$

Then, we have the following regularity result:

Proposition 4.3.20. *If $f \in L^\infty$, Gf is differentiable and for all x,*

$$\nabla Gf(x) = \int_{\mathcal{M}} \nabla_1 G(x,y)f(y)\mathrm{d}y.$$

In particular, $x \mapsto \|\nabla Gf(x)\|_2$ is L^∞.

Proof. Fix a point $x \in \mathcal{M}$. Let $\gamma : [0,1] \to \mathcal{M}$ be any differential curve such that $\gamma(0) = x$ and $\gamma'(0) = v \in T_x\mathcal{M}$ and that $\|v\| = 1$.
Then for any radius $r > 0$, thanks to the smoothness of G and the compactness of $\mathcal{M}$, we have that:

$$\begin{aligned}
&\limsup_{t\to 0^+} \left| t^{-1}(Gf(\gamma(t)) - Gf(x)) - \int_{\mathcal{M}} \langle \nabla_1 G(x,y), v\rangle f(y)\mathrm{d}y \right| \\
\leq& \limsup_{t\to 0^+} \left| \int_{B_{\mathcal{M}}(x,r)} \left(t^{-1}(G(\gamma(t),y) - G(x,y)) - \langle \nabla_1 G(x,y), v\rangle\right) f(y)\mathrm{d}y \right| + \\
&\limsup_{t\to 0^+} \left| \int_{\mathcal{M}\backslash B_{\mathcal{M}}(x,r)} \left(t^{-1}(G(\gamma(t),y) - G(x,y)) - \langle \nabla_1 G(x,y), v\rangle\right) f(y)\mathrm{d}y \right| \\
=& \limsup_{t\to 0^+} \left| \int_{B_{\mathcal{M}}(x,r)} \left(t^{-1}(G(\gamma(t),y) - G(x,y)) - \langle \nabla_1 G(x,y), v\rangle\right) f(y)\mathrm{d}y \right| + 0 \\
\leq& \|f\|_\infty \left(\int_{B_{\mathcal{M}}(x,r)} \|\nabla_1 G(x,y)\|_2 \mathrm{d}y + \limsup_{t\to 0^+} t^{-1} \int_{B_{\mathcal{M}}(x,r)} \left(\int_0^1 \|\nabla_1 G(\gamma(t))\|_2 \mathrm{d}t \right) \mathrm{d}y \right) \\
\leq& \|f\|_\infty \left(\int_{B_{\mathcal{M}}(x,r)} \|\nabla_1 G(x,y)\|_2 \mathrm{d}y + \sup_{x' \in B_{\mathcal{M}}(x,r)} t^{-1} \int_{B_{\mathcal{M}}(x,r)} \|\nabla_1 G(x')\|_2 \mathrm{d}y \right) \\
\leq& 2\|f\|_\infty \times \sup_{x' \in \mathcal{M}} \int_{B_{\mathcal{M}}(x',2r)} \|\nabla_1 G(x',y)\|_2 \mathrm{d}y.
\end{aligned}$$

Besides, thanks to Theorem 4.3.19, for all x',

$$\int_{B_{\mathcal{M}}(x',2r)} \left\|\nabla_1 G(x',y)\right\|_2 \mathrm{d}y \lesssim_{\mathcal{M}} r.$$

Therefore, by taking $r \to 0^+$, we have the desired conclusion.

□

Then, thanks to the above estimation of the gradient of Green function in Theorem 4.3.19, we have the following lemma:

Lemma 4.3.21. *There is a constant C depending on $\mathcal{M}$ such that for all z,*

$$\int_{\mathcal{M}} h^{-d} \|\nabla_1 G(x,y)\| \times \mathbb{1}_{\rho(x,z)\leq h} \mathrm{d}x \leq C \times \Big(h^{1-d}\mathbf{1}_{\rho(y,z)\leq 2h} + \rho(y,z)^{1-d}\mathbf{1}_{\rho(y,z)>2h}\Big). \tag{4.58}$$

Proof of Lemma 4.3.21. By using Theorem 4.3.19, the proof of this lemma is the same as the proof of Lemma 4.2.7. □

Lemma 4.3.22. *For all 1-Lipschitz function $f \in \mathcal{F}$ and $g \in L^\infty$, we have:*

$$\int_{\mathcal{M}} f(x)g(x)\mathrm{d}x \leq \int_{\mathcal{M}} \|\nabla G(g)(x)\|_2 \mathrm{d}x.$$

Proof of Lemma 4.3.3.1 . For any $f \in \mathcal{C}^\infty(\mathcal{M})$ smooth such that $\int_{\mathcal{M}} f = 0$, we have that:

$$\left|\int_{\mathcal{M}} \langle \nabla f(x), \nabla G g(x)\rangle\right| \leq \|\nabla f\|_\infty \int_{\mathcal{M}} \|\nabla G(g)(x)\|_2 \mathrm{d}x, \tag{4.59}$$

where $\|\nabla f\|_\infty := \sup_x \|\nabla f(x)\|_x$.
Besides, by Fubini and Theorem 4.3.19, we have:

$$\begin{aligned} -\int_{\mathcal{M}} \langle \nabla f(x), \nabla G g(x)\rangle &= \int_{\mathcal{M}} \Delta f(x) G g(x) \mathrm{d}x \\ &= \int_{\mathcal{M}\times\mathcal{M}} \Delta f(x) G(x,y) g(y) \mathrm{d}y\mathrm{d}x = \int_{\mathcal{M}} f(y)g(y)\mathrm{d}y. \end{aligned} \tag{4.60}$$

Therefore, for all $f \in \mathcal{C}^\infty(\mathcal{M})$ smooth, we have:

$$\left|\int_{\mathcal{M}} \langle \nabla f(x), \nabla G g(x)\rangle\right| \leq \|\nabla f\|_\infty \int_{\mathcal{M}} \|\nabla G(g)(x)\|_2 \mathrm{d}x.$$

Thus, by the density of $\mathcal{C}^\infty(\mathcal{M})$, we have the desired conclusion. □

4.3.3.2 Proof of the main result

In this section, we give the proof of Theorem 4.3.18.

Proof of Theorem 4.3.18. By Lemma 4.3.3.1, we have that:

$$\mathcal{W}_1(\widehat{\mu}_{n,h}, \widehat{\mu}_h) \leq \int_{\mathcal{M}} |\nabla G(\widehat{p}_{n,h} - \widehat{p}_h)|,$$

where

$$\begin{aligned} \widehat{p}_{n,h}(y) :=& \frac{1}{n}\sum_{i=1}^n h^{-d} K\Big(\frac{\|X_i - y\|_2}{h}\Big), \\ \widehat{p}_h(y) :=& \int_{\mathcal{M}} h^{-d} K\Big(\frac{\|x-y\|_2}{h}\Big) p(x)\mathrm{d}x. \end{aligned}$$

Hence, by Cauchy-Schwarz's inequality,

$$\mathbb{E}(\mathcal{W}_1(\widehat{\mu}_{n,h},\widehat{\mu}_h))^2 \leq \mathrm{vol}(\mathcal{M}) \int_{\mathcal{M}} \mathbb{E}\Big(\|\nabla G(\widehat{p}_{n,h}-\widehat{p}_h)(y)\|_2^2\Big)\mathrm{d}y.$$

Note that $\widehat{p}_{n,h} = \frac{1}{n}\sum_{i=1}^n p_{X_i,h}(y)$, with $p_{X_i,h}(y) := h^{-d}K\Big(\frac{\|X_1-y\|_2}{h}\Big)$. Therefore,

$$\begin{aligned}
&\int_{\mathcal{M}} \mathbb{E}\Big(\|\nabla G(\widehat{p}_{n,h}-\widehat{p}_h)(y)\|_2^2\Big)\mathrm{d}y \\
\leq& \frac{1}{n}\int_{\mathcal{M}} \mathbb{E}\Big(|\nabla G p_{X_1,h}(y)|^2\Big)\mathrm{d}y = \frac{h^{-2d}}{n}\int_{\mathcal{M}} \mathbb{E}\left(\left|\int_{\mathcal{M}} \nabla_1 G(y,z) K\left(\frac{\|X_1-z\|_2}{h}\right)\mathrm{d}z\right|^2\right)\mathrm{d}y \\
\leq& \frac{h^{-2d}}{n}\int_{\mathcal{M}} \mathbb{E}\left[\left(\int_{\mathcal{M}} \|\nabla_1 G(y,z)\|_2 \left|K\left(\frac{\|X_1-z\|_2}{h}\right)\right|\mathrm{d}z\right)^2\right]\mathrm{d}y \\
\leq& \frac{h^{-2d}}{n} \times \|K\|_\infty^2 \times \int_{\mathcal{M}} \mathbb{E}\left[\left(\int_{\mathcal{M}} \|\nabla_1 G(y,z)\|_2 \mathbb{1}_{\rho(X_1,z)\leq Ch}\mathrm{d}z\right)^2\right]\mathrm{d}y
\end{aligned}$$

where C is the constant depending only on $\mathcal{M} \subset \mathbb{R}^m$ such that for all x, y:

$$\rho(x,y) \leq C\|x-y\|_2.$$

Thus, from Lemma 4.3.21, we imply that:

$$\begin{aligned}
&\mathbb{E}(\mathcal{W}_1(\widehat{\mu}_{n,h},\widehat{\mu}_h))^2 \\
\lesssim_{\mathcal{M}}& \frac{1}{n} \times \|K\|_\infty \times \int_{\mathcal{M}} \mathbb{E}\Big(h\mathbf{1}_{\rho(X_1,y)\leq 2h} + h^d \rho(X_1,y)^{1-d}\mathbf{1}_{\rho(X_1,y)>2h}\Big)^2 \mathrm{d}y \\
\leq& \frac{2}{n} \times \|K\|_\infty \times h^{-2d} \times \mathbb{E}\Big[\int_{\mathcal{M}} \Big(h^2\mathbf{1}_{\rho(X_1,y)\leq 2h} + h^{2d}\rho(X_1,y)^{2-2d}\mathbf{1}_{\rho(X_1,y)>2h}\Big)\mathrm{d}y\Big] \\
\lesssim_{\mathcal{M}}& \frac{1}{n}h^{-2d}\Big[h^2 \times h^d + h^{2d}h^{1-d}\Big].
\end{aligned}$$

Therefore, we have the desired conclusion. □

4.3.4 Conclusion

We now can conclude with the proofs of the main results.

Proof of Theorem 4.1.6 and Corollary 4.1.7. The results of Theorem 4.1.6 follow directly from Theorem 4.3.12 and Theorem 4.3.18.
The first part of Corollary 4.1.7 follows directly from Theorem 4.3.18 after choosing $h_n = n^{-1/(s+d/2)}$ and $\tilde{\mu}_n = \frac{1}{\widehat{\mu}_{n,h_n}(\mathcal{M})}\widehat{\mu}_{n,h_n}$. Note that as shown in Lemma 4.3.16, there a deterministic constant C depending only on $\|K\|_\infty$ and $\mathcal{M}$ such that for all n:

$$|\widehat{\mu}_{n,h_n}(\mathcal{M}) - 1| \leq Ch_n^k.$$

For the almost sure convergence, it is sufficient to show that there is constant $C', C'' > 0$ such that for all n and t,

$$\mathbb{P}(|W_1(\widehat{\mu}_{n,h_n}, \widehat{\mu}_{n,h}(\mathcal{M})\mu) - \mathbb{E}(W_1(\widehat{\mu}_{n,h_n}, \widehat{\mu}_{n,h_n}(\mathcal{M})\mu))| \geq t) \leq C''\exp\left(-C'nt^2\right),$$

Because this, with Borel-Cantelli theorem, implies that almost surely,

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n}\ln(n)}|W_1(\widehat{\mu}_{n,h_n}, \widehat{\mu}_{n,h}(\mathcal{M})\mu) - \mathbb{E}(W_1(\widehat{\mu}_{n,h_n}, \widehat{\mu}_{n,h_n}(\mathcal{M})\mu))| = 0.$$

Let X_1' be a copy of X_1 and independent of $X_1, X_2, ..., X_n$. Consider:

$$\widehat{\mu}'_{n,h_n}(\mathrm{d}y) := \frac{h^{-d}}{n}\left(K\left(\frac{\|X_1'-y\|_2}{h}\right) + \sum_{i=2}^{n} K\left(\frac{\|X_i-y\|_2}{h}\right)\right)\mathrm{d}y.$$

We have that:

$$\begin{aligned}
& W_1(\widehat{\mu}_{n,h_n}, \widehat{\mu}_{n,h_n}(\mathcal{M})\mu) - W_1(\widehat{\mu}'_{n,h_n}, \widehat{\mu}'_{n,h_n}(\mathcal{M})\mu) \\
\leq & \frac{h_n^{-d}}{n} \sup_{\substack{f:\mathcal{M}\to\mathbb{R}\,:\,1-\text{Lipschitz} \\ \int_{\mathcal{M}} f(x)\mathrm{d}x=0}} \left(\int_{\mathcal{M}} K\left(\frac{\|X_1-y\|_2}{h_n}\right) f(y)\mathrm{d}y - \int_{\mathcal{M}} K\left(\frac{\|X_1'-y\|_2}{h_n}\right) f(y)\mathrm{d}y\right) \\
\leq & \frac{\|K\|_\infty}{n} \mathrm{diam}(\mathcal{M}) \sup_x A_{h_n}(x),
\end{aligned}$$

where $A_h(x) = h^{-d}\int_{\mathcal{M}} \mathbb{1}_{\|x-y\|_2\leq h}\mathrm{d}y$.
Therefore, by McDiarmid's inequality, we have for all $t \geq 0$

$$\mathbb{P}(|W_1(\widehat{\mu}_{n,h_n}, \widehat{\mu}_{n,h}(\mathcal{M})\mu) - \mathbb{E}(W_1(\widehat{\mu}_{n,h_n}, \widehat{\mu}_{n,h_n}(\mathcal{M})\mu))| \geq t) \leq \exp\left(-C_n n t^2\right),$$

where

$$C_n = \frac{2}{(\|K\|_\infty \mathrm{diam}(\mathcal{M}) \sup_x A_{h_n}(x))^2}.$$

On top of that, $\lim_{h\to 0}\sup_x A_h(x) = \mathrm{vol}(B_{\mathbb{R}^d}(0,1))$. Therefore, we have the desired conclusion for the almost sure convergence. □

Bibliography

[1] Aamari, E., Berenfeld, C., and Levrard, C. Optimal reach estimation and metric learning. The Annals of Statistics 51, 3 (2023), 1086–1108.

[2] Aamari, E., Kim, J., Chazal, F., Michel, B., Rinaldo, A., and Wasserman, L. Estimating the reach of a manifold. Electron. J. Statist. 13, 1 (2019), 1359–1399.

[3] Aamari, E., and Levrard, C. Stability and minimax optimality of tangential Delaunay complexes for manifold reconstruction. Discrete Comput. Geom. 59, 4 (2018), 923–971.

[4] Aamari, E., and Levrard, C. Nonasymptotic rates for manifold, tangent space and curvature estimation. Ann. Statist. 47, 1 (2019), 177–204.

[5] Adler, R. J., and Taylor, J. E. Random fields and geometry. Springer Monographs in Mathematics. Springer, New York, 2007.

[6] Aldous, D. Stopping times and tightness. Ann. Probab. 6 (1978), 335–340.

[7] Alesker, S., and Shelukhin, E. On a uniform estimate for the quaternionic Calabi problem. Isr. J. Math. 197 (2013), 309–327.

[8] Amari, S.-i. Information geometry and its applications, vol. 194 of Applied Mathematical Sciences. Springer, [Tokyo], 2016.

[9] Amari, S.-i., and Nagaoka, H. Methods of information geometry, vol. 191 of Translations of Mathematical Monographs. American Mathematical Society, Providence, RI; Oxford University Press, Oxford, 2000. Translated from the 1993 Japanese original by Daishi Harada.

[10] Arjovsky, M., Chintala, S., and Bottou, L. Wasserstein generative adversarial networks. In Proceedings of the 34th International Conference on Machine Learning (06–11 Aug 2017), D. Precup and Y. W. Teh, Eds., vol. 70 of Proceedings of Machine Learning Research, PMLR, pp. 214–223.

[11] Aubin, T. Nonlinear analysis on manifolds. Monge-Ampère equations, vol. 252 of Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, New York, 1982.

[12] Bakry, D., Gentil, I., and Ledoux, M. Analysis and geometry of Markov diffusion operators, vol. 348 of Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer, Cham, 2014.

[13] Banyaga, A., and Hurtubise, D. E. A proof of the morse-bott lemma. Expositiones Mathematicae 22, 4 (2004), 365–373.

[14] Baudoin, F. Diffusion processes and stochastic calculus. EMS Textb. Math. Zürich: European Mathematical Society (EMS), 2014.

[15] Belkin, M., and Niyogi, P. Towards a theoretical foundation for Laplacian-based manifold methods. J. Comput. Syst. Sci. 74, 8 (2008), 1289–1308.

[16] Belkin, M., Sun, J., and Wang, Y. Constructing Laplace operator from point clouds in $\mathbb{R}^d$. In Proceedings of the 20th annual ACM-SIAM symposium on discrete algorithms, SODA 2009, New York, NY, USA, January 4–6, 2009. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM); New York, NY: Association for Computing Machinery (ACM), 2009, pp. 1031–1040.

[17] Berenfeld, C., Harvey, J., Hoffmann, M., and Shankar, K. Estimating the reach of a manifold via its convexity defect function. Discrete & Computational Geometry 67, 2 (2022), 403–438.

[18] Berenfeld, C., and Hoffmann, M. Density estimation on an unknown submanifold. Electron. J. Stat. 15, 1 (2021), 2179–2223.

[19] Berenfeld, C., and Hoffmann, M. Density estimation on an unknown submanifold. Electron. J. Stat. 15, 1 (2021), 2179–2223.

[20] Billingsley, P. Probability and Measure, 3 ed. John Wiley & Sons, New York, 1995.

[21] Billingsley, P. Convergence of probability measures. Chichester: Wiley, 1999.

[22] Block, A., Jia, Z., Polyanskiy, Y., and Rakhlin, A. Intrinsic dimension estimation using Wasserstein distance. J. Mach. Learn. Res. 23 (2022), Paper No. [313], 37.

[23] Brown, B., Caterini, A., Ross, B., Cresswell, J., and Loaiza-Ganem, G. Verifying the union of manifolds hypothesis for image data. In International Conference on Learning Representations 2023 (2023).

[24] Calder, J. The game theoretic p-Laplacian and semi-supervised learning with few labels. Nonlinearity 32, 1 (2019), 301–330.

[25] Calder, J., and García Trillos, N. Improved spectral convergence rates for graph Laplacians on ε-graphs and k-NN graphs. Appl. Comput. Harmon. Anal. 60 (2022), 123–175.

[26] Chavel, I. Eigenvalues in Riemannian geometry, vol. 115 of Pure and Applied Mathematics. Academic Press, Inc., Orlando, FL, 1984. Including a chapter by Burton Randol, With an appendix by Jozef Dodziuk.

[27] Chavel, I. Riemannian Geometry, 2 ed., vol. 98 of Cambridge studies in advanced mathematics. Cambridge University Press, 2006.

[28] Cheng, X., and Wu, H.-T. Convergence of graph Laplacian with kNN self-tuned kernels. Inf. Inference 11, 3 (2022), 889–957.

[29] Cipriani, A., and v. G. Bart. The discrete Gaussian free field on a compact manifold. Stochastic Processes Appl. 130, 7 (2020), 3943–3966.

[30] Dalalyan, A. Theoretical guarantees for approximate sampling from smooth and log-concave densities. J. R. Stat. Soc. Ser. B. Stat. Methodol. 79, 3 (2017), 651–676.

[31] Dalalyan, A., and Reiss, M. Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case. Probability theory and related fields 137 (2007), 25–47.

[32] del Barrio, E., González-Sanz, A., and Loubes, J.-M. Central limit theorems for general transportation costs. Annales de l'Institut Henri Poincare (B) Probabilites et statistiques 60, 2 (2024), 847–873.

[33] del Barrio, E., González Sanz, A., and Loubes, J.-M. Central limit theorems for semi-discrete wasserstein distances. Bernoulli 30, 1 (2024), 554–580.

[34] del Barrio, E., Sanz, A. G., Loubes, J.-M., and Niles-Weed, J. An improved central limit theorem and fast convergence rates for entropic transportation costs. SIAM Journal on Mathematics of Data Science 5, 3 (2023), 639–669.

[35] Devroye, L., Györfi, L., and Lugosi, G. A probabilistic theory of pattern recognition, vol. 31 of Applications of Mathematics (New York). Springer-Verlag, New York, 1996.

[36] Divol, V. Minimax adaptive estimation in manifold inference. Electronic Journal of Statistics 15, 2 (2021), 5888–5932.

[37] Divol, V. Measure estimation on manifolds: an optimal transport approach. Probab. Theory Relat. Fields 183, 1-2 (2022), 581–647.

[38] Divol, V., Guérin, H., Nguyen, D.-T., and Tran, V. Measure estimation on a manifold explored by a diffusion process, 2024.

[39] Divol, V., Niles-Weed, J., and Pooladian, A.-A. Tight stability bounds for entropic brenier maps. arXiv preprint arXiv:2404.02855 (2024).

[40] do Carmo, M. P. Geometria Riemanniana. Rio de Janeiro: Instituto de Matemática Pura e Aplicada (IMPA), 2015.

[41] do Carmo, M. P. a. Riemannian geometry, portuguese ed. Mathematics: Theory & Applications. Birkhäuser Boston, Inc., Boston, MA, 1992.

[42] Durmus, A., and Moulines, E. Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. Ann. Appl. Probab. 27, 3 (2017), 1551–1587.

[43] Elworthy, K. D. Stochastic differential equations on manifolds, vol. 70 of Lond. Math. Soc. Lect. Note Ser. Cambridge University Press, Cambridge. London Mathematical Society, London, 1982.

[44] Evans, L. Partial differential equations, vol. 19 of Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 1998.

[45] Figalli, A., and Glaudo, F. An invitation to optimal transport, Wasserstein distances, and gradient flows. EMS Textbooks in Mathematics. EMS Press, Berlin, [2021] ©2021.

[46] Fournier, N., and Guillin, A. On the rate of convergence in wasserstein distance of the empirical measure. Probability Theory and Related Fields 162, 3-4 (2015), 707–738.

[47] Frogner, C., Zhang, C., Mobahi, H., Araya-Polo, M., and Poggio, T. A. Learning with a wasserstein loss. In Advances in Neural Information Processing Systems 28 (NIPS 2015) (2015), pp. 2053–2061.

[48] Gall, J.-F. L. Brownian motion, martingales, and stochastic calculus, french ed., vol. 274 of Graduate Texts in Mathematics. Springer, [Cham], 2016.

[49] Gao, F., Guillin, A., and Wu, L. Bernstein-type concentration inequalities for symmetric markov processes. Theory of Probability & Its Applications 58, 3 (2014), 358–382.

[50] Gao, R., Xie, L., Xie, Y., and Xu, H. Robust hypothesis testing using wasserstein uncertainty sets. In Advances in Neural Information Processing Systems 31 (NeurIPS 2018) (2018), pp. 7919–7929.

[51] García Trillos, N., Gerlach, M., Hein, M., and Slepčev, D. Error estimates for spectral convergence of the graph Laplacian on random geometric graphs toward the Laplace-Beltrami operator. Found. Comput. Math. 20, 4 (2020), 827–887.

[52] Genovese, C., Pacifico, M. P., Verdinelli, I., and Wasserman, L. Minimax manifold estimation. Journal of machine learning research 13 (2012), 1263–1291.

[53] Genovese, C., Perone-Pacifico, M., Verdinelli, I., and Wasserman, L. Manifold estimation and singular deconvolution under hausdorff loss. Ann. Statist. 40, 2 (2012), 941–963.

[54] Giné, E., and Guillou, A. Rates of strong uniform consistency for multivariate kernel density estimators. (Vitesse de convergence uniforme presque sûre pour des estimateurs à noyaux de densités multivariées). Ann. Inst. Henri Poincaré, Probab. Stat. 38, 6 (2002), 907–921.

[55] Giné, E., Koltchinskii, V., Li, W., and Zinn, J., Eds. High dimensional probability. Proceedings of the fourth international conference. Many papers based on the presentations a vol. 51. Beachwood, OH: IMS, Institute of Mathematical Statistics, 2006.

[56] Giné, E., and Nickl, R. Mathematical foundations of infinite-dimensional statistical models, vol. 40. Cambridge: Cambridge University Press, 2016.

[57] Giovanni, L. A first course in fractional Sobolev spaces, vol. 229 of Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, [2023] ©2023.

[58] Girolami, M., and Calderhead, B. Riemann manifold langevin and hamiltonian monte carlo methods. Journal of the Royal Statistical Society: Series B 73, 2 (2011), 123–214.

[59] Grigoryan, A. Heat kernel and analysis on manifolds, vol. 47 of AMS/IP Stud. Adv. Math. Providence, RI: American Mathematical Society (AMS); Somerville, MA: International Press, 2009.

[60] Gromov, M. Curvature, diameter and Betti numbers. Comment. Math. Helv. 56, 2 (1981), 179–195.

[61] Guérin, H., Nguyen, D.-T., and Tran, V. Strong uniform convergence of laplacians of random geometric and directed knn graphs on compact manifolds, 2022.

[62] Hall, P., and Patil, P. On wavelet methods for estimating smooth functions.

[63] Hastie, T., Tibshirani, R., and Friedman, J. The elements of statistical learning, second ed. Springer Series in Statistics. Springer, New York, 2009. Data mining, inference, and prediction.

[64] Hebey, E. Sobolev spaces on Riemannian manifolds, vol. 1635. Springer Science & Business Media, 1996.

[65] Hein, M., Audibert, J.-Y., and von Luxburg, U. From graphs to manifolds—weak and strong pointwise consistency of graph Laplacians. In Learning theory, vol. 3559 of Lecture Notes in Comput. Sci. Springer, Berlin, 2005, pp. 470–485.

[66] Hein, M., Audibert, J.-Y., and von Luxburg, U. Graph Laplacians and their convergence on random neighborhood graphs. J. Mach. Learn. Res. 8 (2007), 1325–1368.

[67] Hsu, E. P. Stochastic analysis on manifolds, vol. 38 of Grad. Stud. Math. Providence, RI: American Mathematical Society (AMS), 2002.

[68] Ikeda, N., and Watanabe, S. Stochastic differential equations and diffusion processes., 2nd ed. ed., vol. 24 of North-Holland Math. Libr. Amsterdam etc.: North-Holland; Tokyo: Kodansha Ltd., 1989.

[69] Jacod, J., and Shiryaev, A. Limit theorems for stochastic processes, second ed., vol. 288 of Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 2003.

[70] Jakubowski, A. On the Skorokhod topology. Annales de l'Institut Henri Poincaré 22, 3 (1986), 263–285.

[71] Joffe, A., and Métivier, M. Weak convergence of sequences of semimartingales with applications to multitype branching processes. Adv. Appl. Probab. 18 (1986), 20–65.

[72] K.-J. Engel, K.-J., and Nagel, R. One-parameter semigroups for linear evolution equations, vol. 194 of Grad. Texts Math. Berlin: Springer, 2000.

[73] Kallenberg, O. Foundations of modern probability. Probab. Appl. New York, NY: Springer, 1997.

[74] Kantorovitch, L. On the translocation of masses. C. R. (Doklady) Acad. Sci. URSS (N.S.) 37 (1942), 199–201.

[75] Karatzas, I., and Shreve, S. E. Brownian motion and stochastic calculus., 2nd ed. ed., vol. 113 of Grad. Texts Math. New York etc.: Springer-Verlag, 1991.

[76] Kulik, A. Ergodic behavior of Markov processes, vol. 67 of De Gruyter Studies in Mathematics. De Gruyter, Berlin, 2018. With applications to limit theorems.

[77] Lee, J. Introduction to smooth manifolds, vol. 218 of Graduate Texts in Mathematics. Springer-Verlag, New York, 2003.

[78] Lee, J. M. Introduction to smooth manifolds, second ed., vol. 218 of Graduate Texts in Mathematics. Springer, New York, 2013.

[79] Lee, J. M. Introduction to Riemannian manifolds, vol. 176. Cham: Springer, 2018.

[80] Lofstrom, J. B.-J., and Bergh, J. Interpolation spaces. SpringereVerlag, Newe (1976).

[81] Lott, J., and Villani, C. Hamilton-Jacobi semigroup on length spaces and applications. J. Math. Pures Appl. (9) 88, 3 (2007), 219–229.

[82] M., A., and M., D. Morse theory and Floer homology. Universitext. Springer, London; EDP Sciences, Les Ulis, 2014. Translated from the 2010 French original by Reinie Erné.

[83] M, H., Audibert, J.-Y., and Von Luxburg, U. Graph Laplacians and their convergence on random neighborhood graphs. J. Mach. Learn. Res. 8 (2007), 1325–1368.

[84] Maaten, L. v. d., and Hinton, G. Visualizing data using t-sne. Journal of Machine Learning Research 9, Nov (2008), 2579–2605.

[85] Martens, J. Deep learning via Hessian-free optimization. In Proceedings of the 27th International Conference on Machine Learning (ICML) (2010), pp. 735–742.

[86] Massart, P. About the constants in Talagrand's concentration inequalities for empirical processes. Ann. Probab. 28, 2 (2000), 863–884.

[87] Milnor, J. On manifolds homeomorphic to the 7-sphere. Annals of Mathematics 64, 2 (1956), 399–405.

[88] Monge, G. Mémoire sur la théorie des déblais et des remblais. 1781.

[89] Nguyen, D.-T. Scaling limit of the collision measures of multiple random walks. ALEA Lat. Am. J. Probab. Math. Stat. 20, 2 (2023), 1385–1410.

[90] Niles-Weed, J., and Berthet, Q. Minimax estimation of smooth densities in Wasserstein distance. Ann. Statist. 50, 3 (2022), 1519–1540.

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

[92] Nolan, D., and Pollard, D. U-processes: Rates of convergence. Ann. Stat. 15 (1987), 780–799.

[93] Oksendal, B. Stochastic differential equations. An introduction with applications., 6th ed. ed. Universitext. Berlin: Springer, 2003.

[94] Page, L., Brin, S., Motwani, R., and Winograd, T. The pagerank citation ranking: Bringing order to the web. In Proceedings of the 7th International World Wide Web Conference (Brisbane, Australia, 1998), pp. 161–172.

[95] Pagès, G., and Panloup, F. Ergodic approximation of the distribution of a stationary diffusion: Rate of convergence. Annals of Applied Probability 22, 3 (2012), 1059–1100.

[96] Peyre, R. Comparison between W_2 distance and $\dot{H}^{-1}$ norm, and localization of Wasserstein distance. ESAIM Control Optim. Calc. Var. 24, 4 (2018), 1489–1501.

[97] Pooladian, A.-A., and Niles-Weed, J. Entropic estimation of optimal transport maps. arXiv preprint arXiv:2109.12004 (2021).

[98] Protter, P. E. Stochastic integration and differential equations, 2nd ed. ed., vol. 21 of Appl. Math. (N. Y.). Berlin: Springer, 2004.

[99] Puchkin, N., and Spokoiny, V. Structure-adaptive manifold estimation. Journal of Machine Learning Research 23, 40 (2022), 1–62.

[100] Ramdas, A., García Trillos, N., and Cuturi, M. On Wasserstein two-sample testing and related families of nonparametric tests. Entropy 19, 2 (2017), Paper No. 47, 15.

[101] Revuz, D., and Yor, M. Continuous Martingales and Brownian Motion. Springer-Verlag, Berlin, 1998.

[102] Roberts, G., and Tweedie, R. Exponential convergence of Langevin distributions and their discrete approximations. Bernoulli 2, 4 (1996), 341–363.

[103] Roweis, S. T., and Saul, L. K. Nonlinear dimensionality reduction by locally linear embedding. Science 290, 5500 (2000), 2323–2326.

[104] Rudin, W. Real and complex analysis. New York, NY: McGraw-Hill, 1987.

[105] Rudin, W. Functional Analysis, 2nd ed. McGraw-Hill, New York, 1991.

[106] Santambrogio, F. Optimal transport for applied mathematicians, vol. 87 of Progress in Nonlinear Differential Equations and their Applications. Birkhäuser/Springer, Cham, 2015. Calculus of variations, PDEs, and modeling.

[107] Shubin, M. A. Pseudodifferential operators and spectral theory, second ed. Springer-Verlag, Berlin, 2001. Translated from the 1978 Russian original by Stig I. Andersson.

[108] Singer, A. From graph to manifold Laplacian: the convergence rate. Appl. Comput. Harmon. Anal. 21, 1 (2006), 128–134.

[109] Stéphanovitch, A., Aamari, E., and Levrard, C. Wasserstein generative adversarial networks are minimax optimal distribution estimators. Ann. Statist. 52, 5 (2024), 2167–2193.

[110] Tameling, C., Sommerfeld, M., and Munk, A. Empirical optimal transport on countable metric spaces: Distributional limits and statistical applications.

[111] Tang, R., and Yang, Y. Minimax rate of distribution estimation on unknown submanifolds under adversarial losses. The Annals of Statistics 51, 3 (2023), 1282–1308.

[112] Tao, T. Lecture notes 1 for 247a, 2006. Accessed: 2025-03-16.

[113] Tao, W., and Shi, Z. Convergence of laplacian spectra from random samples. Journal of Computational Mathematics 38, 6 (2020), 952–984.

[114] Tenenbaum, J. B., De Silva, V., and Langford, J. C. A global geometric framework for nonlinear dimensionality reduction. Science 290, 5500 (2000), 2319–2323.

[115] Ting, D., Huang, L., and Jordan, M. An analysis of the convergence of graph laplacians. ICML (2010).

[116] Tsybakov, A. B. Introduction to nonparametric estimation. Springer Series in Statistics. Springer, New York, 2009. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.

[117] Tyagi, H., Vural, E., and Frossard, P. Tangent space estimation for smooth embeddings of riemannian manifolds®. Information and Inference: A Journal of the IMA 2, 1 (2013), 69–114.

[118] Uppal, A., Singh, S., and Poczos, B. Nonparametric density estimation & convergence rates for gans under besov ipm losses. In Advances in Neural Information Processing Systems 32 (2019), pp. 15887–15897.

[119] Villani, C. Optimal transport, vol. 338 of Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 2009. Old and new.

[120] Wang, F.-Y. Analysis for diffusion processes on Riemannian manifolds, vol. 18 of Advanced Series on Statistical Science & Applied Probability. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2014.

[121] Wang, F.-Y., and Zhu, J.-X. Limit theorems in Wasserstein distance for empirical measures of diffusion processes on Riemannian manifolds. Ann. Inst. Henri Poincaré, Probab. Stat. 59, 1 (2023), 437–475.

Table 4.1: Notation Table

Symbol	**Description**
$\Delta_{\mathbb{R}^m}$	Laplace operator of $\mathbb{R}^m$, usually written as Δ for short if there is no ambiguity.
$\mathcal{M}$	A smooth manifold without boundary embedded in $\mathbb{R}^m$.
$\Delta_{\mathcal{M}}$	Laplace (-Beltrami) operator of manifold $\mathcal{M}$, usually written as Δ for short if there is no ambiguity.
∇	Covariant derivative/ Affine connection/ Gradient operator on $\mathcal{M}$.
D_t	Covariant derivative along a curve.
d	Dimension of $\mathcal{M}$.
$\mathbf{d}, \mathbf{d}^*$	Exterior differentiation on $\mathcal{M}$ and its formal adjoint.
$\int Y \circ \mathrm{d}W$	Stratonovich integral of Y against W.
$\mathrm{F}(\mathcal{M})$	Frame bundle of $\mathcal{M}$.
$\mathrm{O}(\mathcal{M})$	Orthonormal frame bundle of $\mathcal{M}$.
$\mathbf{p}$	Bundle projection.
$(e_i)_{i\in\overline{1;d}}$	The standard basis of $\mathbb{R}^d$.
$\mathrm{H}_z(u)$	Horizontal lift of $uz \in T_{\mathbf{p}u}\mathcal{M}$ to u.
$\mathbf{H}_Z$	Horizontal lift of $Z \in \Gamma(T\mathcal{M})$ to a vector field on $\mathrm{F}(\mathcal{M})$.
H_i	H_{e_i}.
$\Gamma(T\mathcal{M})$	Space of all smooth vector fields on $\mathcal{M}$.
$\mathrm{inj}_{\mathcal{M}}$	Injectivity radius of $\mathcal{M}$.
$\mathrm{J}F\cdot$	The absolute value of the Jacobian determinant of function F.
$\exp, \mathcal{E}$	Exponential map of $\mathcal{M}$.
$\Omega^k(\mathcal{M})$	Space of all smooth differential k-forms on $\mathcal{M}$.
$\mathfrak{X}(\mathcal{M})$	Space of all smooth vector fields on $\mathcal{M}$.
$\mathcal{F} = (\mathcal{F}_t)_{t\geq 0}$	Filtration of a filtered probability space.
$\mathbf{B}$	a Banach space
L^p	the space of functions with finite L^p- norm

Index

Abstract and keywords

This thesis studies the connection between probability theory and differential geometry. These two fields have many common ideas that can lead to new insights and applications in areas such as physics, finance, and machine learning.

My work is divided into three main parts. First, we study random operators on smooth, compact, and connected manifolds. We look at graph Laplacians, which are built from points sampled on a manifold. Graph Laplacians act as discrete versions of the Laplace–Beltrami operator, a fundamental object in differential geometry. We extend previous research by relaxing the assumptions on the kernel functions used to construct these graphs. As a result, we are able to prove uniform convergence rates for a wide range of kernel-induced random operators, including those related to the k-nearest neighbor random walk. We then show that as the number of sample points increases, the random walks on these graphs converge to diffusion processes on the manifold. This result helps to explain how discrete models can approximate continuous phenomena.

The second part of the thesis focuses on the long-time behavior of diffusion processes on manifolds. We study the occupation measures of these diffusion processes, which describe how much time these processes spend in different regions of the manifold. By smoothing these measures with an appropriate kernel, we can measure their convergence using the Wasserstein distance. The smoothing improves the convergence rate with respect to the existing results, and we prove that these rates are optimal in a minimax sense. This work is important for applications where one needs to recover the geometric properties of a manifold from observed trajectories. Unlike many previous studies that assume independent samples, our approach takes into account the natural time dependence found in stochastic processes.

The third part revisits the problem of density estimation on manifolds. Density estimation is a key problem in statistics, and it becomes more challenging when the data lie on a curved space. Building on recent work, we extend the known results by proving that the minimax convergence rates for density estimators still hold for a larger class of density functions. In our analysis, we consider densities that are not necessarily bounded away from zero and that may even have unbounded support. We use techniques from optimal transport theory and nonparametric statistics to generalize previous results. This not only improves our theoretical understanding but also has practical implications for data analysis in high-dimensional settings where the data are believed to lie on a low-dimensional manifold.

Keywords: diffusion processes on manifolds, optimal transport, kernel smoothing, minimax rate, limit theorems, random walks, differential geometry.

Résumé et mot-clés

Cette thèse étudie le lien entre la théorie des probabilités et de la géométrie différentielle. Ces deux domaines peuvent interagir pour conduire à de nouvelles idées et applications en physique, finance, et machine learning.

Mon travail est divisé en trois parties. Premièrement, nous étudions les opérateurs aléatoires sur les variétés lisses, compactes et connectées. Nous étudions les Laplaciens de graphe, qui sont construits à partir de points échantillonnés sur une variété. Ces Laplaciens agissent comme des versions discrètes de l'opérateur de Laplace-Beltrami, objet fondamental de la géométrie différentielle. Nous étendons les recherches précédentes en assouplissant les hypothèses sur les noyaux utilisées. Nous sommes ainsi en mesure de prouver des taux de convergence uniformes pour une large classe d'opérateurs aléatoires induits par le noyau, y compris ceux liés à la marche aléatoire du k-plus proches voisins le plus proche. Nous montrons ensuite que lors que le nombre de points de l'échantillon augmente, les marches aléatoires sur ces graphes convergent vers des processus de diffusion sur la variété. Ce résultat permet d'expliquer comment les modèles discrets peuvent approcher des processus de diffusion continus.

La deuxième partie de la thèse se concentre sur le comportement en temps long des processus de diffusion sur les variétés. Nous étudions les mesures d'occupation de ces processus, qui décrivent le temps passé dans différentes régions de la variété. En lissant ces mesures avec un noyau approprié, nous pouvons mesurer leur vitesse de convergence en distance de Wasserstein. Le lissage améliore la vitesse de convergence par rapport aux résultats existants, et nous prouvons que ces taux sont optimaux au sens minimax. Ce travail est important pour les applications où l'on a besoin de récupérer les propriétés géométriques d'une variété à partir de trajectoires l'explorant. Contrairement à de nombreuses études antérieures qui considèrent des échantillons indépendants, notre approche prend en compte la dépendance temporelle naturelle que l'on trouve dans les processus stochastiques.

La troisième partie revisite le problème de l'estimation de la densité sur les variétés. L'estimation de la densité est un problème clé en statistique, et elle devient plus difficile lorsque les données se trouvent sur un espace courbé. Nous étendons les résultats connus en prouvant que les taux de convergence minimax pour les estimateurs de densité restent valables pour une classe plus large de fonctions de densité. Dans notre analyse, nous considérons des densités qui ne sont pas nécessairement minorée par une constante positive et qui peuvent de plus avoir un support non borné. Nous utilisons des techniques de la théorie du transport optimal et de la statistique non paramétrique pour généraliser les résultats précédents. Cela améliore non seulement notre compréhension théorique, mais a également des implications pratiques pour l'analyse des données dans des contextes à haute dimension où les données sont censées se situer sur une variété de basse dimension.

Mot-clés: processus de diffusion sur les variétés, transport optimal, lissage par noyaux, vitesse minimax , théorèmes limites, marches aléatoires, géométrie différentielle.