Adv. Appl. Probab. 1–36 (2026)
*doi:*10.1017/apr.2026.10060

STRONG UNIFORM CONVERGENCE OF LAPLACIANS OF RANDOM GEOMETRIC AND DIRECTED *k*NN GRAPHS ON COMPACT MANIFOLDS

HÉLÉNE GUÉRIN,* *UQAM*
DINH-TOÀN NGUYEN,** *Université Gustave Eiffel and UQAM*
VIET CHI TRAN,*** *Inria*

Abstract

Consider n points independently sampled from a density p of class $\mathcal{C}^2$ on a smooth compact d-dimensional submanifold $\mathcal{M}$ 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. In contrast to known results, our result does not require the kernel K to be continuous, which covers the cases of walks exploring k-nearest neighbor (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.

Keywords: Laplace– Beltrami operator; graph Laplacian; kernel estimator; large-sample approximations; Riemannian geometry; concentration inequalities; k-nearest neighbors; kNN random walk

2020 Mathematics Subject Classification: Primary 60F05; 05C81; 62G05
Secondary 53B20

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 $\mu(\mathrm{d}x)$ the Euclidean distance of $\mathbb{R}^m$, the geodesic distance on $\mathcal{M}$, and the volume measure on $\mathcal{M}$, respectively. Let $(X_i,\, i\in\mathbb{N})$ be a sequence of independent and identically distributed (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 $\mathcal{M}$.

Received 11 December 2024; accepted 27 January 2026.
* Postal address: Département de Mathématiques, UQAM, Canada. Email: guerin.helene@uqam.ca
** Postal address: LAMA, Université Gustave Eiffel, France; Département de Mathématiques, UQAM, Canada. Email: dinhtoan.nguyenvn@gmail.com
*** Postal address: Centre Inria de l'Université de Lille, France. Email: viet-chi.tran@inria.fr

In this paper, we study the limit of the random operators $(\mathcal{A}_{h_n,n},\, n\in\mathbb{N})$:

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

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 $[\![1,n]\!]=\{1,\ldots,n\}$. Such a process jumps from its position x to a 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. Note 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, we show that the limits of the two operators $\mathcal{A}_{h_n,n}$ and $\widetilde{\mathcal{A}}_{h_n,n}$ are the same, due to the equivalence between the geodesic and Euclidean distances (see [13, Proposition 2]). In view of applications to manifold learning, when $\mathcal{M}$ is unknown and when only the sample points X_i are available, using the norm of the ambient space $\mathbb{R}^m$ can be justified.

The operator (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 the machine learning literature as an approximation to the Laplace–Beltrami operator of $\mathcal{M}$ (see, for example, [2, 3, 15, 27, 31, 32]). Most of these results have been obtained for the Gaussian kernel, $K(a)=\mathrm{e}^{-a^2}$, or sufficiently smooth kernels. These assumptions on the kernel are too strong to include the case of ε-geometric graphs or the 'true' k-nearest neighbor (kNN) graphs, and that correspond to choices of indicators for the kernel K. In recent years, much work has been done to relax the regularity of K, giving rise to many interesting papers (e.g., [7, 33]), as discussed in the following.

In the sequel, under a mild assumption on K (weaker than continuity; see Assumption 1) and a condition on the rate of convergence of (h_n), we show that the sequence of operators $(\mathcal{A}_{h_n,n})$ almost surely converges uniformly on $\mathcal{M}$ to the second-order differential operator $\mathcal{A}$ on $\mathcal{M}$ defined as

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

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

$$c_0 := \frac{1}{d}\int_{\mathbb{R}^d} K(\|v\|_2)\,\|v\|_2^2\,\mathrm{d}v = \frac{1}{d}S_{d-1}\int_0^\infty K(a)a^{d+1}\,\mathrm{d}a, \tag{3}$$

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

Assumption 1. *The kernel $K:\mathbb{R}_+\to\mathbb{R}_+$ is a measurable function with $K(\infty)=0$ and of bounded variation H such that*

$$\int_0^\infty a^{d+3}\,\mathrm{d}H(a) < \infty. \tag{4}$$

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

Recall that the total variation H of a kernel K is defined for each non-negative 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 1 is the key to avoiding making continuity hypotheses on the kernel K. This allows, for example, step functions, which are very useful for modeling and inference purposes (beyond applications in machine learning; see, for example, [24, 30]).

Theorem 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 Assumption 1 on the kernel K is satisfied and that*

$$\lim_{n\to+\infty} h_n = 0 \quad \textit{and} \quad \lim_{n\to+\infty} \frac{\log h_n^{-1}}{nh_n^{d+2}} = 0. \tag{5}$$

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

$$\sup_{x\in\mathcal{M}} \left|\mathcal{A}_{h_n,n}(f)(x) - \mathcal{A}(f)(x)\right| = O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}} + h_n\right). \tag{6}$$

Note that the window h_n that optimizes the convergence rate in (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 [18].

An important point in the assumptions of Theorem 1 is that K is not necessarily continuous nor with mass equal to 1. This can allow one to tackle the cases of geometric or kNN graphs, for example.

This theorem extends the convergence given by Giné and Koltchinskii [15, Theorem 5.1]. They consider the kernel $K(a) = \mathrm{e}^{-a^2}$ 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 third order. For such class of functions, the constants on the right-hand side of (6) can be made independent of f and we recover a similar uniform bound.

Condition (5) results from a classical bias–variance trade-off that appears in a similar way in the work of Giné and Koltchinskii [15]. Note that the speed $\sqrt{\log h_n^{-1}/(nh_n^{d+2})}$ is also obtained by these authors under the additional assumption that $nh_n^{d+4}/\log h_n^{-1} \to 0$. We do not make this assumption here. When the additional assumption of Giné and Koltchinskii is satisfied, our rate and their rate coincide as $h_n^2 = o\big(\log h_n^{-1}/(nh_n^{d+2})\big)$. Hein *et al.* [17, 18] extended the results of Giné and Koltchinskii to other kernels K, but requiring in particular that these kernels are twice continuously differentiable and with exponential decay (see, for example, [17, Assumption 2] or [18, Assumption 20]). Singer [31], considering Gaussian kernels, upper-bounds the variance term in a different manner compared with Hein *et al.*, improving the convergence rate when p is the uniform distribution.

To the best of 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.* [33] also worked under the bounded variation assumption on K. Additionally, they had to assume that K is compactly supported. In [7], Calder and García-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 $\exp(-C\delta^2 nh_n^{d+2})$, when $n \to +\infty$. In this paper, thanks to the uniform convergence in Theorem 1, we obtain a similar result with uniformity on the test functions f and the point $x \in \mathcal{M}$.

Corollary 1. *Suppose that the density p on the smooth manifold $\mathcal{M}$ is of class $\mathcal{C}^2$, and that Assumption 1 and (5) are satisfied. Then there exists a constant $C' > 0$ (see (58)), such that for all n and $\delta \in \left[h_n \vee \sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}}, 1\right]$, we have*

$$\mathrm{P}\Big(\sup_{f\in\mathcal{F}}\sup_{x\in\mathcal{M}} |\mathcal{A}_{h_n,n}(f)(x) - \mathcal{A}f(x)| > C'\delta\Big) \leqslant \exp\left(-nh_n^{d+2}\delta^2\right), \tag{7}$$

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

The fact that the convergence in Theorem 1 is uniform has several other applications. First, the uniformity on the class of test functions $\mathcal{F}$ provides the convergence of the generators seen as a sequence of operators, and is related to the convergence of the semi-groups of the associated stochastic processes, as known since the Trotter–Kato theorem (see, for example, [36]). Uniformity over x can be, for example, a step toward studying the spectral convergence for the graph Laplacian using the Courant–Fisher minmax principle (see, for example, [7]). Interestingly, the uniform convergence of the Laplacians is also used to study Gaussian free fields on manifolds [10, 34].

The result of Theorem 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 $\{X_1, \ldots, X_n\}$ 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

$$R_{n,k}(x) = \inf\left\{r \geqslant 0,\ \sum_{i=1}^{n} \mathbf{1}_{\|x-X_i\|_2 \leqslant r} \geqslant k\right\}. \tag{8}$$

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

$$\mathcal{A}_n^{k\mathrm{NN}}(f)(x) := \frac{1}{nR_{n,k_n}^{d+2}(x)} \sum_{i=1}^{n} \mathbf{1}_{[0,1]}\left(\frac{\|X_i - x\|_2}{R_{n,k_n}(x)}\right)(f(X_i) - f(x)). \tag{9}$$

A major difficulty here is that the width of the moving window, $R_{n,k_n}(x)$ is random and depends on $x \in \mathcal{M}$, in contrast 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 has been fully and rigorously considered, first, because of the non-regularity of the kernel K, and second, because the kNN graph is not symmetric, or more precisely, because the fact that 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.* [33] 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 Laplacians, without convergence rate. In the present proof, we do not require such a Taylor-like expansion. Let us mention also the work of Calder and García Trillos [7] where the spectral convergence is established. In other papers such as [9], equation (8) is considered for defining the window width h_n but the kernel K remains continuous.

We prove the following limit theorem for the rescaled kNN Laplacian.

Theorem 2. *Under Assumption 1, if the density $p \in \mathcal{C}^2(\mathcal{M})$ is such that for all $x \in \mathcal{M}$,*

$$0 < p_{min} \leqslant p(x) \leqslant p_{max}, \tag{10}$$

and if

$$\lim_{n\to+\infty} \frac{k_n}{n} = 0 \quad \textit{and} \quad \lim_{n\to+\infty} \frac{1}{n}\left(\frac{k_n}{n}\right)^{-1-2/d} \log\left(\frac{k_n}{n}\right) = 0, \tag{11}$$

then with probability 1, we have

$$\sup_{x\in\mathcal{M}} \left|\mathcal{A}_n^{k\mathrm{NN}}(f)(x) - \mathcal{A}(f)(x)\right| = O\left(\sqrt{\log\left(\frac{n}{k_n}\right)}\frac{1}{\sqrt{k_n}}\left(\frac{n}{k_n}\right)^{1/d} + \left(\frac{k_n}{n}\right)^{1/d}\right). \tag{12}$$

This theorem is proved in Section 6. Note that the important point in assumption (10) is the lower bound, since in our compact manifold case, any continuous function p is bounded. Condition (11) and the rate of convergence in (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 [9]. The proof of Theorem 2 follows the main steps presented in the proof of Theorem 1 with some slight modifications.

Note that assumption (11) is satisfied for

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

for instance. Optimizing the upper bound in (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 this paper, the generator $\mathcal{A}_{h_n,n}$ can be seen as the infinitesimal generator of continuous-time random walks $(X^{(n)})_{n\geqslant 0}$ visiting the points $(X_i)_{i\in[\![1,n]\!]}$. Their trajectories are described by the following stochastic differential equation (SDE):

$$X_t^{(n)} = X_0^{(n)} + \int_0^t \int_{\mathbb{N}} \int_{\mathbb{R}_+} \mathbf{1}_{i\leqslant n} \mathbf{1}_{\theta \leqslant \frac{1}{nh_n^{d+2}} K\left(\frac{\|X_i - X_{s_-}^{(n)}\|_2}{h_n}\right)} \left(X_i - X_{s_-}^{(n)}\right) Q(\mathrm{d}s, \mathrm{d}i, \mathrm{d}\theta), \tag{13}$$

with initial condition $X_0^{(n)}$ and where $Q(\mathrm{d}s, \mathrm{d}i, \mathrm{d}\theta)$ is a Poisson point measure on $\mathbb{R}_+ \times \mathbb{N} \times \mathbb{R}_+$, 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 $\mathbb{N}$ (see, for example, [20] for an introduction on SDEs driven by Poisson point measures). Consider a fixed time $T > 0$. 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

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

processes, embedded with the Skorokhod topology (see [5, 21]), and converge to a diffusion on the manifold $\mathcal{M}$ with generator $\mathcal{A}$.

Theorem 3. *Let $T > 0$ be fixed. Suppose that the density p on the smooth manifold $\mathcal{M}$ is of class $\mathcal{C}^2$ and that Assumption 1 and (5) are satisfied. Assume additionally that the initial conditions $\left(X_0^{(n)}\right)_{n\geqslant 0}$ converge in distribution to a probability measure ν on $\mathcal{M}$. Then the sequence of random walks $(X^{(n)})_{n\geqslant 0}$ 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.*

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

$$
\begin{aligned}
X_t^{(n),k\mathrm{NN}} &= X_0^{(n),k\mathrm{NN}} \\
&+ \int_0^t \int_{\mathbb{N}} \int_{\mathbb{R}_+} \mathbf{1}_{i\leqslant n} \mathbf{1}_{\theta \leqslant \frac{1}{nR_{n,k_n}^{d+2}\left(X_{s-}^{(n),k\mathrm{NN}}\right)}} \mathbf{1}_{\|X_i - X_{s-}^{(n),k\mathrm{NN}}\|_2 \leqslant R_{n,k_n}^{d+2}\left(X_{s-}^{(n),k\mathrm{NN}}\right)} \left(X_i - X_{s-}^{(n)}\right) Q(\mathrm{d}s, \mathrm{d}i, \mathrm{d}\theta).
\end{aligned}
\tag{14}
$$

Theorem 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 Assumption 1, (10), and (11) are satisfied. Assume additionally that the initial conditions $(X_0^{(n),k\mathrm{NN}})_{n\geqslant 0}$ converge in distribution to a probability measure ν on $\mathcal{M}$. Then the sequence of random walks $(X^{(n),k\mathrm{NN}})_{n\geqslant 0}$ 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 convergence established in Theorems 3 and 4 of the above random walks is of particular interest, as certain functionals are continuous with respect to topologies on path spaces. In particular, it implies the convergence of the occupation measures of the random walks toward that of the diffusive process with generator $\mathcal{A}$. This limiting occupation measure has been studied in [12, 35].

The rest of the paper is organized as follows. In Section 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 3, we provide some geometric background that will be useful for the study of the third term, which is treated in Section 4. The two first statistical terms are considered in Section 5, which will end the proof of Theorem 1. Corollary 1 is then proved at the end of this section. In Section 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 7.

Notation 1. *In this paper* $\mathrm{diam}(\mathcal{M})$, $B_{\mathbb{R}^d}(0, r)$, *and* S_{d-1} *denote 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* $\mathbb{R}^d$, *respectively.*

2. Outline of the proof for Theorem 1

First, we focus on the proof of Theorem 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

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

each $h > 0$, $x \in \mathcal{M}, f \in \mathcal{C}^3(\mathcal{M})$:

$$\mathcal{A}_h(f)(x) := \frac{1}{h^{d+2}} \int_{\mathcal{M}} K\left(\frac{\|x-y\|_2}{h}\right)(f(y)-f(x))p(y)\mu(\mathrm{d}y), \tag{15}$$

$$\tilde{\mathcal{A}}_h(f)(x) := \frac{1}{h^{d+2}} \int_{\mathcal{M}} K\left(\frac{\rho(x,y)}{h}\right)(f(y)-f(x))p(y)\mu(\mathrm{d}y). \tag{16}$$

The difference between $\mathcal{A}_h$ and $\tilde{\mathcal{A}}_h$ lies 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 that these two metrics are comparable for close x and y.

Theorem 5 (Approximation inequality for Riemannian distance) [13, Proposition 2]. *There is a constant c such that, for $x, y \in \mathcal{M}$, we have*

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

Let us sketch the proof of Theorem 1. By the classical triangular inequality,

$$\begin{aligned}\left|\mathcal{A}_{h_n,n}(f)(x) - \mathcal{A}(f)(x)\right| \leqslant{}& \left|\mathcal{A}(f)(x) - \widetilde{\mathcal{A}}_{h_n}(f)(x)\right| \\ &+ \left|\widetilde{\mathcal{A}}_{h_n}(f)(x) - \mathcal{A}_{h_n}(f)(x)\right| \\ &+ \left|\mathcal{A}_{h_n}(f)(x) - \mathcal{A}_{h_n,n}(f)(x)\right|. \end{aligned} \tag{17}$$

The first term on the right-hand side of (17) corresponds to the convergence of the kernel-based generator to a continuous diffusion generator on $\mathcal{M}$. The following proposition is proved in Section 4.2.

Proposition 1 (Convergence of averaging kernel operators). *Under Assumption 1, and if p is of class $\mathcal{C}^2$, we have, for all $f \in \mathcal{C}^3(\mathcal{M})$,*

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

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 [15, Theorem 3.1], but with continuous assumptions on K that exclude the kNN cases.

The second term in (17) corresponds to the approximation of the Euclidean distance by the geodesic distance and is dealt with in the following proposition, proved in Section 4.3.

Proposition 2. *Under Assumption 1, 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 on the right-hand side of (17), note that

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

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 5 using Vapnik–Chervonenkis theory.

Proposition 3. *Under Assumption 1 and for a bounded measurable function p, we have, for all $f \in \mathcal{C}^3(\mathcal{M})$,*

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

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

It is worth noting 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.

3. Some geometric background

3.1. Riemannian manifold

Let us recall some facts from differential geometry that will be useful. We refer the reader to [8, 26] 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 with respect to the scalar product $\mathbf{g}(x)$.

Consider a coordinate chart $\Phi = (x^1,\ldots,x^d) : U \to \mathbb{R}^d$ on a neighborhood U of x. Denote by $\left\{ \frac{\partial}{\partial x^1}\Big|_x, \frac{\partial}{\partial x^2}\Big|_x, \ldots, \frac{\partial}{\partial x^d}\Big|_x \right\}$ the natural basis of $T_x\mathcal{M}$ associated with the coordinates $(x^1,\ldots,x^d)$. 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 $\eta \in T_x\mathcal{M}$,

$$\langle\xi,\eta\rangle_{\mathbf{g}} = \sum_{i,j=1}^{d} g_{ij}(x)\xi^i\eta^j, \tag{18}$$

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

$$g_{ij}(x) := \left\langle \frac{\partial}{\partial x^i}\Big|_x, \frac{\partial}{\partial x^j}\Big|_x \right\rangle_{\mathbf{g}}, \tag{19}$$

and $g_{ij} : U \subset M \to \mathbb{R}$ is smooth. For a real function f on $\mathcal{M}$, we will denote by $\widehat{f}$ 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

$$\widehat{g}_{ij} := g_{ij} \circ \Phi^{-1}, \tag{20}$$

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 [26, pp. 131–132] or the remark below for a definition). We denote by $\mathcal{E}_x$ the Riemannian normal parameterization at x, that is, $\mathcal{E}_x^{-1}$ is the corresponding normal coordinate chart.

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

Remark 1 *(Construction of $\mathcal{E}_x^{-1}$)*. For the sake of completeness, we briefly recall the construction of [26]. Let U be an open subset of $\mathcal{M}$. There exists a local orthonormal frame $(E_i)_{i\in[\![1,d]\!]}$ over U (see [26, Proposition 2.8, p. 14]). The tangent bundle TU can be identified with $U \times \mathbb{R}^d$ thanks to the smooth map

$$F: \begin{array}{ccc} U \times \mathbb{R}^d & \rightarrow & TU \\ (x, (v_1, \ldots, v_d)) & \mapsto & v = \sum_{i=1}^d v_i E_i|_x. \end{array} \tag{21}$$

Thus, for each $x \in U$, $F(x, \cdot)$ is an isometry between $\mathbb{R}^d$ and $T_x\mathcal{M}$.

Recall that by [26, Proposition 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 for all $x \in \mathcal{M}$, $\overrightarrow{0}_x \in W$, where $\overrightarrow{0}_x$ is the zero element of $T_x\mathcal{M}$. Then the map $\exp \circ F : (x, v) \mapsto \mathcal{E}_x(v) := \exp \circ F(x, (v_1, \ldots, v_d))$ 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 6 (Derivatives of Riemannian metrics in normal coordinate charts) [26, Proposition 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 $(\widehat{g}_{ij}; 1 \le i, j \le d)$ be the coordinate representation of the Riemannian metric of $\mathcal{M}$ in the local chart $\mathcal{E}_x^{-1}$. Then, for all i,j,*

$$\widehat{g}_{ij}(0) = \delta_{ij}, \quad \widehat{g}'_{ij}(0) = 0, \tag{22}$$

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

$$\rho(x, y) = \|\mathcal{E}_x^{-1}(y)\|_2. \tag{23}$$

Notation 2. *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 bilinear map and trilinear maps that represent the second- and third-order derivatives 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 key to our computations in the sequel.

Theorem 7 (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 $x \in \mathcal{M}$, the following assertions hold.*

(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 $(\widehat{g}^x_{ij}(v); 1 \le i, j \le d)$ 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 $v \in B_{\mathbb{R}^d}(0, c_1)$,*

$$\left| \sqrt{\det \widehat{g}^x_{ij}(v)} - 1 \right| \le c_2 \|v\|_2^2. \tag{24}$$

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

(iii) *We have* $\|\mathcal{E}_x(v) - x\|_2 \leq \|v\|_2$. *In addition, for all* $v \in B_{\mathbb{R}^d}(0, c_1)$,

$$\|\mathcal{E}_x(v) - x - \mathcal{E}_x'(0)(v)\|_2 \leq c_2\|v\|_2^2 \tag{25}$$

and

$$\|\mathcal{E}_x(v) - x - \mathcal{E}_x'(0)(v) - \frac{1}{2}\mathcal{E}_x''(0)(v, v)\|_2 \leq c_2\|v\|_2^3. \tag{26}$$

Proof. Let U be an open domain of a local chart of $\mathcal{M}$. Following Remark 1 and noting 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 note that $\|\mathcal{E}_x(v) - x\|_2 \leq \|v\|_2$ thanks to Theorem 5 and (23).

Restricting U if necessary (by an open subset with compact closure in U), there exist 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, for all $v \in B_{\mathbb{R}^d}(0, c_1)$, expressions (25) and (26) hold by Taylor expansions of $\mathcal{E}_x$. Expression (24) is a consequence of the smoothness of $\mathcal{E}_x$ and of the compactness of $\mathcal{M}$ (see Propositions 5.19 and 5.24 in [26]).

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 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 $\mathcal{E}_x$. □

3.2. Gradient operator and 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 $\nabla_{\mathbb{R}^m}$ and 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

$$\nabla_{\mathcal{M}} f(x) = \sum_{i,j=1}^{d} g^{ij}(x)\frac{\partial f}{\partial x^i}(x)\left.\frac{\partial}{\partial x^j}\right|_x, \tag{27}$$

where $(g^{ij})_{1\leqslant i,j,\leqslant d}$ 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 $\mathcal{C}^1$,

$$\langle\nabla_{\mathcal{M}}(f), \nabla_{\mathcal{M}}(h)\rangle_{\mathbf{g}} = \sum_{i,j=1}^{d} g^{ij}\frac{\partial f}{\partial x^i}\frac{\partial h}{\partial x^j}. \tag{28}$$

The Laplace–Beltrami operator is defined by (see [19, Section 3.1])

$$\Delta_{\mathcal{M}} f := \sum_{i,j=1}^{d} \frac{1}{\sqrt{\det(\mathbf{g})}}\frac{\partial}{\partial x^i}\left(\sqrt{\det(\mathbf{g})}g^{ij}\frac{\partial f}{\partial x^j}\right). \tag{29}$$

When using normal coordinates, the expressions for the Laplacian and the gradient of a smooth function f at a point x match their definitions in $\mathbb{R}^d$.

Proposition 4. *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) $\Delta_{\mathcal{M}} f(x) = \Delta_{\mathbb{R}^d}\hat{f}(0)$,

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

Proof. Recall that $g_{ij}(x) = \widehat{g}_{ij}(0)$. By Theorem 6, we know that $\hat{g}_{ij}(0) = \delta_{ij}$, thus, $\widehat{g}^{ij}(0) = \delta_{ij}$ and assertion (i) is a consequence of (28). For assertion (ii) we use (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 can conclude. □

4. Some kernel-based approximations of $\mathcal{A}$

The aim of this section is to prove the estimates for the two error terms on the right-hand side of (17) and prove Propositions 1 and 2. Both error terms are linked with the geometry of the problem and use the results presented in Section 3. The first deals with the approximation of the Laplace–Beltrami operator by a kernel estimator (see Section 4.2), whereas the second treats the differences between the use of the Euclidean norm of $\mathbb{R}^m$ and the use of the geodesic distance (see Section 4.3).

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, for example, [4]). In particular, the following lemma does not require the compactness of the support of K, which also encompasses the Gaussian kernel function studied in [15], for example.

Lemma 1. *Under Assumption 1, uniformly in $x \in \mathcal{M}$, when h converges to 0, we have*

$$\frac{1}{h^{d+2}} \int_{\mathcal{M}} \mathbf{1}_{\rho(x,y)\geq c_1} K\left(\frac{\rho(x,y)}{h}\right) \mu(\mathrm{d}y) = o(h), \tag{30}$$

$$\frac{1}{h^{d+2}} \int_{\mathcal{M}} \mathbf{1}_{\rho(x,y)\geq c_1} K\left(\frac{\|x-y\|_2}{h}\right) \mu(\mathrm{d}y) = o(h), \tag{31}$$

and there is a generic constant c such that, for all points $x \in \mathcal{M}$ and positive number $h > 0$, we have

$$\frac{1}{h^{d+2}} \int_{\mathcal{M}} K\left(\frac{\rho(x,y)}{h}\right) \|x-y\|_2^3 \mu(\mathrm{d}y) \leq ch, \tag{32}$$

$$\frac{1}{h^{d+2}} \int_{\mathcal{M}} K\left(\frac{\rho(x,y)}{h}\right) \|x-y\|_2^2 \mu(\mathrm{d}y) \leq c, \tag{33}$$

$$\frac{1}{h^{d+2}} \int_{\mathcal{M}} K\left(\frac{\|x-y\|_2}{h}\right) \|x-y\|_2^3 \mu(\mathrm{d}y) \leq ch, \tag{34}$$

$$\frac{1}{h^{d+2}} \int_{\mathcal{M}} K\left(\frac{\|x-y\|_2}{h}\right) \|x-y\|_2^2 \mu(\mathrm{d}y) \leq c. \tag{35}$$

Proof. Using Lemma 4, we have

$$\begin{aligned}
\int_{\mathcal{M}} \mathbf{1}_{\rho(x,y)\geq c_1} K\left(\frac{\rho(x,y)}{h}\right) \mu(\mathrm{d}y) &\leq \mu(\mathcal{M}) \sup_{r\geq c_1} K\left(\frac{r}{h}\right) \\
&\leq \mu(\mathcal{M}) \left[H(\infty) - H\left(\frac{c_1}{h}\right)\right] \\
&= \mu(\mathcal{M}) \int_{(c_1/h,\infty)} \mathrm{d}H(a) \\
&\leq h^{d+3} \frac{\mu(\mathcal{M})}{c_1^{d+3}} \int_{(c_1/h,\infty)} a^{d+3} \mathrm{d}H(a).
\end{aligned} \tag{36}$$

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

Thanks to the boundedness of $\int_0^\infty a^{d+3}\,\mathrm{d}H(a)$, we obtain (30). Then, as a consequence of (30), by the compactness of $\mathcal{M}$, we readily observe that uniformly in $x \in \mathcal{M}$, when h converges to 0,

$$\frac{1}{h^{d+2}} \int_{\mathcal{M}} \mathbf{1}_{\rho(x,y)\geq c_1} K\left(\frac{\rho(x,y)}{h}\right) \|x-y\|_2^3 \mu(\mathrm{d}y) = o(h).$$

Thus, to prove inequality (32), it remains to prove that uniformly in x, when h converges to 0,

$$I := \frac{1}{h^{d+3}} \int_{\mathcal{M}} \mathbf{1}_{\rho(x,y)<c_1} K\left(\frac{\rho(x,y)}{h}\right) \|x-y\|_2^3 \mu(\mathrm{d}y) = O(1). \tag{37}$$

Recall that in Theorem 7, 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 (23). Thus, the term I on the left-hand side of (37) can be rewritten 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 Theorems 5 and 7(ii) and (iii),

$$\begin{aligned}
I &\leq \frac{c_2}{h^{d+3}} \int_{B_{\mathbb{R}^d}(0,c_1)} K\left(\frac{\|v\|_2}{h}\right) \|v\|_2^3 \left(1 + c_2\|v\|_2^2\right)\,\mathrm{d}v \\
&\leq \frac{c_2}{h^{d+3}} \int_{B_{\mathbb{R}^d}(0,c_1)} K\left(\frac{\|v\|_2}{h}\right) \|v\|_2^3 \left(1 + c_2 c_1^2\right)\mathrm{d}v \\
&\leq \frac{c_2}{h^{d+3}} \int_{\mathbb{R}^d} K\left(\frac{\|v\|_2}{h}\right) \|v\|_2^3 \left(1 + c_2 c_1^2\right)\mathrm{d}v \\
&= c_2\big(1 + c_2 c_1^2\big) \int_{\mathbb{R}^d} K(\|v\|_2)\|v\|_2^3\,\mathrm{d}v.
\end{aligned} \tag{38}$$

Using the spherical coordinate system when $d \geqslant 2$,

$$\begin{aligned}
I &\leqslant c_2\big(1 + c_2 c_1^2\big) \left[\int_0^\infty K(a)a^3 \times a^{d-1}\,\mathrm{d}a\right] \\
&\quad \times \left[\int_{[0,2\pi]\times[0,\pi]^{d-2}} \sin^{d-2}(\theta_1)\sin^{d-3}(\theta_2)\cdots\sin(\theta_{d-2})\,\mathrm{d}\theta\right] \\
&= c_2\big(1 + c_2 c_1^2\big)\, S_{d-1} \left[\int_0^\infty K(a)a^{d+2}\,\mathrm{d}a\right].
\end{aligned}$$

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 4 in the Appendix and by Fubini's theorem, we have

$$\begin{aligned}
I &\leq c_2\big(1 + c_2 c_1^2\big) S_{d-1} \int_0^\infty (H(\infty) - H(a))\, a^{d+2}\,\mathrm{d}a \\
&= c_2\big(1 + c_2 c_1^2\big)(d+3)^{-1} S_{d-1} \int_{[0,\infty]} b^{d+3}\,\mathrm{d}H(b) < \infty.
\end{aligned} \tag{39}$$

Therefore, inequality (32) is proved. The proof of inequality (33) is similar.

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

For inequalities (31), (34), and (35), we observe that they are indeed consequences of (30), (32), and (33). Consider, for example, (31), again using Lemma 4 and Theorem 5:

$$\begin{aligned}
&\frac{1}{h^{d+2}} \int_{\mathcal{M}} \mathbf{1}_{\rho(x,y)\geqslant c_1} K\left(\frac{\|x-y\|_2}{h}\right) \mu(\mathrm{d}y) \\
&\leq \frac{1}{h^{d+2}} \int_{\mathcal{M}} \mathbf{1}_{\rho(x,y)\geqslant c_1} \left[H(\infty) - H\left(\frac{\rho(x,y)}{h(1+c_3\|x-y\|_2^2)}\right)\right] \mu(\mathrm{d}y) \\
&\leq \frac{1}{h^{d+2}} \int_{\mathcal{M}} \mathbf{1}_{\rho(x,y)\geqslant c_1} \left[H(\infty) - H\left(\frac{\rho(x,y)}{h(1+c_3\mathrm{diam}(\mathcal{M})^2)}\right)\right] \mu(\mathrm{d}y) \\
&= \frac{1}{h^{d+2}} \int_{\mathcal{M}} \mathbf{1}_{\rho(x,y)\geqslant c_1} \tilde{K}\left(\frac{\rho(x,y)}{h}\right) \mu(\mathrm{d}y),
\end{aligned} \tag{40}$$

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 the fact that H is a non-decreasing function. Thus, inequality (31) corresponds to inequality (30), where K is replaced with $\tilde{K}$. Clearly, the function $\tilde{K}$ is of bounded variation and satisfies Assumption 1, which concludes the proof for (31). The arguments are similar for (34) and (35).

4.2. Proof of Proposition 1

We now prove Proposition 1, dealing with the approximation of the Laplace–Beltrami operator by a kernel operator. In the course of the proof, some quantities involving gradients and Laplacians will appear repetitively. The next lemma will be useful in dealing with these expressions and its proof is postponed to Appendix C.

Lemma 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 that $G : \mathbb{R}_+ \to \mathbb{R}$ is a locally bounded measurable function. Then, for all $c > 0$,*

$$\begin{aligned}
&\int_{B_{\mathbb{R}^d}(0,c)} G(\|v\|_2) \langle \nabla_{\mathbb{R}^m} f(x), k'(0)(v)\rangle \langle \nabla_{\mathbb{R}^m} h(x), k'(0)(v)\rangle \,\mathrm{d}v \\
&\qquad = \langle \nabla_{\mathbb{R}^d}(f\circ k)(0), \nabla_{\mathbb{R}^d}(h\circ k)(0)\rangle \left(\frac{1}{d}\int_{B_{\mathbb{R}^d}(0,c)} G(\|v\|_2)\|v\|_2^2 \,\mathrm{d}v\right)
\end{aligned} \tag{41}$$

and

$$\begin{aligned}
&\int_{B_{\mathbb{R}^d}(0,c)} G\left(\|v\|_2\right) \left[\left\langle \nabla_{\mathbb{R}^m} f(x), k'(0)(v) + \frac{1}{2}k''(0)(v,v)\right\rangle \right. \\
&\qquad \left. + \frac{1}{2} f''(x)(k'(0)(v), k'(0)(v))\right] \mathrm{d}v \\
&\qquad = \frac{1}{2}\Delta_{\mathbb{R}^d}(f\circ k)(0) \left(\frac{1}{d}\int_{B_{\mathbb{R}^d}(0,c)} G(\|v\|_2)\|v\|_2^2 \,\mathrm{d}v\right).
\end{aligned} \tag{42}$$

We are now in a position to prove Proposition 1. As $\mathcal{M}$ is compact, it is a properly embedded submanifold of $\mathbb{R}^m$ (see [25, 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 [25, Lemma 5.34]). Thus, without loss of generality, assume f and p are $\mathcal{C}^3$ and $\mathcal{C}^2$ functions on $\mathbb{R}^m$ with compact supports, respectively.

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

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) and (16). Thanks to Lemma 1, it is sufficient to prove that

$$\left|\mathcal{A}(f)(x) - \frac{1}{h^{d+2}} \int_{\mathcal{M}} \mathbf{1}_{\rho(y,x)<c_1} K\left(\frac{\rho(x,y)}{h}\right)(f(y) - f(x))p(y)\mu(\mathrm{d}y)\right| = O(h),$$

uniformly in $x \in \mathcal{M}$, to conclude the proof of Proposition 1, where $c_1 > 0$ is introduced in Theorem 7.

In addition, 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 $x, y \in \mathcal{M}$,

$$\begin{aligned}\Bigg|(f(y) - f(x))p(y) - \left(\langle \nabla_{\mathbb{R}^m} f(x), y - x\rangle + \frac{1}{2} f''(x)(y - x, y - x)\right)p(x) \\ - \langle \nabla_{\mathbb{R}^m} f(x), y - x\rangle \langle \nabla_{\mathbb{R}^m} p(x), y - x\rangle\Bigg| \leq c_4 \|x - y\|_2^3. \end{aligned} \tag{43}$$

Hence, by inequality (32), it is sufficient to prove that uniformly in x,

$$\begin{aligned} J_1 &:= \frac{1}{h^{d+2}} \int_{\mathcal{M}} \mathbf{1}_{\rho(y,x)<c_1} K\left(\frac{\rho(x,y)}{h}\right) \langle \nabla_{\mathbb{R}^m} f(x), y - x\rangle \langle \nabla_{\mathbb{R}^m} p(x), y - x\rangle \mu(\mathrm{d}y) \\ &= c_0 \langle \nabla_{\mathcal{M}}(f)(x), \nabla_{\mathcal{M}}(p)(x)\rangle_{\mathbf{g}} + O(h) \end{aligned} \tag{44}$$

and

$$\begin{aligned} J_2 &:= \frac{1}{h^{d+2}} \int_{\mathcal{M}} \mathbf{1}_{\rho(y,x)<c_1} K\left(\frac{\rho(x,y)}{h}\right) \left[\langle \nabla_{\mathbb{R}^m} f(x), y - x\rangle + \frac{1}{2} f''(x)(y - x, y - x)\right] \mu(\mathrm{d}y) \\ &= \frac{1}{2} c_0 \Delta_{\mathcal{M}}(f)(x) + O(h). \end{aligned} \tag{45}$$

The proof is similar to the study of I given by (37) in the proof of Lemma 1. We rewrite the integrals considered 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 7 we have

$$\begin{aligned} &\left| J_1 - \int_{B_{\mathbb{R}^d}(0,c_1)} K\left(\frac{\|v\|_2}{h}\right) \langle \nabla_{\mathbb{R}^m} f(x), \mathcal{E}_x(v) - x\rangle \langle \nabla_{\mathbb{R}^m} p(x), \mathcal{E}_x(v) - x\rangle \, \mathrm{d}v \right| \\ &\quad \leq \frac{c_2}{h^{d+2}} \|\nabla_{\mathbb{R}^m} f(x)\|_2 \|\nabla_{\mathbb{R}^m} p(x)\|_2 \int_{B_{\mathbb{R}^d}(0,c_1)} K\left(\frac{\|v\|_2}{h}\right) \|v\|_2^2 \|\mathcal{E}_x(v) - x\|_2^2 \, \mathrm{d}v \\ &\quad \leq \frac{c_2^3}{h^{d+2}} \|\nabla_{\mathbb{R}^m} f(x)\|_2 \|\nabla_{\mathbb{R}^m} p(x)\|_2 \int_{B_{\mathbb{R}^d}(0,c_1)} K\left(\frac{\|v\|_2}{h}\right) \|v\|_2^4 \, \mathrm{d}v \\ &\quad \leq \frac{c_2^3 c_1}{h^{d+2}} \|\nabla_{\mathbb{R}^m} f(x)\|_2 \|\nabla_{\mathbb{R}^m} p(x)\|_2 \int_{B_{\mathbb{R}^d}(0,c_1)} K\left(\frac{\|v\|_2}{h}\right) \|v\|_2^3 \, \mathrm{d}v. \end{aligned}$$

As in the proof of Lemma 1, we deduce that the latter is bounded by $O(h)$.

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

In addition, again using property (iii) in Theorem 7, we have that, uniformly in x,

$$\left|\frac{1}{h^{d+2}}\int_{B_{\mathbb{R}^d}(0,c_1)} K\left(\frac{\|v\|_2}{h}\right)\langle\nabla_{\mathbb{R}^m}f(x),\,\mathcal{E}_x(v)-x\rangle\langle\nabla_{\mathbb{R}^m}p(x),\,\mathcal{E}_x(v)-x\rangle\,\mathrm{d}v - J_{11}\right|$$
$$= O(h), \tag{46}$$

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} with the first term of the generator $\mathcal{A}$. Using equation (41) of Lemma 2, with $G(||v||_2)=K\Big(\frac{||v||_2}{h}\Big)$, $k=\mathcal{E}_x$, and Proposition 4, we have

$$\begin{aligned}
J_{11} &= \frac{1}{h^{d+2}}\left(\frac{1}{d}\int_{B_{\mathbb{R}^d}(0,c_1)} K\left(\frac{\|v\|_2}{h}\right)\|v\|_2^2\,\mathrm{d}v\right)\langle\nabla_{\mathbb{R}^d}(f\circ\mathcal{E}_x)(0),\,\nabla_{\mathbb{R}^d}(p\circ\mathcal{E}_x)(0)\rangle\\
&= \left(\frac{1}{d}\int_{B_{\mathbb{R}^d}(0,c_1/h)} K\left(\|v\|_2\right)\|v\|_2^2\mathrm{d}v\right)\langle\nabla_{\mathcal{M}}f(x),\,\nabla_{\mathcal{M}}p(x)\rangle_{\mathbf{g}}\\
&= \left(\frac{1}{d}\int_{\mathbb{R}^d} K\left(\|v\|_2\right)\|v\|_2^2\mathrm{d}v\right)\langle\nabla_{\mathcal{M}}f(x),\,\nabla_{\mathcal{M}}p(x)\rangle_{\mathbf{g}} + o(h),
\end{aligned}$$

where the last estimation is uniform in $x\in\mathcal{M}$ and comes from the second estimation in (B2) in Lemma 4. Thus, we have proved Equation (44) for J_1.

The proof for J_2, given by (45), is similar to what we have done for J_1. To identify the Laplace–Beltrami operator in the last step of the proof, we use (42) of Lemma 2 and the second point of Proposition 4. Therefore, we have proved Proposition 1.

4.3. Proof of Proposition 2

Let us now prove Proposition 2. This proposition deals with the difference between the geodesic distance on $\mathcal{M}$ and the Euclidean norm of $\mathbb{R}^m$.

By inequalities (30) and (31) of Lemma 1, we know that uniformly in x, when h converges to 0,

$$\int_{\mathcal{M}}\left(K\left(\frac{\|x-y\|_2}{h}\right)+K\left(\frac{\rho(x-y)}{h}\right)\right)\mathbf{1}_{\rho(x,y)\geq c_1}\mu(\mathrm{d}y)=o(h^{d+3}).$$

Thus, by regularity of f, boundedness of p and compactness of $\mathcal{M}$, uniformly in x, when h converges to 0,

$$\int_{\mathcal{M}}\left(K\left(\frac{\|x-y\|_2}{h}\right)+K\left(\frac{\rho(x-y)}{h}\right)\right)|f(x)-f(y)|p(y)\mathbf{1}_{\rho(x,y)\geq c_1}\mu(\mathrm{d}y)=o(h^{d+3}).$$

Thus, we only have to prove that, uniformly in x,

$$\int_{\mathcal{M}}\left|K\left(\frac{\rho(x,y)}{h}\right)-K\left(\frac{\|x-y\|_2}{h}\right)\right|\,|f(x)-f(y)|p(y)\mathbf{1}_{\rho(x,y)<c_1}\mu(\mathrm{d}y)=O(h^{d+3}),$$

or equivalently, using the change of variables $v=\mathcal{E}_x^{-1}(y)$ and $\rho(x,y)=\|\mathcal{E}_x^{-1}(y)\|_2$ by (23),

$$\int_{B_{\mathbb{R}^d(0,c_1)}}\left|K\left(\frac{\|v\|_2}{h}\right)-K\left(\frac{\|\mathcal{E}_x(v)-x\|_2}{h}\right)\right|\,|f\circ\mathcal{E}_x(v)-f(x)|\,p\circ\mathcal{E}_x(v)\sqrt{\det\widehat{g}^x_{ij}(v)}\,\mathrm{d}v$$
$$=O(h^{d+3}).$$

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

In addition, 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 7, the function $v \mapsto \det \widehat{g}^x_{ij}(v)$ 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 7). By (B1) in Lemma 4, we have

$$\begin{aligned} I &\leqslant \int_{B_{\mathbb{R}^d(0,c_1)}} \left(\int_{\left(\frac{\|\mathcal{E}_x(v)-x\|_2}{h}, \frac{\|v\|_2}{h}\right]} \mathrm{d}H(a) \right) \|v\|_2 \mathrm{d}v \\ &= \int_{B_{\mathbb{R}^d(0,c_1)}} \left(\int_{\mathbb{R}_+} \mathbf{1}_{\|\mathcal{E}_x(v)-x\|_2 < ah \leq \|v\|_2} \mathrm{d}H(a) \right) \|v\|_2 \mathrm{d}v. \end{aligned}$$

Also by Theorem 5, there exists a constant c_3 such that, for all $x, y \in \mathcal{M}$,

$$\rho(x, y) \leqslant c_3 \|x-y\|_2^3 + \|x-y\|.$$

In addition, the polynomial function $z \mapsto z + c_3 z^3$ is an increasing bijective function and we denote by φ its inverse. Thus, for all $x, y \in \mathcal{M}$, $\varphi(\rho(x, y)) \leq \|x-y\|_2$. Then, using that $\varphi(\|v\|_2) = \varphi(\rho(x, \mathcal{E}_x(v)))$, we deduce that $\varphi(\|v\|_2) \leq \|\mathcal{E}_x(v)-x\|_2$. Therefore,

$$\begin{aligned} I &\leq \int_{B_{\mathbb{R}^d(0,c_1)}} \left(\int_{\mathbb{R}_+} \mathbf{1}_{\varphi(\|v\|_2) < ah \leq \|v\|_2} \mathrm{d}H(a) \right) \|v\|_2 \mathrm{d}v \\ &= \int_{\mathbb{R}^+} \left(\int_{B_{\mathbb{R}^d(0,c_1)}} \|v\|_2 . \mathbf{1}_{ah \leq \|v\|_2 < ah + c_3(ah)^3} \mathrm{d}v \right) \mathrm{d}H(a), \end{aligned}$$

by Fubini's theorem. Finally, using the spherical coordinate system as in the proof of Lemma 1, we see that

$$\begin{aligned} I &\leqslant S_{d-1} \int_{\mathbb{R}^+} \left(\int_0^{c_1} r^d \mathbf{1}_{ah \leq r < ah + c_3(ah)^3} \mathrm{d}r \right) \mathrm{d}H(a) \\ &\leq S_{d-1} \int_{\mathbb{R}^+} \left(\mathbf{1}_{ah \leq c_1} \times \int_{ah}^{ah + c_3(ah)^3} r^d \, \mathrm{d}r \right) \mathrm{d}H(a) \\ &\leq S_{d-1} \int_{\mathbb{R}^+} \left(\mathbf{1}_{ah \leq c_1} \times c_3 (ah)^3 \left[ah + c_3 (ah)^3 \right]^d \right) \mathrm{d}H(a) \\ &\leq S_{d-1} \int_{\mathbb{R}_+} c_3 (ah)^{d+3} \left(1 + c_3 c_1^2 \right)^d \mathrm{d}H(a) \\ &= S_{d-1} c_3 \left(1 + c_3 c_1^2 \right)^d h^{d+3} \int_{\mathbb{R}^+} a^{d+3} \, \mathrm{d}H(a). \end{aligned}$$

This concludes the proof of Proposition 2.

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

5. Approximations by random operators

In this section, we study the statistical error and prove Proposition 3.

Notation 3. *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 multilinear maps, that is,*

$$\|f''\|_\infty = \sup_{\substack{x\in\mathcal{M},(v,w)\in(\mathbb{R}^m)^2\\ \|v\|_2\le 1,\|w\|_2\le 1}} \left|f''(x)(v, w)\right|.$$

Recall that for $\alpha \in [\![1, m]\!]$ and $x \in \mathbb{R}^m$, we denote by x^α the αth coordinate of x.

Let us consider the following collection $\mathcal{F}$ of $\mathcal{C}^3$-functions:

$$\mathcal{F} := \{f \in \mathcal{C}^3(\mathcal{M}) : \|f\|_\infty \le 1,\ \|f'\|_\infty \le 1,\ \|f''\|_\infty \le 1,\ \|f'''\|_\infty \le 1\}. \tag{47}$$

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 $(Z_n, n \in \mathbb{N})$:

$$\begin{aligned}
Z_n &:= \sup_{f\in\mathcal{F}} \sup_{x\in\mathcal{M}} \left|\mathcal{A}_{h_n,n}(f)(x) - \mathrm{E}[\mathcal{A}_{h_n,n}(f)(x)]\right| \\
&= \frac{1}{nh_n^{d+2}} \sup_{f\in\mathcal{F}} \sup_{x\in\mathcal{M}} \left|\sum_{i=1}^n \left(K\left(\frac{\|X_i - x\|_2}{h_n}\right)(f(X_i) - f(x))\right.\right. \\
&\qquad\qquad \left.\left. - \mathrm{E}\left[K\left(\frac{\|X - x\|_2}{h_n}\right)(f(X) - f(x))\right]\right)\right|.
\end{aligned}$$

Recall that for all functions f and points x, $\mathrm{E}[\mathcal{A}_{h_n,n}(f)(x)] = \mathcal{A}_{h_n}(f)(x)$. We want to prove that with probability 1,

$$Z_n = O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}} + h_n\right). \tag{48}$$

The general idea in proving this estimation is that instead of directly proving this convergence speed for (Z_n), we show that its expectation also has this speed of convergence, that is,

$$\limsup_{n\to\infty} \left[\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}} + h_n\right)^{-1} \mathrm{E}(Z_n)\right] < \infty. \tag{49}$$

Then (48) will follow easily from Talagrand's inequality (see Corollary 1 in the Appendix) and the Borel–Cantelli lemma, as explained in Section 5.4. The detailed plan for the proof of (48) is as follows.

I. Use Taylor's expansion to divide each Z_n into many simpler terms.

II. Use Vapnik–Chervonenkis theory and Theorem 8 to bound the expectation of each term.

III. Use Talagrand's inequality to conclude.

Having used Talagrand's inequality, we will have a non-asymptotic estimation of

$$\mathrm{P}\left(\sup_{f\in\mathcal{F}} \sup_{x\in\mathcal{M}} \left|\mathcal{A}_{h_n,n}(f)(x) - \mathrm{E}[\mathcal{A}_{h_n,n}(f)(x)]\right| \ge \delta\right)$$

for some suitable constant δ and will be able to prove the Corollary 1 at the end of this section.

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

5.1. On Vapnik–Chervonenkis theory

Before starting the proof, we recall the main definitions and an important result from Vapnik–Chervonenkis theory for the Borelian space $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ we will need. Other useful results are given in Appendix A. For more details on the Vapnik–Chervonenkis theory, we refer the reader to [11, 16, 29]. In this section, we will recall upper bounds that exist for

$$\sup_{f\in\mathcal{F}} \mathrm{E}\left[\left|\sum_{i=1}^{n} (f(X_i) - \mathrm{E}[f(X_i)])\right|\right]$$

when the functions f range over certain Vapnik–Chervonenkis 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

$$\begin{aligned} N(\varepsilon, T, d) := \inf\{&N \in \mathbb{N} \cup \{+\infty\} : \text{there are } N \text{ points in } T \\ &\text{such that they form an } \varepsilon\text{-cover of } (T, d)\}. \end{aligned}$$

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 an *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 Vapnik–Chervonenkis classes of functions (see Definition 3.6.10 in [16]). 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 1 *(Vapnik-Chervonenkis class of functions).* A class of measurable functions $\mathcal{F}$ is of Vapnik–Chervonenkis 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 write

$$N(\varepsilon, \mathcal{F}) := \sup_Q N(\varepsilon, \mathcal{F}, L^2(Q)).$$

We now present a version of the useful inequality (2.5) of Giné and Guillou in [14] that gives a bound for the expected concentration rate. For a class of function $\mathcal{F}$, let us define for any real valued function $\varphi : \mathcal{F} \to \mathbb{R}$,

$$\|\varphi(f)\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} |\varphi(f)|.$$

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

Theorem 8 (See [14, 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 Vapnik–Chervonenkis-type class of functions on $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$. We introduce two positive real numbers, σ^2 and U, such that*

$$\sigma^2 \geq \sup_{f\in\mathcal{F}} \mathrm{Var}(f(X_1)), \quad U \geq \sup_{f\in\mathcal{F}} \|f\|_\infty, \quad \text{and} \quad 0 < \sigma \leq 2U.$$

Then there exists a constant R, depending only on the Vapnik–Chervonenkis parameters A, v of $\mathcal{F}$ and on U, such that

$$\mathrm{E}\left[\left\|\sum_{i=1}^n (f(X_i) - \mathrm{E}[f(X_i)])\right\|_{\mathcal{F}}\right] \leq R\left(\sqrt{n}\sigma\sqrt{|\log\sigma|} + |\log\sigma|\right).$$

Note that there also exists a formulation of the previous result in terms of deviation probability (see, for example, [28, Theorem 3]) that would lead to results similar to the ones established in [7].

5.2. Step I: Decomposition of Z_n

We first upper-bound the quantity Z_n with a sum of simpler terms.

Lemma 3. *There is a constant $c > 0$ such that, for all $n \geqslant 1$,*

$$nh_n^{d+2} Z_n \leq \sum_{\alpha=1}^m Y_n^\alpha + \sum_{\alpha,\beta=1}^m Y_n^{\alpha,\beta} + Y_n^{(3)} + 2nch_n^{d+3}, \tag{50}$$

where

$$Y_n^\alpha := \sup_{x\in\mathcal{M}} \left|\sum_{i=1}^n \left[K\left(\frac{\|X_i - x\|_2}{h_n}\right)(X_i^\alpha - x_i^\alpha) - \mathrm{E}\left(K\left(\frac{\|X - x\|_2}{h_n}\right)(X^\alpha - x^\alpha)\right)\right]\right|,$$

$$Y_n^{\alpha,\beta} := \sup_{x\in\mathcal{M}} \left|\sum_{i=1}^n \left[K\left(\frac{\|X_i - x\|_2}{h_n}\right)(X_i^\alpha - x^\alpha)(X_i^\beta - x^\beta) - \mathrm{E}\left(K\left(\frac{\|X - x\|_2}{h_n}\right)(X^\alpha - x^\alpha)(X^\beta - x^\beta)\right)\right]\right|,$$

$$Y_n^{(3)} := \sup_{x\in\mathcal{M}} \left|\sum_{i=1}^n K\left(\frac{\|X_i - x\|_2}{h_n}\right)\|X_i - x\|_2^3 - \mathrm{E}\left[K\left(\frac{\|X - x\|_2}{h_n}\right)\|X - x\|_2^3\right]\right|.$$

Proof. Since, for any $f \in \mathcal{F}$, the differentials up to third order have operator norms bounded by 1, we have, by the Taylor expansion theorem, for any $(x, y) \in (\mathbb{R}^m)^2$,

$$f(y) - f(x) = f'(x)(y - x) + \frac{1}{2}f''(x)(y - x, y - x) + \tau_f(y;x),$$

where τ_f is some function satisfying

$$\sup_{f\in\mathcal{F}} |\tau_f(y, x)| \leq \|f'''\|_\infty \|x - y\|_2^3 = \|x - y\|_2^3. \tag{51}$$

Thus, using the notation of the lemma, we deduce

$$nh_n^{d+2} Z_n \leq \sum_{\alpha=1}^m Y_n^\alpha + \sum_{\alpha=1,\beta=1}^m Y_n^{\alpha,\beta} + Y_n^r,$$

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

with

$$Y_n^r := \sup_{\substack{f\in\mathcal{F}\\ x\in\mathcal{M}}} \left| \sum_{i=1}^{n} \left(K\left(\frac{\|X_i - x\|_2}{h_n}\right) \tau_f(X_i, x) - \mathrm{E}\left[K\left(\frac{\|X - x\|_2}{h_n}\right) \tau_f(X, x)\right]\right)\right|.$$

Using (51), we now control Y_n^r by $Y_n^{(3)}$ as follows:

$$\begin{aligned} Y_n^r &\le \sup_{x\in\mathcal{M}} \left| \sum_{i=1}^{n} K\left(\frac{\|X_i - x\|_2}{h_n}\right) \|X_i - x\|_2^3 \right| + n \sup_{x\in\mathcal{M}} \mathrm{E}\left[K\left(\frac{\|X - x\|_2}{h_n}\right) \|X - x\|_2^3\right] \\ &\le Y_n^{(3)} + 2n \sup_{x\in\mathcal{M}} \mathrm{E}\left[K\left(\frac{\|X - x\|_2}{h_n}\right) \|X - x\|_2^3\right]. \end{aligned}$$

Since the function p is bounded on the compact $\mathcal{M}$, using (34) of Lemma 1, we deduce that $Y_n^r \leqslant Y_n^{(3)} + 2nch_n^{d+3}$, which concludes the proof. □

5.3. Step II: Application of the Vapnik–Chervonenkis theory

5.3.1. *Control the first-order term* $\mathrm{E}[Y_n^\alpha]$ 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 $\mathcal{M}$:

$$\begin{aligned} \mathcal{G} &:= \{\varphi_{h,y,z} : y, z \in \mathcal{M}, h > 0\}, \\ \mathcal{G}_1 &:= \{\psi_{h,y} : y \in \mathcal{M}, h > 0\} \\ \mathcal{G}_2 &:= \{\zeta_y(x) : y \in \mathcal{M}\}, \end{aligned}$$

with

$$\begin{aligned} \varphi_{h,y,z} : x &\longmapsto K\left(\tfrac{\|x-y\|_2}{h}\right)(x^\alpha - z^\alpha), \\ \psi_{h,y} : x &\longmapsto K\left(\tfrac{\|x-y\|_2}{h}\right), \\ \zeta_y : x &\longmapsto x^\alpha - y^\alpha. \end{aligned}$$

Since K is of bounded variation, by [29, Lemma 22], $\mathcal{G}_1$ is of Vapnik–Chervonenkis type with respect to a constant envelope. Since $\mathcal{M}$ is a compact manifold, by Lemma 3, $\mathcal{G}_2$ is also of Vapnik–Chervonenkis type with respect to a constant envelope. Thus, using Lemma 2, we deduce that $\mathcal{G}$ is a Vapnik–Chervonenkis-type class of functions because $\mathcal{G} = \mathcal{G}_1 \cdot \mathcal{G}_2$. Thus, by Definition 1, there exist real values $A \geq 6$, $v \geq 1$ depending only on the Vapnik–Chervonenkis characteristics of $\mathcal{G}_1$ and $\mathcal{G}_2$ such that, for all $\varepsilon \in (0, 1)$,

$$N(\varepsilon, \mathcal{G}) \le \left(\frac{A}{2\varepsilon}\right)^v.$$

Now let us consider the following sequence of families of real functions on $\mathcal{M}$:

$$\mathcal{H}_n = \left\{\varphi_{n,y} : y \in \mathcal{M}\right\}, \quad \text{with } \varphi_{n,y} : x \longmapsto K\left(\frac{\|x - y\|_2}{h_n}\right)(x^\alpha - y^\alpha).$$

Proposition 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* $\mathcal{M}$.

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

Then, if the kernel K satisfies Assumption 1 and the sequence $(h_n)_{n\geqslant 0}$ *satisfies* (5)*, then we have*

$$\frac{1}{nh_n^{d+2}}\mathrm{E}\left[\left\|\sum_{i=1}^{n}(f(X_i)-\mathrm{E}[f(X)])\right\|_{\mathcal{H}_n}\right]=O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}}\right). \tag{52}$$

Proof. Since $\mathcal{H}_n\subset\mathcal{G}$, by Lemma 1, for all n, we have $N(\varepsilon,\mathcal{H}_n)\leq\left(\frac{A}{\varepsilon}\right)^{v}$. Hence, by Theorem 8, there exists a constant R, depending only on A, v, and U, such that

$$\mathrm{E}\left[\left\|\sum_{i=1}^{n}f(X_i)-\mathrm{E}(f(X))\right\|_{\mathcal{H}_n}\right]\leq R\left(\sqrt{n}\sigma\sqrt{|\log\sigma|}+|\log\sigma|\right),$$

where U is a constant such that $U\geq\sup_{f\in\mathcal{H}_n}\|f\|_\infty$, and σ is a constant such that $4U^2\geq\sigma^2\geq\sup_{f\in\mathcal{H}_n}\mathrm{E}[f^2(X)]$.

Since $\mathcal{H}_n\subset\mathcal{G}$, we can choose U to be the constant envelope of $\mathcal{G}$ (thus, independent of n). Furthermore, we see that

$$\sup_{f\in\mathcal{H}_n}\mathrm{E}\big[f^2(X)\big]\leq\|K\|_\infty\sup_{x\in\mathcal{M}}\int_{\mathcal{M}}K\left(\frac{\|x-y\|}{h_n}\right)(x^\alpha-y^\alpha)^2p(y)\mu(\mathrm{d}y).$$

By (33) of Lemma 1, we deduce that there is $c>0$ such that

$$\sup_{f\in\mathcal{H}_n}\mathrm{E}\big[f^2(X)\big]\leq\|K\|_\infty\|p\|_\infty\mu(\mathcal{M})ch_n^{d+2}, \tag{53}$$

which goes to 0 when $n\to+\infty$. Choose $\sigma^2:=\sigma_n^2=\|K\|_\infty\|p\|_\infty\mu(\mathcal{M})ch_n^{d+2}$. For n large enough, $\sigma_n\leq 2U$. Hence, using assumption (5) on the sequence $(h_n)_{n\geqslant 1}$, we deduce

$$\begin{aligned}\frac{1}{nh_n^{d+2}}\mathrm{E}\left[\left\|\sum_{i=1}^{n}f(X_i)-\mathrm{E}[f(X)]\right\|_{\mathcal{H}_n}\right]&=O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}}+\frac{\log h_n^{-1}}{nh_n^{d+2}}\right)\\&=O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}}\right).\end{aligned}$$

This concludes the proof. □

The conclusion of the above proposition means that

$$\frac{1}{nh_n^{d+2}}\mathrm{E}\big[Y_n^\alpha\big]=O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}}\right).$$

5.3.2. *Control the second-order term* $\mathrm{E}[Y_n^{\alpha,\beta}]$ 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 Vapnik–Chervonenkis-type family of functions:

$$\mathcal{I}_n:=\left\{\xi_{n,y,z}:x\mapsto K\left(\frac{\|x-y\|}{h_n}\right)(x^\alpha-y^\alpha)(x^\beta-q^\beta):y\in\mathcal{M},z\in\mathcal{M},\right\}.$$

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

We note that, for any random variable X,

$$\mathrm{E}\Big[\sup_{g\in\mathcal{I}_n}\big|g^2(X)\big|\Big]\leqslant \mathrm{diam}(\mathcal{M})^2\mathrm{E}\Big[\sup_{f\in\mathcal{H}_n}\big|f^2(X)\big|\Big].$$

Using (53), we deduce that

$$\sup_{g\in\mathcal{I}_n}\mathrm{E}[g^2(X)]=O\big(h_n^{d+2}\big)$$

and

$$\frac{1}{nh_n^{d+2}}\mathrm{E}\sup_{g\in\mathcal{I}_n}\left|\sum_{i=1}^n(g(X_i)-\mathrm{E}[g(X_i)])\right|=O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}}\right).$$

Therefore, we conclude that

$$\frac{1}{nh_n^{d+2}}\mathrm{E}\big[Y_n^{\alpha,\beta}\big]=O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}}\right).$$

5.3.3. *Control the third-order term* $\mathrm{E}[Y_n^{(3)}]$ This step is essentially the same as the two previous steps, except that the family of functions considered is a little bit different:

$$\mathcal{K}_n:=\left\{x\mapsto K\left(\frac{\|x-y\|}{h_n}\right)\|x-y\|^3:y\in\mathcal{M}\right\}.$$

With the same arguments as before, we obtain

$$\frac{1}{nh_n^{d+2}}\mathrm{E}\big[Y_n^{(3)}\big]=O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}}\right).$$

Thanks to Step I, Step II,and Lemma 3, we have shown that

$$\mathrm{E}[Z_n]=O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}}+h_n\right). \tag{54}$$

5.4. Step III: Conclusion

Recall that the set of functions $\mathcal{F}$ is defined by (47). Since p is bounded on $\mathcal{M}$, by (35) of Lemma 1, there exists $c>0$ such that for all $f\in\mathcal{F}$, for all $x\in\mathcal{M}$,

$$\begin{aligned}\mathrm{E}&\left[K\left(\frac{\|X-x\|}{h_n}\right)^2(f(X)-f(x))^2\right]\\&\leq\|K\|_\infty\,\mathrm{E}\left[K\left(\frac{\|X-x\|}{h_n}\right)\|X-x\|^2\right]\leq\|K\|_\infty ch_n^{d+2}.\end{aligned}$$

In other words,

$$\sup_{f\in\mathcal{F}}\sup_{x\in\mathcal{M}}\mathrm{E}\left[K\left(\frac{\|X-x\|}{h_n}\right)^2(f(X)-f(x))^2\right]\leq\|K\|_\infty ch_n^{d+2}.$$

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

Thus, by choosing $\sigma := \sigma_n = \sqrt{\|K\|_\infty c h_n^{d+2}}$, and using Massart's version of the Talagrand inequality (cf. Corollary 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 $1 - \mathrm{e}^{-t_n}$,

$$\sup_{f\in\mathcal{F}} \sup_{y\in\mathcal{M}} nh_n^{d+2} |\mathcal{A}_{h_n,n}(f)(x) - \mathrm{E}[\mathcal{A}_{h_n,n}(f)(x)]| \le 9\big(nh_n^{d+2}\mathrm{E}[Z_n] + \sigma_n\sqrt{nt_n} + bt_n\big), \tag{55}$$

where, in this case, the constant envelope b is equal to

$$b := \|K\|_\infty \mathrm{diam}\mathcal{M}.$$

Choose $t_n = 2\log n$. By the Borel–Cantelli lemma, with probability 1,

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

Furthermore, under assumption (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,

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

This concludes the proof of Proposition 3. Hence, Theorem 1 is proved.

5.5. Proof of Corollary 1

Using the results of the above sections, we can now prove Corollary 1. First, we see that by the proofs of Propositions Proof of Proposition 1, Proof of Proposition 2, and (54), there is a constant $C > 0$ such that for all $h > 0$, $n \in \mathbb{N}$,

$$\sup_{f\in\mathcal{F}} \sup_{x\in\mathcal{M}} \left|\mathrm{E}[\mathcal{A}_{h,n}(f)(x)] - \mathcal{A}(f)(x)\right| \le Ch \tag{56}$$

and

$$\mathrm{E}[Z_n] \le C\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}} + h_n\right). \tag{57}$$

Then, by choosing $t_n := \delta^2 nh_n^{d+2}$ in (55) with

$$\delta \in \left[h_n \vee \sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}}, 1\right],$$

we know that, with probability at least $1 - \mathrm{e}^{-\delta^2 nh_n^{d+2}}$,

$$\sup_{f\in\mathcal{F}} \sup_{y\in\mathcal{M}} \left|\mathcal{A}_{h_n,n}(f)(x) - \mathrm{E}\big[\mathcal{A}_{h_n,n}(f)(x)\big]\right| \le \frac{9\left(nh_n^{d+2}\mathrm{E}[Z_n] + \sigma_n\sqrt{nt_n} + bt_n\right)}{nh_n^{d+2}}.$$

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

In addition, by (57), we have

$$\begin{aligned}
\frac{nh_n^{d+2}\mathrm{E}[Z_n]+\sigma_n\sqrt{nt_n}+bt_n}{nh_n^{d+2}} &\le C\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}}+h_n\right)+\frac{\sigma_n\sqrt{nt_n}+bt_n}{nh_n^{d+2}}\\
&= C\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}}+h_n\right)+\sqrt{\|K\|_\infty c}\delta+\|K\|_\infty(\dim\mathcal{M})\delta^2\\
&\le \big(2C+\sqrt{\|K\|_\infty c}+\|K\|_\infty(\dim\mathcal{M})\big)\delta.
\end{aligned}$$

In addition, from (56), we have

$$\begin{aligned}
\sup_{f\in\mathcal{F}}\sup_{x\in\mathcal{M}}\big|\mathcal{A}_{h_n,n}(f)(x)-\mathcal{A}(f)(x)\big| &\le \sup_{f\in\mathcal{F}}\sup_{x\in\mathcal{M}}\big|\mathcal{A}_{h_n,n}(f)(x)-\mathrm{E}[\mathcal{A}_{h_n,n}(f)(x)]\big|\\
&\quad+\sup_{f\in\mathcal{F}}\sup_{x\in\mathcal{M}}\big|\mathrm{E}[\mathcal{A}_{h_n,n}(f)(x)]-\mathcal{A}(f)(x)\big|\\
&\le \sup_{f\in\mathcal{F}}\sup_{x\in\mathcal{M}}\big|\mathcal{A}_{h_n,n}(f)(x)-\mathrm{E}[\mathcal{A}_{h_n,n}(f)(x)]\big|+Ch_n\\
&\le \sup_{f\in\mathcal{F}}\sup_{x\in\mathcal{M}}\big|\mathcal{A}_{h_n,n}(f)(x)-\mathrm{E}[\mathcal{A}_{h_n,n}(f)(x)]\big|+C\delta.
\end{aligned}$$

Therefore, by letting

$$C' := 9[2C+\sqrt{\|K\|_\infty c}+\|K\|_\infty(\dim\mathcal{M})]+C, \tag{58}$$

where C is the constant appearing in (56) and (57), we have

$$\begin{aligned}
&\mathbf{P}\Big(\sup_{f\in\mathcal{F}}\sup_{x\in\mathcal{M}}\big|\mathcal{A}_{h_n,n}(f)(x)-\mathcal{A}(f)(x)\big|>C'\delta\Big)\\
&\le \mathbf{P}\Big(\sup_{f\in\mathcal{F}}\sup_{x\in\mathcal{M}}\big|\mathcal{A}_{h_n,n}(f)(x)-\mathrm{E}[\mathcal{A}_{h_n,n}(f)(x)]\big|>C'\delta-C\delta\Big)\\
&= \mathbf{P}\Big(\sup_{f\in\mathcal{F}}\sup_{x\in\mathcal{M}}\big|\mathcal{A}_{h_n,n}(f)(x)-\mathrm{E}[\mathcal{A}_{h_n,n}(f)(x)]\big|>9\left(2C+\sqrt{\|K\|_\infty c}+\|K\|_\infty(\dim\mathcal{M})\right)\delta\Big)\\
&\le \exp\big(-\delta^2 nh_n^{d+2}\big).
\end{aligned}$$

This proves Corollary 1.

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 $n\in\mathbb{N}$, $k\in\{1,\dots,n\}$, and $x\in\mathcal{M}$, the distance between x and its k-nearest neighbor is defined in (8) and that the Laplacian of the kNN graph is given by, for $x\in\mathcal{M}$,

$$\mathcal{A}_n^{k\mathrm{NN}}(f)(x) := \frac{1}{nR_{n,k_n}^{d+2}(x)}\sum_{i=1}^n \mathbf{1}_{[0,1]}\left(\frac{\|X_i-x\|_2}{R_{n,k_n}(x)}\right)(f(X_i)-f(x)). \tag{59}$$

Note here that the width of the moving window, $R_{n,k_n}(x)$, is random and depends on $x\in\mathcal{M}$, in contrast to h_n in the previous generator $\mathcal{A}_{h_n,n}$ defined by (1).

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

To overcome this difficulty, we use the result of Cheng and Wu [9, Theorem 2.3], with $h = \mathbf{1}_{[0,1]}$, which allows us to control the randomness and locality of the window.

Theorem 9 (Cheng–Wu [9, Theorem 2.3]). *Under Assumption 1, if the density p satisfies (10) and if*

$$\lim_{n\to+\infty} \frac{k_n}{n} = 0 \quad \text{and} \quad \lim_{n\to+\infty} \frac{k_n}{\log(n)} = +\infty,$$

then, with probability greater than $1 - n^{-10}$,

$$\sup_{x\in\mathcal{M}} \left| \frac{R_{n,k_n}(x)}{V_d^{1/d} p^{-1/d}(x) \left(\frac{k_n}{n}\right)^{1/d}} - 1 \right| = O\left(\left(\frac{k_n}{n}\right)^{2/d} + \frac{3\sqrt{13}}{d} \sqrt{\frac{\log n}{k_n}} \right), \tag{60}$$

where V_d is the volume of the unit d-ball.

As a corollary to Theorem 9, we deduce that the distance $R_{n,k_n}(x)$ is, uniformly in x and with large probability, of the order of h_n:

$$\mathrm{P}\big(\forall x \in \mathcal{M},\ R_{n,k_n}(x) \in [h_n(x) - \gamma_n, h_n(x) + \gamma_n]\big) \geqslant 1 - n^{-10}, \tag{61}$$

with

$$h_n(x) = V_d^{1/d} p^{-1/d}(x) \left(\frac{k_n}{n}\right)^{1/d} \quad \text{and} \quad \gamma_n = 2\left(\left(\frac{k_n}{n}\right)^{2/d} + \frac{3\sqrt{13}}{d} \sqrt{\frac{\log n}{k_n}} \right). \tag{62}$$

We will then derive Theorem 2 for the rescaling of the kNN Laplacian using the following result.

Theorem 10. *Suppose that the density of points p on the compact smooth manifold $\mathcal{M}$ is of class $\mathcal{C}^2$. Suppose that Assumption 1 on the kernel K is satisfied and that $(h_n, n \in \mathbb{N})$ satisfies (5), that is,*

$$\lim_{n\to+\infty} h_n = 0 \quad \text{and} \quad \lim_{n\to+\infty} \frac{\log h_n^{-1}}{n h_n^{d+2}} = 0. \tag{63}$$

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

$$\sup_{\kappa^{-1} h_n \le r \le \kappa h_n} \sup_{x\in\mathcal{M}} \left| \mathcal{A}_{r,n}(f)(x) - \mathcal{A}(f)(x) \right| = O\left(\sqrt{\frac{\log h_n^{-1}}{n h_n^{d+2}}} + h_n \right), \tag{64}$$

where $\mathcal{A}_{r,n}$ and $\mathcal{A}$ are defined by (1) (replacing h_n with r) and (2), respectively.

Proof of Theorem 2. Assume that Theorem 10 is proved, and consider k_n as in the statement of Theorem 2. For h_n and γ_n defined in (62), we know by (61) 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 the Borel–Cantelli lemma, 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

$$\left| \mathcal{A}_n^{k\mathrm{NN}}(f)(x) - \mathcal{A}(f)(x) \right| \leqslant \sup_{r\in[a_n,b_n]} \left| \mathcal{A}_{r,n}(f)(x) - \mathcal{A}(f)(x) \right|$$

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

with

$$a_n = V_d^{1/d} p_{max}^{-1/d} \left(\frac{k_n}{n}\right)^{1/d} - \gamma_n \quad \text{and} \quad b_n = V_d^{1/d} p_{min}^{-1/d} \left(\frac{k_n}{n}\right)^{1/d} + \gamma_n. \tag{65}$$

Note that for n large enough, a_n will be positive. Choosing $h'_n = b_n$ 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 from Theorem 10. In Theorem 2, the choice of number of neighbors k_n in (11) comes from (63) with our choice of h'_n. The rate of convergence given in (12) results from (64). □

Proof of Theorem 10. The proof for the above theorem is essentially the same as the proof we presented for Theorem 1, except for some necessary modifications. Decomposing the error term as in (17), we have to deal with similar terms. The approximations involving the geometry and corresponding to Propositions Proof of Proposition 1 and Proof of Proposition 2 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 (47). Following the computations in Section 5, we introduce the following sequence of random variables $(\tilde{Z}_n, n \in \mathbb{N})$, where here $K = \mathbf{1}_{[0,1]}$:

$$\begin{aligned}
\tilde{Z}_n &:= \sup_{f \in \mathcal{F}} \sup_{\kappa^{-1} h_n \le r \le \kappa h_n} \sup_{x \in \mathcal{M}} \left| \mathcal{A}_{r,n}(f)(x) - \mathrm{E}[\mathcal{A}_{r,n}(f)(x)] \right| \\
&= \frac{1}{n h_n^{d+2}} \sup_{f \in \mathcal{F}} \sup_{\kappa^{-1} h_n \le r \le \kappa h_n} \sup_{x \in \mathcal{M}} \left| \sum_{i=1}^{n} \left(K\left(\frac{\|X_i - x\|_2}{r}\right) (f(X_i) - f(x)) \right.\right. \\
&\qquad\qquad \left.\left. - \mathrm{E}\left[K\left(\frac{\|X - x\|_2}{r}\right) (f(X) - f(x)) \right] \right) \right|.
\end{aligned}$$

Similar to what we did in Section 5.2, we can show that there is a constant c independent of n such that

$$n h_n^{d+2} \tilde{Z}_n \le \sum_{\alpha=1}^{m} \tilde{Y}_n^{\alpha} + \sum_{\alpha,\beta=1}^{m} \tilde{Y}_n^{\alpha,\beta} + \tilde{Y}_n^{(3)} + 2nch_n^{d+3}, \tag{66}$$

where

$$\begin{aligned}
\tilde{Y}_n^{\alpha} &:= \sup_{\kappa^{-1} h_n \le r \le \kappa h_n} \sup_{x \in \mathcal{M}} \left| \sum_{i=1}^{n} \left[K\left(\frac{\|X_i - x\|_2}{r}\right) (X_i^{\alpha} - x_i^{\alpha}) \mathrm{E}\left(K\left(\frac{\|X - x\|_2}{r}\right) (X^{\alpha} - x^{\alpha}) \right) \right] \right| \\
\tilde{Y}_n^{\alpha,\beta} &:= \sup_{\kappa^{-1} h_n \le r \le \kappa h_n} \sup_{x \in \mathcal{M}} \left| \sum_{i=1}^{n} \left[K\left(\frac{\|X_i - x\|_2}{r}\right) (X_i^{\alpha} - x^{\alpha})(X_i^{\beta} - x^{\beta}) \right.\right. \\
&\qquad\qquad \left.\left. - \mathrm{E}\left(K\left(\frac{\|X - x\|_2}{r}\right) (X^{\alpha} - x^{\alpha})(X^{\beta} - x^{\beta}) \right) \right] \right|, \\
\tilde{Y}_n^{(3)} &:= \sup_{\kappa^{-1} h_n \le r \le \kappa h_n} \sup_{x \in \mathcal{M}} \left| \sum_{i=1}^{n} K\left(\frac{\|X_i - x\|_2}{r}\right) \|X_i - x\|_2^3 - \mathrm{E}\left[K\left(\frac{\|X - x\|_2}{r}\right) \|X - x\|_2^3 \right] \right|.
\end{aligned}$$

We now treat these terms by applying the Vapnik–Chervonenkis theory. Let us start with control of the first-order terms $\mathrm{E}[\tilde{Y}_n^{\alpha}]$.

In Section 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 Vapnik–Chervonenkis class of functions, and that there exist real values $A \geq 6$, $v \geq 1$ such that, for all $\varepsilon \in (0, 1)$, $N(\varepsilon, \mathcal{G}) \leq \left(A/2\varepsilon\right)^v$.

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 Vapnik–Chervonenkis class for which we can use the Talagrand inequality (8). The latter can deal with the additional supremum with respect to the window width. Similarly to what we did in the proof of Proposition 5, we obtain that:

$$\frac{1}{nh_n^{d+2}} \mathrm{E}\left[\left\| \sum_{i=1}^n (f(X_i) - \mathrm{E}[f(X)]) \right\|_{\tilde{\mathcal{H}}_n}\right] = O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}}\right),$$

which means that as $n \to \infty$,

$$\frac{1}{nh_n^{d+2}} \mathrm{E}[\tilde{Y}_n^\alpha] = O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}}\right).$$

Control of the second- and third-order terms is done as in Sections 5.3.1 and 5.3.2, 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 gives

$$\frac{1}{nh_n^{d+2}} \mathrm{E}[\tilde{Y}_n^{\alpha,\beta}] = O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}}\right) \quad \text{and} \quad \frac{1}{nh_n^{d+2}} \mathrm{E}[\tilde{Y}_n^{(3)}] = O\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}}\right). \tag{67}$$

Therefore, we can deduce the conclusion by using the argument presented in Section 5.4. □

7. Tightness and convergence of random walks

7.1. Proof of Theorem 3

Let K be a kernel that satisfies Assumption 1. 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 $\{X_1, \ldots, X_n\}$, which solves

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

the SDE

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

with initial condition $X_0^{(n)}$ and where $Q(\mathrm{d}s, \mathrm{d}i, \mathrm{d}\theta)$ is a Poisson point measure on $\mathbb{R}_+ \times \mathbb{N} \times \mathbb{R}_+$ 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 $\mathbb{N}$.

Remark 2. The initial distribution of $X_0^{(n)}$ can have support on the sample points $\{X_1, \ldots, X_n\}$, but not necessarily. It can be any distribution on the manifold $\mathcal{M}$. In any case, the random walk reaches the sample points $\{X_1, \ldots, X_n\}$, and stays in this set, after the first jump.

Proposition 6. *For a fixed* $n \geqslant 1$, *a random variable* $X_0^{(n)}$ *and a Poisson point measure* $Q(\mathrm{d}s, \mathrm{d}i, \mathrm{d}\theta)$*, there exists a unique strong solution of the SDE* (13).

For any real-valued measurable bounded function f on $\mathcal{M}$, *we have that*

$$M_t^{n,f} = f\left(X_t^{(n)}\right) - f\left(X_0^{(n)}\right) - \int_0^t \mathcal{A}_{h_n,n}(f)\left(X_s^{(n)}\right) \mathrm{d}s \tag{68}$$

is a square-integrable martingale with predictable quadratic variation:

$$\langle M^{n,f} \rangle_t = \int_0^t \frac{1}{nh_n^{d+2}} \sum_{i=1}^n K\left(\frac{\|X_i - X_s^{(n)}\|_2}{h_n}\right) \left(f(X_i) - f\left(X_s^{(n)}\right)\right)^2 \mathrm{d}s. \tag{69}$$

Proof. The proof is straightforward as the process takes its values in the compact manifold $\mathcal{M}$. The jump rate therefore remains 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 [20, Theorem 5.1, p. 66]). □

Let $T > 0$ be a positive fixed time. Note that when we have Feller processes, the convergence of the generators implies the convergence of the stochastic processes (see, for example, [23, Theorem 17.25]). However, without a continuity assumption on the kernel K, the Feller property may not be satisfied for our random walks and we follow a more classical line of proof. The proof of Theorem 3 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., [5]), the sequence is therefore sequentially relatively compact. The convergence of the generators $\mathcal{A}_{h_n,n}$ to $\mathcal{A}$ defined in (2) will yield that any limiting value is a solution of the martingale problem associated with $\mathcal{A}$, which is well posed.

Lemma 4. *Under the hypotheses of Theorem 3, 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's criteria for tightness (see, for example, [1, 22]). Because $\mathcal{M}$ is compact, only (T2) needs to be considered. Thanks to the Markov inequality, it is sufficient to show that for any $\varepsilon > 0$, there exists $\delta > 0$ such that for any

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

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

$$\mathrm{P}\big(\rho\big(X^{(n)}_{S_n}, X^{(n)}_{T_n}\big) > \varepsilon\big) \leqslant \varepsilon. \tag{70}$$

By the Markov inequality and Theorem 5, it is sufficient to study $\mathrm{E}\big[\big\|X^{(n)}_{S_n} - X^{(n)}_{T_n}\big\|_2^2\big]$. We observe that

$$\begin{aligned}
\mathrm{E}\Big[\Big\|X^{(n)}_{S_n} - X^{(n)}_{T_n}\Big\|_2^2\Big] &\leqslant \mathrm{E}\left[\int_{S_n}^{T_n}\sum_{i=1}^n \frac{1}{nh_n^{d+2}}K\left(\frac{\|X_i - X^{(n)}_s\|_2}{h_n}\right)\Big\|X_i - X^{(n)}_s\Big\|_2^2 \mathrm{d}s\right]\\
&\leqslant \delta\,\mathrm{E}\left[\sup_{x\in\mathcal{M}}\sum_{i=1}^n\frac{1}{nh_n^{d+2}}K\left(\frac{\|X_i - x\|_2}{h_n}\right)\|X_i - x\|_2^2\right]\\
&\leqslant \delta\ \sup_{x\in\mathcal{M}}\frac{1}{h_n^{d+2}}\int K\left(\frac{\|x-y\|_2}{h_n}\right)\|x-y\|_2^2 p(y)\mu(\mathrm{d}y)\\
&\quad + \delta\frac{1}{h_n^{d+2}}\mathrm{E}\left[\sup_{x\in\mathcal{M}}\left|\frac{1}{n}\sum_{i=1}^n K\left(\frac{\|X_i - x\|_2}{h_n}\right)\|X_i - x\|_2^2 - \mathrm{E}\left[K\left(\frac{\|X_1 - x\|_2}{h_n}\right)\|X_1 - x\|_2^2\right]\right|\right].
\end{aligned}$$

Since p is bounded on the compact $\mathcal{M}$ by continuity, using (35) of Lemma 1, and using the estimate of $Y_n^{(\alpha,\beta)}$ in Section 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

$$\mathrm{E}\Big[\Big\|X^{(n)}_{S_n} - X^{(n)}_{T_n}\Big\|_2^2\Big] \leqslant C\delta. \tag{71}$$

Choosing $\delta = \varepsilon^2/C$ yields (70). □

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 the Skorokhod representation theorem (see [6, Theorem 29.6 p. 399]), we can assume that this convergence holds almost surely, and with an abuse of notation we write $(X^{(n)})_{n\geqslant 0}$ for the subsequence converging to Y.

Lemma 5. *Under the hypothesis of Theorem 3, 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

$$\mathrm{E}\left[\left(\prod_{i=0}^k g_i(Y_{t_i})\right)\left(f(Y_t) - f(Y_s) - \int_s^t \mathcal{A}f(Y_u)\,\mathrm{d}u\right)\right] = 0.$$

Let us denote by Ψ the application

$$\Psi : y\in\mathbb{D}([0, T], \mathcal{M})\mapsto \left(\prod_{i=0}^k g_i\big(y_{t_i}\big)\right)\left(f(y_t) - f(y_s) - \int_s^t \mathcal{A}f(y_u)\,\mathrm{d}u\right).$$

We have

$$\begin{aligned}\left|\mathrm{E}\big[\Psi\big(X^{(n)}\big)\big]\right| &\leqslant \left|\mathrm{E}\left[\left(\prod_{i=0}^{k} g_i\big(X_{t_i}^{(n)}\big)\right)\left(f\big(X_t^{(n)}\big)-f\big(X_s^{(n)}\big)-\int_s^t \mathcal{A}_{h_n,n}f(X_u^{(n)})\,\mathrm{d}u\right)\right]\right| \\ &\quad + \left|\mathrm{E}\left[\left(\prod_{i=0}^{k} g_i\big(X_{t_i}^{(n)}\big)\right)\int_s^t \big(\mathcal{A}_{h_n,n}f(X_u^{(n)})-\mathcal{A}f(X_u^{(n)})\,\mathrm{d}u\right]\right| \\ &= O\left((t-s)\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}}+h_n\right)\end{aligned}$$

by Proposition 6 and Theorem 1. The application Ψ is not continuous on $\mathbb{D}([0,T],\mathcal{M})$ in general, but since the limiting process is continuous almost surely, we have that $\Psi(X^{(n)})$ converges to $\Psi(Y)$ almost surely. We can then conclude the proof as $\mathrm{E}[\Psi(Y)]=0$ by using the dominated convergence theorem. □

Proof of Theorem 3. From Lemmas Proof of Lemma 4 and Proof of Lemma 5, 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 [19]. 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 3. □

7.2. Convergence of the kNN random walk

We prove Theorem 4. For the sake of notation, the random walk $X^{(n),k\mathrm{NN}}$ is now denoted by $X^{(n)}$. Let us recall its SDE:

$$X_t^{(n)}=X_0^{(n)}+\int_0^t\int_{\mathbb{N}}\int_{\mathbb{R}_+}\mathbf{1}_{i\leqslant n}\mathbf{1}_{\theta\leqslant\frac{1}{nR_{n,k_n}^{d+2}(X_{s_-}^{(n)})}}\mathbf{1}_{[0,1]}\left(\frac{\|X_i-X_{s_-}^{(n)}\|_2}{R_{n,k_n}\big(X_{s^-}^{(n)}\big)}\right)\big(X_i-X_{s_-}^{(n)}\big)\,Q(\mathrm{d}s,\mathrm{d}i,\mathrm{d}\theta),$$

with

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

The related martingale is thus

$$M_t^{n,f}=f\big(X_t^{(n)}\big)-f\big(X_0^{(n)}\big)-\int_0^t \mathcal{A}_n^{k\mathrm{NN}}(f)(X_s^{(n)})\,\mathrm{d}s$$

with predictable quadratic variation

$$\langle M^{n,f}\rangle_t=\int_0^t\sum_{i=1}^{n}\mathbf{1}_{[0,1]}\left(\frac{\|X_i-X_s^{(n)}\|_2}{R_{n,k_n}\big(X_s^{(n)}\big)}\right)\big(f(X_i)-f\big(X_s^{(n)}\big)\big)^2\,\mathrm{d}s.$$

As in the proof of Lemma 4, to obtain the tightness of the distribution, thanks to Aldous's 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

$$\mathrm{P}\big(\rho\big(X_{S_n}^{(n)},X_{T_n}^{(n)}\big)>\varepsilon\big)\leqslant\varepsilon.$$

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

We observe that, using the quantities $h_n(x)$ and γ_n defined by (62) in Section 6,

$$\begin{aligned}\mathrm{P}\Big(\rho\big(X^{(n)}_{S_n}, X^{(n)}_{T_n}\big) > \varepsilon\Big) \leqslant &\mathrm{P}\Big(\rho\big(X^{(n)}_{S_n}, X^{(n)}_{T_n}\big) > \varepsilon,\ \sup_{x\in\mathcal{M}} \big|R_{n,k_n}(x) - h_n(x)\big| \leqslant \gamma_n\Big)\\ &+ \mathrm{P}\Big(\sup_{x\in\mathcal{M}} \big|R_{n,k_n}(x) - h_n(x)\big| > \gamma_n\Big)\\ \leqslant &\frac{1}{\varepsilon^2}\mathrm{E}\bigg(\Big\|X^{(n)}_{S_n} - X^{(n)}_{T_n}\Big\|_2^2 \mathbf{1}_{\sup_{x\in\mathcal{M}} |R_{n,k_n}(x)-h_n(x)|\leqslant\gamma_n}\bigg) + n^{-10}.\end{aligned}$$

Introducing (a_n, b_n) given by (65), 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 $\big\{\sup_{x\in\mathcal{M}} \big|R_{n,k_n}(x) - h_n(x)\big| \leqslant \gamma_n\big\}$,

$$\begin{aligned}&\Big\|X^{(n)}_{T_n} - X^{(n)}_{S_n}\Big\|_2\\ &\quad\leqslant \int_{S_n}^{T_n}\int_{\mathbb{N}}\int_{\mathbb{R}_+} \mathbf{1}_{i\leqslant n}\mathbf{1}_{\theta\leqslant\frac{1}{nR^{d+2}_{n,k_n}(X^{(n)}_{s-})}}\mathbf{1}_{[0,1]}\left(\frac{\|X_i - X^{(n)}_{s-}\|_2}{R_{n,k_n}\Big(X^{(n)}_{s^-}\Big)}\right)\Big\|X_i - X^{(n)}_{s-}\Big\|_2\, Q(\mathrm{d}s, \mathrm{d}i, \mathrm{d}\theta)\\ &\quad\leqslant \int_{S_n}^{T_n}\int_{\mathbb{N}}\int_{\mathbb{R}_+} \mathbf{1}_{i\leqslant n}\mathbf{1}_{\theta\leqslant\frac{1}{na_n^{d+2}}}\mathbf{1}_{\|X_i-X^{(n)}_{s-}\|_2\leqslant b_n}\Big\|X_i - X^{(n)}_{s-}\Big\|_2\, Q(\mathrm{d}s, \mathrm{d}i, \mathrm{d}\theta).\end{aligned}$$

We deduce that

$$\begin{aligned}&\mathrm{E}\bigg[\Big\|X^{(n)}_{S_n} - X^{(n)}_{T_n}\Big\|_2^2 \mathbf{1}_{\sup_{x\in\mathcal{M}} |R_{n,k_n}(x)-h_n(x)|\leqslant\gamma_n}\bigg]\\ &\quad\leqslant \mathrm{E}\bigg[\int_{S_n}^{T_n}\sum_{i=1}^n \frac{1}{na_n^{d+2}}\mathbf{1}_{\|X_i-X^{(n)}_{s-}\|_2\leqslant b_n}\Big\|X_i - X^{(n)}_{s}\Big\|_2^2 \mathrm{d}s\bigg]\\ &\quad\leqslant \frac{\delta}{na_n^{d+2}}\,\mathrm{E}\bigg[\sup_{x\in\mathcal{M}}\sum_{i=1}^n \mathbf{1}_{\|X_i-x\|_2\leqslant b_n}\|X_i - x\|_2^2\bigg]\\ &\quad\leqslant \frac{\delta}{a_n^{d+2}}\int_{\mathcal{M}} \mathbf{1}_{\|x-y\|_2\leqslant b_n}\|x-y\|_2^2\, p(y)\mu(\mathrm{d}y)\\ &\qquad + \frac{\delta}{a_n^{d+2}}\,\mathrm{E}\bigg[\sup_{x\in\mathcal{M}}\bigg|\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\|X_i-x\|_2\leqslant b_n}\|X_i - x\|_2^2 - \mathrm{E}\big[\mathbf{1}_{\|X_1-x\|_2\leqslant b_n}\|X_1 - x\|_2^2\big]\bigg|\bigg].\end{aligned}$$

As in the proof of Lemma 4, we use for the first term on the right-hand side that p is bounded on the compact $\mathcal{M}$, together with (35) of Lemma 1. For the second term we use the estimate of $\tilde{Y}_n^{(\alpha,\beta)}$ in Section 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 or ε, such that

$$\mathrm{E}\Big[\Big\|X^{(n)}_{S_n} - X^{(n)}_{T_n}\Big\|_2^2\Big] \leqslant C\delta. \tag{72}$$

This shows that the sequence of random walks is tight. The identification of the limiting martingale problem follows the proof of Theorem 3 using Theorem 2.

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

Appendix A. Some concentration inequalities

A.1. Talagrand's concentration inequality

As a corollary of Talagrand's inequality presented in Massart [28, Theorem 3], where for simplicity we choose $\varepsilon = 8$, we have the following deviation inequality.

Corollary 1 (Simplified version of Massart's inequality). *Consider n independent random variables $\xi_1, \ldots, \xi_n$ 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, $\|f\|_\infty \le b$ for every $f \in \mathcal{F}$. Furthermore, given*

$$Z := \sup_{f\in\mathcal{F}} \left| \sum_{i=1}^{n} \Big(f(\xi_i) - \mathrm{E}\left[f(\xi_i)\right] \Big) \right|,$$

then with $\sigma^2 = \sup_{f\in\mathcal{F}} \mathrm{Var}(f(\xi_1))$, and for any positive real number x,

$$\mathrm{P}\big(Z \ge 9(\mathrm{E}[Z] + \sigma\sqrt{nx} + bx)\big) \le \mathrm{e}^{-x}.$$

A.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 $N(\varepsilon, S, d) \le N(\varepsilon, T, d)$ because of the constraints that the centers x_i should belong to S. However, we can bound the covering number of S by T as follows.

Lemma 1. *If $S \subset T$ is a subspace of the metric space (T, d), then for any positive number ε,*

$$N(2\varepsilon, S, d) \le N(\varepsilon, T, d).$$

Proof. Let $\{x_1, \ldots, x_N\}$ be a ε-cover of T and, for any $i \in [\![1, N]\!]$, let us define $K_i := \{x \in T : d(x, x_i) \le \varepsilon\}$. Of course, K_i may not intersect S, hence, without loss of generality, assume that for a natural number $0 < m \le N$ we have that $K_i \cap S \ne \emptyset$ if and only if $i \le m$. Let y_i be any point in $K_i \cap S$ for $i \in [\![1, m]\!]$. Since $\{x_1, \ldots, x_N\}$ is a ε cover of T, for any $y \in S$, there exists an $i \le m$ such that $y \in K_i \cap S$. Hence, $d(y, y_i) \le 2\varepsilon$. Consequently, $y_1, \ldots, y_m$ is a 2ε-cover of (S, d). □

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 (Bound on ε-covering numbers). *Let $\mathcal{F}, \mathcal{G}$ be two bounded collections of measurable functions, that is, there are two constants c_1, c_2 such that*

$$\|f\|_\infty \le c_1 \quad \textit{and} \quad \|g\|_\infty \le c_2, \quad \textit{for all } f \in \mathcal{F},\ g \in \mathcal{G}.$$

Then for any probability measure Q,

$$N(2\varepsilon c_1 c_2, \mathcal{F}\cdot\mathcal{G}, L^2(Q)) \le N(\varepsilon c_1, \mathcal{F}, L^2(Q)) N(\varepsilon c_2, \mathcal{G}, L^2(Q)).$$

Proof. If $f_1, f_2, \ldots, f_n$ is a εc_1-cover of $(\mathcal{F}, L^2(Q))$ and $g_1, g_2, \ldots, g_m$ is a εc_2-cover of $(\mathcal{G}, L^2(Q))$, then, for any $(f, g) \in \mathcal{F} \times \mathcal{G}$, we have

$$|f(x)g(x) - f_i(x)g_j(x)| \le |f(x) - f_i(x)|c_2 + c_1|g(x) - g_i(x)|,$$

which implies that $\{f_i g_j : 1 \le i \le n \text{ and } 1 \le j \le m\}$ is a $2\varepsilon c_1 c_2$ -cover of $\mathcal{F}\cdot\mathcal{G}$. □

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

The following lemma is just a simplified version of the result of the theory of Vapnik–Chervonenkis Hull class of functions [16, Section 3.6.3].

Lemma 3. *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 of Vapnik–Chervonenkis type with respect to a constant envelope.

Proof. Let $N = [\frac{b-a}{\varepsilon}]$, and let $f_i = f + i\varepsilon$ 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 measures Q on $\mathbb{R}^m$, we have $\|g - f_i\|_{L^2(Q)} \leq \varepsilon$, which makes $\mathcal{H} := \{f_i : i \in [\![1, N]\!]\}$ an 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 Vapnik–Chervonenkis-type class of functions with $A = b - a$, $v = 1$, and $F = \max(1, \|f\|_\infty + |a|, \|f\|_\infty + |b|)$. □

Appendix B. Some estimates using the total variation

Lemma 4. *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 $a \leqslant b$,*

$$|K(b) - K(a)| \leq H(b) - H(a). \tag{B1}$$

Furthermore, if K satisfies Assumption 1, then, when b goes to infinity,

$$K(b)b^{d+3} = o(1) \quad \text{and} \quad \int_b^\infty K(a)a^{d+1}\mathrm{d}a = o(1/b). \tag{B2}$$

Proof of Lemma 4. Inequality (B1) 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 1,

$$\lim_{b\to+\infty} b^{d+3}(H(\infty) - H(b)) = 0.$$

Then, as

$$K(b)b^{d+3} \leq b^{d+3}(H(\infty) - H(b)),$$

we have proved the first estimation in (B2).

For the second estimation, we see that

$$\begin{aligned}(d+2)\int_b^\infty bK(a)a^{d+1}\mathrm{d}a &\leq (d+2)\int_b^\infty b(H(\infty) - H(a))a^{d+1}\mathrm{d}a \\ &= -b^{d+3}(H(\infty) - H(b)) + b\int_b^\infty a^{d+2}\mathrm{d}H(a) \\ &\leq -b^{d+3}(H(\infty) - H(b)) + \int_b^\infty a^{d+3}\mathrm{d}H(a).\end{aligned}$$

and we are done. □

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

Appendix C. Proof of Lemma 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, the left-hand side of (41) is equal to

$$\begin{aligned} &\left[\frac{1}{d}\int_{B_{\mathbb{R}^d}(0,c)} G(\|v\|_2)\|v\|_2^2 \,\mathrm{d}v\right]\left[\sum_{i=1}^{d}\left\langle \nabla_{\mathbb{R}^m} f(x), \frac{\partial k}{\partial x^i}(0)\right\rangle\left\langle \nabla_{\mathbb{R}^m} h(x), \frac{\partial k}{\partial x^i}(0)\right\rangle\right] \\ &= \left[\frac{1}{d}\int_{B_{\mathbb{R}^d}(0,c)} G(\|v\|_2)\|v\|_2^2 \,\mathrm{d}v\right]\left[\sum_{i=1}^{d}\frac{\partial (f\circ k)}{\partial x^i}(0)\,\frac{\partial (h\circ k)}{\partial x^i}(0)\right] \\ &= \left[\frac{1}{d}\int_{B_{\mathbb{R}^d}(0,c)} G(\|v\|_2)\|v\|_2^2 \,\mathrm{d}v\right]\left\langle \nabla_{\mathbb{R}^d}(f\circ k)(0), \nabla_{\mathbb{R}^d}(h\circ k)(0)\right\rangle. \end{aligned}$$

Hence, we have (41).

For (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, the left-hand side of (42) is equal to

$$\left[\left\langle \nabla_{\mathbb{R}^m} f(x), \frac{1}{2}\sum_{i=1}^{d}\frac{\partial^2 k}{\partial x^i \partial x^i}(0)\right\rangle + \frac{1}{2}\sum_{i=1}^{d} f''(x)\left(\frac{\partial k}{\partial x^i}(0), \frac{\partial k}{\partial x^i}(0)\right)\right]\frac{1}{d}\int_{B_{\mathbb{R}^d}(0,c)} G(\|v\|_2)\|v\|_2^2 \,\mathrm{d}v.$$

Furthermore, 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 concludes the proof of Lemma 2.

Acknowledgements

The authors thank Frédéric Rochon for useful discussions. They also thank the anonymous referee for their comments and corrections that improved the paper.

https://doi.org/10.1017/apr.2026.10060 Published online by Cambridge University Press

Funding information

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.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] ALDOUS, D. (1978). Stopping times and tightness. *Ann. Probab.* **6**, 335–340.

[2] BELKIN, M. AND NIYOGI, P. (2008). Towards a theoretical foundation for Laplacian-based manifold methods. *J. Comput. Syst. Sci.* **74**(8), 1289–1308.

[3] BELKIN, M., SUN, J. AND WANG, Y. (2009). 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.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Association for Computing Machinery (ACM), New York, pp. 1031–1040.

[4] BERENFELD, C. AND HOFFMANN, M. (2021). Density estimation on an unknown submanifold. *Electron. J. Stat.* **15**(1), 2179–2223.

[5] BILLINGSLEY, P. (1999). *Convergence of Probability Measures*. Wiley, Chichester.

[6] BILLINGSLEY, P. (2012). *Probability and Measure*, anniversary ed., Wiley Series in Probability and Statistics. John Wiley & Sons, Hoboken, NJ.

[7] CALDER, J. AND GARCA TRILLOS, N. (2022). Improved spectral convergence rates for graph Laplacians on ε-graphs and k-NN graphs. *Appl. Comput. Harmon. Anal.* **60**, 123–175.

[8] CHAVEL, I. (2006). *Riemannian Geometry*, 2nd edn., Cambridge Studies in Advanced Mathematics, vol. 98. Cambridge University Press, Cambridge.

[9] CHENG, X. AND WU, H.-T. (2022). Convergence of graph Laplacian with kNN self-tuned kernels. *Inform. Inference* **11**(3), 889–957.

[10] CIPRIANI, A. AND VAN GINKEL, B. (2020). The discrete Gaussian free field on a compact manifold. *Stoch. Process. Appl.* **130**(7), 3943–3966.

[11] DEVROYE, L., GYÖRFI, L. AND LUGOSI, G. (1996). *A probabilistic Theory of Pattern Recognition*, Applications of Mathematics (New York), vol. 31. Springer-Verlag, New York.

[12] DIVOL, V., GUÉRIN, H., NGUYEN, D.-T. AND TRAN, V. (2025). Measure estimation on a manifold explored by a diffusion process. *Probab. Theory Relat. Fields.*

[13] GARCA TRILLOS, N., GERLACH, M., HEIN, M. AND Slepcčev, D. (2020). Error estimates for spectral convergence of the graph Laplacian on random geometric graphs toward the Laplace-Beltrami operator. *Found. Comput. Math.* **20**(4), 827–887.

[14] GINÉ, E. AND GUILLOU, A. (2002). 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), 907–921.

[15] GINÉ, E., KOLTCHINSKII, V., LI, W. AND ZINN, J., Eds. (2006). High Dimensional Probability. Proceedings of the Fourth International Conference. Many Papers based on the Presentations at the Conference, Santa Fe, NM, USA, June 20–24, 2005, vol. 51. IMS, Institute of Mathematical Statistics, Beachwood, OH.

[16] GINÉ, E. AND NICKL, R. (2016). Mathematical Foundations of Infinite-Dimensional Statistical Models. In *Cambridge Series in Statistical and Probabilistic Mathematics*, vol. 40. Cambridge University Press, Cambridge.

[17] HEIN, M., AUDIBERT, J.-Y. AND VON LUXBURG, U. (2005). From graphs to manifolds—weak and strong pointwise consistency of graph Laplacians. In *Learning Theory*, Lecture Notes in *Computer Science*, vol. 3559. Springer, Berlin, pp. 470–485.

[18] HEIN, M., AUDIBERT, J.-Y. AND VON LUXBURG, U. (2007). Graph Laplacians and their convergence on random neighborhood graphs. *J. Mach. Learn. Res.* **8**, 1325–1368.

[19] HSU, E. P. (2002). *Stochastic Analysis on Manifolds*, *Graduate Studies in Mathematics*, vol. 38. American Mathematical Society, Providence, RI.

[20] IKEDA, N. AND WATANABE, S. (1989). *Stochastic Differential Equations and Diffusion Processes*, 2nd ed., North-Holland Mathematical Library, vol. 24. North-Holland, Amsterdam; Kodansha Ltd., Tokyo.
[21] JAKUBOWSKI, A. (1986). On the Skorokhod topology. *Ann. Inst. Henri Poincaré* **22**(3), 263–285.
[22] JOFFE, A. AND MÉTIVIER, M. (1986). Weak convergence of sequences of semimartingales with applications to multitype branching processes. *Adv. Appl. Probab.* **18**, 20–65.
[23] KALLENBERG, O. (2021). *Foundations of Modern Probability*, 3rd edn., Probability Theory and Stochastic Modelling, vol. 99. Springer, New York.
[24] LAMBERT, R., TULEAU-MALOT, C., BESSAIH, T., RIVOIRARD, V., BOURET, Y., LERESCHE, N. AND REYNAUD-BOURET, P. (2018). Reconstructing the functional connectivity of multiple spike trains using Hawkes models. *J. Neurosci. Methods* **297**, 9–21.
[25] LEE, J. M. (2013). *Introduction to Smooth Manifolds*, 2nd edn., Graduate Texts in Mathematics, vol. 218. Springer, New York.
[26] LEE, J. M. (2018). *Introduction to Riemannian Manifolds*, Graduate Texts in Mathematics, vol. 176. Springer, Cham.
[27] M, H., AUDIBERT, J.-Y. AND VON LUXBURG, U. (2007). Graph Laplacians and their convergence on random neighborhood graphs. *J. Mach. Learn. Res.* **8**, 1325–1368.
[28] MASSART, P. (2000). About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.* **28**(2), 863–884.
[29] NOLAN, D. AND POLLARD, D. (1987). U-processes: Rates of convergence. *Ann. Stat.* **15**, 780–799.
[30] SART, M. (2017). Estimating the conditional density by histogram type estimators and model selection. *ESAIM: PS* **21**, 34–55.
[31] SINGER, A. (2006). From graph to manifold Laplacian: The convergence rate. *Appl. Comput. Harmon. Anal.* **21**(1), 128–134.
[32] TAO, W. AND SHI, Z. (2020). Convergence of Laplacian spectra from random samples. *J. Comput. Math.* **38**(6), 952–984.
[33] TING, D., HUANG, L. AND JORDAN, M. (2010). An analysis of the convergence of graph Laplacians. In *ICML*.
[34] VAN GINKEL, B. AND REDIG, F. (2020). Hydrodynamic limit of the symmetric exclusion process on a compact Riemannian manifold. *J. Stat. Phys.* **178**(1), 75–116.
[35] WANG, F.-Y. AND ZHU, J.-X. (2023). Limit theorems in Wasserstein distance for empirical measures of diffusion processes on Riemannian manifolds. *Ann. Inst. Henri Poincaré, Probab. Stat.* **59**(1), 437–475.
[36] YOSIDA, K. (1980). *Functional Analysis*, 6th edn., Classics in Mathematics. Springer, Berlin.