

Nonparametric Regression on Random Geometric Graphs Sampled from Submanifolds

Paul Rosa

PAUL.ROSA@STATS.OX.AC.UK

Department of Statistics

University of Oxford

24-29 St Giles', Oxford OX1 3LB, United Kingdom

Judith Rousseau

ROUSSEAU@CEREMADE.DAUPHINE.FR

Department of Statistics

University of Oxford

24-29 St Giles', OX1 3LB Oxford, United Kingdom

and CEREMADE, CNRS

Université Paris-Dauphine, PSL University

Place du Maréchal de Lattre de Tassigny, 75775 Paris Cedex 16, France

Editor: Tong Zhang

Abstract

We consider the nonparametric regression problem when the covariates are located on an unknown compact submanifold of a Euclidean space. Under defining a random geometric graph structure over the covariates we analyse the asymptotic frequentist behaviour of the posterior distribution arising from Bayesian priors designed through random basis expansion in the graph Laplacian eigenbasis. Under Hölder smoothness assumption on the regression function and the density of the covariates over the submanifold, we prove that the posterior contraction rates of such methods are minimax optimal (up to logarithmic factors) for any positive smoothness index.

Keywords: graph Laplacian, Bayesian nonparametrics, manifold hypothesis, nonparametric regression, dimensionality reduction

1. Introduction

When the data is high dimensional, it is common practice to assume a lower dimensional structure. This idea dates back at least to Pearson's principal component analysis (Pearson, 1901) and the Johnson-Lindenstrauss lemma (Johnson, Lindenstrauss, and Schechtman, 1986; Ailon and Chazelle, 2009) where a linear projection onto a lower dimensional subspace of variables is performed while still retaining much of the statistical information. However, a linear constraint on the possible intrinsic shape of the data can be restrictive, and non-linear dimensionality reduction has received much attention during the past two decades. Examples include kernel PCA (Schölkopf, Smola, and Müller, 1998), Isomap (Tenenbaum, de Silva, and Langford, 2000), locally linear embeddings (Roweis and Saul, 2000), Laplacian eigenmaps and diffusion maps (Coifman and Lafon, 2006). These techniques typically take into account intrinsic geometric properties of the data.

Inference of intrinsic geometric properties of data is often believed to be a powerful way to gain insight in complex high dimensional statistical problems. Examples include its support (Genovese, Perone-Pacifico, Verdinelli, and Wasserman, 2012; Aamari and Levrard, 2019) under the Hausdorff metric, differential quantities (Aamari and Levrard, 2019; Aamari, Berenfeld, and Levrard, 2023), the geodesic distance (Aamari, Berenfeld, and Levrard, 2023; Bernstein, de Silva, Langford, and Tenenbaum), the intrinsic dimension (Kim, Rinaldo, and Wasserman, 2019; Denti, Doimo, Laio, and Mira, 2022) or even the Laplace-Beltrami operator eigenpairs (García Trillos, Gerlach, Hein, and Slepčev, 2020; Calder and Trillos, 2020; Dunson, Wu, and Wu, 2021; Wormell and Reich, 2021). In the same way, inferring the intrinsic topological structure in the data has also received a lot of interest since the modern development of topological data analysis as field within data science, with in particular the study of the persistence diagrams and modules (Chazal and Michel, 2021; Chazal, de Silva, Glisse, and Steve, 2016; Divol and Lacombe, 2021; Divol and Chazal, 2019; Loiseaux, Carrière, and Blumberg, 2023; Loiseaux, Scoccola, Carrière, Botnan, and Oudot, 2024).

The Laplacian eigenmaps, introduced initially by Belkin and Niyogi (2001, 2003) is a popular non linear method for dimensionality reduction, together with a tool for inference in high dimension setups. It has also been used successfully in the context of spectral clustering (Liu and Han, 2014; von Luxburg, 2007), supervised and semisupervised learning (Green, Balakrishnan, and Tibshirani, 2021; Shi, Balasubramanian, and Polonik, 2023; Sanz-Alonso and Yang, 2022) : in this paper we consider the use of Laplacian eigenmaps in the context of the reknown nonparametric regression problem where

$$y_i = f(x_i) + \varepsilon_i, \quad \varepsilon_i \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2), \quad i \leq n; \quad (1)$$

with covariates $x_i \in \mathbb{R}^D$ and D is possibly large. Then, a way to formalize a low dimensional structure to the problem is to assume that the covariates X_i 's belong to some low dimensional submanifold \mathcal{M} of \mathbb{R}^D , also known as the manifold hypothesis. This is the setup we are considering. We also consider both the supervised learning setup where the observations consist of $(y_i, x_i)_{i=1}^n$, under model 1; and the semi-supervised setup where in addition another sample $(x_i)_{i=n+1}^N$ of covariates are observed without their labels y_i .

Laplacian eigenmaps approaches are based on the construction of a graph $G = (V, E)$ whose vertices are the covariates $(x_i)_{i=1}^N$ and where an edge between two vertices x_i and x_j exists if and only if $\|x_i - x_j\| \leq h$ for some predefined threshold h . From that the Laplacian of the graph $L : L^2(V) \rightarrow L^2(V)$, is constructed together with its spectral decomposition $L = \sum_{j=1}^N \lambda_j \langle u_j | \cdot \rangle_{L^2(V)} u_j$, $0 \leq \lambda_1 \leq \dots \leq \lambda_N$ for some appropriate Euclidian space structure on V to be defined in Section 2. Dimensionality reduction is then obtained by considering the projection operator $\sum_{j=1}^J \langle u_j | \cdot \rangle_{L^2(V)} u_j$ for some truncation level J . This dimension reduction can be used in different inferential contexts

In the context of the regression problem 1, the vector $f_N = (f(x_1), \dots, f(x_N))^T$ is modelled as an element of $\text{span}(u_1, \dots, u_J)$. This construction has been studied with variants in the definition of the graph Laplacian L , for instance by Green, Balakrishnan, and Tibshirani (2021); Shi, Balasubramanian, and Polonik (2023), where the PCR-LE estimator \hat{f}_N , i.e. it is the minimizer of $\|Y - f_N\|^2$ over $\Sigma^J := \text{span}(u_1, \dots, u_J)$ (in the fully supervised case $N = n$) while Sanz-Alonso and Yang (2022); Fichera, Borovitskiy, Krause, and Billard

(2024); Dunson, Wu, and Wu (2022) are using a Bayesian estimation procedure based on Gaussian processes defined via L .

In this paper we are interested in understanding the capacity of graph Laplacian regression methods of capturing both the unknown low dimensional submanifold support \mathcal{M} of the design, together with the unknown smoothness β of the regression functions.

1.1 Related works :

The asymptotic behaviour of least square estimators has been studied in Green, Balakrishnan, and Tibshirani (2021); Shi, Balasubramanian, and Polonik (2023) where they obtain minimax convergence rates when the support \mathcal{M} of the covariates is an open set of \mathbb{R}^D and under a Sobolev regularity condition. Extension to the case where \mathcal{M} is a d dimensional submanifold of \mathbb{R}^D , with $d < D$, is also considered in Green, Balakrishnan, and Tibshirani (2021) however only for regression functions which have Sobolev smoothness restricted to the set $\{1, 2, 3\}$.

Bayesian estimators of f based on the graph Laplacian have been considered in the literature as well for instance by Sanz-Alonso and Yang (2022); Fichera, Borovitskiy, Krause, and Billard (2024); Dunson, Wu, and Wu (2022), however the asymptotic properties in terms of posterior contraction rates around the true regression function have only been derived in some restricted cases. Dunson, Wu, and Wu (2022) show that in the case $N = n$ and when the true regression vector $f_{0,N} = (f_0(x_i))_{i=1}^N$ belongs to Σ^J for $J \lesssim \left(\frac{n}{\ln n}\right)^{\frac{d}{2q+2d}}$ for some $q \geq \frac{d}{2}$, the posterior contraction rate (with respect to the empirical L^2 norm) is upper bounded by a multiple of $\left(\frac{\ln n}{n}\right)^{\frac{d}{2q+2d}}$, under some kind of separability assumption between the eigenvalues of the Laplace-Beltrami operator of \mathcal{M} . Assuming that $f_{0,N} \in \Sigma^J$ avoids studying the approximation of $f_{0,N}$ by an element in Σ^J , which is a nontrivial problem where one only has access to smoothness assumptions on the function $f_0 : \mathcal{M} \rightarrow \mathbb{R}$. A step in this direction has been made by Sanz-Alonso and Yang (2022), where they established that if the number of unlabeled covariates $\{x_{n+1}, \dots, x_N\}$ is large enough, i.e at least $N \geq n^{2d}$, then posterior contraction at a minimax optimal rate (up to logarithmic factors) of the true regression vector $f_{0,N}$ is possible, provided that its smoothness β (more precisely : f_0 belongs to the Besov space $B_{\infty\infty}^\beta(\mathcal{M})$, see Coulhon, Kerkycharian, and Petrushev (2012)) satisfies $\beta > d - \frac{1}{2}$ using a graph Matérn process. Notably, as is pointed out in the paper, their assumptions rule out the fully supervised setting and every other intermediate cases. Furthermore, adaptation to the smoothness β is not discussed. In both papers Dunson, Wu, and Wu (2022) and Sanz-Alonso and Yang (2022) control the contraction of the posterior distribution on f_N by controlling the convergence of the graph Laplacian (or the associated heat kernel) to the Laplacian-Beltrami operator of \mathcal{M} . As we will see, it is not fundamentally needed to study the nonparametric regression problem; this is also the point of view taken by Green, Balakrishnan, and Tibshirani (2021); Shi, Balasubramanian, and Polonik (2023).

Thus, the question the approximation of $f_{0,N}$ by elements of Σ^J for $J = o(N)$, when f_0 is a smooth function on an unknown manifold \mathcal{M} remains unanswered. In this paper we therefore aim to bridge this gap in the literature.

1.2 Our contributions :

In this paper, our main aim is to study posterior contraction rates for priors based on the graph Laplacian for the recovery of the regression vector $f_{0,N}$. To do that we first show that any β Hölder function on \mathcal{M} , $\beta > 0$ can be well approximated (in some sense to be made precise later in the paper) by a function in Σ^J for $J \geq n^{d/(2\beta+d)}$ (up to $\log n$ terms), where d is the intrinsic dimension of the submanifold \mathcal{M} . To do that we use a novel approximation argument by first constructing an approximation of $f_{0,N}$ by a vector of the form $e^{-tL/h^2} f_t$ for some well chosen $t > 0$ and $f_t \in \mathbb{R}^V$. Thanks to the exponential factor we can then control well the difference between $e^{-tL/h^2} f_t$ and its projection on Σ^J . This approximation argument is valid for any $\beta > 0$. This result has an interest in its own right; in particular we apply it to the frequentist PCR-LE estimator.

We also propose general classes of priors on f_N supported on Σ^J for which we characterize the posterior contraction rates, both for the empirical loss $\|f - f_0\|_n^2 = \sum_{i=1}^n (f(x_i) - f_0(x_i))^2/n$ and for the global loss $\|f - f_0\|_N^2 = \sum_{i=1}^N (f(x_i) - f_0(x_i))^2/N$, as functions of the truncation level J and the connectivity parameter h .

Finally by considering hyperpriors on h and J we propose a novel class of priors that achieve minimax adaptive posterior contraction rates (up to $\ln n$ terms), both under the $\|\cdot\|_n$ and the $\|\cdot\|_N$ (pseudo-)distances.

Hence, contrarywise to Sanz-Alonso and Yang (2022) we do not need a very large number of unlabelled covariates which were used to ensure an accurate discrete to continuum approximation of the graph Laplacian eigenpairs and compared to Green, Balakrishnan, and Tibshirani (2021); Shi, Balasubramanian, and Polonik (2023); Sanz-Alonso and Yang (2022) we significantly weaken the smoothness assumptions on f_0 (our assumptions actually match those considered in the continuous setting (see van der Vaart and van Zanten, 2011; Rosa et al., 2023)). Also note that we do not require separability assumptions on the eigenvalues of the Laplace-Beltrami operator on \mathcal{M} , as opposed to Dunson, Wu, and Wu (2022).

1.3 Outline :

We start by describing our geometrical and statistical model as well as our notations in Section 2. We then describe our priors and the associated posterior contraction results in Section 3; with non adaptive results in Section 3.2.1 and adaptive results in Section 3.2.2. Convergence rates for the PCR-LE estimator are provided in Section 3.2.1. The approximation theory for $f_{0,N}$ by Σ^J is provided in Section 3.3. Main proofs are provided in Section 4. Finally, the appendix contains additional proofs.

2. Model and notations

In this paper we consider the semi-supervised nonlinear regression model where we observe labelled data $(y_i, x_i)_{i \leq n}$ together with unlabelled data $(x_i)_{i=n+1}^N$ (in particular $N = N_n \geq n$ depends on n). Note that $N = n$ corresponds to the supervised case. In this work we make the following growth assumption on N : there exists $b > 0$ such that $N_n \leq n^b$ for all n . Note that in particular we always have the inequality $\ln n \leq \ln N \leq b \ln n$. The nonparametric

regression model is then

$$\begin{aligned} y_i &= f_0(x_i) + \varepsilon_i, \quad \varepsilon_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2), \quad i = 1, \dots, n, \\ x_i &\stackrel{i.i.d.}{\sim} \mu_0, \quad i = 1, \dots, N \end{aligned} \quad (2)$$

where $\sigma > 0$, $f_0 \in \mathcal{C}^\beta(\mathcal{M})$ with \mathcal{M} an unknown compact connected submanifold of regularity $\alpha \geq (\beta + 3) \vee 6$ of \mathbb{R}^D (see appendix A for a precise definition of α and the space $\mathcal{C}^\beta(\mathcal{M})$) and μ_0 is a probability distribution on \mathcal{M} (the connectedness assumption could actually be dropped by working on each connected component separately). We assume that μ_0 is absolutely continuous with respect to the volume measure on \mathcal{M} , with density p_0 and f_0, \mathcal{M}, p_0 are unknown quantities. Without loss of generality we consider the case where σ is known, since the case of unknown σ can be easily derived from the results here, using the approach of Naulet and Barat (2018). Moreover, it would also be possible to handle non Gaussian errors as done by Ghosal and van der Vaart (2017, chap. 8.5.2).

We consider an unweighted random geometric graph $G = G^{(h)} = (V, E^{(h)})$ based on $x_{1:N} = (x_i)_{i=1}^N$ generated as follows: the vertex set is $V = \{x_1, \dots, x_N\}$ and the edges are defined by $E_{ij} = E_{ij}^{(h)} = 1$ if and only if $\|x_i - x_j\| < h$, for some $h > 0$, where $\|\cdot\|$ denotes the Euclidean distance in \mathbb{R}^D . Hereafter we denote $x_i \sim x_j \iff E_{ij} = 1$.

We define the *normalized graph Laplacian*

$$L = L^{(h)} = I - D^{-1}A$$

where $D = D^{(h)} = \text{diag}(\mu)$ is the *degrees matrix* $D_{xy} = \delta_x^y \mu_x$,

$$\mu_x = \mu_x^{(h)} = \#\{i : \|x - x_i\| < h\}$$

and $A = A^{(h)}$ is the *adjacency matrix* $A_{xy} = A_{xy}^{(h)} = \mu_{xy} = \mathbb{1}_{x \sim y}$. We also define the *normalized degree measure* $\nu = \nu^{(h)} = \frac{\mu}{\mu(V)}$. The graph Laplacian L is a nonnegative self-adjoint operator from $L^2(\nu)$ to itself, where the inner product is given by

$$\forall f, g \in L^2(\nu), \langle f | g \rangle_{L^2(\nu)} = \sum_{y \in V} f(y)g(y)\nu_y$$

and L is given explicitly by the formula

$$\forall f \in \mathbb{R}^V, (Lf)(x) = \frac{1}{\mu_x} \sum_{y \sim x} (f(x) - f(y))$$

We will often use the following identity

$$\forall f, g \in L^2(\nu), \langle f | Lg \rangle_{L^2(\nu)} = \frac{1}{2\mu(V)} \sum_{x \sim y} (f(x) - f(y))(g(x) - g(y))$$

The negative graph Laplacian $-L$ satisfies

$$1. \quad \forall x \in V, -L_{xx} = \frac{\mu_{xy}}{\mu_x} - 1 \leq 0$$

2. $\forall x \neq y \in V, -L_{xy} = \frac{\mu_{xy}}{\mu_x} \geq 0$
3. $\forall x \in V, \sum_{y \in V} -L_{xy} = \sum_{y \in V} \frac{\mu_{xy}}{\mu_x} - 1 = 0$

As a consequence e^{-tL} is a row stochastic matrix for any $t \geq 0$ (see Norris, 1997, chap. 2). It is the transition matrix in time t of the continuous time Markov chain $(\tilde{W}_t)_{t \geq 0}$ going from vertex x to vertex y with rate $\frac{\mu_{xy}}{\mu_x}$. This chain has a constant unit speed and a jump matrix given by $D^{-1}A$: equivalently $\tilde{W}_t = Y_{N_t}$ where N is a Poisson process with unit rate and $(Y_n)_{n \geq 0}$ is an independent Markov chain with transition matrix $D^{-1}A$. We will denote by $\mathbb{P}_x^{(h)}$ the probability distribution of the corresponding Markov chains (in either discrete or continuous times) starting from $x \in V$. For convenience in what follows we will define $\mathcal{L} = h^{-2}L$ and $W_t = \tilde{W}_{t/h^2}$.

It is shown for instance by García Trillos, Gerlach, Hein, and Slepčev (2020); Calder and Trillos (2020) that under appropriate conditions the graph Laplacian \mathcal{L} converges in some sense to the true Laplace-Beltrami operator Δ of the limiting manifold \mathcal{M} (up to a proportionality constant). Since W is a continuous time Markov chain with transition matrix $e^{-t\mathcal{L}}$, this shows that W can actually be seen in a way as a numerical approximation of a Brownian motion on \mathcal{M} , i.e an \mathcal{M} -valued continuous time Markov process with infinitesimal generator $-\Delta$.

For each $x \in V$, $t \geq 0$ we define $p_t(x, \cdot) = p_t^{(h)}(x, \cdot)$ as the density of $e^{-t\mathcal{L}}(x, \cdot)$ with respect to ν

$$\forall y \in V, \quad p_t(x, y) = \frac{e^{-t\mathcal{L}}(x, y)}{\nu_y}$$

and we call the matrix p_t the *heat kernel of \mathcal{L} with respect to ν* . If $(\lambda_j, u_j)_{j=1}^N = (\lambda_j^{(h)}, u_j^{(h)})_{j=1}^N$ is an orthonormal eigendecomposition of \mathcal{L} in $L^2(\nu)$, i.e.

$$\forall j, l \in \{1, \dots, N\}, \quad \mathcal{L}u_j = \lambda_j u_j, \quad \langle u_j | u_l \rangle_\nu = \delta_j^l$$

(the existence of which is guaranteed by the finite dimensional spectral theorem) then we have

$$\forall t \geq 0, \quad e^{-t\mathcal{L}} = \sum_{j=1}^N e^{-t\lambda_j} \langle u_j | \cdot \rangle_\nu u_j = \frac{1}{\mu(V)} \sum_{j=1}^N e^{-t\lambda_j} u_j u_j^T$$

Hence

$$\forall t \geq 0, \quad x, y \in V, p_t(x, y) = \sum_{j=1}^N e^{-t\lambda_j} u_j(x) u_j(y)$$

In particular p_t is a symmetric matrix.

In addition to the $L^2(\nu)$ structure we will also use the uniform norm

$$\forall f \in \mathbb{R}^V, \quad \|f\|_{L^\infty(\nu)} = \max_{1 \leq i \leq N} |f(x_i)|$$

As well as the associated operator norm

$$\|A\|_{L^\infty(\nu)} = \sup_{\|f\|_{L^\infty(\nu)} \leq 1} \|Af\|_{L^\infty(\nu)}$$

defined for any linear endomorphism $A : L^\infty(\nu) \rightarrow L^\infty(\nu)$.

2.1 Notations

We denote by \mathbb{P}_0 the frequentist probability distribution of $\mathbb{X}^n := (x_{1:N}, y_{1:n})$, i.e

$$\mathbb{P}_0(dx_{1:N}, dy_{1:n}) = \prod_{i=1}^N p_0(x_i) \mu(dx_i) \prod_{i=1}^n \mathcal{N}(y_i | f_0(x_i), \sigma^2) dy_i$$

as well as $P_f = \bigotimes_{i=1}^n \mathcal{N}(f_i, \sigma^2)$ for any $f \in \mathbb{R}^n$, with the abuse of notation $P_{f_0} = P_{(f_0(x_i))_{i=1}^n}$. We will also denote the probability distribution of the infinite sequences $(x_i)_{i \geq 1}$, $(y_i)_{i \geq 1}$ by P_0^∞ . We denote the geodesic metric on \mathcal{M} by $\rho : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}_+$. For any $x \in \mathbb{R}^D$, $r > 0$ we define $B_{\mathbb{R}^D}(x, r) = \{y \in \mathbb{R}^D : \|x - y\| < r\}$ the Euclidean ball and $B_\rho(x, r) = \{y \in \mathcal{M} : \rho(x, y) < r\}$ the geodesic ball. For any measurable subset A of \mathbb{R}^d , $d \geq 1$, $\text{vol}(A)$ will denote its d -dimensional Lebesgue measure.

For any open subset \mathcal{U} of \mathbb{R}^d , $d, k \geq 1$ and $f : \mathcal{U} \rightarrow \mathbb{R}$ we define its $\mathcal{C}^k(\mathcal{U})$ norm by $\|f\|_{\mathcal{C}^k(\mathcal{U})} = \max_{\alpha \in \mathbb{N}^d, |\alpha| \leq k} \sup_{x \in \mathcal{U}} \left| \frac{\partial^{|\alpha|} f}{\partial x^\alpha}(x) \right|$. We also define $\mathcal{C}^\infty(\mathcal{U}) = \bigcap_{k \geq 0} \mathcal{C}^k(\mathcal{U})$. For non integer $\beta \in (k, k+1)$, $k \in \mathbb{N}$ we define the Hölder class $\mathcal{C}^\beta(\mathcal{U})$ class as the space of all functions $f \in \mathcal{C}^k(\mathcal{U})$ satisfying

$$\|f\|_{\mathcal{C}^\beta(\mathcal{U})} := \|f\|_{\mathcal{C}^k(\mathcal{U})} + \max_{\alpha \in \mathbb{N}^d, |\alpha|=k} \sup_{x \neq y \in \mathcal{U}} \frac{\left| \frac{\partial^{|\alpha|} f}{\partial x^\alpha}(x) - \frac{\partial^{|\alpha|} f}{\partial x^\alpha}(y) \right|}{\|x - y\|^{\beta-k}} < +\infty$$

If V is a finite dimensional vector space and $\mathcal{U} \subset \mathbb{R}^d$, $d \geq 1$, we say that a mapping $f : \mathcal{U} \rightarrow V$ is of class $\mathcal{C}^\beta(\mathcal{U})$ if $T \circ f : \mathcal{U} \rightarrow \mathbb{R} \in \mathcal{C}^\beta(\mathcal{U})$ for any $T \in V^*$ (linear form on V). Equivalently, each coordinate function of f (in any choice of basis) is of class \mathcal{C}^β .

It is immediate to check that any $f \in \mathcal{C}^\beta(\mathcal{U})$, $k < \beta \leq k+1$ has a Taylor development of the form $f(y) = \sum_{l=0}^k \frac{d^l f(x) \cdot (y-x)^l}{l!} + R_k(x, y)$, $|R_k(x, y)| \leq C_f \|x - y\|^\beta$ with $C_f = \|f\|_{\mathcal{C}^\beta(\mathcal{U})}/k!$.

We write

$$\|f - f_0\|_n^2 = \frac{1}{n} \sum_{i=1}^n (f(x_i) - f_0(x_i))^2 \text{ and } \|f - f_0\|_N^2 = \frac{1}{N} \sum_{i=1}^N (f(x_i) - f_0(x_i))^2.$$

In the following the symbols \gtrsim , \lesssim refer to inequalities up to constants that depend only on \mathcal{M} , p_0 and f_0 . We will also write $a \asymp b$ if both $a \lesssim b$ and $b \lesssim a$ hold. We will also often write $C((\lambda)_{\lambda \in \Lambda})$ or $C_{(\lambda)_{\lambda \in \Lambda}}$ generically to denote a positive constant depending only on the parameters $\lambda \in \Lambda$ whose value can change from line to line.

3. Main results

In the nonparametric regression model 2, the goal is to estimate the regression function f_0 evaluated on the covariates x_i , $i = 1, \dots, N$. Hence from a Bayesian point of view the goal is to design a prior distribution on an element $f : V \rightarrow \mathbb{R}$ leading to good frequentist guarantees a posteriori. More precisely, following Ghosal, Ghosh, and van der Vaart (2000); Ghosal and van der Vaart (2017) our goal is to derive *posterior contraction rates* for our

suggested Bayesian methodologies : we wish to identify a sequence ε_n of positive real numbers satisfying

$$\Pi[\|f - f_0\|_n > \varepsilon_n | \mathbb{X}^n] \xrightarrow[n \rightarrow \infty]{P_0^\infty} 0$$

Deriving such rates typically requires to prove preliminary approximation results which is the object of section 3.3.

3.1 Priors

As done by Sanz-Alonso and Yang (2022); Fichera, Borovitskiy, Krause, and Billard (2024) we design our priors by random basis expansion in the graph Laplacian eigenbasis. Our construction depends on 2 degrees of freedom : the number of basis functions allowed in the sum, and the graph connectivity parameter h . This leads to a first family of priors :

Prior 1 Fixed J, h case:

We consider the prior $\Pi[\cdot | J, h]$ defined by the probability law of the random vector

$$f = \sum_{j=1}^J Z_j u_j, \quad Z_j \stackrel{i.i.d}{\sim} \Psi \quad (3)$$

for some $h > 0$, $J \in \{1, \dots, N\}$ and Ψ a probability distribution having a positive and continuous density ψ on \mathbb{R} satisfying $\int_{|x| > z} \psi(x) dx \leq e^{-b_1 z^{b_2}}$, for all $z \geq z_0$ for some $b_1, b_2, z_0 > 0$. In particular, Ψ can be Gaussian, leading to the construction of a Gaussian vector f on V , as done by Dunson, Wu, and Wu (2022); Sanz-Alonso and Yang (2022).

As we shall see in section 3.2.1, this prior leads to explicit posterior contraction rates which happen to be minimax optimal (up to logarithmic factors) under appropriate choice of the hyperparameters $J = J_n, h = h_n$ that depends on the unknown regularity β of f_0 together with the dimension d of \mathcal{M} . In this sense the proposed prior is *non adaptive*.

It is then of interest to choose the hyperparameters J, h in a data driven way. From the Bayesian point of view this can be done by putting an additional prior layer over J and h . For practical reasons, since for each value of h a new Laplacian \mathcal{L} needs to be computed, together with its eigenpairs (u_j, λ_j) we only consider discrete priors on h , i.e. we enforce $h \in \mathcal{H}$ where \mathcal{H} is a discrete subset of \mathbb{R}_+ , and $\pi_h(\cdot | J)$ denotes the prior probability mass function of h given J .

Prior 2 Adaptive prior:

We consider the prior Π defined by the probability law of the random vector

$$J \sim \pi_J, \quad h | J \sim \pi_h(\cdot | J), \quad f | J, h \sim \Pi[\cdot | J, h], \quad \text{denote } \mathcal{H}_J = \text{supp}(\pi_h(\cdot | J)) \subset \mathcal{H}, \quad (4)$$

where π_h is a probability distribution supported on \mathcal{H}_J (that may depends on J) and π_J is a probability distribution on \mathbb{N}^* satisfying

$$\forall j \geq j_0, \quad e^{-a_1 j^{L_j}} \leq \pi_J(j) \leq e^{a_2 j^{L_j}} \text{ where } L_j = 1 \text{ or } L_j = \ln j, \text{ and } \#\mathcal{H}_J \leq K_1 e^{K_2 J \ln n}$$

for some constants $a_1, a_2, K_1, K_2 > 0$.

Note that the joint prior on (J, h) can be of very different forms. For instance if $J \sim \mathcal{P}(\lambda)$ then the condition on π_J holds with $L_j = \log j$ while if π_J is a Geometric distribution,

then it holds with $L_j = 1$. Also we can choose for instance the support of $\pi_h(\cdot|J)$ to be $\mathcal{H}_J = \left\{ \frac{J^{-1/d}}{\ln^\kappa N} \right\}$ for some $\kappa > 2$ and $\mathcal{H} = \cup_J \mathcal{H}_J$, where here the dimension d of \mathcal{M} is assumed to be known. We will see that this leads to adaptive nearly minimax estimation rates. Alternatively and for more flexibility we can choose $\mathcal{H} = \{h_l := 2^l h_* : l = 0, \dots, L\}$ with $h_* = \left(\frac{\ln n}{n}\right)^{1/d}$, $\kappa > 2$ and $L \in \mathbb{N}$ such that $2^L h_* \leq 1 < 2^{L+1} h_*$, together with

$$\pi_h(h_l) \geq b_1 e^{-b_2 h_l^{-d}} \quad \forall h \in \mathcal{H}$$

for some constants $b_1, b_2 > 0$. We will detail the precise assumptions on Π_J , $\pi_h(\cdot|J)$ in Theorems 6 & 7.

Remark 1. Both **prior 1** and **prior 2**, depend on the covariates $x_{1:N}$ and possibly on the sample size n , but we keep the notation Π (and not $\Pi_{n,x}$) to avoid cumbersome notations.

Having defined the two types of priors we now study their posterior contraction properties.

3.2 Posterior contraction rates

We present here results on posterior contraction for **prior 1** and **prior 2**. We first present the non adaptive posterior contraction rate as it highlights the role of J and h and then we present the posterior contraction rate derived from prior 2. The proof is based on the general prior mass and testing theorem (see Ghosal and van der Vaart, 2007, 2017; Rousseau, 2016). A key aspect of the proof is a new scheme to approximate f_0 by functions in the span of u_1, \dots, u_J . This allows in particular to extend the results of Green, Balakrishnan, and Tibshirani (2021) to non linear manifolds when the regularity of f_0 not restricted to $\{1, 2, 3\}$. The approximation result is interesting in its own right and it allows also to derive minimax convergence rates for the frequentist estimator of Green, Balakrishnan, and Tibshirani (2021); hence it is presented in Section 3.3 and the convergence rate of the frequentist estimator is provided in Section 3.2.1. Moreover, the proof is very different from Green, Balakrishnan, and Tibshirani (2021) and is valid even when $N > n$.

3.2.1 NON ADAPTIVE RATES: PRIOR 1 AND PCR-LE ESTIMATOR

Throughout this section we assume that the chosen connectivity parameter h_n is not too small :

Assumption 2. The connectivity parameter h_n satisfies $h_n \rightarrow 0$ and $\frac{N h_n^d}{\ln N} \rightarrow \infty$.

Our first main result identifies a posterior contraction rate associated with prior 3 if h satisfies assumption 2 and J is not too small. In particular, we recover minimax optimal rates up to logarithmic factors for appropriate choices of parameters :

Theorem 3. Let (J_n, h_n) be a sequence of truncation and connectivity parameters satisfying $J_n \in \{1, \dots, N\}$, $J_n \geq \ln^\kappa N$, $\kappa > d$ and h_n satisfying assumption 2. Consider a prior on f belonging to the **Prior 1** family. Then there exists $C > 0$ such that

$$\Pi[\|f - f_0\|_n > C \varepsilon_n | \mathbb{X}^n, J_n, h_n] \xrightarrow[n \rightarrow \infty]{P_0^\infty} 0$$

where

$$\varepsilon_n = \varepsilon_n(J_n, h_n) = \sqrt{\frac{J_n \ln N}{n}} + \ln N^{\lceil \beta/2 \rceil} \max \left(1, \frac{\ln N}{J_n^{2/d} h_n^2} \right)^{\lceil \beta/2 \rceil} \left(h_n^\beta + \mathbb{1}_{\beta > 1} \left(\frac{\ln N}{N h_n^d} \right)^{1/2} h_n \right) \quad (5)$$

In particular, for any $\tau > d/2$ and

$$h_n = n^{-\frac{1}{2\beta+d}} (\ln n)^{-\frac{1-\tau-2(1+2\tau/d)\lceil \beta/2 \rceil}{2\beta+d}}, \quad J_n = \frac{h_n^{-d}}{\ln^\tau N}$$

then

$$\varepsilon_n \asymp (\ln n)^{\frac{(2\tau+d)\lceil \beta/2 \rceil + (1-\tau)\beta}{2\beta+d}} n^{-\frac{\beta}{2\beta+d}}$$

The proof is given in section D.1. Theorem 3 highlights the trade - off between the complexity of the prior (term $\sqrt{J \ln N/n}$) and the approximation error bounded by

$$\ln N^{\lceil \beta/2 \rceil} \max \left(1, \frac{\ln N}{J_n^{2/d} h_n^2} \right)^{\lceil \beta/2 \rceil} \left[h_n^\beta + \mathbb{1}_{\beta > 1} h_n \left(\frac{\ln N}{N h_n^d} \right)^{1/2} \right].$$

The first term of the approximation error has a bias flavour while the second, which appears only when $\beta > 1$ comes, roughly speaking, from stochastic deviations of $\mathcal{L}f_0 - \mathbb{E}(\mathcal{L}f_0)$, see Lemma 17. More discussion is given after Theorem 11.

Notice that theorem 3 only guarantees a posterior contraction with respect to the empirical L^2 norm $\|f - f_0\|_n$. Thus, it is not clear whether or not the posterior distribution correctly extrapolates outside of the design $\{x_1, \dots, x_n\}$ in order to estimate the values $\{f_0(x_{n+1}), \dots, f_0(x_N)\}$. This is the object of theorem 4 below, which requires stronger assumptions on J_n, h_n and β .

Theorem 4. Assume that $\beta > d/2$ and let h_n satisfying $h_n \geq (\ln N)^s n^{-\frac{1}{2\beta+d}}$ for some $s \in \mathbb{R}$ as well as $J = J_n \in \{1, \dots, N\}$, $\frac{h_n^{-d}}{\ln^{\tau_1} N} \leq J_n \leq \frac{h_n^{-d}}{\ln^{\tau_2} N}$, $\tau_1 \geq \tau_2 > 2d$. Then there exists $C > 0$ such that

$$\Pi[\|f - f_0\|_N > C\varepsilon_n | \mathbb{X}^n, J_n, h_n] \xrightarrow[n \rightarrow \infty]{P_0^\infty} 0$$

where

$$\varepsilon_n = \sqrt{\frac{J_n \ln N}{n}} + \left(\frac{J_n^{-2/d} \ln N}{h_n^2} \right)^{\lceil \beta/2 \rceil} \left(h_n^\beta + \mathbb{1}_{\beta > 1} \left(\frac{\ln N}{N h_n^d} \right)^{1/2} h_n \right)$$

In particular, choosing $J_n = \frac{h_n^{-d}}{\ln^\tau N}$, $\tau > 2d$ and $h_n = n^{-\frac{1}{2\beta+d}} (\ln N)^{-\frac{1-\tau-2\lceil \beta/2 \rceil(1+2\tau/d)}{2\beta+d}}$ leads to a posterior contraction rate of order

$$\varepsilon_n \asymp (\ln N)^{\frac{2\lceil \beta/2 \rceil(2\tau+d)+2\beta(1-\tau)}{2\beta+d}} n^{-\frac{\beta}{2\beta+d}}$$

The proof is provided in Section 4.4 and relies on Theorem 3 together with a concentration inequality to upper bound $\|f - f_0\|_N$ by (a multiple of) $\|f - f_0\|_n$.

The main limitation of **prior 1**, with deterministic J, h is that the choice of J, h crucially impacts the asymptotic behaviour of the posterior distribution. In particular to obtain optimal behaviour (with respect to $\|\cdot\|_n$ or $\|\cdot\|_N$) one needs to choose them as a function of β and d . While there exist simple and consistent estimators of d (see Berenfeld and Hoffmann, 2021), it is very difficult to estimate β .

It is interesting to compare our results with those of Sanz-Alonso and Yang (2022). In this work a Matérn type Gaussian process of the form $f = \sum_{j=1}^{J_N} (1 + \lambda_j)^{-\frac{\beta+d/2}{2}} Z_j u_j$, $Z_j \stackrel{i.i.d}{\sim} \mathcal{N}(0, 1)$ (also for some deterministic values of J_n, h_n) is used in place of prior 3. In Sanz-Alonso and Yang (2022) the author obtain the minimax rate of convergence with respect to the empirical L^2 norm $\|\cdot\|_n$ by choosing J, N, h accordingly but under a more restrictive assumption on N, β , implying in particular $N \gg n^{2d}$ and $\beta > d - 1/2$. Also, their proof technique is very different from ours as it is based on the convergence of the (discrete) Matérn Gaussian process prior to its continuous counterpart (associated to the Laplace-Beltrami operator on \mathcal{M}).

In the fully supervised $N = n$ case, our result can also be compared to the (frequentist) PCR-LE estimator of Green, Balakrishnan, and Tibshirani (2021). To be completely clear the estimator proposed in Green, Balakrishnan, and Tibshirani (2021) relies on the unnormalized graph Laplacian rather than the normalized one, but we believe our proof could be applied for the unnormalized graph Laplacian as well. The PCR-LE estimator is defined as

$$\hat{f} = \operatorname{argmin}_{f \in \operatorname{span}(u_1, \dots, u_J)} \|Y - f\|_2, \quad Y = (y_1, \dots, y_n), \quad f = (f(x_1), \dots, f(x_n))$$

Green, Balakrishnan, and Tibshirani (2021) prove that when f_0 belongs to a β -Sobolev class with $\beta \in \{1, 2, 3\}$ (where the Sobolev space is defined using the Laplacian operator weighted according to the covariates density), then \hat{f} converges to f_0 at the rate $n^{-\beta/(2\beta+d)}$, if J and h are chosen accordingly. In comparison we obtain a near minimax rate for any β -Hölder regularity assumption on f_0 , $\beta > 0$ (for the $\|\cdot\|_n$ loss) or $\beta > d/2$ (for the $\|\cdot\|_N$ loss). It should be noted that the restriction $\beta > d/2$ also appears in the continuous case when extending outside of the labelled points (see van der Vaart and van Zanten, 2011; Rosa et al., 2023).

As mentioned earlier, a key component of the proof of Theorems 3 & 4 is to show that f_0 can be well approximated by elements in the linear span of (u_1, \dots, u_J) , see Theorem 11. A consequence of Theorem 11 is thus that the PCR-LE estimator of Green, Balakrishnan, and Tibshirani (2021) achieves near minimax convergence rate over Hölder classes of any order β , thus extending the results of Green, Balakrishnan, and Tibshirani (2021).

Theorem 5. *Performance of PCR-LE for the random walk graph Laplacian*

Assume that $N = n$ and that the sequence (J_n, h_n) is such that $J_n \in \{1, \dots, n\}$, $J_n \geq \ln^\kappa N$, $\kappa > d$ and that $h_n > 0$ satisfies assumption 2. Define

$$\hat{f} = \sum_{j=1}^{J_n} \langle u_j | Y \rangle_{L^2(\nu)} u_j$$

Then for any $q \geq 0$ there exists $C > 0$ such that

$$\mathbb{E}_0 \left[\left\| \hat{f} - f_0 \right\|_n^q \right] \leq C \varepsilon_n^q$$

where

$$\varepsilon_n = \sqrt{\frac{J_n}{n}} + (\ln n)^{[\beta/2]} \max \left(1, \left(\frac{J_n^{-2/d} \ln n}{h_n^2} \right)^{[\beta/2]} \left(h_n^\beta + \mathbb{1}_{\beta > 1} \left(\frac{\ln n}{n h_n^d} \right)^{1/2} h_n \right) \right)$$

In particular for the choice $J_n = \frac{h_n^{-d}}{\ln^\tau n}$, $\tau > d/2$, $h_n = n^{-\frac{1}{2\beta+d}} (\ln n)^\rho$, $\rho \in \mathbb{R}$ then for any $q > 0$ there exists $C > 0$ such that

$$\mathbb{E}_0 \left[\left\| \hat{f} - f_0 \right\|_n^q \right] \leq C \varepsilon_n^q$$

where

$$\varepsilon_n = n^{-\frac{\beta}{2\beta+d}} (\ln n)^{(2[\beta/2](1+\tau/d)+\beta\rho) \vee (-\rho d/2-\tau/2)}$$

The proof of Theorem 5 is given in Section C.3. Corollary 5 hence shows that in the case $N = n$ the PCR-LE estimator of Green, Balakrishnan, and Tibshirani (2021) can achieve optimal rates for particular choice of hyperparameters (up to logarithmic factors) for any smoothness index $\beta > 0$ if one is willing to work on the Hölder spaces scale, and not just for β -Sobolev functions, $\beta \in \{1, 2, 3\}$: the limitations of the manifold adaptivity was therefore only an artefact of the proof, as the non-negligible discrepancy between geodesic and Euclidean distances can be circumvented by considering a different approximation technique. Extending 5 to an adaptive estimator by Lepski's method as done by Shi, Balasubramanian, and Polonik (2023) in the Euclidean domain could also be of interest, which then would provide a frequentist alternative to our adaptive Bayesian prior 4.

In the case $N > n$, our theorems 11 & 13 can also potentially be used to design frequentist estimators in alternative to our priors 3 & 4. Indeed, as an example Theorem 11 can be reformulated as a high probability upper bound on $\left\| f_0 - Q_t^{(k)} f_0 \right\|_{L^\infty(\nu)}$ where $Q_t^{(k)}$ is the kernel operator

$$\forall f : V \rightarrow \mathbb{R}, \left(Q_t^{(k)} f \right) (x) = \sum_{y \in V} \underbrace{\sum_{l=0}^k \frac{(t\mathcal{L})^l e^{-t\mathcal{L}}(x, y)}{l!}}_{=: Q_t^{(k)}(x, y) \nu_y} f(y)$$

This can be generalized by

$$\chi_t^{(k)}(x, y) = \sum_{l=0}^k \frac{(-t\mathcal{L})^l \chi^{(l)}(t\mathcal{L})(x, y)}{l!} = \sum_{l=0}^k \frac{1}{l!} \sum_{j=1}^N (-t\lambda_j)^l \chi^{(l)}(t\lambda_j) u_j(x) u_j(y)$$

for an arbitrary $\chi \in \mathcal{C}^k(\mathbb{R}_+, \mathbb{R})$ with $\chi(0) = 1$ which motivates the definition of kernel regression estimators of the form $\hat{f} = \sum_{i=1}^n \chi_t^{(k)}(\cdot, x_i) y_i \nu_{x_i}$: this is always tractable even when $N > n$ and analyzing the asymptotic properties of such estimators (even on other graph models) is therefore an interesting research direction.

3.2.2 RESULTS FOR THE ADAPTIVE PRIORS

Having proved Theorems 3 & 4 it is straightforward to derive adaptive posterior contraction rates under the **prior 2**. This results in a Bayesian method able to achieve minimax optimal contraction rates (up to logarithmic factors) with a data driven choice for J, h .

Theorem 6. *Consider the prior Π on f belonging to the class **Prior 2** as defined by (4) and assume that for some $K_1, K_2 > 0$ and some $j_0, a_1, a_2 > 0$*

$$\forall j \geq j_0, \quad e^{-a_1 j L_j} \leq \pi_J(j) \leq e^{-a_2 j L_j}, \quad \text{where } L_j = 1 \quad \text{or} \quad L_j = \ln j, \quad \text{and } \#\mathcal{H}_J \leq K_1 e^{K_2 J \ln n} \quad (6)$$

Assume in addition that for some $\tau > d/2$, $J_0, h_0 > 0$ and

$$J_n = J_0 (\ln N)^{\frac{2d(1+2\tau/d)[\beta/2]}{2\beta+d}} n^{\frac{d}{2\beta+d}}$$

$$h_n \in \mathcal{H}_{J_n} \cap \left[\frac{h_0 J_n^{-1/d}}{2 \ln^{\tau/d} N}, \frac{h_0 J_n^{-1/d}}{\ln^{\tau/d} N} \right] \text{ and } c > 0 \text{ such that}$$

$$\pi_h(h_n | J_n) \geq e^{-c h_n^{-d}}, \quad (7)$$

Then there exists $C > 0$ such that

$$\Pi [\|f - f_0\|_n > C \varepsilon_n |\mathbb{X}^n] \xrightarrow[n \rightarrow \infty]{P_0^\infty} 0, \quad \text{with } \varepsilon_n = n^{-\frac{\beta}{2\beta+d}} (\ln N)^{\frac{(2\tau+d)[\beta/2] - \beta(\tau+2\beta/d)}{2\beta+d}}.$$

For clarity, we remind the reader that in Equation 7, $\pi_h(h_n | J_n)$ denotes the prior probability mass function of h given $J = J_n$ evaluated at h_n .

The proof of Theorem 6 is given in Section D.2. Notice that even if some particular values of J_n, h_n are part of a condition that **prior 2** (4), must satisfy theoretically, the precise values of J_n, h_n are not part of the definition of the prior itself.

As discussed above in section 4, there are many priors on J, h which satisfy (7). Consider two cases of particular interest

- If J follows a Poisson or a Geometric prior and, given J , $h = \frac{J^{-1/d}}{\ln^{\tau/d} N}$, $\tau > d/2$ then condition (7) is fulfilled (with $\mathcal{H}_J = \left\{ \frac{J^{-1/d}}{\ln^{\tau/d} N} \right\}$).
- A more flexible alternative is to choose a discrete prior on $\mathcal{H} = \{2^l h_* : l = 0, \dots, L\}$ for $h_* = \left(\frac{\ln n}{n}\right)^{1/D}$ and L such that $2^L h_* \leq 1 < 2^{L+1} h_*$, together with a probability mass function of the order of $e^{-\lambda h^{-1}} h^a$ for arbitrary $\lambda > 0$ and $a \geq 0$. For instance, such a prior can be constructed by first considering $\tilde{h} \sim IG(a, \lambda)$, an inverse Gamma random variable and then defining $h = \sum_{l=0}^L 2^l h_* \mathbb{1}_{I_l}(\tilde{h})$, $I_0 = [0, h_*]$, $I_l =]2^l h_*, 2^{l+1} h_*]$ for $1 \leq l < L$ and $I_L =]2^L h_*, +\infty[$. In particular this prior is fully adaptive as it depends neither on β nor on d .

Similarly to the non adaptive prior, it is possible to obtain a posterior contraction rate in terms of the loss $\|\cdot\|_N$ from the result of Theorem 6, under additional assumptions on β and the prior on J, h .

Theorem 7. *Under the same conditions as in theorem 6 and if in addition we have $\mathcal{H}_J \subset [h_*, \frac{J^{-1/d}}{\ln^{\tau/d} N}]$ Π -almost surely for some $\tau > 2d$, $h_* > 0$ satisfying assumption 2 and if $\beta > d/2$ then there exists $M > 0$ such that*

$$\Pi [\|f - f_0\|_N > M\varepsilon_n | \mathbb{X}^n] \xrightarrow[n \rightarrow \infty]{P_0^\infty} 0$$

for the rate

$$\varepsilon_n = n^{-\frac{\beta}{2\beta+d}} (\ln N)^{\frac{(2\tau+d)[\beta/2]-\beta(\tau+2\beta/d)}{2\beta+d}}$$

Remark 8. *Theorem 7 thus shows that we can construct (near) minimax adaptive Bayesian procedures based on graph Laplacian decompositions, where the adaptivity is related to the (typically unknown) smoothness β . Similarly to theorem 6, many priors satisfy the assumptions of Theorem 7, for instance a Poisson or geometric prior on J can be chosen together with $h = \frac{J^{-1/d}}{\ln^{\tau/d} n}$, $\tau > d/2$. Alternatively, as discussed below Theorem 6, we can choose a discrete prior on $\mathcal{H}_J := \{2^l h_* : l = 0, \dots, L\}$ with $h_* = m_n \left(\frac{\ln n}{n}\right)^{1/D}$ and L such that $2^L h_* \leq \frac{J^{-1/d}}{\ln^{\tau/d} N} < 2^{L+1} h_*$, together with a probability mass function of the form of $e^{-\lambda h^{-d}} h^a$ for arbitrary $m_n \rightarrow \infty$, $\lambda > 0$ and $a \geq 0$.*

Compared to Theorem 6, we require $\mathcal{H}_J \subset [h_, \frac{J^{-1/d}}{\ln^{\tau/d} N}]$ for some h_* satisfying assumption 2 in order to apply Lemma 36 a posteriori (and importantly with constants that do not depend on h, J), see the detailed proof in Section 4.4. In fact the condition could be weakened to $\mathcal{H}_J \subset [h_*, \frac{J^{-1/d}}{\ln^{\tau/d} N}]$ on a set of posterior probability tending to 1, but this is not necessary for the examples of priors we provide.*

Remark 9. *The same condition $\mathcal{H}_J \subset [h_*, \frac{J^{-1/d}}{\ln^{\tau/d} N}]$ prevents us to design priors adaptive to d , contrariwise to Theorem 6. This condition is crucial in our setting in order to use Lemma 36. Whether or not such a restriction is necessary or is an artefact of our proof technique is therefore an interesting question. To design a procedure that is adaptive both to β and d one could plug in a consistent estimator of d (see Berenfeld and Hoffmann, 2021)—the procedure then becomes an empirical Bayes procedure—and our theoretical results remain unchanged since the estimator only depends on x and not on y .*

Remark 10. *Theorems 6 & 7 hold for discrete priors on h , however we believe that this is not a restriction since a continuous prior would imply that for every value of h , one would need to compute the eigenpairs (u_j, λ_j) of the Laplacian corresponding to h . Obtaining a result which would hold for a continuous prior on h , or for empirical Bayes procedure, where optimization with respect to h is performed, is not straightforward since it would necessitate to prove the concentration lemma 17 uniformly over $h \in (h_0, h_1) \varepsilon_n^{1/\beta}$.*

Notice that our posterior contraction rates in Theorems 3 4 6 5 & 7 are all impacted by a power of $\ln N$: this can be large, for instance if N grows exponentially with n . However under the mild assumption $N \leq n^B$, $B > 0$ this factor become a power of $\ln n$ which is typically considered small from an asymptotic viewpoint, ignoring the multiplicative constant. Extending our results to the case $N \gg n^B$ for all $B > 0$ would require more work, but we believe that in this case the proof strategy of Sanz-Alonso and Yang (2022) could be adapted.

3.3 Approximation result

All three theorems 3, 4 5, 7 & 6 are proved using approximation results of f_0 in the graph spectral domain, but since the smoothness assumption $f_0 \in \mathcal{C}^\beta(\mathcal{M})$ is only formulated at the continuous level this is a non trivial task. To prove that f_0 can be well approximated by an element of $\text{span}(u_1, \dots, u_J)$ we first find an approximation of f_0 (as a function on V) by an element of the form $e^{-t\mathcal{L}}f_t$ for some explicit f_t , with high probability with respect to the sampling distribution of the covariates $x_{1:N}$. Then we show that the projection of $e^{-t\mathcal{L}}f_t$ onto $\text{span}(u_1, \dots, u_J)$ is close to $e^{-t\mathcal{L}}f_t$ and thus to f_0 .

Theorem 11. *Let $\beta > 0$, $f_0 \in \mathcal{C}^\beta(\mathcal{M})$, assume that h_n satisfies assumption 2 and define for $t > 0$*

$$f_t = \sum_{l=0}^k \frac{1}{l!} (t\mathcal{L})^l f_0, \quad f_t : V \rightarrow \mathbb{R},$$

where $k = \lceil \beta/2 \rceil - 1$. Then

1. *If $0 < \beta \leq 1$, there exists a constant $C(\beta, f_0)$ we have*

$$\|f_0 - e^{-t\mathcal{L}}f_t\|_{L^\infty(\nu)} \leq C(\beta, f_0) t h_n^{\beta-2}$$

2. *if $\beta > 1$, $p_0 \in \mathcal{C}^{\beta-1}(\mathcal{M})$ then for any $H > 0$ there exists a constant $C(H, \beta, f_0, p_0, \mathcal{M})$, such that*

$$\mathbb{P}_0 \left(\|f_0 - e^{-t\mathcal{L}}f_t\|_{L^\infty(\nu)} > C(H, \beta, f_0, p_0, \mathcal{M}) \left(\frac{t}{h_n^2} \right)^{k+1} \left[h_n^\beta + \left(\frac{\ln N}{N h_n^d} \right)^{1/2} h_n \right] \right) \leq N^{-H}$$

The proof of Theorem 11 is given in Section 4.1.

Interestingly, when $\beta > 1$,

$$h_n^\beta \geq h_n \sqrt{\frac{\ln N}{N h_n^d}}, \quad \Leftrightarrow h_n \geq \left(\frac{\ln N}{N} \right)^{\frac{1}{2(\beta-1)+d}}$$

which is—up to the $\ln N$ term—the condition required by Theorem 6 in Green, Balakrishnan, and Tibshirani (2021) in the case $N = n$. Although the proof techniques are very different, in both cases this condition is used to bound deviations of $\mathcal{L}f$ around its mean.

Remark 12. *The proof of Theorem 11 relies heavily on the control of quantities of the form $\|\mathcal{L}^k f_0\|_{L^\infty(\nu)}$, $k < \lceil \beta/2 \rceil$, which we obtain by smooth local parametrisations of \mathcal{M} . Extensions to possibly non smooth compact Dirichlet spaces as in Coulhon et al. (2012) might be possible but would require other arguments to control $\mathcal{L}^k f_0$ assuming as in Coulhon et al. (2012) a Besov $B_{\infty\infty}^\beta$ smoothness on f_0 , which is a natural extension of our Hölder classes $\mathcal{C}^\beta(\mathcal{M})$ for Dirichlet spaces.*

To prove that f_0 can be well approximated by a function of the form $f = \sum_{j=1}^{J_n} z_j u_j$ for some choice of truncation J_n , we now show that $e^{-t\mathcal{L}}f_t$ is close to $p_{J_n}(e^{-t\mathcal{L}}f_t)$, its $L^2(\nu)$ projection onto $\text{span}(u_1, \dots, u_{J_n})$. Since we are able to lower bound the eigenvalues of \mathcal{L} with high probability, thanks to the exponential factor $e^{-t\mathcal{L}}$ this is actually straightforward.

Theorem 13. *Let $\beta > 0$, $f_0 \in \mathcal{C}^\beta(\mathcal{M})$, h_n satisfying assumption 2 and $J_n \in \{1, \dots, N\}$ satisfying $J_n \geq \ln^\kappa N$, $\kappa > d$. Let $f_t : V \rightarrow \mathbb{R}$ be defined by*

$$f_t = \sum_{l=0}^k \frac{1}{l!} (t\mathcal{L})^l f_0,$$

where $k = \lceil \beta/2 \rceil - 1$. Then there exists $c > 0$ such that with $t = c\lambda_{J_n}^{-1} \ln N$ and $\tilde{\varepsilon}_n(J_n, h_n)$ the rate

$$\tilde{\varepsilon}_n(J_n, h_n) = (\ln N)^{\lceil \beta/2 \rceil} \max \left(1, \frac{J_n^{-2/d} \ln N}{h_n^2} \right)^{k+1} \left(h_n^\beta + \mathbb{1}_{\beta > 1} \left(\frac{\ln N}{N h_n^d} \right)^{1/2} h_n \right)$$

we have

1. If $0 < \beta \leq 1$: there exists a constant $C(\beta, f_0, \rho)$ such that

$$\|f_0 - p_{J_n}(e^{-t\mathcal{L}} f_t)\|_{L^\infty(\nu)} \leq C(\beta, f_0) \tilde{\varepsilon}_n(J_n, h_n)$$

2. If $\beta > 1$, $p_0 \in \mathcal{C}^{\beta-1}(\mathcal{M})$, $H > 0$: for some constant $C(H, \beta, p_0, f_0, \mathcal{M}, \rho)$

$$\mathbb{P}_0 \left(\|f_0 - p_{J_n}(e^{-t\mathcal{L}} f_t)\|_{L^\infty(\nu)} \geq C(H, \beta, p_0, f_0, \mathcal{M}) \tilde{\varepsilon}_n(J_n, h_n) \right) \leq N^{-H}$$

The proof of Theorem 13 is given in Section 4.2.

Remark 14. *As in Remark 12, extending Theorem 13 to more general spaces \mathcal{M} such as Dirichlet spaces is of interest and would require an alternative proof on the lower bound of the eigenvalues of \mathcal{L} (see Lemma 34).*

Both theorem 11 & 13 are useful results in their own rights, and can be used to design and study the convergence properties of frequentist estimators, as we have done in Theorem 5.

4. Main proofs

In this Section we present the proofs of Theorems 11 and 13. To control the approximation of f_0 by $e^{-t\mathcal{L}} f_t$ we introduce a deterministic approximation of the operator $\mathcal{L}f$. More precisely, recall that

$$\mathcal{L}f = \frac{1}{\mu_x h^2} \sum_{y \sim x} f(x) - f(y).$$

By analogy this leads us to the following definition :

Definition 15. *For every $f \in \mathcal{C}(\mathcal{M})$ and $h > 0$ we define*

$$T_h f(x) = \frac{1}{h^2 P_0(B_{\mathbb{R}^D}(x, h))} \int_{B_{\mathbb{R}^D}(x, h) \cap \mathcal{M}} (f(x) - f(y)) p_0(y) \mu(dy)$$

where $P_0(B_{\mathbb{R}^D}(x, h)) = \int_{B_{\mathbb{R}^D}(x, h) \cap \mathcal{M}} p_0(y) \mu(dy)$.

T_h acts as a second order nonlocal differential operator and is a deterministic approximation of the operator \mathcal{L} as shown in Lemma 17 below.

4.1 Proof of theorem 11

By a Taylor expansion of $t \mapsto e^{-\lambda t}$ (with $\lambda \geq 0$) we have

$$1 = e^{-\lambda t} \sum_{l=0}^k \frac{(\lambda t)^l}{l!} + R_k(t)$$

where the remainder is given by

$$R_k(t) = \frac{1}{k!} \int_0^1 (\lambda x t)^{k+1} e^{-\lambda x t} \frac{dx}{x}$$

Hence by (finite dimensional) functional calculus we find

$$f_0 = e^{-t\mathcal{L}} f_t + \frac{1}{k!} \int_0^1 (xt\mathcal{L})^{k+1} e^{-xt\mathcal{L}} f_0 \frac{dx}{x}$$

Therefore, for all $x \in V$, using $\|e^{-t\mathcal{L}}\|_{L^\infty(\nu)} = 1$, we bound

$$\begin{aligned} \|f_0 - e^{-t\mathcal{L}} f_t\|_{L^\infty(\nu)} &\leq \frac{1}{k!} \int_0^1 \left\| (xt\mathcal{L})^{k+1} e^{-xt\mathcal{L}} f_0 \right\|_{L^\infty(\nu)} \frac{dx}{x} = \frac{1}{k!} \int_0^t \left\| (s\mathcal{L})^{k+1} e^{-s\mathcal{L}} f_0 \right\|_{L^\infty(\nu)} \frac{ds}{s} \\ &\leq \frac{1}{k!} \int_0^t \left\| (s\mathcal{L})^{k+1} f_0 \right\|_{L^\infty(\nu)} \frac{ds}{s} = \frac{\left\| (t\mathcal{L})^{k+1} f_0 \right\|_{L^\infty(\nu)}}{(k+1)!}. \end{aligned}$$

We now bound $\left\| (t\mathcal{L})^{k+1} f_0 \right\|_{L^\infty(\nu)}$, considering the three cases: $\beta \leq 1$, $1 < \beta \leq 2$ and $\beta > 2$.

- If $\beta \leq 1$ then the result is trivial as

$$|(t\mathcal{L}) f_0(x)| = t \left| \frac{1}{h_n^2 \mu_x} \sum_{y \sim x} f_0(x) - f_0(y) \right| \lesssim \|f_0\|_{\mathcal{C}^\beta(\mathcal{M})} t h_n^{\beta-2}.$$

- If $1 < \beta \leq 2$, we have

$$\|t\mathcal{L}f_0\|_{L^\infty(\nu)} \leq t \|\mathcal{L}f_0 - T_{h_n}f_0\|_{L^\infty(\nu)} + t \|T_{h_n}f_0\|_{L^\infty(\nu)}.$$

Then lemma 16 implies

$$\|T_{h_n}f_0\|_{L^\infty(\mathcal{M})} \leq C(\beta, p_0) \|f_0\|_{\mathcal{C}^\beta(\mathcal{M})} h_n^{\beta-2}$$

and Lemma 17 implies that for any $H > 0$ there exists $M_0 > 0$ such that

$$\mathbb{P}_0 \left(\|\mathcal{L}f_0 - T_{h_n}f_0\|_{L^\infty(\nu)} > M_0 \left(\frac{\ln N}{N} \right)^{1/2} h_n^{-(1+d/2)} \right) \leq N^{-H}.$$

Hence, for some $C(H, \beta, f_0, p_0) > 0$

$$\mathbb{P}_0 \left(\|h_n^2 \mathcal{L}f_0\|_{L^\infty(\nu)} > C(H, \beta, p_0, f_0) \left(\left(\frac{\ln N}{N} \right)^{1/2} h_n^{1-d/2} + h_n^\beta \right) \right) \leq N^{-H}.$$

- When $\beta > 2$ we use an inductive argument. We have shown that for $k = \lceil \beta/2 \rceil - 1 = 0$, if $f \in \mathcal{C}^\beta$,

$$\|h_n^2 \mathcal{L}f\|_{L^\infty(\nu)} \leq C_0(H, \beta, f_0, p_0, \mathcal{M})[h_n^\beta + \Delta_n(f)]$$

where $\mathbb{P}_0 \left(|\Delta_n(f)| > M_0 \left(\frac{\ln N}{N h_n^d} \right)^{1/2} h_n \right) \leq N^{-H}$. Assume that for all $k' < k$, if $f \in \mathcal{C}^{\beta'}$ with $2k' < \beta' \leq 2(k' + 1)$, then

$$\|(h_n^2 \mathcal{L}f)^{k'+1}\|_{L^\infty(\nu)} \leq C_{k'}(H, \beta, p_0, f_0, \mathcal{M})[h_n^{\beta'} + \Delta_{n,k'}(f)]$$

where $\mathbb{P}_0 \left(|\Delta_{n,k'}(f)| > M_0 \left(\frac{\ln N}{N h_n^d} \right)^{1/2} h_n \right) \leq N^{-H}$. We prove that the same holds for k . Let $f \in \mathcal{C}^\beta$ with $2k < \beta \leq 2(k + 1)$. Since $\beta > 2$ by lemma 16 we find f_1, \dots, f_k (that depend on h_n) such that for all $l \in \{1, \dots, k\}$

$$\|f_l\|_{\mathcal{C}^{\beta-2l}(\mathcal{M})} \lesssim \|f\|_{\mathcal{C}^\beta(\mathcal{M})}, \quad \left\| T_{h_n} f - \sum_{l=1}^k h_n^{2(l-1)} f_1^{(l)} \right\|_{L^\infty(\mathcal{M})} \leq C_l(p_0, \mathcal{M}) \|f\|_{\mathcal{C}^\beta(\mathcal{M})} h_n^{\beta-2}$$

Using $\|h_n^2 \mathcal{L}\|_{L^\infty(\nu)} \leq 2$ which holds since for any $f \in \mathbb{R}^V$ and $x \in V$ we have

$$|(h_n^2 \mathcal{L}f)(x)| = \left| \frac{1}{\mu_x} \sum_{y \sim x} f(x) - f(y) \right| \leq \frac{2 \|f\|_{L^\infty(\nu)}}{\mu_x} \# \{y : y \sim x\} = 2 \|f\|_{L^\infty(\nu)},$$

we get

$$\begin{aligned} & \left\| (h_n^2 \mathcal{L})^{k+1} f_0 \right\|_{L^\infty(\nu)} \\ &= h_n^2 \left\| (h_n^2 \mathcal{L})^k (T_{h_n} f_0 + \mathcal{L}f_0 - T_{h_n} f_0) \right\|_{L^\infty(\nu)} \\ &\leq h_n^2 \left\| (h_n^2 \mathcal{L})^k T_{h_n} f_0 \right\|_{L^\infty(\nu)} + h_n^2 \left\| (h_n^2 \mathcal{L})^k (\mathcal{L}f_0 - T_{h_n} f_0) \right\|_{L^\infty(\nu)} \\ &\lesssim h_n^2 \left\| (h_n^2 \mathcal{L})^k T_{h_n} f_0 \right\|_{L^\infty(\nu)} + h_n^2 \|\mathcal{L}f_0 - T_{h_n} f_0\|_{L^\infty(\nu)} \\ &= h_n^2 \left\| (h_n^2 \mathcal{L})^k \left(\sum_{l=1}^k h_n^{2(l-1)} f_l + T_{h_n} f_0 - \sum_{l=1}^k h_n^{2(l-1)} f_l \right) \right\|_{L^\infty(\nu)} + h_n^2 \|\mathcal{L}f_0 - T_{h_n} f_0\|_{L^\infty(\nu)} \\ &\leq h_n^2 \sum_{l=1}^k h_n^{2(l-1)} \left\| (h_n^2 \mathcal{L})^k f_l \right\|_{L^\infty(\nu)} + h_n^{2k} C_l(p_0, \mathcal{M}) \|f\|_{\mathcal{C}^\beta(\mathcal{M})} h_n^{\beta-2} + \Delta_n(f_0). \end{aligned}$$

Now,

$$\begin{aligned} \left\| (h_n^2 \mathcal{L})^k f_l \right\|_{L^\infty(\nu)} &\leq 2^{l-1} \left\| (h_n^2 \mathcal{L})^{k-l+1} f_l \right\|_{L^\infty(\nu)} \\ &\leq C_{k-l}(H, \beta, p_0, f_0, \mathcal{M})[h_n^{\beta-2l} + \Delta_{n,k-l}(f)] \end{aligned}$$

so that for all $f \in \mathcal{C}^\beta$

$$\begin{aligned} \left\| (h_n^2 \mathcal{L})^{k+1} f \right\|_{L^\infty(\nu)} &\leq h_n^2 \sum_{l=1}^k C_{k-l} (\|f\|_{\mathcal{C}^\beta}, p_0, \mathcal{M}) h_n^{2(l-1)} h_n^{\beta-2l} \\ &\quad + \sum_{l=0}^k C_{k-l} (\|f\|_{\mathcal{C}^\beta}, p_0, \mathcal{M}) \Delta_{n,k-l}(f_0) \\ &\leq C'(\beta, f_0, p_0, \mathcal{M}) [h_n^\beta + \Delta_{n,k}(f)] \end{aligned}$$

where $\mathbb{P}_0 \left(|\Delta_{n,k}(f)| > M_0 \left(\frac{\ln N}{N h_n^d} \right)^{1/2} h_n \right) \leq N^{-H}$ for M_0 large enough. This concludes the inductive argument. Therefore

$$\|f_0 - e^{-t\mathcal{L}} f_t\|_{L^\infty(\nu)} \leq C'(H, \beta, f_0, p_0, \mathcal{M}) \left(\frac{t}{h_n^2} \right)^{k+1} [h_n^\beta + \Delta_{n,k}(f_0)],$$

where $\mathbb{P}_0 \left(|\Delta_{n,k}(f_0)| > M_0 \left(\frac{\ln N}{N h_n^d} \right)^{1/2} h_n \right) \leq N^{-H}$.

4.2 Proof of Theorem 13

We bound $\|f_0 - p_{J_n}(e^{-t\mathcal{L}} f_t)\|_{L^\infty(\nu)}$ by showing that $\|e^{-t\mathcal{L}} f_t - p_{J_n}(e^{-t\mathcal{L}} f_t)\|_{L^\infty(\nu)}$ is small, using a lower bound on the eigenvalues λ'_j s given by lemma 34 : since h_n satisfies assumption 2, choosing $h_- = h_n$ in the definition of A_N , together with Lemma 34, we have that $\lambda_j^{(h_n)} \geq b_3 j^{2/d}$ for all $b_1 \ln^d N \leq j \leq t_0(h_n) = b \frac{h_n^{-d}}{\ln^{d/2} N}$, so that since $\ln^\kappa N \leq J$ with $\kappa > d$, we have $\lambda_J = \lambda_J^{(h_n)} \geq b_3 J^{2/d}$ when $J \leq b \frac{h_n^{-d}}{\ln^{d/2} N} := \bar{J}_n$.

Moreover, because the λ'_j s are ordered we have

$$\forall J \geq \bar{J}_n, \quad \lambda_J \geq \lambda_{\lfloor \bar{J}_n \rfloor} \geq b_3 b^{2/d} \frac{h_n^{-2}}{\ln N}$$

Therefore

$$\begin{aligned} \|f_0 - p_J(e^{-t\mathcal{L}} f_t)\|_{L^\infty(\nu)} &\leq \|f_0 - e^{-t\mathcal{L}} f_t\|_{L^\infty(\nu)} + \left\| \sum_{j>J} \langle u_j | e^{-t\mathcal{L}} f_t \rangle_{L^2(\nu)} u_j \right\|_{L^\infty(\nu)} \\ &= \|f_0 - e^{-t\mathcal{L}} f_t\|_{L^\infty(\nu)} + \left\| \sum_{j>J} e^{-t\lambda_j} \langle u_j | f_t \rangle_{L^2(\nu)} u_j \right\|_{L^\infty(\nu)} \\ &\leq \|f_0 - e^{-t\mathcal{L}} f_t\|_{L^\infty(\nu)} + \sum_{j>J} e^{-t\lambda_j} \left| \langle u_j | f_t \rangle_{L^2(\nu)} \right| \|u_j\|_{L^\infty(\nu)} \\ &\leq \|f_0 - e^{-t\mathcal{L}} f_t\|_{L^\infty(\nu)} + \|f_t\|_{L^2(\nu)} e^{-t\lambda_J} \sum_{j>J} \|u_j\|_{L^\infty(\nu)} \end{aligned}$$

By lemma 20 we have $\max_{1 \leq j \leq N} \|u_j\|_{L^\infty(\nu)} \leq N$ and therefore

$$\|f_0 - p_J(e^{-t\mathcal{L}} f_t)\|_{L^\infty(\nu)} \lesssim \|f_0 - e^{-t\mathcal{L}} f_t\|_{L^\infty(\nu)} + N^2 \|f_t\|_{L^2(\nu)} e^{-t\lambda_J}$$

Moreover, since $f_t = \sum_{l=0}^k \frac{1}{l!} (t\mathcal{L})^l f_0$ and $\|h_n^2 \mathcal{L}\|_{\mathcal{L}(L^\infty(\nu))} \leq 2$ we get

$$\|f_t\|_{L^\infty(\nu)} \leq e^2 \left(\frac{t}{h_n^2} \right)^k \|f_0\|_{L^\infty(\mathcal{M})},$$

which implies

$$\|f_0 - p_J(e^{-t\mathcal{L}} f_t)\|_{L^\infty(\nu)} \lesssim \|f_0 - e^{-t\mathcal{L}} f_t\|_{L^\infty(\nu)} + N^2 e^2 \left(\frac{t}{h_n^2} \right)^k e^{-t\lambda_J}.$$

Thus for any $K > 0$, choosing $t = c\lambda_J^{-1} \ln N$ for $c > 0$ large enough yields

$$\|f_0 - p_J(e^{-t\mathcal{L}} f_t)\|_{L^\infty(\nu)} \lesssim \|f_0 - e^{-t\mathcal{L}} f_t\|_{L^\infty(\nu)} + N^{-K}.$$

Finally we bound $\|f_0 - e^{-t\mathcal{L}} f_t\|_{L^\infty(\nu)}$ using Theorem 11 which gives us for $K > 0$ large enough and arbitrarily large $H > 0$

$$\mathbb{P}_0(\|f_0 - p_J(e^{-t\mathcal{L}} f_t)\| > C(H, \beta, f_0, p_0, \mathcal{M}) \tilde{\varepsilon}_n) \leq N^{-H}$$

where

$$\begin{aligned} \tilde{\varepsilon}_n &= \left(\frac{t}{h_n^2} \right)^{[\beta/2]} \left(h_n^\beta + \mathbb{1}_{\beta > 1} \left(\frac{\ln N}{N h_n^d} \right)^{1/2} h_n \right) \lesssim \left(\frac{\lambda_J^{-1} \ln N}{h_n^2} \right)^{[\beta/2]} \left(h_n^\beta + \mathbb{1}_{\beta > 1} \left(\frac{\ln N}{N h_n^d} \right)^{1/2} h_n \right) \\ &\lesssim \begin{cases} \left(\frac{J^{-2/d} \ln N}{h_n^2} \right)^{[\beta/2]} \left(h_n^\beta + \mathbb{1}_{\beta > 1} \left(\frac{\ln N}{N h_n^d} \right)^{1/2} h_n \right) & \text{if } J \leq \bar{J}_n \\ (\ln N)^{2[\beta/2]} \left(h_n^\beta + \mathbb{1}_{\beta > 1} \left(\frac{\ln N}{N h_n^d} \right)^{1/2} h_n \right) & \text{if } J > \bar{J}_n \end{cases} \end{aligned}$$

which concludes the proof.

4.3 Lemmas 16 and 17

Below we present two Lemmas which control the terms $\Delta_n(f)$ together with the behaviour of $T_h f$.

Lemma 16. *Let $a_{\mathcal{M}}, C_0, h_+$ the constants defined in Appendix A and B. Then*

1. *If $0 < \beta \leq 2$, $h \leq \frac{\pi a_{\mathcal{M}}}{C_0}$, $f \in \mathcal{C}^\beta(\mathcal{M})$ (with the additional assumption $p_0 \in \mathcal{C}^{\beta-1}(\mathcal{M})$ if $1 < \beta \leq 2$), then $\|T_h f\|_{L^\infty(\mathcal{M})} \lesssim \|f\|_{\mathcal{C}^\beta(\mathcal{M})} h^{\beta-2}$*
2. *If $\beta > 2$, $f \in \mathcal{C}^\beta(\mathcal{M})$, $p_0 \in \mathcal{C}^{\beta-1}(\mathcal{M})$, $h \leq h_+$, then there exists $\left(g_h^{(l)}\right)_{l=1}^k$, $g_h^{(l)} \in \mathcal{C}^{\beta-2l}(\mathcal{M})$ such that*

$$\left\| g_h^{(l)} \right\|_{\mathcal{C}^{\beta-2l}(\mathcal{M})} \lesssim \|f\|_{\mathcal{C}^\beta(\mathcal{M})}, \quad \left\| T_h f - \sum_{l=1}^k h^{2(l-1)} g_h^{(l)} \right\|_{L^\infty(\mathcal{M})} \lesssim \|f\|_{\mathcal{C}^\beta(\mathcal{M})} h^{\beta-2}$$

Lemma 17. *Let $h_n > 0$ satisfying assumption 2. Then*

- If $f \in \mathcal{C}^\beta(\mathcal{M})$, $0 < \beta \leq 1$ we have

$$\|\mathcal{L}f\|_{L^\infty(\nu)} \lesssim h_n^{\beta-2}, \|T_{h_n}f\|_{L^\infty(\mathcal{M})} \lesssim h_n^{\beta-2}$$

- If $f \in \mathcal{C}^1(\mathcal{M})$ then any $H > 0$ there exists $M_0 > 0$ such that

$$\mathbb{P}_0 \left(\|\mathcal{L}f - T_{h_n}f\|_{L^\infty(\nu)} > M_0 \left(\frac{\ln N}{N} \right)^{\frac{1}{2}} h_n^{-(1+d/2)} \right) \leq N^{-H}$$

The proofs of both lemmas are provided in Sections C.1 and C.2 respectively.

4.4 Proof of Theorems 4 and 7

We now show how to extend the posterior contraction rates in terms of $\|f - f_0\|_n$ obtained in Theorems 3 and 6 to rates in terms of $\|f - f_0\|_N$.

As in van der Vaart and van Zanten (2011); Rosa, Borovitskiy, Terenin, and Rousseau (2023) where a posterior contraction rate with respect to the empirical $\|f - f_0\|_n$ norm is used to show a contraction rate with respect to the continuous L^2 norm, we will use a concentration inequality : indeed, by exchangeability the variables $(x_i)_{i=1}^n$ can be considered as sampled uniformly without replacement from the full sample $(x_i)_{i=1}^N$, and therefore the quantity $\|f - f_0\|_n^2 = \frac{1}{n} \sum_{i=1}^n (f(x_i) - f_0(x_i))^2$ is close to its mean $\frac{1}{N} \sum_{i=1}^N (f(x_i) - f_0(x_i))^2 = \|f - f_0\|_N^2$. We make this informal argument rigorous below.

We start by the non adaptive case, i.e the proof of theorem 4. Since $h_n \geq (\ln N)^s n^{-\frac{1}{2\beta+d}}$, it satisfies assumption 2, and since $\frac{h_n^{-d}}{\ln^{\tau_1} N} \leq J_n \leq \frac{h_n^{-d}}{\ln^{\tau_2} N}$, $\tau_1 \geq \tau_2 > 2d$, Theorem 3 implies that, for some $C > 0$

$$\Pi[\|f - f_0\|_n > C\varepsilon_n(J_n, h_n) | \mathbb{X}^n, J_n, h_n] \xrightarrow[n \rightarrow \infty]{P_0^\infty} 0$$

where $\varepsilon_n(J_n; h_n)$ is given by (5). Note that the choice of J_n implies that

$$\varepsilon_n(J_n, h_n) = \sqrt{\frac{J_n \ln N}{n}} + \left(\frac{J_n^{-2/d} \ln N}{h_n^2} \right)^{[\beta/2]} \left(h_n^\beta + \mathbb{1}_{\beta > 1} \left(\frac{\ln N}{N h_n^d} \right)^{1/2} h_n \right)$$

Let

$$f_n = p_{J_n}(e^{-t_n \mathcal{L}} f_{t_n}), \quad f_{t_n} = \sum_{l=0}^{[\beta/2]-1} \frac{(t_n \mathcal{L})^l f_0}{l!}, \quad t_n = c \lambda_{J_n}^{-1} \ln N, c > 0.$$

Then Theorem 13 implies that for some $c > 0$ and any $H > 0$, choosing C large enough,

$$\mathbb{P}_0 \left(\|f_n - f_0\|_{L^\infty(\nu)} \leq C\varepsilon_n(J_n, h_n) \right) = 1 + O(N^{-H}).$$

Now if $M > C$, then on $\{\|f_n - f_0\|_{L^\infty(\nu)} \leq C\varepsilon_n\} = V_n$,

$$\Pi[\|f - f_0\|_N > M\varepsilon_n(J_n, h_n) | \mathbb{X}^n, J_n, h_n] \tag{8}$$

$$\begin{aligned} &\leq \Pi[\|f - f_n\|_N > (M - C)\varepsilon_n(J_n, h_n) | \mathbb{X}^n, J_n, h_n] \\ &\leq \Pi[\|f - f_n\|_n > (C + 1)\varepsilon_n(J_n, h_n) | \mathbb{X}^n, J_n, h_n] \end{aligned} \tag{9}$$

$$+ \Pi \left[\|f - f_n\|_N > (M - C)\varepsilon_n(J_n, h_n) \geq \frac{M - C}{C + 1} \|f - f_n\|_n | \mathbb{X}^n, J_n, h_n \right]. \tag{10}$$

Since

$$\|f - f_n\|_n \leq \|f - f_0\|_n + \|f_0 - f_n\|_{L^\infty(\nu)} \leq \|f - f_0\|_n + C\varepsilon_n(J_n, h_n),$$

we have on V_n

$$\Pi[\|f - f_n\|_n > (C+1)\varepsilon_n(J_n, h_n) | \mathbb{X}^n, J_n, h_n] \leq \Pi[\|f - f_0\|_n > \varepsilon_n(J_n, h_n) | \mathbb{X}^n, J_n, h_n] = o_{\mathbb{P}_0}(1).$$

Hence it remains to bound the second term of the right hand side of (8). For this notice that a consequence of the proof of Theorem 37 (see Ghosal and van der Vaart, 2007) is that there exists $c' > 0$ such that

$$\mathbb{P}_0 \left[D_n \leq e^{-c'n\varepsilon_n(J_n, h_n)^2} \right] = o(1), \quad D_n = \int_{\Sigma^{J_n}} e^{\ell_n(f) - \ell_n(f_0)} d\Pi(f | J_n, h_n)$$

where $\ell_n(f)$ is the log-likelihood at f . As a consequence,

$$\begin{aligned} & \mathbb{E}_0 \left[\Pi \left[\|f - f_n\|_N \geq \frac{M-C}{C+1} \|f - f_n\|_n \mid \mathbb{X}^n, J_n, h_n \right] \right] \\ & \leq o(1) + e^{c'n\varepsilon_n^2(J_n, h_n)} \mathbb{E}_0 \left[\int_{\Sigma^{J_n}} \mathbb{1}_{\|f - f_n\|_N > \frac{M-C}{C+1} \|f - f_n\|_n} \mathbb{1}_{\|f - f_n\|_n < (C+1)\varepsilon_n} e^{\ell_n(f) - \ell_n(f_0)} d\Pi(f | J_n, h_n) \right] \\ & \leq o(1) + e^{c'n\varepsilon_n^2(J_n, h_n)} \mathbb{E}_0 \int_{f \in \Sigma^{J_n}} \mathbb{1}_{\|f - f_n\|_N > \frac{M-C}{C+1} \|f - f_n\|_n} \Pi(df) \end{aligned}$$

where the last expectation is now only with respect to the distribution of $x_{1:N}$. Notice that by exchangeability of the x_i 's, for any permutation $\tau \in \mathcal{S}_N$ the set of permutations of $\{1, \dots, N\}$ we have $(x_i)_{i=1}^N \stackrel{(d)}{=} (x_{\tau(i)})_{i=1}^N$ and therefore, writing $\|f - f_n\|_{\tau, n}^2 = \frac{1}{n} \sum_{i=1}^n (f(x_{\tau(i)}) - f_n(x_{\tau(i)}))^2$,

$$\mathbb{E}_0 \int_{f \in \Sigma^{J_n}} \mathbb{1}_{\|f - f_n\|_N > \frac{M-C}{C+1} \|f - f_n\|_n} \Pi(df) = \mathbb{E}_0 \int_{f \in \Sigma^{J_n}} \mathbb{1}_{\|f - f_n\|_N > \frac{M-C}{C+1} \|f - f_n\|_{\tau, n}} \Pi(df).$$

Hence

$$\begin{aligned} & \mathbb{E}_0 \int_{f \in \Sigma^{J_n}} \mathbb{1}_{\|f - f_n\|_N > \frac{M-C}{C+1} \|f - f_n\|_n} \Pi(df) \\ & = \frac{1}{N!} \sum_{\tau \in \mathcal{S}_N} \mathbb{E}_0 \int_{f \in \Sigma^{J_n}} \mathbb{1}_{\|f - f_n\|_N > \frac{M-C}{C+1} \|f - f_n\|_{\tau, n}} \Pi(df) \\ & = \mathbb{E}_{\tau \sim \mathcal{U}(\mathcal{S}_N)} \mathbb{E}_0 \int_{f \in \Sigma^{J_n}} \mathbb{1}_{\|f - f_n\|_N > \frac{M-C}{C+1} \|f - f_n\|_{\tau, n}} \Pi(df) \\ & = \mathbb{E}_0 \int_{f \in \Sigma^{J_n}} \mathbb{P}_{\tau \sim \mathcal{U}(\mathcal{S}_N)} \left(\|f - f_n\|_N > \frac{M-C}{C+1} \|f - f_n\|_{\tau, n} \right) \Pi(df). \end{aligned}$$

Define for $i = 1, \dots, N$, $W_i = (f(x_i) - f_n(x_i))^2$ and $W'_i = (f(x_{\tau(i)}) - f_n(x_{\tau(i)}))^2$. Then given $(x_i)_{i=1}^N$, $(W'_i)_{i=1}^n$ is a uniform sampling without replacement from $(W_i)_{i=1}^N$. Hoeffding's lemma for random variables sampled without replacement from a finite population

(see Hoeffding, 1963) then yields, for $M = 3C + 2 > C$

$$\begin{aligned}
 & \mathbb{P}_{\tau \sim \mathcal{U}(\mathcal{S}_N)} \left(\|f - f_n\|_N > \frac{M - C}{C + 1} \|f - f_n\|_{\tau, n} \right) \\
 &= \mathbb{P}_{\tau \sim \mathcal{U}(\mathcal{S}_N)} \left(\frac{1}{N} \sum_{i=1}^N W_i^2 > \frac{4}{n} \sum_{i=1}^n W_i'^2 \right) \\
 &= \mathbb{P}_{\tau \sim \mathcal{U}(\mathcal{S}_N)} \left(\frac{-3}{4} \frac{1}{N} \sum_{i=1}^N W_i^2 > \frac{1}{n} \sum_{i=1}^n W_i'^2 - \frac{1}{N} \sum_{i=1}^N W_i^2 \right) \\
 &= \mathbb{P}_{\tau \sim \mathcal{U}(\mathcal{S}_N)} \left(\frac{-1}{n} \sum_{i=1}^n W_i'^2 + \frac{1}{N} \sum_{i=1}^N W_i^2 > \frac{3}{4} \frac{1}{N} \sum_{i=1}^N W_i^2 \right) \\
 &\leq \exp \left(-\frac{3n \|f - f_n\|_N^2}{2 \|f - f_n\|_{L^\infty(\nu)}^2} \right).
 \end{aligned}$$

Using theorem 36 yields, on $V_n \cap A_N$ (which is an event of probability tending to 1), with $h_- = h_n$ and if $b_4 \leq J_n \leq b_5 \frac{t_0^{-d/2}}{\ln^{3d/2} N}$ (where $t_0 = 64h_n^2 \ln(2c_- N h_n^d)$)

$$\forall f \in \Sigma^{J_n}, \quad \frac{\|f - f_n\|_N^2}{\|f - f_n\|_{L^\infty(nu)}^2} \geq e^{-1} a_3^{-1} b_6^{-d/2} J_n^{-1} \ln^{-3d/2} N.$$

Note that since by assumption $\frac{h_n^{-d}}{\ln^{\tau_1} N} \leq J_n \leq \frac{h_n^{-d}}{\ln^{\tau_1} N}$, $\tau_1 \geq \tau_2 > 2d$, then $b_4 \leq J_n \leq b_5 \frac{t_0^{-d/2}}{\ln^{3d/2} N}$. All in all we get

$$\begin{aligned}
 & \mathbb{E}_0 \left[\Pi \left[\|f - f_n\|_N > (M - C) \varepsilon_n \geq \frac{M - C}{C + 1} \|f - f_n\|_n \mid \mathbb{X}^n, J_n, h_n \right] \right] \\
 &\leq o(1) + e^{c' n \varepsilon_n^2(J_n, h_n)} \mathbb{E}_0 \left[\mathbb{1}_{A_N \cap V_n} \int_{f \in \Sigma^{J_n}} \exp \left(-\frac{3n \|f - f_n\|_N^2}{2 \|f - f_n\|_{L^\infty(\nu)}^2} \right) \Pi(df) \right] \\
 &\leq o(1) + e^{c' n \varepsilon_n^2(J_n, h_n)} \mathbb{E}_0 \int_{f \in \Sigma^{J_n}} \exp \left(-\frac{3n J_n^{-1}}{2 e a_3 b_6^{d/2} \ln^{3d/2} N} \right) \Pi(df) \\
 &= o(1) + \exp \left(c' n \varepsilon_n^2(J_n, h_n) - \frac{3n J_n^{-1}}{2 e a_3 b_6^{d/2} \ln^{3d/2} N} \right).
 \end{aligned}$$

Moreover

$$J_n^{-1} \ln^{-3d/2} N \geq h_n^d \ln^{\tau_2 - 3d/2} N$$

and by definition of $\varepsilon_n(J_n, h_n)$ we have

$$\varepsilon_n^2(J_n, h_n) \lesssim \frac{J_n \ln N}{n} + \left(\frac{J_n^{-2/d} \ln N}{h_n^2} \right)^{2[\beta/2]} h_n^{2\beta} \leq \frac{h_n^{-d} \ln^{1-\tau_2} N}{n} + h_n^{2\beta} \ln^{2[\beta/2](1+2\tau_1/d)} N.$$

But since by assumption $h_n \geq (\ln N)^s n^{-\frac{1}{2\beta+d}}$ for some $s \in \mathbb{R}$ and $\beta > d/2$, this actually implies $n J_n^{-1} \ln^{-3d/2} N \gg n \varepsilon_n^2(J_n, h_n)$, which concludes the proof.

The proof of the adaptive case, i.e of theorem 7, is fairly similar. Again, we use Theorem 6, so that there exists $C > 0$

$$\mathbb{E}_0 (\Pi [\|f - f_0\|_n > C\varepsilon_n | \mathbb{X}^n]) = o(1).$$

where $\varepsilon_n = n^{-\beta/(2\beta+d)} (\ln N)^{\frac{(2\tau+d)\lfloor \beta/2 \rfloor - \beta(\tau+2\beta/d)}{2\beta+d}}$.

Moreover, since by assumption on Π_J for any J_n, k we have $\Pi [J > kJ_n] \lesssim e^{-a_2 k J_n}$, the remaining mass Theorem (see Ghosal and van der Vaart, 2017, chap. 8) together with our prior thickness result 38 yields, for some $k > 0$

$$\Pi [J > kJ_n | X^n] \xrightarrow[n \rightarrow \infty]{P_0^\infty} 0$$

for $J_n = n\varepsilon_n^2$.

Therefore, a similar proof as in the non adaptive case above leads to the upper bound: on $V_n \cap A_N$,

$$\begin{aligned} & \Pi [\|f - f_0\|_N > M\varepsilon_n | \mathbb{X}^n] \\ & \leq o_{\mathbb{P}_0}(1) + e^{c'n\varepsilon_n^2(J_n, h_n)} \sum_{J \leq kJ_n} \mathbb{E}_0 \int_{f \in \Sigma^J} \exp \left(-\frac{3n \|f - f_n\|_N^2}{2 \|f - f_n\|_{L^\infty(\nu)}^2} \right) d\Pi(f|F) \Pi_J(J) \end{aligned}$$

for some $c' > 0$. Moreover, since by construction Π -almost surely we have $J \leq \frac{h^{-d}}{\ln^{\tau/d} N}$, $\tau > 2d$, we can still apply theorem 36 to get on $V_n \cap A_N$,

$$\begin{aligned} & \Pi [\|f - f_0\|_N > M\varepsilon_n | \mathbb{X}^n] \\ & \leq o_{\mathbb{P}_0}(1) + e^{c'n\varepsilon_n^2} \sum_{J \leq kJ_n} \mathbb{E}_0 \int_{f \in \Sigma^J} \exp \left(-\frac{3nJ^{-1}}{2ea_3b_6^{d/2} \ln^{3d/2} N} \right) d\Pi(f|J) \Pi_J(J) \\ & \leq o_{\mathbb{P}_0}(1) + e^{c'n\varepsilon_n^2} \sum_{J \leq kJ_n} \exp \left(-\frac{3nJ^{-1}}{2ea_3b_6^{d/2} \ln^{3d/2} N} \right) \\ & \leq o_{\mathbb{P}_0}(1) + J_n \exp \left(c'n\varepsilon_n^2 - \frac{3nJ_n^{-1}}{2ea_3b_6^{d/2} \ln^{3d/2} N} \right) \\ & = o_{\mathbb{P}_0}(1) + \exp \left(c'n\varepsilon_n^2 - \frac{3}{2eka_3b_6^{d/2}} \varepsilon_n^{-2} \ln^{-3d/2} N \right) \end{aligned}$$

Given the expression of ε_n and since $\beta > d/2$, the right hand side is $o_{\mathbb{P}_0}(1)$, which concludes the proof of the adaptive case.

Acknowledgments

The project leading to this work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 834175).

Appendix A. Regularity assumption on the manifold and definition of the Hölder spaces

In this section we give details on the regularity of the submanifold \mathcal{M} and the definition of the Hölder spaces. It is usually assumed in Riemannian geometry that the manifolds are \mathcal{C}^∞ -smooth but here our results also apply for manifolds with finite Hölder regularity if the latter is large enough, hence we precise our notations and recall some general facts. We assume that \mathcal{M} to be a d -dimensional \mathcal{C}^α compact and connected submanifold of \mathbb{R}^D in the following sense : we can find a family $(\varphi_i, \mathcal{U}_i)_{i \in I}$ such that $(\mathcal{U}_i)_{i \in I}$ is an open cover of \mathcal{M} (that is, each \mathcal{U}_i is an open set of \mathcal{M} and $\mathcal{M} = \cup_{i \in I} \mathcal{U}_i$), $\varphi_i : \varphi_i^{-1}(\mathcal{U}_i) \subset \mathbb{R}^d \rightarrow \mathcal{U}_i \subset \mathcal{M} \subset \mathbb{R}^D$ is injective, of class \mathcal{C}^α , $\alpha \geq 1$ and its differential $d\varphi_i : \mathbb{R}^d \rightarrow \mathbb{R}^D$ is everywhere injective (hence φ_i is a \mathcal{C}^α immersion). Since \mathcal{M} is compact we can assume without loss of generality that I is finite, but the regularity results below stay true in the general case, at least locally.

The geodesics on the submanifold are defined as the curves $\gamma(t)$ on \mathcal{M} satisfying the geodesics equation : in local coordinates $y = \varphi^{-1}(\gamma)$ (we drop the subscript i for convenience) it takes the form

$$\forall k = 1, \dots, d, \quad \ddot{y}^k + \Gamma_{ij}^k \dot{y}^i \dot{y}^j = 0$$

where the Christoffel symbols are defined by

$$\Gamma_{ij}^k = \frac{1}{2} g^{kl} (\partial_i g_{jl} + \partial_j g_{il} - \partial_l g_{ij})$$

Here both the metric tensor g , its inverse (g^{kl}) and the Christoffel symbols depend on the point $\gamma(t) = \varphi(y(t))$. Moreover, with $\left(\partial_i := \frac{\partial \varphi}{\partial y_i}\right)_{i=1}^d$ the coordinates vector fields and $\varphi = (\varphi^l)_{l=1}^D$, the metric tensor is given by the Gram matrix

$$\forall p \in \mathcal{M}, \quad g_{ij}(p) = \langle \partial_i | \partial_j \rangle_{T_p \mathcal{M}} = \sum_{l=1}^D \frac{\partial \varphi^l}{\partial y^i}(\varphi^{-1}(p)) \frac{\partial \varphi^l}{\partial y^j}(\varphi^{-1}(p)) = \left(\frac{\partial \varphi}{\partial y_i} \right)^T \left(\frac{\partial \varphi}{\partial y_j} \right) (\varphi^{-1}(p))$$

and at $p = \gamma(t)$

$$\partial_l g_{ij} = \frac{\partial (g_{ij} \circ \varphi)}{\partial y^l}(y(t)) = \left[\frac{\partial}{\partial y^l} \sum_{k=1}^D \frac{\partial \varphi^k}{\partial y^i} \frac{\partial \varphi^k}{\partial y^j} \right] (y(t))$$

Hence for each i, j, k , Γ_{ij}^k is a $\mathcal{C}^{\alpha-2}$ Hölder function of $y(t)$. Therefore by standard ODE theory, if $\alpha \geq 3$ existence and uniqueness of geodesics $t \mapsto \gamma_{x,v}(t)$ defined on maximal intervals and such that $\varphi(y(0)) = \gamma_{x,v}(0) = x$, $\frac{\partial \varphi}{\partial y}(\varphi^{-1}(x)) \dot{y}(0) = \gamma'_{x,v}(0) = v$ for arbitrary $x \in \mathcal{M}$, $v \in T_x \mathcal{M}$ is guaranteed. Moreover the flow $(x, v, t) \mapsto (\gamma_{x,v}(t), \gamma'_{x,v}(t))$ is of class $\mathcal{C}^{\alpha-2}$, in the sense that

$$(y_0, \dot{y}_0, t) \mapsto \left(\gamma_{\varphi(y_0), \frac{\partial \varphi}{\partial y}(y_0) \dot{y}_0}(t), \gamma'_{\varphi(y_0), \frac{\partial \varphi}{\partial y}(y_0) \dot{y}_0}(t) \right)$$

is of class $\mathcal{C}^{\alpha-2}$ on its domain. In particular the exponential map defined as $\exp_x(v) = \gamma_{x,v}(1)$ for all $\|v\|_{T_x \mathcal{M}} < r_x$ (the injectivity radius of \mathcal{M} at x) is in this sense of class $\mathcal{C}^{\alpha-2}$ on its

domain, as seen as a function with values in \mathbb{R}^D . Since \mathcal{M} is compact it has a positive global injectivity radius $r_{\mathcal{M}} := \inf_{x \in \mathcal{M}} r_x$ and therefore for any $x \in \mathcal{M}$ the exponential map \exp_x is well defined from $B_{T_x \mathcal{M}}(0, r_{\mathcal{M}})$ to \mathcal{M} and is a diffeomorphism onto its image. Moreover, by compactness of \mathcal{M} the Hopf-Rinow theorem implies that for each $x \in \mathcal{M}$ the exponential map \exp_x is actually defined on the whole of $T_x \mathcal{M}$ (while obviously no longer being guaranteed to be injective). Hence by the above, for any $i \in I$ the following map

$$\begin{pmatrix} \varphi_i^{-1}(\mathcal{U}_i) \times \mathbb{R}^d & \rightarrow & \mathbb{R}^D \\ (y_0, \dot{y}_0) & \mapsto & \exp_{\varphi_i(y_0)} \left(\frac{\partial \varphi_i}{\partial y}(y_0) \dot{y}_0 \right) \end{pmatrix}$$

is of class $\mathcal{C}^{\alpha-2}$.

Finally let us define the different Hölder spaces over \mathcal{M} used in the paper. Let $(\chi_i)_{i \in I}$ be a partition of the unity subordinated to the open cover $(\mathcal{U}_i)_{i \in I}$. We say that $f : \mathcal{M} \rightarrow \mathbb{R}$ is of class \mathcal{C}^β (β -Hölder) on \mathcal{M} if $f \circ \varphi_i : \varphi_i^{-1}(\mathcal{U}_i) \rightarrow \mathbb{R}$ is of class \mathcal{C}^β for any $i \in I$. For a finite dimensional vector space V we say that $f : \mathcal{M} \rightarrow V$ is of class \mathcal{C}^β (β -Hölder) on \mathcal{M} if $T \circ f \circ \varphi_i : \varphi_i^{-1}(\mathcal{U}_i) \rightarrow \mathbb{R}$ is of class \mathcal{C}^β for any $i \in I$ and linear form $T \in V^*$. Equivalently the coordinates of f in any choice of basis of V are all of class \mathcal{C}^β . We also define $\mathcal{C}^\infty(\mathcal{M}) = \cap_{\beta \geq 0} \mathcal{C}^\beta(\mathcal{M})$. Furthermore we define the \mathcal{C}^β norm of f by $\|f\|_{\mathcal{C}^\beta(\mathcal{M})} = \max_{i \in I} \|(\chi_i f) \circ \varphi_i^{-1}\|_{\mathcal{C}^\beta(\mathcal{U}_i)}$. The definition of the Hölder space $\mathcal{C}^\beta(\mathcal{M}, V)$ can be seen to be independent of the chosen family $(\varphi_i, \mathcal{U}_i, \chi_i)_{i \in I}$ as long as $\alpha \geq \beta \vee 1$, with equivalence of the resulting Hölder norms.

Recall that we can always define regular local orthonormal frames on $T\mathcal{M}$: on \mathcal{U}_i simply apply the Gram-Schmidt orthonormalization algorithm to the coordinates vector fields $x \mapsto \frac{\partial \varphi}{\partial y_i}$, which gives a family $(e_i^j)_{i \in I, j=1, \dots, d}$ of orthonormal vector fields whose components in local coordinates are rational functions of the components of the coordinates vector fields $\frac{\partial \varphi_i}{\partial y_i}$. As $\varphi_i : \varphi_i^{-1}(\mathcal{U}_i) \subset \mathbb{R}^d \mapsto \mathcal{U}_i \subset \mathbb{R}^D$ is $\mathcal{C}^{\alpha-1}$, the maps: $x \in \varphi_i^{-1}(\mathcal{U}_i) \subset \mathbb{R}^d \mapsto e_i^j(x) \in \mathcal{U}_i \subset \mathbb{R}^D$ are all $\mathcal{C}^{\alpha-1}$. Moreover by construction the function $\omega \mapsto \psi_i(x, \omega) := \sum_j \omega_j e_i^j(x)$ is a linear isometry from \mathbb{R}^d to $T_x \mathcal{M}$ for each $x \in \mathcal{U}_i$ where \mathbb{R}^d is equipped with its standard Euclidean structure and $T_x \mathcal{M}$ its inner product induced by the metric.

In particular if $\alpha - 2 \geq \beta \vee 1 \iff \alpha \geq (\beta + 2) \vee 3$ then any $f \in \mathcal{C}^\beta(\mathcal{M})$ satisfies

$$\begin{aligned} \sup_{x \in \mathcal{U}_i} \|f \circ \exp_x(\psi_i(x, \cdot))\|_{\mathcal{C}^\beta(B_{\mathbb{R}^d}(0, r_{\mathcal{M}}))} &= \sup_{x \in \mathcal{U}_i} \left\| \sum_j (\chi_j f) \circ \exp_x(\psi_i(x, \cdot)) \right\|_{\mathcal{C}^\beta(B_{\mathbb{R}^d}(0, r_{\mathcal{M}}))} \\ &\leq \sum_j \sup_{x \in \mathcal{U}_i} \|(\chi_j f) \circ \exp_x(\psi_i(x, \cdot))\|_{\mathcal{C}^\beta(B_{\mathbb{R}^d}(0, r_{\mathcal{M}}))} \\ &= \sum_j \sup_{x \in \mathcal{U}_i} \left\| ((\chi_j f) \circ \varphi_j) \circ (\varphi_j^{-1} \circ \exp_x(\psi_i(x, \cdot))) \right\|_{\mathcal{C}^\beta(B_{\mathbb{R}^d}(0, r_{\mathcal{M}}))} \\ &\lesssim \max_j \|(\chi_j f) \circ \varphi_j\|_{\mathcal{C}^\beta(\mathcal{U}_j)} \sum_j \sup_{x \in \mathcal{U}_i} \left\| \varphi_j^{-1} \circ \exp_x(\psi_i(x, \cdot)) \right\|_{\mathcal{C}^\beta(B_{\mathbb{R}^d}(0, r_{\mathcal{M}}), \varphi_j^{-1}(\mathcal{U}_j))} \lesssim \|f\|_{\mathcal{C}^\beta(\mathcal{M})}, \end{aligned}$$

the last step using resulting from the compactness of \mathcal{M} , finiteness of I and the fact that $\alpha - 2 \geq \beta$. We will use this property in our approximation results.

We recall that we assume $\alpha \geq (\beta + 3) \vee 6$. This is for technical reasons that will appear later in the proofs.

We finish this section by showing a differential geometry lemma that will be useful for the proofs of the results in section 3 : very roughly speaking, this lemma states that for small h we can find a function $t = t_h(x, v)$ of $x \in \mathcal{M}, v \in T_x \mathcal{M}$ such that $\|\exp_x(thv) - x\|_{\mathbb{R}^D} = h$, and that moreover this function can be written as $t_h = 1 + h^2 s_h$ where s_h is Hölder regular and has bounded derivatives of order less than $\alpha - 5$, even when h is close to 0. First, recall that the radius of curvature (see Bernstein et al.) $a_{\mathcal{M}}$ of \mathcal{M} is positive by compactness of \mathcal{M} .

Lemma 18. *There exists $r'_{\mathcal{M}} > 0$ such that for each $x \in \mathcal{M}, v \in T_x \mathcal{M}, \|v\| = 1$ the function $r \mapsto \|\exp_x(rv) - x\|$ is increasing on $[0, r'_{\mathcal{M}}]$ and satisfies $\|\exp_x(rv) - x\|^2 \geq \frac{r^2}{4}, r \in [0, r'_{\mathcal{M}}]$.*

As a consequence for any $h \in [0, \frac{r'_{\mathcal{M}}}{2}]$ there is a unique $t = t_h(x, v) \in [0, h^{-1}r'_{\mathcal{M}}]$ satisfying $\|\exp_x(thv) - x\| = h$. Moreover $t_h(x, v) \in [1, 1 + \frac{\pi^3 h^2}{192 a_{\mathcal{M}}^2}]$ and there exists $h_+ > 0$ such that

$$\max_{\mathbf{k} \in \mathbb{N}^d, |\mathbf{k}| \leq \alpha - 5} \sup_{\substack{i \in I \\ y \in \varphi_i^{-1}(\mathcal{U}_i) \\ 0 < h < h'_+ \\ \|v\|_{T_{\varphi_i(y)} \mathcal{M}} = 1}} \left| \frac{\partial^{|\mathbf{k}|}}{\partial y^{\mathbf{k}}} \left(\frac{t_h(\varphi_i(y), v) - 1}{h^2} \right) \right| < +\infty$$

Proof We have for $r < r_{\mathcal{M}}$,

$$\begin{aligned} & \frac{\partial}{\partial r} \|\exp_x(rv) - x\|_{\mathbb{R}^D}^2 \\ &= 2 \langle d \exp_x(rv) v | \exp_x(rv) - x \rangle_{\mathbb{R}^D} \\ &= 2 \langle v + d \exp_x(rv) v - v | rv + \exp_x(rv) - x - rv \rangle_{\mathbb{R}^D} \\ &= 2 \langle v | rv \rangle_{\mathbb{R}^D} + 2 \langle d \exp_x(rv) v - v | rv \rangle_{\mathbb{R}^D} + 2 \langle \exp_x(rv) - x - rv | v \rangle_{\mathbb{R}^D} \\ & \quad + 2 \langle d \exp_x(rv) v - v | \exp_x(rv) - x - rv \rangle_{\mathbb{R}^D} \\ &= 2r + \mathcal{O}(r^2) \end{aligned}$$

where the last equality uses

$$\begin{aligned} \exp_x(rv) - x - rv &= r d \exp_x(0)v - rv + \mathcal{O}(r^2) = \mathcal{O}(r^2), \\ d \exp_x(rv) v - v &= d \exp_x(0)v - v + \mathcal{O}(r^2) = \mathcal{O}(r^2). \end{aligned}$$

By compactness of \mathcal{M} and since $(x, v) \mapsto \exp_x(v) \in \mathcal{C}^{\alpha-3}, \alpha \geq 3$ this implies that

$$\sup_{\substack{0 < r < r_{\mathcal{M}} \\ x \in \mathcal{M} \\ \|v\|_{T_x \mathcal{M}} = 1}} \frac{\left| \frac{\partial}{\partial r} \|\exp_x(rv) - x\|_{\mathbb{R}^D}^2 - 2r \right|}{r^2} < +\infty,$$

In particular there exists $r = r'_{\mathcal{M}} > 0$ such that for any $r \in [0, r'_{\mathcal{M}}]$ we have

$$\forall x \in \mathcal{M}, \quad \forall v \in T_x \mathcal{M}, \quad \|v\|_{T_x \mathcal{M}} = 1, \quad \frac{\partial}{\partial r} \|\exp_x(rv) - x\|_{\mathbb{R}^D}^2 \geq \frac{r}{2}.$$

As a consequence, the function $r \mapsto \|\exp_x(rv) - x\|_{\mathbb{R}^D}^2$ is increasing on $[0, r'_{\mathcal{M}}]$ and so is $r \mapsto \|\exp_x(rv) - x\|_{\mathbb{R}^D}$. Moreover by integrating we get

$$\forall r \in [0, r'_{\mathcal{M}}], \quad \|\exp_x(rv) - x\|_{\mathbb{R}^D}^2 \geq \frac{r^2}{4},$$

so that, if $h \in [0, \frac{r'_{\mathcal{M}}}{2}]$, then $r = 2h \in [0, r'_{\mathcal{M}}]$ and $\|\exp_x(rv) - x\|_{\mathbb{R}^D}^2 \geq \frac{(2h)^2}{4} = h^2$. By continuity and strict monotonicity there exists a unique $r = r_h(x, v) \in [0, 2h]$ solution of $\|\exp_x(rv) - x\| = h$ on $[0, r'_{\mathcal{M}}]$. Defining $t_h(x, v) = h^{-1}r_h(x, v)$ gives the first part of the statement.

Let $i \in I$, $x \in \mathcal{U}_i$, $\omega \in \mathbb{R}^d$, $\|\omega\| = 1$ and consider the Taylor expansion (justified since $\alpha \geq 5$ and $\exp \in \mathcal{C}^{\alpha-2}$)

$$\begin{aligned} & \exp_x(th\psi_i(x, \omega)) \\ &= x + th\psi_i(x, \omega) + \frac{t^2 h^2}{2} d^2 \exp_x(0) \cdot \psi_i(x, \omega) + h^3 \int_0^t \frac{(t-s)^2}{2} d^3 \exp_x(sh\psi_i(x, \omega)) \psi_i(x, \omega)^3 ds. \end{aligned}$$

Changing of variables with $y \in \varphi_i^{-1}(\mathcal{U}_i)$, $x = \varphi_i(y)$, we then have

$$\begin{aligned} & F(y, t) \\ &:= h^{-2} \|\exp_x(th\psi_i(x, \omega)) - x\|^2 - 1 \\ &= h^{-2} \left\| th\psi_i(x, \omega) + \frac{t^2 h^2}{2} d^2 \exp_x(0) \cdot \psi_i(x, \omega) + h^3 \int_0^t \frac{(t-s)^2}{2} d^3 \exp_x(sh\psi_i(x, \omega)) \psi_i(x, \omega)^3 ds \right\|^2 - 1 \\ &= \left\| t\psi_i(x, \omega) + \frac{t^2 h}{2} d^2 \exp_x(0) \cdot \psi_i(x, \omega) + h^2 \int_0^t \frac{(t-s)^2}{2} d^3 \exp_x(sh\psi_i(x, \omega)) \psi_i(x, \omega)^3 ds \right\|^2 - 1 \end{aligned}$$

Therefore we have

$$F(y, t) = t^2 - 1 + h^2 G(y, t)$$

where, with

$$A = t\psi_i(x, \omega), \quad B = \frac{t^2}{2} d^2 \exp_x(0) \cdot \psi_i(x, \omega), \quad C = \int_0^t \frac{(t-s)^2}{2} d^3 \exp_x(sh\psi_i(x, \omega)) \psi_i(x, \omega)^3 ds$$

we have defined

$$G(y, t) = G_{i,h,\omega}(y, t) = 2h^{-1} \langle A|B \rangle + 2 \langle A|C \rangle + \|B\|^2 + 2h \langle B|C \rangle + h^2 \|C\|^2$$

Moreover, since $\|t\psi_i(x, \omega)\|^2 = t^2$ and $d^2 \exp_x(0) \cdot \psi_i(x, \omega) \in N_x \mathcal{M}$ (as the second derivative at time 0 of the geodesic $t \mapsto \exp_x(t\psi_i(x, \omega))$) this implies $\langle A|B \rangle = 0$ and therefore

$$G(y, t) = 2 \langle A|C \rangle + \|B\|^2 + 2h \langle B|C \rangle + h^2 \|C\|^2$$

Furthermore since the exponential map is of class $\mathcal{C}^{\alpha-2}$ we see that G is of class $\mathcal{C}^{\alpha-5}$ and that

$$M := \max_{\substack{i \in I \\ (\mathbf{k}, l) \in \mathbb{N}^{d+1} \\ |\mathbf{k}| + l \leq \alpha - 5}} \sup_{\substack{y \in \varphi_i^{-1}(\mathcal{U}_i) \\ 1 \leq t \leq 1 + \frac{\pi}{2} \\ 0 < h < r'_{\mathcal{M}}/2 \\ \|\omega\| = 1}} \left| \frac{\partial^{|\mathbf{k}|+l} G}{\partial y^{|\mathbf{k}|} \partial t^l} \right| < +\infty.$$

Differentiating yields

$$\frac{\partial F}{\partial y} = h^2 \frac{\partial G}{\partial y}, \frac{\partial F}{\partial t} = 2t + h^2 \frac{\partial G}{\partial t}.$$

Also, notice that $1 \leq t_h(x, v) \leq 1 + \frac{\pi^3 h^2}{192 a_{\mathcal{M}}^2} \wedge \frac{\pi}{2}$: indeed, without loss of generality $r'_{\mathcal{M}} \leq \pi a_{\mathcal{M}}$ and therefore if $h \leq r'_{\mathcal{M}}/2$ then for $t = t_h(x, v) \in [0, 2]$ we have $th = \rho(x, \exp_x(th\psi_i(v))) \leq r'_{\mathcal{M}} \leq \pi a_{\mathcal{M}}$ and

$$\begin{aligned} \frac{2}{\pi} th &= \frac{2}{\pi} \rho(x, \exp_x(th\psi_i(x, \omega))) \leq \|\exp_x(th\psi_i(x, \omega)) - x\| = h \leq \rho(x, \exp_x(th\psi_i(x, \omega))) = th \\ &\implies 1 \leq t_h(x, v) \leq \frac{\pi}{2}. \end{aligned}$$

Reusing this gives

$$\begin{aligned} &th - \frac{(th)^3}{24a_{\mathcal{M}}^2} \\ &= \rho(x, \exp_x(th\psi_i(x, \omega))) - \frac{\rho(x, \exp_x(th\psi_i(x, \omega)))^3}{24a_{\mathcal{M}}^2} \\ &\leq \|\exp_x(th\psi_i(x, \omega)) - x\| \\ &= h \leq \rho(x, \exp_x(th\psi_i(x, \omega))) \\ &= th, \end{aligned}$$

which implies

$$1 \leq t_h(x, v) \leq 1 + \frac{t^3 h^2}{24a_{\mathcal{M}}^2} \leq 1 + \frac{\pi^3 h^2}{192a_{\mathcal{M}}^2}.$$

In particular, using $\alpha \geq 6$

$$\begin{aligned} &\sup_{\substack{i \in I \\ y \in \varphi_i^{-1}(\mathcal{U}_i) \\ 0 < h < r'_{\mathcal{M}}/2 \\ \|v\|_{T_{\varphi_i^{-1}(y)} \mathcal{M}} = 1}} \frac{\left| \frac{\partial F}{\partial t}(\varphi_i^{-1}(y), t_h(\varphi_i^{-1}(y), v)) - 2t_h(\varphi_i^{-1}(y), v) \right|}{h^2} \\ &= \sup_{\substack{i \in I \\ y \in \varphi_i^{-1}(\mathcal{U}_i) \\ 0 < h < r'_{\mathcal{M}}/2 \\ \|v\|_{T_{\varphi_i^{-1}(y)} \mathcal{M}} = 1}} \left| \frac{\partial G}{\partial t}(y, t_h(\varphi_i^{-1}(y), v)) \right| \\ &\leq M < +\infty \end{aligned}$$

Since $2t_h(x, v) \geq 2$, there exists $0 < h_+ \leq r'_{\mathcal{M}}/2$ such that for any $i \in I$, $h \leq h_+$, $x \in \mathcal{U}_i$, $\|v\|_{T_x \mathcal{M}} = 1$ we have $\frac{\partial F}{\partial t} \neq 0$. Hence the implicit function theorem implies that $y \mapsto t_h(\varphi_i(y), v)$ is smooth on $\varphi_i^{-1}(\mathcal{U}_i)$ with differential given by

$$\frac{\partial}{\partial y} t_h(\varphi_i(y), v) = -\frac{\frac{\partial F}{\partial y}}{\frac{\partial F}{\partial t}} = -\frac{h^2 \frac{\partial G}{\partial y}}{2t_h(\varphi_i(y), v) + h^2 \frac{\partial G}{\partial t}} \in \mathcal{L}(\mathbb{R}^d, \mathbb{R})$$

To conclude notice first that since $t_h(x, v) \in [1, 1 + \frac{\pi^3 h^2}{192 a_{\mathcal{M}}^2}]$ we have, setting $s_h(x, v) = \frac{t_h(x, v) - 1}{h^2}$

$$\sup_{\substack{x \in \mathcal{M} \\ 0 < h < r'_{\mathcal{M}}/2 \\ \|v\|_{T_x \mathcal{M}} = 1}} |s_h(x, v)| \leq \frac{\pi^3}{192 a_{\mathcal{M}}^2}$$

Moreover for the derivatives we get, for any $y \in \varphi_i^{-1}(\mathcal{U}_i)$

$$\begin{aligned} \frac{\partial}{\partial y} s_h(\varphi_i(y), v) &= h^{-2} \frac{\partial}{\partial y} t_h(\varphi_i(y), v) = -h^{-2} \frac{h^2 \frac{\partial G}{\partial y}}{2t_h(\varphi_i(y), v) + h^2 \frac{\partial G}{\partial t}} \\ &= -\frac{\frac{\partial G}{\partial y}}{2t_h(\varphi_i(y), v) + h^2 \frac{\partial G}{\partial t}} = -\frac{\frac{\partial G}{\partial y}}{2 + 2h^2 s_h(\varphi_i(y), v) + h^2 \frac{\partial G}{\partial t}}. \end{aligned}$$

Hence for any $y \in \varphi_i^{-1}(\mathcal{U}_i)$, $\|v\|_{T_{\varphi_i(y)} \mathcal{M}} = 1$, $0 < h < h'_+ = h_+ \wedge M^{-1/2}$

$$\forall 1 \leq j \leq d, \quad \left| \frac{\partial}{\partial y_j} s_h(\varphi_i(y), v) \right| = \frac{\left| \frac{\partial G}{\partial y_j} \right|}{\left| 2t_h(\varphi_i(y), v) + h^2 \frac{\partial G}{\partial t} \right|} \leq \frac{M}{2 - h^2 M} \leq M$$

In the same way by induction we can prove that

$$\max_{\alpha \in \mathbb{N}^d, |\alpha| \leq \alpha - 5} \sup_{\substack{i \in I \\ y \in \varphi_i^{-1}(\mathcal{U}_i) \\ 0 < h < h'_+ \\ \|v\|_{T_{\varphi_i(y)} \mathcal{M}} = 1}} \left| \frac{\partial^{|\alpha|}}{\partial y^\alpha} s_h(\varphi_i(y), v) \right| < +\infty$$

■

Remark 19. We have stated our lemma with $r'_{\mathcal{M}}$ such that $\|\exp_x(rv) - x\|^2 \geq \frac{r^2}{4}$ for any $r \in [0, r'_{\mathcal{M}}]$, but more generally for any $c \in (0, 1)$ we could have found $r'_{\mathcal{M}, c} > 0$ such that $r \in [0, r'_{\mathcal{M}, c}]$ implies $\|\exp_x(rv) - x\|^2 \geq c^2 r^2$.

Appendix B. Geometrical and analytical properties of the random graph

In this section we prove/recall useful facts on the graph coming from geometrical properties of the manifold.

We start by a simple but useful property satisfied by the graph Laplacian eigenvectors :

Proposition 20. For any $h > 0$, $j \in \{1, \dots, N\}$ we have $\|u_j\|_{L^\infty(\nu)} \leq N$.

Proof For all $x \in V$ we have $\nu_x = \frac{\mu_x}{\sum_{y \in V} \mu_y} \geq \frac{1}{N^2}$, which implies

$$\forall x \in V, u_j(x)^2 \leq \frac{1}{\nu_x} \sum_{y \in V} u_j(y)^2 \nu_y = \frac{1}{\nu_x} \|u_j\|_{L^2(\nu)}^2 = \frac{1}{\nu_x} \leq N^2$$

■

As we will see below, we actually have $\nu_x \asymp N^{-1}$ on a high probability event, so that the result of proposition 20 could actually be improved to $\max_{1 \leq j \leq N} \|u_j\|_{L^\infty(\nu)} \lesssim \sqrt{N}$, but this would not improve the final rates anyway.

In the next subsections we will be discussing various properties of the random geometric graph, namely volume regularity, on diagonal heat kernel bounds, Weyl type upper and lower bounds on the graph Laplacian eigenvalues and finally a norm comparison theorem. These properties will be used in the proof of Theorems 3,4,6 & 7. More precisely, the proof of the Kullback-Leibler prior mass condition (Lemma 38, which is then subsequently used in the proof of Theorems 3 & 6) uses a lower bound on the eigenvalues of the graph Laplacian $\mathcal{L}^{(h_n)}$ for a determined h_n (the one appearing in the statements of Theorems 3 & 6). To prove contraction rates with respect to $\|\cdot\|_N$ (Theorems 4 & 7), we combine the results for $\|\cdot\|_n$ together with Lemma 36, which states an inequality of the form : with \mathbb{P}_0 -probability going to 1

$$\forall h_- \leq h \leq h_0, \quad J_0 \leq J \leq J_1 h^{-d}, \quad f \in \text{span}\{u_1, \dots, u_J\}, \quad \|f\|_{L^\infty(\nu)} \lesssim \sqrt{J} \|f\|_{L^2(\nu)} \quad (11)$$

for some constants $h_0, J_0, J_1 > 0$, up to logarithmic factors, as long as h_- satisfies assumption 2. To prove theorem 7 (adaptive posterior contraction rates with respect to $\|\cdot\|_N$) we use inequality 11 for any h in the support of the posterior distribution. Thus in this section we analyse the properties of the random graph associated to the connectivity parameter h for h belonging to $[h_-, h_0]$. Note that these properties will also be used in the proof of the other results but for a specific value h_n in $[h_-, h_0]$.

B.1 Volume regularity

In this subsection we establish a volume regularity property for the random geometric graph : roughly speaking, with high probability, for every suitable r, h and geodesic ball of radius r we have $\nu^{(h)}(B) \asymp r^d$. To start with notice that since \mathcal{M} is compact it has a positive radius of curvature $a_{\mathcal{M}}$ and in particular by lemma 3 in Bernstein, de Silva, Langford, and Tenenbaum for any $x, y \in \mathcal{M}$ with $\rho(x, y) \leq \pi a_{\mathcal{M}}$ we find

$$\frac{2}{\pi} \rho(x, y) \leq 2a_{\mathcal{M}} \sin(\rho(x, y)/2ra_{\mathcal{M}}) \leq \|x - y\| \leq \rho(x, y) \quad (12)$$

and, using $\sin(\rho/2a_{\mathcal{M}}) \geq \frac{\rho}{2ra_{\mathcal{M}}} - \frac{1}{6} \left(\frac{\rho}{2a_{\mathcal{M}}}\right)^3$,

$$\rho(x, y) - \frac{\rho(x, y)^3}{24a_{\mathcal{M}}^2} \leq \|x - y\| \leq \rho(x, y).$$

Moreover the set $\mathcal{M}^2 \setminus \{(x, y) \in \mathcal{M}^2 : \rho(x, y) < \pi a_{\mathcal{M}}\}$ is compact with respect to the topology inherited from the Euclidean distance or equivalently the geodesic distance. Hence the function $(x, y) \mapsto \frac{\rho(x, y)}{\|x - y\|}$ is bounded on $\mathcal{M}^2 \setminus \{(x, y) \in \mathcal{M}^2 : \rho(x, y) < \pi a_{\mathcal{M}}\}$ by some constant C_0 , while it satisfies $\frac{\rho(x, y)}{\|x - y\|} \leq \frac{\pi}{2}$ when $\rho(x, y) \leq \pi a_{\mathcal{M}}$. Combining the two cases and assuming without loss of generality $C_0 \geq \frac{\pi}{2}$ we find the existence of $C_0 > 0$ such that $\|x - y\| \leq \rho(x, y) \leq C_0 \|x - y\|$ on \mathcal{M} .

Theorem 21. Let h_- satisfy Assumption 2, i.e. $Nh_-^d \gg \ln N$, and A_N be the event

$$A_N = \left\{ \forall i = 1, \dots, N, \forall h_- \leq r \leq \text{diam}_\rho(\mathcal{M})/C_0, c_- r^d \leq \frac{\mu_{X_i}^{(r)}}{N} \leq c_+ r^d \right\},$$

where $\mu_{X_i}^{(r)} = \mu(B(X_i, r))$. Then, for some $c_-, c_+, c > 0$ we have $\mathbb{P}(A_N) \geq 1 - e^{-cNh_-^d}$. Moreover, on A_N we have $\frac{c_-/c_+}{N} \leq \nu_y^{(h)} \leq \frac{c_+/c_-}{N}$ for any $h_- \leq h \leq \text{diam}_\rho(\mathcal{M})$ and $y \in V$.

A consequence of Theorem 21 is that on A_N for all $r \geq h_-$, $\nu^{(h)}(B(X_i, r)) \geq r^d c_-^2 / c_+$ for all $i \leq N$.

Proof Using

$$\forall x \in \mathcal{M}, r > 0, B_\rho(x, r) \subset B_{\mathbb{R}^D}(x, r) \subset B_\rho(x, C_0 r),$$

Theorem 3.8 in Göbel and Blanchard (2020) implies that for some $b_-, b_+, c > 0$, for all $h_- \leq r \leq \text{diam}_\rho(\mathcal{M})/C_0$, $\text{diam}_\rho(\mathcal{M}) := \max_{x, y \in \mathcal{M}} \rho(x, y)$, $i = 1, \dots, N$, with probability at least $1 - e^{-cNh_-^d}$

$$b_- r^d \leq \frac{\mu_{X_i}^{(r)} - 1}{N - 1} \leq b_+ r^d.$$

Using $\frac{Nh_-^d}{\ln N} \rightarrow \infty$ and a union bound we find that the last inequality holds simultaneously for all $i \in \{1, \dots, N\}$ with probability at least $1 - e^{-cNh_-^d}$ for some $c > 0$. To make the result uniform in r , define $r_k = 2^k h_-$, $k_+ = \lfloor \frac{\ln \text{diam}(\mathcal{M})}{\ln 2} \rfloor \leq c_{\mathcal{M}}^{-1} \ln N$ (without loss of generality for small $c_{\mathcal{M}}$, because $h_- \gg (\frac{\ln N}{N})^{1/d}$). Using a union bound we find that with probability $1 - (k_+ + 1)e^{-c_{\mathcal{M}}Nh_-^d} \geq 1 - c_{\mathcal{M}}^{-1} \ln N e^{-cNh_-^d}$, for every $i = 1, \dots, N$ and $k = 0, \dots, k_+ + 1$,

$$b_- r_k^d \leq \frac{\mu_{X_i}^{(r_k)} - 1}{N - 1} \leq b_+ r_k^d.$$

But then, for every $k = 0, \dots, k_+ + 1$, $r_k \leq r \leq r_{k+1}$ and $i \in \{1, \dots, N\}$ we have (assuming $N \geq 2$)

$$\begin{aligned} & \frac{N - 1}{N} \frac{\mu_{X_i}^{(r)} - 1}{N - 1} \leq \frac{\mu_{X_i}^{(r)}}{N} = \frac{1}{N} + \frac{\mu_{X_i}^{(r)} - 1}{N - 1} \frac{N - 1}{N} \\ \implies & \frac{N - 1}{N} \frac{\mu_{X_i}^{(r_k)} - 1}{N - 1} \leq \frac{\mu_{X_i}^{(r)}}{N} \leq \frac{1}{N} + \frac{\mu_{X_i}^{(r_{k+1})} - 1}{N - 1} \frac{N - 1}{N} \\ \implies & \frac{1}{2} \frac{\mu_{X_i}^{(r_k)} - 1}{N - 1} \leq \frac{\mu_{X_i}^{(r)}}{N} \leq \frac{1}{N} + \frac{\mu_{X_i}^{(r_{k+1})} - 1}{N - 1} \\ \implies & \frac{1}{4} b_- r_{k+1}^d = \frac{1}{2} b_- r_k^d \leq \frac{\mu_{X_i}^{(r)}}{N} \leq \frac{1}{N} + b_+ r_{k+1}^d = \frac{1}{N} + 2b_+ r_k^d \\ \implies & \frac{1}{4} b_- r^d \leq \frac{\mu_{X_i}^{(r)}}{N} \leq h^d + 2b_+ r^d \implies \frac{1}{4} b_- r^d \leq \frac{\mu_{X_i}^{(r)}}{N} \leq (2b_+ + 1) r^d. \end{aligned}$$

Therefore, using again $\frac{Nh_-^d}{\ln N} \rightarrow \infty$ with probability $1 - e^{-cNh_-^d}$, for every $i = 1, \dots, N$ and $h_- \leq r \leq \text{diam}_\rho(\mathcal{M})$,

$$\frac{1}{4}b_-r^d \leq \frac{\mu_{X_i}^{(r)}}{N} \leq (2b_+ + 3)r^d,$$

which proves the first statement with $c_- = \frac{1}{4}b_-$, $c_+ = 2b_+ + 3$. The second statement is an immediate consequence of the first one. Indeed, using again

$$\forall x \in \mathcal{M}, \quad r > 0, \quad B_\rho(x, r) \subset B_{\mathbb{R}^D}(x, r) \subset B_\rho(x, C_0 r)$$

we get for any $y \in V$ and $h_- \leq h \leq \text{diam}_\rho(\mathcal{M})/C_0$,

$$\frac{c_-/c_+}{N} = \frac{c_-Nr^d}{c_+N^2r^d} \leq \nu_y^{(h)} = \frac{\mu_y^{(h)}}{\sum_{x \in V} \mu_x^{(h)}} \leq \frac{c_+Nr^d}{c_-N^2r^d} = \frac{c_+/c_-}{N}.$$

■

In what follows we take $h_0 \leq \text{diam}_\rho(\mathcal{M})$, $r_0 \leq \text{diam}_\rho(\mathcal{M})/C_0$ (that we will actually reduce along the proof) and we fix h_- satisfying assumption 2 in order to work on the corresponding A_N .

B.2 Heat kernel bounds

The goal of this section is to show the following result : we consider A_N and h_- as defined in Theorem 21.

Theorem 22. *There exist constants $a_0, a_1, a_2, a_3, h_0 > 0$ (that also depend on \mathcal{M}, p_0) such that, on A_N , for all $h_- \leq h \leq h_0$ and $t_0(h) := a_0 h^2 \ln(Nh^d) \leq t \leq \frac{a_1}{\ln^2 N}$*

$$\forall x \in V, \quad a_2 \frac{t^{-d/2}}{\ln^d N} \leq p_t^{(h)}(x, x) \leq a_3 t^{-d/2}.$$

This result is a very weak form (it is often called an "on diagonal" upper bound) of the heat kernel bound used in Castillo, Kerkycharian, and Picard (2014); Coulhon, Kerkycharian, and Petrushev (2012) over continuous spaces, but fortunately will be enough for our purposes. It is potentially possible to apply further techniques in Barlow (2017) and get full off diagonal bounds (i.e, bounds on $p_t^{(h)}(x, y), x \neq y$) yielding a Gaussian type behaviour of the heat kernel, but we won't need this. It should be noted that the restriction $t \gtrsim h^2$ (up to a logarithmic factor) for the heat kernel bound to be valid substantially complicates the analysis of the situation. Indeed, if instead we could prove the bound for every $0 < t \leq \frac{a_1}{\ln^2 N}$ then we could conduct our analysis by combining the techniques from Coulhon and Sikora (2008); Coulhon, Kerkycharian, and Petrushev (2012), our approximation results and the proof strategy of Castillo, Kerkycharian, and Picard (2014). Since we require $t \gtrsim h^2 \ln(Nh^d)$ with h not arbitrarily small (otherwise our approximation results 1113 become vacuous) this is not possible and we have to prove things differently.

As in section 2, we emphasize the dependence of the graph (and in particular the graph Laplacian, the heat kernel and its eigendecomposition) with respect to h with a (h) exponent. While the notation become heavier, this is important to keep in mind as theorem 22

deals with a high probability control of the heat kernels associated with different values of h *simultaneously*.

B.2.1 ON DIAGONAL UPPER BOUND UP TO t_1

Lemma 23. *With A_N, c_-, c_+ the event and constant defined in Theorem 21 we have, on A_N*

$$\forall h_- \leq h \leq \text{diam}_\rho(\mathcal{M}), \quad t \geq 16h^2 \ln(2c_-Nh^d), \quad x, y \in V, \quad p_t^{(h)}(x, y) \leq 2\frac{c_+}{c_-}h^{-d}$$

Proof Recall that for all $x \in V, \quad t > 0, \quad \nu_y p_t^{(h)}(x, y) = \mathbb{P}_x^{(h)}(W_t = y)$. On the event A_N we have $\nu_y^{(h)} \geq \frac{c_-}{c_+N}$ for any $h_- \leq h \leq h_0$, which implies

$$\forall x, y \in V, \quad p_t^{(h)}(x, y) \leq \frac{c_+}{c_-}N\mathbb{P}_x^{(h)}(W_t = y) = \frac{c_+}{c_-}N \sum_{l \geq 0} \mathbb{P}(N_{t/h^2} = l) \mathbb{P}_x^{(h)}(Y_l = y)$$

where the last equality uses the independence of N and Y . Moreover, using Lemma 5.13 (e) from Barlow (2017) we find

$$\forall \rho \geq 2, \quad \mathbb{P}\left(N_\rho \leq \frac{\rho}{2}\right) \leq \mathbb{P}\left(|N_\rho - \rho| \geq \frac{\rho}{2}\right) \leq 2e^{-\rho/16}$$

which implies for any $x, y \in V$,

$$\begin{aligned} p_t^{(h)}(x, y) &\leq \frac{c_+}{c_-}N\mathbb{P}_x^{(h)}(W_t = y) \\ &\leq \frac{c_+}{c_-}N \left\{ 2e^{-t/16h^2} + \sum_{l \geq t/2h^2} \mathbb{P}(N_{t/h^2} = l) \mathbb{P}_x^{(h)}(Y_l = y) \right\} \\ &\leq \frac{c_+}{c_-}N \left\{ 2e^{-t/16h^2} + \sup_{l \geq 1} \mathbb{P}_x^{(h)}(Y_l = y) \right\} \end{aligned}$$

For each $l \geq 1$ we have $\mathbb{P}_x^{(h)}(Y_l = y) = \mathbb{E}_x^{(h)}[\mathbb{P}_x^{(h)}(Y_l = y | Y_{l-1})]$, and given Y_{l-1} the probability of going from Y_{l-1} to $Y_l = y$ is by definition 0 (if Y_{l-1} and y are not neighbours) or $\frac{1}{\mu_x^{(h)}}$ (if Y_{l-1} and y are neighbours). In either case on A_N we have $\mu_x^{(h)} \geq c_-Nh^d$ which implies

$$\forall x, y \in V, \quad p_t^{(h)}(x, y) \leq \frac{c_+}{c_-}N \left\{ 2e^{-t/16h^2} + \frac{1}{c_-Nh^d} \right\}.$$

All in all on A_N we get

$$\forall h_- \leq h \leq h_0, \quad t \geq 16h^2 \ln(2c_-Nh^d), \quad x, y \in V, \quad p_t^{(h)}(x, y) \leq 2\frac{c_+}{c_-}h^{-d}.$$

■

The goal is now, roughly speaking, to refine the inequality $p_t^{(h)}(x, y) \lesssim h^{-d}$ to $p_t^{(h)}(x, y) \lesssim t^{-d/2}$ (up to a logarithmic factor) by using the approach of Barlow (2017). This requires a local Poincare inequality, i.e an inequality of the form:

$$\forall x \in V, \quad r \in [r_-, r_+], \quad B = B_\rho(x, r), \quad f : B \rightarrow \mathbb{R},$$

$$\sum_{y \in B} (f(y) - f_B) \nu_y^{(h)} \leq Cr^2 \sum_{x, y \in B, y \sim x} (f(x) - f(y))^2, \quad \text{where } f_B = \frac{\sum_{y \in B} f(y) \nu_y^{(h)}}{\nu^{(h)}(B)}.$$

where the constants r_-, r_+, C need to be controlled. Göbel and Blanchard (2020) provide a way to prove such an inequality using the construction of random Hamming paths, however we give below a detailed application of their results to our setting in order to control the dependence of the constant C of the inequality in the parameters N, h .

Theorem 24. *Poincare inequality*

Let $A'_N = \{G^{(h_-)} \text{ is connected}\}$. Then there exist constants

$$r_0 = r_0(\mathcal{M}), \quad h_0 = h_0(\mathcal{M}), \quad C = C(\mathcal{M}) > 0$$

such that on A_N , for any $x \in V$, $0 < r < r_0$, $h_- \leq h < h_0$ and $f : B = B_\rho(x, r) \rightarrow \mathbb{R}$, on $A_N \cap A'_N$ we have

$$\sum_{y \in B} (f(y) - f_B) \nu_y^{(h)} \leq C \frac{(r/h)^2}{N^2 h^d} \sum_{z \in B, y \in B, y \sim x} (f(z) - f(y))^2$$

where $f_B = \sum_{y \in B} f(y) \nu_y^{(h)} / \nu^{(h)}(B)$. Moreover, $\mathbb{P}_0(A_N \cap A'_N) \geq 1 - e^{-cNh_-^d}$ for some $c > 0$.

Proof Let $h_- > 0$ and A_N, A'_N defined accordingly. The event $A_N \cap A'_N$ satisfies $\mathbb{P}_0(A_N \cap A'_N) \geq 1 - e^{-cNh_-^d}$ for some $c > 0$, see remark 25 below. For any $h \geq h_-$, $x \in V$, $B = B_\rho(x, r)$, Corollary 5.6 in Göbel and Blanchard (2020) shows that

$$\forall f : B \rightarrow \mathbb{R}, \quad \sum_{y \in B} (f(y) - f_B) \nu_y^{(h)} \leq \tilde{\kappa}_B r^2 \sum_{z \in B, y \in B, y \sim x} (f(z) - f(y))^2$$

where

$$\tilde{\kappa}_B = \frac{\max_{z \in B} \nu^{(h)}(y)^2}{2\nu^{(h)}(B)} l_{\max}(\tilde{\Gamma}_B) b_{\max}(\tilde{\Gamma}_B)$$

and $l_{\max}(\tilde{\Gamma}_B), b_{\max}(\tilde{\Gamma}_B)$ are quantities defined in Göbel and Blanchard (2020). Notice that Göbel and Blanchard (2020) requires connectedness of the subgraph restricted to B . This is shown in remark 25 below.

By following the proof of Corollary 5.13 in Göbel and Blanchard (2020) we find the existence of $C = C(\mathcal{M}, p_0)$, $r_0 = r_0(\mathcal{M})$, $h_0 = h_0(\mathcal{M}) > 0$ such that on A_N , for any $0 < r < r_0$, $x \in V$, $B = B_\rho(x, r)$

$$l_{\max}(\tilde{\Gamma}_B) \leq Ch^{-1}, b_{\max}(\tilde{\Gamma}_B) \leq Ch^{-(d+1)}$$

(in Göbel and Blanchard (2020) the result is proved with high probability over a single ball, but an inspection of the proof shows that the only thing needed is a control of the number of points in the balls on the graph for different values of r and h , which is precisely what A_N is for as shown in lemma 21, hence the uniformity). This shows the desired inequality

$$\sum_{y \in B} (f(y) - f_B) \nu_y^{(h)} \lesssim \frac{(r/h)^2}{N^2 h^d} \sum_{z \in B, y \in B, x \sim y} (f(z) - f(y))^2$$

■

Remark 25. We show the following simple result : if h_- satisfies assumption 2 then the probability that the resulting random geometric graph $G^{(h_-)}$ is connected (i.e the probability of the event A'_N) converges to 1. In particular, by monotonicity the probability that all random geometric graphs with connectivity parameter $h \geq h_-$ are connected converges to 1 as well.

Indeed, since \mathcal{M} is connected, consider $\{y_1, \dots, y_p\}$ an $h/8$ -net of \mathcal{M} with respect to the Euclidean distance $\|\cdot\|$. By standard arguments and compactness of \mathcal{M} we can assume that $p = \mathcal{O}(h^{-d})$. Then the event

$$A'_n = \{\forall l, \exists i_l, \|y_l - x_{i_l}\| < h/8\}$$

satisfies

$$\mathbb{P}((A'_n)^c) \lesssim p \left(1 - ch^d\right)^N = \mathcal{O}\left(\frac{1}{h^d} e^{-c8^{-d}Nh^d}\right) \lesssim e^{-c'Nh^d}$$

for some $c' > 0$ by assumption 2. Here $c > 0$ is small enough such that $\int_{B_{\mathbb{R}^D}(x,h)} p_0(x) \mu(dx) \geq ch^d$ for all $x \in \mathcal{M}$ (c exists since $p_0 \geq p_{\min} > 0$).

Moreover by connectedness of \mathcal{M} (which implies path connectedness since \mathcal{M} is a submanifold of \mathbb{R}^D) for each $i, j \in \{1, \dots, N\}$ there exists a continuous path $c : [0, 1] \rightarrow \mathcal{M}$ with $c(0) = x_i, c(1) = x_j$. Hence for each $t \in [0, 1]$ there exists $l(t) \in \{1, \dots, p\}$ such that $\|c(t) - y_{l(t)}\| < h/8$, and therefore on A'_n for each $t \in [0, 1]$ there exists $i_t = i_{l(t)}$ such that

$$\|c(t) - x_{i_t}\| \leq \|c(t) - y_{l(t)}\| + \|y_{l(t)} - x_{i_{l(t)}}\| < h/4$$

By continuity of c , this implies the existence of $K \geq 1$ and $1 \leq i_k \leq N$ for $k = 1, \dots, K$ such that $\|x_{i_k} - x_{i_{k+1}}\| \leq h/2$ for any $1 \leq k < K$ and such that $c(0) = x_{i_1}, c(1) = x_j = x_{i_K}$. Therefore the path $x_{i_1}, \dots, x_{i_{K-1}}, x_j$ connects x_{i_1} and x_j in the graph which implies connectedness on the event A'_n .

Notice that this is a very basic and coarse result about connectivity and more generally connected components of random geometric graphs. More details and results can be found in Penrose (2003).

With the Poincare inequality 24 we can now apply the proof strategy of Barlow (2017). We first prove an upper bound of the form

$$\forall x, y \in V, \quad p_t^{(h)}(x, y) \lesssim t^{-d/2}$$

for $t \gtrsim h^2 \ln(Nh^d)$ (up to a logarithmic factor) but less than a random time $t_1 = t_1(x, h)$ that we shall lower bound adequately later.

Lemma 26. *There exists $h_0 = h_{0|\mathcal{M}, p_0}$, $K = K_{\mathcal{M}, p_0}$, $K' = K'_{\mathcal{M}, p_0} > 0$ such that, on A_N , for every $h_- \leq h \leq h_0$, $x_0 \in V$, with the random times*

$$t_1(x, h) = \inf \left\{ t > 0 : p_{2t}^{(h)}(x, x) \leq K' \right\}$$

and

$$t_1(h) = \min_{x \in V} t_1(x, h),$$

we have, with $t_0(h) = 64h^2 \ln(2c_-Nh^d)$,

$$\forall h_- \leq h < h_0, \quad t \in [t_0(h), 2t_1(h)], \quad x, y \in V, \quad p_t^{(h)}(x, y) \leq Kt^{-d/2}.$$

Proof First consider the case $x = y$. Take $x_0 \in V$ and let $\phi(t) = p_{2t}^{(h)}(x_0, x_0) = \sum_{x \in V} f_t(x)^2 \nu_x^{(h)}$, $f_t(x) = p_t^{(h)}(x_0, x)$. Then since $\nu_y^{(h)} \mathcal{L}_{yx} = \nu_x^{(h)} \mathcal{L}_{xy}$, $\frac{d}{dt} f_t = -\mathcal{L} f_t$ we have

$$-\phi'(t) = 2 \langle f_t | \mathcal{L} f_t \rangle_{L^2(\nu)} = \frac{1}{\mu(V) h^2} \sum_{x \sim y} (f_t(x) - f_t(y))^2 \geq \frac{1}{c_+ N^2 h^{d+2}} \sum_{x \sim y} (f_t(x) - f_t(y))^2,$$

where we have used theorem 21 to bound $\mu(V) \leq c_+ N^2 h^d$. This proves that ϕ is non-increasing. Taking a covering of V by balls $B_i = B_\rho(x_i, r)$, $h_- \leq r \leq \text{diam}_\rho(\mathcal{M})/C_0$ given by lemma 27 below, since each x_i belongs to at most $M = 3^d (c_+/c_-)^3$ of the balls we have

$$-\phi'(t) \geq \frac{1}{c_+ N^2 h^{d+2}} \sum_{x \sim y} (f_t(x) - f_t(y))^2 \geq \frac{1}{c_+ M N^2 h^{d+2}} \sum_i \sum_{x \sim y \in B_i} (f_t(x) - f_t(y))^2.$$

On A_N , if $h_- < r < r_0$, $h_- \leq h < h_0$ and using Theorem 24 in each of the balls we find

$$-\phi'(t) \geq \frac{r^{-2}}{M C c_+} \sum_i \sum_{x \in B_i} (f_t(x) - f_{t|B_i})^2 \nu_x^{(h)}.$$

Defining $\nu_{B_i}^{(h)} = \frac{\nu_{|B_i}^{(h)}}{\nu^{(h)}(B_i)}$ the normalised restriction of $\nu^{(h)}$ to B_i and $\text{Var}(X) = EX^2 - (EX)^2$ we get

$$\begin{aligned} -\phi'(t) &\geq \frac{r^{-2}}{M C c_+} \sum_i \sum_{x \in B_i} (f_t(x) - f_{t|B_i})^2 \nu_x^{(h)} \\ &= \frac{r^{-2}}{M C c_+} \sum_i \nu^{(h)}(B_i) \sum_{x \in B_i} (f_t(x) - f_{t|B_i})^2 \nu_{B_i}^{(h)}(x) \\ &= \frac{r^{-2}}{M C c_+} \sum_i \nu^{(h)}(B_i) \left\{ \sum_{x \in B_i} f_t(x)^2 \nu_{B_i}^{(h)}(x) - \left(\sum_{x \in B_i} f_t(x) \nu_{B_i}^{(h)}(x) \right)^2 \right\} \\ &= \frac{r^{-2}}{M C c_+} \sum_i \left\{ \sum_{x \in B_i} f_t(x)^2 \nu_x - \nu^{(h)}(B_i) \left(\sum_{x \in B_i} f_t(x) \nu_{B_i}^{(h)}(x) \right)^2 \right\}. \end{aligned}$$

But the balls form a covering of V , therefore

$$-\phi'(t) \geq \frac{r^{-2}}{MCc_+} \left\{ \sum_{x \in V} f_t(x)^2 \nu_x - \sum_i \nu^{(h)}(B_i) \left(\sum_{x \in B_i} f_t(x) \nu_{B_i}^{(h)}(x) \right)^2 \right\}.$$

We have

$$\sum_{x \in V} f_t(x)^2 \nu_x^{(h)} = \phi(t)$$

and, still on A_N

$$\begin{aligned} \sum_i \nu^{(h)}(B_i) \left(\sum_{x_j \in B_i} f_t(x_j) \nu_{B_i}^{(h)}(x_j) \right)^2 &= \sum_i \nu^{(h)}(B_i)^{-1} \left(\sum_{x_j \in B_i} f_t(x_j) \nu_{x_j}^{(h)} \right)^2 \\ &\leq \sum_i \frac{c_+}{c_-^2} r^{-d} \left(\sum_{x_j \in B_i} f_t(x_j) \nu_{x_j}^{(h)} \right)^2 \leq \frac{c_+}{c_-^2} r^{-d} \left(\sum_i \sum_{x_j \in B_i} f_t(x_j) \nu_{x_j}^{(h)} \right)^2 \\ &= \frac{c_+}{c_-^2} r^{-d} \left(\sum_{x_j \in V} f_t(x_j) \nu_{x_j}^{(h)} \sum_{i: x_j \in B_i} 1 \right)^2 \\ &\leq \frac{c_+}{c_-^2} r^{-d} \left(\sum_{x_j \in V} f_t(x_j) \nu_{x_j}^{(h)} M \right)^2 = \frac{M^2 c_+}{c_-^2} r^{-d} \end{aligned}$$

where $h_- \leq r \leq r_0$, $h_- \leq h \leq h_0$ and we have used Theorem 21 to bound $\nu^{(h)}(B_i) \geq r^d c_-^2 / c_+$. Hence

$$-\phi'(t) \geq \frac{r^{-2}}{MCc_+} \left\{ \phi(t) - \frac{c_+ M^2}{c_-^2} r^{-d} \right\},$$

as long as $h_- \leq r \leq r_0$, $h_- \leq h \leq h_0$. Let $r = r(t) = \left(\frac{\phi(t)}{K} \right)^{-1/d}$, $K = 2c_+ M^2 / c_-^2 \vee 2c_+ / c_-$. Define $t_1(x, h)$ the random time

$$t_1(x, h) = \inf \{ t \geq 0 : r(t) \geq r_0 \} = \inf \left\{ t \geq 0 : \phi(t) \leq K r_0^{-d} \right\}, \quad t_1(h) = \inf_x t_1(x, h);$$

now, using Lemma 23 and since $K \geq 2c_+ / c_-$

$$\forall t \geq t_0(h)/2, \quad \phi(t) \leq \frac{2c_+}{c_-} h^{-d} \leq K h^{-d} \leq K h_-^{-d},$$

which in turns implies that $r(t) \geq h_-$. Moreover by definition of $t_1(h)$ we have

$$\forall t \leq t_1, \quad \phi(t) \geq K r_0^{-d} \quad \text{so that} \quad r(t) \leq r_0.$$

Therefore $h_- \leq r(t) \leq r_0$ whenever $t_0(h)/2 \leq t \leq t_1(h)$.

Then on A_N we obtain

$$\forall t \in [t_0(h)/2, t_1(h)], \quad -\phi'(t) \geq \frac{(\phi(t)/K)^{2/d}}{MCc_+} \frac{1}{2} \phi(t) = \frac{1}{2MCK^{2/d}c_+} \phi(t)^{1+2/d}.$$

Integrating yields, for any $t_1(h) \geq t \geq t_0(h)/2$

$$\left[\frac{d}{2} \phi^{-2/d} \right]_{t_0(h)/2}^t \geq \frac{1}{2MCK^{2/d}c_+^2} (t - t_0(h)/2) \geq \frac{t}{4MCK^{2/d}c_+^2}$$

which implies in particular

$$\frac{d}{2} \phi(t)^{-2/d} \geq \frac{d}{2} \phi(t_0(h)/2)^{-2/d} + \frac{t}{4MCK^{2/d}c_+^2} \geq \frac{t}{4MCK^{2/d}c_+^2}$$

i.e

$$\forall t \in [t_0(h)/2, t_1(h)], \quad \phi(t) \leq \left(\frac{MCc_+^2 K^{2/d}}{dt} \right)^{d/2}.$$

To conclude we apply Cauchy-Schwarz inequality to find

$$\begin{aligned} \forall x, y \in V, \quad p_t(x, y) &= \sum_{j=1}^N e^{-t\lambda_j} u_j(x) u_j(y) \\ &\leq \left(\sum_{j=1}^N e^{-t\lambda_j} u_j(x)^2 \right)^{1/2} \left(\sum_{j=1}^N e^{-t\lambda_j} u_j(y)^2 \right)^{1/2} \\ &= \sqrt{p_t(x, x) p_t(y, y)} \\ &\leq \max_{x \in V} p_t(x, x) \end{aligned}$$

■

Lemma 27. *For all $h_- \leq h \leq h_0$, $h \leq r \leq r_0$, on A_N there exists a covering of V by balls $B_\rho(x_i, r)$ such that the balls $B_\rho(x_i, r/2)$ are disjoint and each x_i belongs to at most $3^d (c_+/c_-)^3$ balls $B_\rho(x_j, r)$.*

Proof Take a maximal packing $(B_\rho(x_i, r/2))$. Then the balls $B_\rho(x_i, r/2)$ cover V : indeed, if there exists $x_j \in V$ such that $\rho(x_i, x_j) \geq r$ for all i , then this contradicts the maximal packing property by considering the ball $B_\rho(x_j, r/2)$. Moreover if $y \in V$ belong to M of the balls $B_\rho(x_i, r)$ (and we may assume without loss of generality that these balls are those corresponding to $i = 1, \dots, M$), using theorem 21 we have

$$\begin{aligned} \frac{Mc_-^2}{c_+N} (r/2)^d &\leq M \min_i \nu(B_\rho(x_i, r/2)) \leq \sum_{i=1}^M \nu(B_\rho(x_i, r/2)) = \nu \left(\bigsqcup_{i=1}^M B_\rho(x_i, r/2) \right) \\ &\leq \nu \left(\bigsqcup_{i=1}^M B_\rho(y, 3/2r) \right) = \nu(B_\rho(y, 3/2r)) \leq \frac{c_+^2}{c_-N} (3/2r)^d, \end{aligned}$$

which implies the result. ■

B.2.2 ESCAPE FROM THE ORIGIN FOR THE MARKOV PROCESS W_t

In this section we find an upper bound on the quantities $m(t, x) = \mathbb{E}_x^{(h)} [\rho(x, W_t)]$. Let $q(t, x) = \mathbb{E}_x^{(h)} [\ln p_t^{(h)}(x, W_t)] = -\sum_{y \in V} p_t^{(h)}(x, y) \ln p_t^{(h)}(x, y) \nu_y^{(h)}$ as well as $M(s, x) = m(st_0, x)/\sqrt{t_0}$, $Q(s, x) = q(st_0, x)$, $t_0(h) = 64h^2 \ln(2c_- N h^d)$.

Lemma 28. *For all $x \in V$,*

1.

$$\forall t > 0, \quad q'(t, x) \geq \frac{1}{4C_0^2} m'(t, x)^2, \quad \text{and} \quad \forall s \geq 1, \quad Q'(s, x) \geq \frac{1}{4C_0^2} M'(s)^2$$

2. On A_N we have

$$\forall 1 \leq s \leq t_1(h)/t_0(h), \quad Q(s, x) \geq -\left(\ln K - \frac{d}{2} \ln(st_0(h))\right) = \frac{d}{2} \ln s + \frac{d}{2} \ln t_0(h) - \ln K$$

3.

$$1 + M(s, x) \geq e^{-\frac{e+1}{ed}} e^{Q(s)/d}$$

Proof

1. Let $f_t(y) = p_t^{(h)}(x, y)$, $b_t(y, z) = f_t(y) + f_t(z)$. Throughout the proof we write $m(t), M(t), q(t), Q(t)$ in place of $m(t, x), M(t, x), q(t, x), Q(t, x)$ and $t_0 = t_0(h)$ for shortness sake. We have

$$\begin{aligned} & |m'(t)| \\ &= \left| \frac{\partial}{\partial t} \sum_{y \in V} \rho(x, y) f_t(y) \nu_y^{(h)} \right| = \left| \sum_{y \in V} \rho(x, y) (\mathcal{L} f_t)(y) \nu_y^{(h)} \right| \\ &= \left| \langle \rho(x, \cdot) | \mathcal{L} f_t \rangle_{L^2(\nu)} \right| = \left| \frac{1}{2h^2 \mu(V)} \sum_{y \sim z} (\rho(x, y) - \rho(x, z)) (f_t(y) - f_t(z)) \right| \\ &\leq \frac{C_0}{h\mu(V)} \sum_{y \sim z} |f_t(y) - f_t(z)| = \frac{C_0}{h\mu(V)} \sum_{y \sim z} \frac{|f_t(y) - f_t(z)|}{|f_t(y) + f_t(z)|^{1/2}} (f_t(y) + f_t(z))^{1/2} \\ &\leq \frac{C_0}{h\mu(V)} \left\{ \sum_{y \sim z} \frac{(f_t(y) - f_t(z))^2}{f_t(y) + f_t(z)} \right\}^{1/2} \left\{ \sum_{y \sim z} (f_t(y) + f_t(z)) \right\}^{1/2} \\ &= \frac{C_0}{h\mu(V)} \left\{ \sum_{y \sim z} \frac{(f_t(y) - f_t(z))^2}{f_t(y) + f_t(z)} \right\}^{1/2} \{2\mu(V)\}^{1/2} = \frac{C_0 \sqrt{2}}{h\mu(V)^{1/2}} \left\{ \sum_{y \sim z} \frac{(f_t(y) - f_t(z))^2}{f_t(y) + f_t(z)} \right\}^{1/2}. \end{aligned}$$

Now using the inequality (page 157 in Barlow (2017))

$$\forall u, v > 0, \quad \frac{(u - v)^2}{u + v} \leq (u - v) (\ln u - \ln v)$$

We find

$$m'(t)^2 \leq \frac{2C_0^2}{h^2 \mu(V)} \sum_{y \sim z} (f_t(y) - f_t(z)) (\ln f_t(y) - \ln f_t(z))$$

On the other hand

$$\begin{aligned}
q'(t) &= \frac{\partial}{\partial t} \sum_{y \in V} -f_t(y) \ln f_t(y) \nu_y^{(h)} \\
&= - \sum_{y \in V} \left\{ (-\mathcal{L}f_t)(y) \ln f_t(y) + f_t(y) \frac{(-\mathcal{L}f_t)(y)}{f_t(y)} \right\} \nu_y^{(h)} \\
&= \sum_{y \in V} (\mathcal{L}f_t)(y) \{1 + \ln f_t(y)\} \nu_y^{(h)} \\
&= \frac{1}{2h^2 \mu(V)} \sum_{y \sim z} (f_t(y) - f_t(z)) (\ln f_t(y) - \ln f_t(z))
\end{aligned}$$

Hence

$$q'(t) \geq \frac{1}{4C_0^2} m'(t)^2$$

To conclude we use the fact that $Q'(s) = t_0 q'(st_0)$, $M'(s) = m'(st_0) \sqrt{t_0}$ which implies via $t = st_0$

$$q'(t) \geq \frac{1}{4C_0^2} m'(t)^2 \quad \text{and} \quad Q'(s) \geq \frac{t_0}{4C_0^2} [M'(s)/\sqrt{t_0}]^2 = \frac{1}{4C_0^2} M'(s)^2$$

2. Using lemma 26, for any $h_- \leq h \leq h_0$, $x, y \in V$

$$\forall 1 \leq s \leq t_1(h)/t_0, \quad p_{st_0}(x, y) \leq K (st_0)^{-d/2}.$$

Therefore

$$\forall 1 \leq s \leq t_1(h)/t_0, \quad Q(s) \geq - \left(\ln K - \frac{d}{2} \ln(st_0) \right) = \frac{d}{2} \ln s + \frac{d}{2} \ln t_0 - \ln K$$

3. Let $a \leq 1$, $b \geq 0$. Then

$$\begin{aligned}
-Q(s) + aM(s) + b &= \mathbb{E}_x \left[\ln p_{st_0}^{(h)}(x, W_{st_0}) + a\rho(x, W_{st_0})/\sqrt{t_0} + b \right] \\
&= \sum_{y \in V} \nu_y^{(h)} p_{st_0}^{(h)}(x, y) \left\{ \ln p_{st_0}^{(h)}(x, y) + a\rho(x, y)/\sqrt{t_0} + b \right\}
\end{aligned}$$

Using the inequality $u(\ln u + \lambda) \geq -e^{-1-\lambda}$ and setting $a = (1 + M(s))^{-1}$, $b = d \ln(1 + M(s))$ we find

$$\begin{aligned}
-Q(s) + aM(s) + b &= \sum_{y \in V} \nu_y^{(h)} p_{st_0}^{(h)}(x, y) \left\{ \ln p_{st_0}^{(h)}(x, y) + a\rho(x, y)/\sqrt{t_0} + b \right\} \\
&\geq - \sum_{y \in V} \exp \left\{ -1 - a\rho(x, y)/\sqrt{t_0} - b \right\} \nu_y^{(h)} \\
&= -e^{-1-b} \sum_{y \in V} \exp \left\{ -a\rho(x, y)/\sqrt{t_0} \right\} \nu_y^{(h)} \\
&\geq -e^{-1-b} \sum_{y \in V} \nu_y^{(h)} = -e^{-1-b} \geq -e^{-1}.
\end{aligned}$$

Hence

$$-Q(s) + 1 + d \ln(1 + M(s)) \geq -Q(s) + \frac{M(s)}{1 + M(s)} + d \ln(1 + M(s)) \geq -e^{-1}$$

and rearranging yields

$$1 + M(s) \geq e^{-\frac{e+1}{ed}} e^{Q(s)/d}$$

■

Lemma 29. *There exists $c = c_{\mathcal{M}, p_0, d}$ such that, on A_N we have, for every $h_- \leq h \leq h_0$ and $t_0(h) \leq t \leq t_1(h)$*

$$m(t, x) \leq c \ln N \sqrt{t}, \quad \forall x \in V$$

Proof We follow the proof of Lemma 6.13 in Barlow (2017) (it is stated for $s \geq 1$, but the same proof remains valid for $1 \leq s \leq t_1(h)/t_0$). As done in the previous computations we drop x in the notations $m(t, x), M(t, x), q(t, x), Q(t, x)$. From Lemma 28 we bound $M'(s) \leq 2C_0 \sqrt{Q'(s)}$ and defining $r(s) = \frac{1}{d} (Q(s) + \ln K - \frac{d}{2} \ln(st_0(h))) \geq 0$, we bound $\forall 1 \leq s \leq t_1(h)/t_0(h)$,

$$\begin{aligned} e^{-\frac{e+1}{ed}} K^{-1/d} \sqrt{st_0(h)} e^{r(s)} &\leq 1 + M(s) \\ &\leq 1 + M(1) + 2C_0 \sqrt{d} \int_1^s \left(r'(\sigma) + \frac{1}{2\sigma} \right)^{1/2} d\sigma \\ &\leq 1 + C_0 \sqrt{t_0(h)}/h + 2C_0 \sqrt{2ds} (1 + r(s)) \\ &= 1 + 8C_0 \sqrt{\ln(2c_- N h^d)} + 2C_0 \sqrt{2ds} (1 + r(s)) \end{aligned}$$

where we have used the bound

$$m(t_0) = \mathbb{E}_x^{(h)}[\rho(x, W_{t_0(h)})] \leq C_0 h \mathbb{E}(N_{t_0(h)}/h^2) = \frac{C_0 t_0(h)}{h}.$$

Dividing by $\sqrt{s} \geq 1$ we get

$$\left\{ e^{-\frac{e+1}{ed}} K^{-1/d} \sqrt{t_0} \right\} e^{r(s)} \leq \frac{1 + M(s)}{\sqrt{s}} \leq \left(1 + C_0 \sqrt{64 \ln(2c_- N h^d)} \right) + 2C_0 \sqrt{2d} (1 + r(s))$$

In particular, weakening the constants we find c that depend on d, K, C_0, c_- such that

$$\begin{aligned} e^{r(s)} \leq c t_0^{-1/2} \left(1 + \sqrt{\ln(N h^d) + r(s)} \right) &\iff \frac{e^{1 + \sqrt{\ln(N h^d) + r(s)}}}{1 + \sqrt{\ln(N h^d) + r(s)}} \leq c e t_0^{-1/2} \exp \left(\sqrt{\ln(N h^d)} \right) \\ &\iff \frac{e^{\frac{1}{2} (1 + \sqrt{\ln(N h^d) + r(s)})}}{\underbrace{\frac{1}{2} \left(1 + \sqrt{\ln(N h^d) + r(s)} \right)}_{\geq 2e^{1/2}}} \frac{1}{2} e^{\frac{1}{2} (1 + \sqrt{\ln(N h^d) + r(s)})} \\ &\leq c e t_0^{-1/2} \exp \left(\sqrt{\ln(N h^d)} \right), \end{aligned}$$

which implies

$$e^{\frac{1}{2}(1+\sqrt{\ln(Nh^d)+r(s)})} \leq ce^{1/2}t_0(h)^{-1/2} \exp\left(\sqrt{\ln(Nh^d)}\right).$$

In particular

$$r(s) \leq 2 \left\{ \ln\left(ce^{1/2}\right) + \ln\left(1/\sqrt{t_0(h)}\right) + \sqrt{\ln(Nh^d)} \right\}$$

and using $t_0(h) = 64h^2 \ln(2c_-Nh^d)$ and $h_- \leq h \leq h_0$ we find, for some $c = c_{d,c_-,K,C_0}$

$$r(s) \leq c \left(\ln \frac{1}{h_-} \vee \sqrt{\ln N} \right)$$

Since $h_- \gg \left(\frac{\ln N}{N}\right)^{1/d}$ this implies $\ln \frac{1}{h_-} \leq \sqrt{\ln N}$, and therefore

$$M(s) \leq c \ln N \sqrt{s}$$

■

B.2.3 ON DIAGONAL LOWER BOUND

In this section we set again $t_0 = t_0(h) = 64h^2 \ln(2c_-Nh^d)$.

Definition 30. For every $x \in V$ and $r > 0$ we define the exit time $\tau(x, r)$ of the ball $B_\rho(x, r)$ by

$$\tau(x, r) = \inf \{t \geq 0 : \rho(x, W_t) > r\} = \inf \{t \geq 0 : W_t \notin B_\rho(x, r)\}$$

Lemma 31. There exists $c = c_{\mathcal{M}, p_0, d}$ such that on A_N and with $2t_0 \leq t \leq t_1(h)$, $r = c \ln N \sqrt{t}$ we have

$$\mathbb{P}_x^{(h)}(\tau(x, r) \leq t) \leq \frac{1}{2}$$

Proof We have, with $\tau = \tau(x, r)$, $t_0 \leq t \leq t_1(h)$, $r > 0$

$$\begin{aligned} r\mathbb{P}_x^{(h)}(\tau \leq t) &= r\mathbb{E}_x^{(h)}[\mathbb{1}_{\tau \leq t}] \\ &\leq \mathbb{E}_x^{(h)}[r\mathbb{1}_{\rho(x, W_{t \wedge \tau}) > r}] \\ &\leq \mathbb{E}_x^{(h)}[\rho(x, W_{t \wedge \tau})] \\ &\leq \mathbb{E}_x^{(h)}[\rho(x, W_{2t})] + \mathbb{E}_x^{(h)}[\rho(W_{2t}, W_{t \wedge \tau})] \\ &= \mathbb{E}_x^{(h)}[\rho(x, W_{2t})] + \mathbb{E}_x^{(h)}\left[\mathbb{E}_x^{(h)}[\rho(W_{2t}, W_{t \wedge \tau}) | W_{t \wedge \tau}]\right] \\ &\leq \mathbb{E}_x^{(h)}[\rho(x, W_{2t})] + \sup_{z \in V, s \leq t} \mathbb{E}_z^{(h)}[\rho(z, W_{2t-s})] \end{aligned}$$

Now using lemma 29 we find, as long as $t_0(h) \leq t \leq t_1(h)$

$$\begin{aligned} r\mathbb{P}_x^{(h)}(\tau \leq t) &\leq m(2t, x) + \sup_{z \in V, s \leq t} m(2t-s, z) \\ &\leq c \ln N \left(\sqrt{t} + \sqrt{2t-s} \right) \leq c \ln N \sqrt{t} \end{aligned}$$

the value of c changing from line to line. Hence by setting $r = r(t) = 2c \ln N \sqrt{t}$ we have

$$\mathbb{P}_x^{(h)}(\tau(x, r) \leq t) \leq \frac{1}{2}.$$

■

Lemma 32. *There exists $c = c_{\mathcal{M}, p_0, d}$ such that on A_N , for any $t_0(h) \leq t \leq t_1(h) \wedge \left(\frac{r_0}{c \ln N}\right)^2$ we have*

$$\forall h_- \leq h \leq h_0, \forall x \in V, p_t^{(h)}(x, x) \geq c \frac{t^{-d/2}}{\ln^d N}$$

Proof We have, with $\tau = \tau(x, r)$, $t_0/2 \leq t \leq t_1(h)$, $r = r(t)$ from Lemma 31 and $B = B_\rho(x, r)$

$$\mathbb{P}_x^{(h)}(W_t \notin B) \leq \mathbb{P}_x^{(h)}(\tau(x, r) \leq t) \leq \frac{1}{2}$$

Hence

$$\begin{aligned} \frac{1}{2} &\leq \mathbb{P}_x^{(h)}(W_t \in B) = \sum_{y \in B} p_t^{(h)}(x, y) \nu_y^{(h)} \\ &\leq \left(\nu^{(h)}(B)\right)^{1/2} \left(\sum_{y \in B} p_t^{(h)}(x, y)^2 \nu_y^{(h)}\right)^{1/2} \leq \left(\nu^{(h)}(B)\right)^{1/2} \left(\sum_{y \in V} p_t^{(h)}(x, y)^2 \nu_y^{(h)}\right)^{1/2} \\ &= \left(\nu^{(h)}(B)\right)^{1/2} p_{2t}^{(h)}(x, x)^{1/2}, \end{aligned}$$

and therefore we have the inequality

$$p_{2t}^{(h)}(x, x) \geq \frac{1}{4\nu^{(h)}(B)}.$$

We now want to conclude using Theorem 21. For this we need to guarantee that

$$h_- \leq r(t) \leq r_0 \iff \left(\frac{h_-}{c \ln N}\right)^2 \leq t \leq \left(\frac{r_0}{c \ln N}\right)^2.$$

Since $2t \geq t_0 = 64h_-^2 \ln(2c_- N h_-^d) \geq 64h_-^2 \ln(2c_- N h_-^d)$ and h_- satisfies 2, we have $t \geq \left(\frac{h_-}{c \ln N}\right)^2$. Since by assumption $t \leq \left(\frac{r_0}{c \ln N}\right)^2$, we conclude that $h_- \leq r(t) \leq r_0$. Hence theorem 21 implies

$$\nu^{(h)}(B) \leq \frac{c_+^2}{c_-} r(t)^d$$

and therefore

$$p_{2t}^{(h)}(x, x) \geq \frac{c_-}{4c_+^2} r(t)^{-d} = \frac{c_- c^{-d}}{4c_+^2} \frac{t^{-d/2}}{\ln^d N}.$$

■

Lemma 33. *There exists $c = c_{\mathcal{M}, p_0, d}$ such that on A_N we have, for any $h_- \leq h \leq h_0$*

$$t_1(h) \geq \frac{c}{\ln^2 N}$$

Proof If $t_1(h) < \left(\frac{r_0}{c \ln N}\right)^2$ then for any $t_0/2 \leq t \leq t_1(h)$ using lemma 32 we have

$$p_{2t}^{(h)}(x, x) \geq \frac{c_- c^{-d}}{4c_+^2} \frac{t^{-d/2}}{\ln^d N}.$$

In particular

$$K' = p_{2t_1(h)}^{(h)}(x, x) \geq \frac{c_- c^{-d}}{4c_+^2} \frac{t_1(h)^{-d/2}}{\ln^d N},$$

i.e.

$$t_1(h) \geq \left(\frac{c_- c^{-d}}{4c_+^2} \frac{K'^{-1}}{\ln^d N} \right)^{2/d}.$$

Hence

$$t_1(h) \geq \left(\frac{r_0}{c \ln N} \right)^2 \wedge \left(\frac{c_- c^{-d}}{4c_+^2} \frac{K'^{-1}}{\ln^d N} \right)^{2/d},$$

which concludes the proof. ■

B.3 Control of the Laplacian spectrum

Recall that here $t_0 = t_0(h) = 64h^2 \ln(2c_- N h^d)$ and $h_0 = h_{0|\mathcal{M}, p_0, d}$ is a theoretical constant.

Lemma 34. *There exists constants $b_1, b_2, b_3 > 0$ such that, on A_N , for any $h_- \leq h \leq h_0$*

$$\forall b_1 \ln^d N \leq j \leq b_2 t_0^{-d/2}, \quad \lambda_j^{(h)} \geq b_3 j^{2/d}$$

Proof Using the inequality $\mathbb{1}_{\lambda \leq \Lambda} \leq e^{1-\lambda/\Lambda}$ we find

$$\begin{aligned} \# \left\{ 1 \leq j \leq N : \lambda_j^{(h)} \leq \Lambda \right\} &= \sum_{j=1}^N \mathbb{1}_{\lambda_j^{(h)} \leq \Lambda} \\ &= \sum_{j=1}^N \mathbb{1}_{\lambda_j^{(h)} \leq \Lambda} \|u_j\|_N^2 \\ &= \sum_{j=1}^N \mathbb{1}_{\lambda_j^{(h)} \leq \Lambda} \sum_{x \in V} u_j(x)^2 \nu_x \\ &= \sum_{x \in V} \left\{ \sum_{j=1}^N \mathbb{1}_{\lambda_j^{(h)} \leq \Lambda} u_j(x)^2 \right\} \nu_x \leq \sum_{x \in V} \left\{ \sum_{j=1}^N e^{1-\lambda_j^{(h)}/\Lambda} u_j(x)^2 \right\} \nu_x \\ &= \sum_{x \in V} e p_{\Lambda^{-1}}^{(h)}(x, x) \nu_x \\ &\leq e \max_{x \in V} p_{\Lambda^{-1}}^{(h)}(x, x) \end{aligned}$$

We now use theorem 22 with $t = \Lambda^{-1}$: there exist constants $a_2, a_3, h_0 > 0$ (that also depend on \mathcal{M}, p_0) such that, on A_N , for all $h_- \leq h \leq h_0$ and $a_1^{-1} \ln^2 N \leq \Lambda \leq t_0^{-1}$

$$\forall x \in V, \quad p_t^{(h)}(x, x) \leq a_3 \Lambda^{d/2}$$

which implies

$$\# \{1 \leq j \leq N : \lambda_j \leq \Lambda\} \leq e a_3 \Lambda^{d/2}$$

In particular $j = 2e a_3 \Lambda^{d/2} \iff \Lambda = (j/2e a_3)^{2/d}$ yields

$$\lambda_j^{(h)} > \Lambda = (j/2e a_3)^{2/d}$$

This is valid if

$$a_1^{-1} \ln^2 N \leq (j/2e a_3)^{2/d} \leq t_0^{-1}$$

which gives the result by inverting the inequality. ■

Lemma 35. *There exist constants $b_4, b_5, b_6 > 0$ such that on A_N , for any $h_- \leq h \leq h_0$*

$$\forall b_4 \leq j \leq b_5 \frac{t_0^{-d/2}}{\ln^{3d/2} N}, \quad \lambda_j^{(h)} \leq b_6 j^{2/d} \ln^3 N$$

Proof As in in the proof of Lemma 3.19 in Coulhon, Kerkycharian, and Petrushev (2012) we start by the following inequality

$$e^{-t\lambda} = \mathbb{1}_{[0, \Lambda]}(\lambda) e^{-t\lambda} + \sum_{l \geq 0} \mathbb{1}_{[2^l \Lambda, 2^{l+1} \Lambda]}(\lambda) e^{-t\lambda} \leq \mathbb{1}_{[0, \Lambda]}(\lambda) + \sum_{l \geq 0} e^{-t2^l \Lambda} \mathbb{1}_{[2^l \Lambda, 2^{l+1} \Lambda]}(\lambda)$$

which holds for any $t, \Lambda > 0$. Hence

$$\begin{aligned} p_t^{(h)}(x, x) &= \sum_{j=1}^N e^{-t\lambda_j} u_j(x)^2 \\ &\leq \sum_{j=1}^N \mathbb{1}_{[0, \Lambda]}(\lambda_j^{(h)}) u_j(x)^2 + \sum_{l \geq 0} e^{-t2^l \Lambda} \sum_{j=1}^N \mathbb{1}_{[2^l \Lambda, 2^{l+1} \Lambda]}(\lambda_j^{(h)}) u_j(x)^2 \end{aligned}$$

Since for any $h > 0$, $j = 1, \dots, N$ we have $\lambda_j^{(h)} \leq 2h^{-2}$ (because $\|\mathcal{L}\|_{L^\infty(\nu)} \leq 2h^{-2}$), we can stop the sum at $L \in \mathbb{N}$ such that $2^L \Lambda \leq 2h^{-2} \leq 2^{L+1} \Lambda$:

$$p_t^{(h)}(x, x) \leq \sum_{j=1}^N \mathbb{1}_{[0, \Lambda]}(\lambda_j^{(h)}) u_j(x)^2 + \sum_{l=0}^L e^{-t2^l \Lambda} \sum_{j=1}^N \mathbb{1}_{[2^l \Lambda, 2^{l+1} \Lambda]}(\lambda_j^{(h)}) u_j(x)^2.$$

Using theorem 22 we find, on A_N and for $t_0 \leq t \leq \frac{a_1}{\ln^2 N}$

$$p_t^{(h)}(x, x) \geq a_2 \frac{t^{-d/2}}{\ln^d N} \implies a_2 \frac{t^{-d/2}}{\ln^d N} \leq \sum_{j=1}^N \mathbb{1}_{[0, \Lambda]}(\lambda_j^{(h)}) u_j(x)^2 + \sum_{l=0}^L e^{-t2^l \Lambda} \sum_{j=1}^N \mathbb{1}_{[2^l \Lambda, 2^{l+1} \Lambda]}(\lambda_j^{(h)}) u_j(x)^2$$

We now define $K \in \mathbb{N}$ such that $2^{K+1}\Lambda \leq t_0^{-1} \leq 2^{K+2}\Lambda$. If $a_1^{-1} \ln^2 N \leq \Lambda \leq t_0^{-1}$ then for $1 \leq l \leq K$ we have $a_1^{-1} \ln^2 N \leq \Lambda \leq 2^{l+1}\Lambda \leq t_0^{-1}$, and therefore

$$\begin{aligned} \sum_{l=0}^K e^{-t2^l\Lambda} \sum_{j=1}^N \mathbb{1}_{[2^l\Lambda, 2^{l+1}\Lambda]} \left(\lambda_j^{(h)} \right) u_j(x)^2 &\leq \sum_{l=0}^K e^{-t2^l\Lambda} \sum_{j=1}^N \mathbb{1}_{[0\Lambda, 2^{l+1}\Lambda]} \left(\lambda_j^{(h)} \right) u_j(x)^2 \\ &\leq \sum_{l=0}^K e^{-t2^l\Lambda} \sum_{j=1}^N e^{1-2^{-(l+1)}\Lambda^{-1}\lambda_j^{(h)}} u_j(x)^2 \\ &= ep_{2^{-(l+1)}\Lambda^{-1}}^{(h)}(x, x) \leq ea_3 2^{\frac{d(l+1)}{2}} \Lambda^{d/2}, \end{aligned}$$

which implies

$$\begin{aligned} a_2 \frac{t^{-d/2}}{\ln^d N} &\leq \sum_{j=1}^N \mathbb{1}_{[0, \Lambda]} \left(\lambda_j^{(h)} \right) u_j(x)^2 + ea_3 2^{d/2} \sum_{l \geq 0} e^{-t2^l\Lambda} \left(2^l \Lambda \right)^{d/2} + \sum_{j=1}^N \sum_{l=K+1}^L e^{-t2^l\Lambda} u_j(x)^2 \\ &\leq \sum_{j=1}^N \mathbb{1}_{[0, \Lambda]} \left(\lambda_j^{(h)} \right) u_j(x)^2 + ea_3 2^{d/2} \sum_{j=1}^N \left(2^l \Lambda \right)^{d/2} + L e^{-t2^{K+1}\Lambda} \sum_{j=1}^N u_j(x)^2. \end{aligned}$$

Taking the expectation of this quantity with respect to $x \sim \nu$ and using $\sum_{x \in V} u_j(x)^2 \nu_x = 1$ then yields

$$a_2 \frac{t^{-d/2}}{\ln^d N} \leq \# \left\{ j : \lambda_j^{(h)} \leq \Lambda \right\} + ea_3 2^{d/2} t^{-d/2} \sum_{l \geq 0} e^{-t2^l\Lambda} \left(2^l t \Lambda \right)^{d/2} + L N e^{-t2^{K+1}\Lambda}$$

Using an integral comparison (see Barlow, 2017, Appendix. A) we find, for any $x \geq 1$

$$\sum_{l \geq 0} e^{-2^l x} \left(2^l x \right)^{d/2} \leq c_d x^{d/2-1} e^{-x/2} \leq c_d c'_d e^{-x/4}, \quad c'_d = \sup_{x \geq 1} x^{d/2-1} e^{-x/4} < +\infty.$$

In particular, if $t \geq \Lambda^{-1}$, with $c = c_d c'_d$

$$\sum_{l \geq 0} e^{-t2^l\Lambda} \left(2^l t \Lambda \right)^{d/2} \leq c e^{-t\Lambda/4}.$$

Therefore we obtain : for any $a_1^{-1} \ln^2 N \leq \Lambda^{-1} \leq t \leq t_0^{-1}$

$$\begin{aligned} \# \left\{ j : \lambda_j^{(h)} \leq \Lambda \right\} &\geq t^{-d/2} \left(a_2 \frac{1}{\ln^d N} - c e a_3 2^{d/2} e^{-t\Lambda/4} - L N t^{d/2} e^{-t2^{K+1}\Lambda} \right) \\ &\geq t^{-d/2} \left(a_2 \frac{1}{\ln^d N} - c e a_3 2^{d/2} e^{-t\Lambda/4} - L N t^{d/2} e^{-t\Lambda} \right) \end{aligned}$$

Let $t = \kappa \Lambda^{-1} \ln N, \kappa > 0$. Then

$$\# \left\{ j : \lambda_j^{(h)} \leq \Lambda \right\} \geq \frac{\Lambda^{d/2}}{\ln^{d/2} N} \left(a_2 \frac{1}{\ln^d N} - c e a_3 2^{d/2} N^{-\kappa/4} - L N \Lambda^{d/2} \kappa^{-d/2} (\ln N)^{-d/2} N^{-\kappa} \right)$$

But notice that $L \leq \frac{\ln(2h^{-1}\Lambda^{-1})}{\ln 2} \leq \frac{\ln(2h^{-1}a_1(\ln N)^{-2})}{\ln 2} \lesssim \ln N$, which implies that for $\kappa > 0$ sufficiently large we find

$$\# \left\{ j : \lambda_j^{(h)} \leq \Lambda \right\} \geq \frac{a_2 \Lambda^{d/2}}{2 \ln^{3d/2} N}$$

To conclude we only need to guarantee the conditions on t, Λ, κ :

$$t_0 \leq \Lambda^{-1} \leq t = \kappa \Lambda^{-1} \ln N \leq \frac{a_1}{\ln^2 N} \iff \kappa a_1^{-1} \ln^3 N \leq \Lambda \leq t_0^{-1}$$

i.e we have obtained

$$\forall \kappa a_1^{-1} \ln^3 N \leq \Lambda \leq t_0^{-1}, \quad \# \left\{ j : \lambda_j^{(h)} \leq \Lambda \right\} \geq \frac{a_2 \Lambda^{d/2}}{2 \ln^{3d/2} N}$$

In particular with $j = \frac{a_2 \Lambda^{d/2}}{2 \ln^{3d/2} N}$ so that $\Lambda = (2j/a_2)^{2/d} \ln^3 N$ we get

$$\forall \frac{a_2}{2} (\kappa a_1^{-1})^{d/2} \leq j \leq \frac{a_2}{2} \frac{t_0^{-d/2}}{\ln^{3d/2} N}, \quad \lambda_j^{(h)} \leq (2j/a_2)^{2/d} \ln^3 N$$

This concludes the proof. ■

B.4 $L^\infty(\nu) - L^2(\nu)$ comparison theorem

Again recall that here $t_0 = t_0(h) = 32h^2 \ln(2c_- N h^d)$.

Lemma 36. *On A_N , for any $h_- \leq h \leq h_0$ with*

$$\Sigma_\Lambda = \text{span} \left\{ u_j^{(h)} : \lambda_j^{(h)} \leq \Lambda \right\}$$

we have

$$\forall a_1^{-1} \ln^2 N \leq \Lambda \leq t_0^{-1}, \quad f \in \Sigma_\Lambda, \quad \|f\|_{L^\infty(\nu)}^2 \leq e a_3 \Lambda^{d/2} \|f\|_{L^2(\nu)}^2$$

Moreover, with

$$\Sigma^J = \text{span} \left\{ u_j^{(h)} : j \leq J \right\}$$

we have

$$\forall b_4 \leq J \leq b_5 \frac{t_0^{-d/2}}{\ln^{3d/2} N}, \quad f \in \Sigma^J, \quad \|f\|_{L^\infty(\nu)}^2 \leq e a_3 b_6^{d/2} J \ln^{3d/2} N \|f\|_{L^2(\nu)}^2$$

Proof Since $f \in \Sigma_\Lambda$ we have $\mathbb{1}_{[0, \Lambda]}(\mathcal{L})(f) = f$, therefore, with $\mathbb{1}_{[0, \Lambda]}(\mathcal{L})(x, y)$ the kernel of $\mathbb{1}_{[0, \Lambda]}(\mathcal{L})$ with respect to $\nu^{(h)}$ is given by

$$\mathbb{1}_{[0, \Lambda]}(\mathcal{L})(x, y) = \sum_{j=1}^N \mathbb{1}_{[0, \Lambda]} \left(\lambda_j^{(h)} \right) u_j^{(h)}(x) u_j^{(h)}(y).$$

Then for any $x \in V$,

$$\begin{aligned}
 f(x)^2 &= [\mathbb{1}_{[0,\Lambda]}(\mathcal{L})(f)](x)^2 = \left[\sum_y \mathbb{1}_{[0,\Lambda]}(\mathcal{L})(x, y) f(y) \nu_y^{(h)} \right]^2 \\
 &\leq \left[\sum_y \mathbb{1}_{[0,\Lambda]}(\mathcal{L})(x, y)^2 \nu_y \right] \left[\sum_y f(y)^2 \nu_y^{(h)} \right] \leq \|f\|_N^2 \max_y \mathbb{1}_{[0,\Lambda]}(\mathcal{L})(x, y)^2 \\
 &= \|f\|_N^2 \max_y \left[\sum_{j=1}^N \mathbb{1}_{[0,\Lambda]}(\lambda_j) u_j^{(h)}(x) u_j^{(h)}(y) \right]^2 \\
 &\leq \|f\|_N^2 \max_y \left[\sum_{j=1}^N \mathbb{1}_{[0,\Lambda]}(\lambda_j) u_j^{(h)}(x)^2 \right] \left[\sum_{j=1}^N \mathbb{1}_{[0,\Lambda]}(\lambda_j) u_j^{(h)}(y)^2 \right] \\
 &= \|f\|_N^2 \max_y \mathbb{1}_{[0,\Lambda]}(\mathcal{L})(x, x) \mathbb{1}_{[0,\Lambda]}(\mathcal{L})(y, y) \leq \|f\|_N^2 \max_x \mathbb{1}_{[0,\Lambda]}(\mathcal{L})(x, x)^2 \\
 &= \|f\|_N^2 \max_x \sum_{j=1}^N \mathbb{1}_{[0,\Lambda]}(\lambda_j) u_j^{(h)}(x)^2 \\
 &\leq \|f\|_N^2 \max_x \sum_{j=1}^N e^{1-\lambda_j/\Lambda} u_j^{(h)}(x)^2 = \|f\|_N^2 \max_x ep_{\Lambda^{-1}}^{(h)}(x, x)
 \end{aligned}$$

Using theorem 22 we thus find

$$\forall a_1^{-1} \ln^2 N \leq \Lambda \leq t_0^{-1}, \quad \forall f \in \Sigma_\Lambda, \quad \|f\|_{L^\infty(\nu)}^2 \leq ea_3 \Lambda^{d/2} \|f\|_{L^2(\nu)}^2.$$

Now if $b_4 \leq J \leq b_5 \frac{t_0^{-d/2}}{\ln^{3d/2} N}$, lemma 35 implies $f \in \Sigma_\Lambda$, $\Lambda = b_6 J^{2/d} \ln^3 N$, which concludes the proof. \blacksquare

Appendix C. Proofs of Lemmas 16 and 17

C.1 Proof of Lemma 16

We first bound $T_h f$ for $\beta \leq 1$, then we treat the case $1 < \beta \leq 2$ followed by the case $\beta > 2$. Recall that $h \leq h_0 \wedge \pi r_0 / C_0$.

Case $\beta \leq 1$. In this case the result is trivial since $\forall x \in \mathcal{M}$,

$$\begin{aligned}
 |T_h f(x)| &= \left| \frac{1}{h^2 P_0(B_{\mathbb{R}^D}(x, h))} \int_{B_{\mathbb{R}^D}(x, h) \cap \mathcal{M}} (f(x) - f(y)) p_0(y) \mu(dy) \right| \\
 &\leq h^{-2} \sup_{y \in B_{\mathbb{R}^D}(x, h) \cap \mathcal{M}} |f(x) - f(y)| \leq h^{-2} \sup_{y \in B_{\mathbb{R}^D}(x, h)} \sum_{i \in I} |\chi_i f(x) - \chi_i f(y)| \\
 &\leq h^{-2} \max_{i \in I} \sup_{y \in B_{\mathbb{R}^D}(x, h)} |(\chi_i f \circ \phi_i^{-1})(\phi_i(y)) - (\chi_i f \circ \phi_i^{-1})(\phi_i(x))| \\
 &\leq h^{-2} \|f\|_{C^\beta(\mathcal{M})} \max_{i \in I} \sup_{y \in B_{\mathbb{R}^D}(x, h)} \|\phi_i(y) - \phi_i(x)\|^\beta
 \end{aligned}$$

But since $h \leq h_0$, for any $y \in B_{\mathbb{R}^D}(x, h) \cap \mathcal{M}$ we have $\rho(x, y) \leq C_0 \|x - y\| \leq C_0 h$. Hence since the functions ϕ_i are Lipschitz

$$|T_h f(x)| \leq h^{-2} \|f\|_{\mathcal{C}^\beta(\mathcal{M})} \max_{i \in I} \sup_{y \in B_\rho(x, h)} \|\phi_i(y) - \phi_i(x)\|^\beta \lesssim \|f\|_{\mathcal{C}^\beta(\mathcal{M})} h^{\beta-2}$$

Case $1 < \beta \leq 2$. Recall that C_0 is such that $\rho(x, y) \leq C_0 \|x - y\|$ for any $x, y \in \mathcal{M}$. Recall that $f_0 \in \mathcal{C}^\beta(\mathcal{M})$, so that using a Taylor expansion of $f_x = f \circ \phi_x$, with $\phi_x = \exp_x(\psi(x, \cdot))$ at some $x \in V$ and for any $\|x - y\| < h$ we have $\rho(x, y) \leq C_0 \|x - y\| \leq C_0 h$ and therefore

$$f_0(y) = f(x) + df_x(0) \cdot \omega + R(x, y)$$

where $\exp_x(\psi(x, \omega)) = y$ and

$$\max_{x \neq y \in \mathcal{M}} \frac{|R(x, y)|}{\rho(x, y)^\beta} \lesssim \|f\|_{\mathcal{C}^\beta(\mathcal{M})}.$$

This, together with the fact that when $x \sim y$ then $\|x - y\| < h$, implies that

$$\left| T_h f(x) - \frac{1}{h^2 P_0(B_{\mathbb{R}^D}(x, h))} \int_{\exp_x \circ \phi_x^{-1} B_{\mathbb{R}^D}(x, h)} df_x(0) \cdot \omega p_x(\omega) J_x(\omega) d\omega \right| \lesssim \frac{\|f\|_{\mathcal{C}^\beta} h^\beta}{h^2}.$$

where $p_x = p_0 \circ \phi_x$, $J_x = |g(\exp_x(\psi(x, \omega)))|^{1/2}$. It remains to bound

$$\frac{1}{h^2 P_0(B_{\mathbb{R}^D}(x, h))} \int_{\exp_x \circ \phi_x^{-1} B_{\mathbb{R}^D}(x, h) \cap \mathcal{M}} df_x(0) \cdot \omega p_x(\omega) J_x(\omega) d\omega$$

For this notice that since $p_0 \in \mathcal{C}^{\beta-1}(\mathcal{M})$ and $J_x(\omega) = 1 + \mathcal{O}(\|\omega\|^2)$ we have

$$p_x(\omega) J(\omega) = \left(p_0(x) + \mathcal{O}(h^{\beta-1}) \right) (1 + \mathcal{O}(h^2)) = p_0(x) + \mathcal{O}(h^{\beta-1})$$

Therefore

$$\begin{aligned} \frac{\int_{\exp_x \circ \phi_x^{-1} B_{\mathbb{R}^D}(x, h)} df_x(0) \cdot \omega p_x(\omega) J_x(\omega) d\omega}{P_0(B_{\mathbb{R}^D}(x, h))} &= \frac{\int_{\exp_x \circ \phi_x^{-1} (B_{\mathbb{R}^D}(x, h))} df_x(0) \cdot \omega (p_0(x) + \mathcal{O}(h^{\beta-1})) d\omega}{P_0(B_{\mathbb{R}^D}(x, h))} \\ &= p_0(x) \frac{\int_{\exp_x \circ \phi_x^{-1} (B_{\mathbb{R}^D}(x, h))} df_x(0) \cdot \omega d\omega}{P_0(B_{\mathbb{R}^D}(x, h))} + \mathcal{O}(\|f\|_{\mathcal{C}^\beta(\mathcal{M})} h^\beta) \end{aligned}$$

Note that since $\exp_x \circ \phi_x^{-1} (B_\rho(x, h)) = \{\omega \in T_x \mathcal{M} : \|\omega\| < h\}$ is a symmetric set, we get

$$\int_{\exp_x \circ \phi_x^{-1} (B_\rho(x, h))} df_x(0) \cdot \omega d\omega = 0,$$

therefore to control the first term of the right hand side of the above term, it is enough to show that $B_{\mathbb{R}^D}(x, h)$ is close to $B_\rho(x, h)$.

Since $h \leq \frac{\pi a_{\mathcal{M}}}{C_0}$, for all $x, y \in \mathcal{M}$, $\|x - y\| < h$ we have $\rho(x, y) \leq C_0 \|x - y\| \leq C_0 h \leq \pi a_{\mathcal{M}}$. Hence, using (12), $\rho(x, y) - \frac{\rho(x, y)^3}{24a_{\mathcal{M}}^2} \leq \|x - y\| \leq \rho(x, y)$. Therefore if

$y \in \mathcal{M}$, $\|x - y\| < h$ but $\rho(x, y) \geq h$ then $\rho(x, y) \leq \|x - y\| + \frac{\rho(x, y)^3}{24a_{\mathcal{M}}^2} \leq h + \frac{C_0^3}{24a_{\mathcal{M}}^2}h^3$. This implies

$$\begin{aligned} \text{vol}(\{\omega : \exp_x \circ \phi_x(\omega) \in B_\rho(x, h)^c \cap B_{\mathbb{R}^D}(x, h)\}) &\leq \text{vol}\left(\left\{\omega : h \leq \|\omega\| < h + \frac{C_0^3}{24a_{\mathcal{M}}^2}h^3\right\}\right) \\ &= \text{vol}(\{\omega : \|\omega\| \leq 1\}) \left(\left(h + \frac{C_0^3}{24a_{\mathcal{M}}^2}h^3\right)^d - h^d\right) \lesssim h^d \left(\left(1 + \frac{C_0^3}{24a_{\mathcal{M}}^2}h^2\right)^d - 1\right) \\ &\leq C_{\mathcal{M}}h^{d+2} \end{aligned}$$

where the constant $C_{\mathcal{M}}$ depends on \mathcal{M} through $a_{\mathcal{M}}, h_0, C_0$ only. Therefore,

$$\begin{aligned} &\int_{\exp_x \circ \phi_x^{-1}(B_{\mathbb{R}^D}(x, h))} df_x(0) \cdot \omega d\omega \\ &= \int_{\exp_x \circ \phi_x^{-1}(B_\rho(x, h))} df_x(0) \cdot \omega d\omega + \int_{\exp_x^{-1}(B_{\mathbb{R}^D}(x, h) \cap B_\rho(x, h)^c)} df_x(0) \cdot \omega d\omega \\ &= \mathcal{O}\left(\|f\|_{\mathcal{C}^\beta(\mathcal{M})} h^{d+3}\right) \leq \mathcal{O}\left(\|f\|_{\mathcal{C}^\beta(\mathcal{M})} h^{d+\beta}\right) \end{aligned}$$

where we have used that $\beta \leq 2 \leq 3$. Finally since $P_0(B(x, h)) \gtrsim h^d$ we obtain

$$\frac{1}{h^2 P_0(B(x, h))} \int_{\exp_x \circ \phi_x^{-1} B_{\mathbb{R}^D}(x, h)} df_x(0) \cdot \omega p_x(\omega) J_x(\omega) d\omega = \mathcal{O}\left(\|f\|_{\mathcal{C}^\beta(\mathcal{M})} h^{\beta-2}\right),$$

which in turns implies that

$$\|T_h f_0\|_{L^\infty(\mathcal{M})} = \mathcal{O}\left(\|f\|_{\mathcal{C}^\beta(\mathcal{M})} h^{\beta-2}\right).$$

Case $\beta > 2$. Let $k = \lceil \beta/2 \rceil - 1 \geq 1$. Since $f \in \mathcal{C}^\beta(\mathcal{M})$ and $p_0 \in \mathcal{C}^{\beta-1}(\mathcal{M})$, then for any $i \in I$, $x \in \mathcal{U}_i$, $y \in B_\rho(x, r_0)$ (with $I, (\mathcal{U}_i)_{i \in I}, (\psi_i)_{i \in I}, (\varphi_i)_{i \in I}$ the objects defined in section A)

$$f_i(y) = f_i(x) + \sum_{l=1}^k \left\{ \frac{d^{2l-1} f_x(0) \cdot \exp_x^{-1}(y)^{2l-1}}{(2l-1)!} + \frac{d^{2l} f_x(0) \cdot \exp_x^{-1}(y)^{2l}}{(2l)!} \right\} + R_{k,1}(x, y)$$

$$\begin{aligned} p_{0,i}(y) &= p_{0,i}(x) + \sum_{l=1}^{k-1} \left\{ \frac{d^{2l-1} p_x(0) \cdot \exp_x^{-1}(y)^{2l-1}}{(2l-1)!} + \frac{d^{2l} p_x(0) \cdot \exp_x^{-1}(y)^{2l}}{(2l)!} \right\} \\ &\quad + \frac{d^{2k-1} p_x(0) \cdot \exp_x^{-1}(y)^{2k-1}}{(2k-1)!} + R_{k,2}(x, y) \end{aligned}$$

$$\begin{aligned} \sqrt{|g|}_i(y) &= \sqrt{|g|}_i(x) + \sum_{l=1}^{k-1} \left\{ \frac{d^{2l-1} J_x(0) \cdot \exp_x^{-1}(y)^{2l-1}}{(2l-1)!} + \frac{d^{2l} J_x(0) \cdot \exp_x^{-1}(y)^{2l}}{(2l)!} \right\} \\ &\quad + \frac{d^{2k-1} p_x(0) \cdot \exp_x^{-1}(y)^{2k-1}}{(2k-1)!} + R_{k,3}(x, y) \end{aligned}$$

where

$$\begin{aligned} f_i &= \chi_i f, p_{0,i} = \chi_i p_0, \quad f_x = (\chi_i f) \circ \exp_x \circ \psi_i(x, \cdot) \\ p_x &= (\chi_i p_0) \circ \exp_x \circ \psi_i(x, \cdot), \quad J_x = \left(\chi_i \sqrt{|g|} \right) \circ \exp_x \circ \psi_i(x, \cdot) \end{aligned}$$

(we drop the index i for convenience), and the remainders satisfy

$$|R_{k,1}(x, y)| \leq \|f\|_{\mathcal{C}^\beta(\mathcal{M})} \rho(x, y)^\beta,$$

$$|R_{k,2}(x, y)| \leq \|p_0\|_{\mathcal{C}^{\beta-1}(\mathcal{M})} \rho(x, y)^{\beta-1},$$

$$|R_{k,3}(x, y)| \leq C_{\mathcal{M}} \rho(x, y)^{\beta-1},$$

where $C_{\mathcal{M}}$ depends on the Hölder constant of the exponential map. Thus, by Riemannian change of variables, there exist coefficients

$$A_{i,m} \in \mathcal{C}^{\beta-m} \left(\mathcal{U}_i, \mathcal{L} \left(\left(\mathbb{R}^d \right)^{\otimes m}, \mathbb{R} \right) \right)$$

and remainders

$$|B_{i,k}(x, \omega)| \leq C(\beta, \mathcal{M}, p_0) \|f\|_{\mathcal{C}^\beta(\mathcal{M})} \|\omega\|^\beta$$

such that

$$\begin{aligned} & \int_{B_{\mathbb{R}^D}(x, h)} (f_0(x) - f_0(y)) p_0(y) \mu(dy) \\ &= \int_{\psi_i(x, \cdot)^{-1} \circ \exp_x^{-1}(B_{\mathbb{R}^D}(x, h))} (f_x(0) - f_x(\omega)) p_x(\omega) J_x(\omega) d\omega \\ &= \sum_{i \in I} \int_{\psi_i(x, \cdot)^{-1} \circ \exp_x^{-1}(B_{\mathbb{R}^D}(x, h))} \left\{ \sum_{m=1}^{2k} A_{i,m}(x) \omega^m + B_{i,k}(x, \omega) \right\} d\omega \\ &= \sum_{i \in I} \left\{ \sum_{m=1}^{2k} A_{i,m}(x) \int_{\psi_i(x, \cdot)^{-1} \circ \exp_x^{-1}(B_{\mathbb{R}^D}(x, h))} \omega^m d\omega + \int_{\psi_i(x, \cdot)^{-1} \circ \exp_x^{-1}(B_{\mathbb{R}^D}(x, h))} B_{i,k}(x, \omega) d\omega \right\} \\ &= \sum_{i \in I} \sum_{m=1}^{2k} A_{i,m}(x) \int_{\psi_i(x, \cdot)^{-1} \circ \exp_x^{-1}(B_{\mathbb{R}^D}(x, h))} \omega^m d\omega + \mathcal{O} \left(\|f\|_{\mathcal{C}^\beta(\mathcal{M})} h^{d+\beta} \right). \end{aligned}$$

We thus get

$$T_h f(x) = \sum_{i \in I} \sum_{m=1}^{2k} A_{i,m}(x) \frac{h^{-(d+2)} \int_{\psi_i(x, \cdot)^{-1} \circ \exp_x^{-1}(B_{\mathbb{R}^D}(x, h))} \omega^m d\omega}{h^{-d} P_0(B_{\mathbb{R}^D}(x, h))} + \mathcal{O} \left(\|f\|_{\mathcal{C}^\beta(\mathcal{M})} h^{\beta-2} \right)$$

By lemma 18 we have, with $s_h(x, v) := \frac{t_h(x, v) - 1}{h^2}$

$$\psi_i(x, \cdot)^{-1} \circ \exp_x^{-1}(B_{\mathbb{R}^D}(x, h)) = \left\{ thv : v \in \mathbb{R}^d, \|v\| = 1, 0 \leq t \leq 1 + h^2 s_h(x, v) \right\}$$

And using the spherical coordinates,

$$\begin{aligned}
 & h^{-(d+2)} \int_{\psi_i(x, \cdot)^{-1} \circ \exp_x^{-1}(B_{\mathbb{R}^D}(x, h))} \omega^m d\omega \\
 &= h^{-(d+2)} \int_0^\pi \dots \int_0^\pi \int_0^{2\pi} \int_0^{1+h^3 s_h(x, v)} (r e_\theta)^m r^{d-1} \sin(\theta_1)^{d-2} \sin(\theta_2)^{d-3} \dots \sin(\theta_{d-1}) dr d\theta_1 \dots d\theta_{d-1} \\
 &= h^{-(d+2)} \int_0^\pi \dots \int_0^\pi \int_0^{2\pi} \int_0^{1+h^2 s_h(x, v)} (t e_\theta)^m (t h)^{d-1} \sin(\theta_1)^{d-2} \sin(\theta_2)^{d-3} \dots \sin(\theta_{d-1}) h dt d\theta_1 \dots d\theta_{d-1} \\
 &= h^{m-2} \int_0^\pi \dots \int_0^\pi \int_0^{2\pi} \int_0^{1+h^2 s_h(x, v)} (t e_\theta)^m t^{d-1} \sin(\theta_1)^{d-2} \sin(\theta_2)^{d-3} \dots \sin(\theta_{d-1}) dt d\theta_1 \dots d\theta_{d-1} \\
 &= h^{m-2} \int_{\mathbb{A}^d} \int_0^{1+h^2 s_h(x, v)} (t e_\theta)^m t^{d-1} K(\theta) dt d\theta \\
 &= \frac{h^{m-2}}{m+d} \int_{\mathbb{A}^d} e_\theta^m (1 + h^2 s_h(x, e_\theta))^{m+d} K(\theta) d\theta
 \end{aligned}$$

where $\mathbb{A}^d = (0, \pi)^{d-2} \times (0, 2\pi)$ and

$$e_\theta = \begin{bmatrix} \cos(\theta_1) \\ \sin(\theta_1) \cos(\theta_2) \\ \sin(\theta_1) \sin(\theta_2) \cos(\theta_3) \\ \vdots \\ \sin(\theta_1) \sin(\theta_2) \dots \sin(\theta_{d-2}) \cos(\theta_{d-1}) \\ \sin(\theta_1) \sin(\theta_2) \dots \sin(\theta_{d-2}) \sin(\theta_{d-1}) \end{bmatrix}, K(\theta) = \sin(\theta_1)^{d-2} \sin(\theta_2)^{d-3} \dots \sin(\theta_{d-1})$$

Moreover if m is odd, by symmetry

$$h^{-(d+2)} \int_{\psi_i(x, \cdot)^{-1} \circ \exp_x^{-1}(B_{\mathbb{R}^D}(x, h))} \omega^m d\omega = h^{-(d+2)} \int_{\psi_i(x, \cdot)^{-1} \circ \exp_x^{-1}(B_{\mathbb{R}^D}(x, h) \setminus B_\rho(0, h))} \omega^m d\omega$$

and therefore the same computations show that

$$\begin{aligned}
 h^{-(d+2)} \int_{\psi_i(x, \cdot)^{-1} \circ \exp_x^{-1}(B_{\mathbb{R}^D}(x, h))} \omega^m d\omega &= \frac{h^{m-2}}{m+d} \int_{\mathbb{A}^d} e_\theta^m \left((1 + h^2 s_h(x, e_\theta))^{m+d} - 1 \right) K(\theta) d\theta \\
 &= \frac{h^{m-2}}{m+d} \int_{\mathbb{A}^d} e_\theta^m \left(\sum_{r=1}^{m+d} \binom{m+d}{r} h^{2r} s_h(x, e_\theta)^r \right) K(\theta) d\theta \\
 &= \frac{h^m}{m+d} \sum_{r=1}^{m+d} \binom{m+d}{r} h^{2r-2} \int_{\mathbb{A}^d} e_\theta^m s_h(x, e_\theta)^r K(\theta) d\theta
 \end{aligned}$$

All in all for each $l \in \{1, \dots, k\}$, $m = 2l - 1$ or $2l$ we have

$$\begin{aligned}
 h^{-(d+2)} \int_{\psi_i(x, \cdot)^{-1} \circ \exp_x^{-1}(B_{\mathbb{R}^D}(x, h))} \omega^m d\omega &= \frac{h^{2(l-1)}}{m+d} \sum_{r=1}^{m+d} \binom{m+d}{r} h^{2r-1} \int_{\mathbb{A}^d} e_\theta^m s_h(x, e_\theta)^r K(\theta) d\theta \text{ if } m = 2l - 1 \\
 h^{-(d+2)} \int_{\psi_i(x, \cdot)^{-1} \circ \exp_x^{-1}(B_{\mathbb{R}^D}(x, h))} \omega^m d\omega &= h^{2(l-1)} \int_{\mathbb{A}^d} e_\theta^m (1 + h^2 s_h(x, e_\theta))^{m+d} K(\theta) d\theta \text{ if } m = 2l
 \end{aligned}$$

This gives

$$T_h f(x) = \frac{1}{h^{-d} P_0(B(x, h))} \sum_{i \in I} \sum_{l=1}^k h^{2(l-1)} \{A_{i,2l-1}(x) V_{i,h,2l-1}(x) + A_{i,2l}(x) V_{i,h,2l}(x)\} + \mathcal{O}(\|f\|_{\mathcal{C}^\beta(\mathcal{M})} h^{\beta-2})$$

where by lemma 18, composition and integration we have

$$\max_{\alpha \in \mathbb{N}^d, |\alpha| \leq \alpha-5} \sup_{\substack{i \in I \\ y \in \varphi_i^{-1}(\mathcal{U}_i) \\ 0 < h < h_+ \\ \|v\|_{T_{\varphi_i(y)} \mathcal{M}} = 1}} \left\| \frac{\partial^{|\alpha|}}{\partial x^\alpha} V_{i,h,m}(\varphi_i(y)) \right\|_{(\mathbb{R}^d)^{\otimes m}} < +\infty$$

To conclude it suffices to show that $x \mapsto h^{-d} P_0(B_{\mathbb{R}^D}(x, h))$ is lower bounded and satisfies

$$\sup_{0 < h < h_+} \left\| h^{-d} P_0(B_{\mathbb{R}^D}(\cdot, h)) \right\|_{\mathcal{C}^{\beta-1}(\mathcal{M})} < +\infty$$

Since p_0 is lower bounded and $\mu(B_{\mathbb{R}^D}(x, h)) \asymp h^d$ (as can be seen by the Riemannian change of variables formula), we get

$$h^{-d} \int_{B_{\mathbb{R}^D}(x, h)} p_0(y) \mu(dy) \geq \inf_{\mathcal{M}} p_0 \times h^{-d} \mu(B_{\mathbb{R}^D}(x, h)) \gtrsim \inf_{\mathcal{M}} p_0$$

hence $x \mapsto h^{-d} P_0(B_{\mathbb{R}^D}(x, h))$ is indeed lower bounded. For the derivatives, doing the same computations as above

$$\begin{aligned} h^{-d} P_0(B_{\mathbb{R}^D}(x, h)) &= \sum_{i \in I} h^{-d} \int_{B_{\mathbb{R}^D}(x, h)} p_0(y) \mu(dy) \\ &= h^{-d} \int_{\varphi_i^{-1} \circ \exp_x^{-1}(B_{\mathbb{R}^D}(x, h))} (\chi_i p_0) \circ \exp_x(\psi_i(x, \omega)) \sqrt{|g(\exp_x(\psi_i(x, \omega)))|} d\omega \\ &= h^{-d} \int_{\varphi_i^{-1} \circ \exp_x^{-1}(B_{\mathbb{R}^D}(x, h))} p_x(\omega) J_x(\omega) d\omega \\ &= h^{-d} \int_{\mathbb{A}^d} \int_0^{1+h^2 s_h(x, v)} p_x(the_\theta) J_x(the_\theta) (th)^{d-1} K(\theta) h dt d\theta \\ &= \int_{\mathbb{A}^d} \int_0^{1+h^2 s_h(x, v)} p_x(the_\theta) J_x(the_\theta) t^{d-1} K(\theta) dt d\theta \end{aligned}$$

And again, under this form, by composition and integration, $x \mapsto h^{-d} P_0(B_{\mathbb{R}^D}(x, h))$ has indeed the regularity of p_0 , i.e is $\mathcal{C}^{\beta-1}(\mathcal{M})$ with uniform bound on the derivatives as $h \rightarrow 0$.

This concludes the proof by setting

$$g_h^{(l)} = \frac{1}{h^{-d} P_0(B_{\mathbb{R}^D}(x, h))} \sum_{i \in I} h A_{i,2l-1}(x) V_{i,h,2l-1}(x) + A_{i,2l}(x) V_{i,h,2l}(x)$$

which is $\mathcal{C}^{\beta-2l}$ since by assumption $\alpha \geq \beta + 3$ which implies $\alpha - 5 \geq \beta - 2 \geq \beta - 2l$ for any $l \in \{1, \dots, k\}$. ■

C.2 Proof of Lemma 17

First if $\beta \leq 1$, $\|T_{h_n} f\|_{L^\infty(\mathcal{M})} \lesssim h_n^{\beta-2}$ follows from lemma 16. For $\|\mathcal{L}f\|_{L^\infty(\nu)}$ notice that for each $x \in V$ we have

$$|(\mathcal{L}f)(x)| = \frac{1}{h_n^2 \mu_x} \left| \sum_{i=1}^N (f(x) - f(x_i)) \mathbb{1}_{x_i \sim x} \right| \lesssim \|f\|_{\mathcal{C}^\beta(\mathcal{M})} h_n^{\beta-2}.$$

Now, if $\beta > 1$,

$$\forall x \in V, h_n^2 \mu_x \mathcal{L}f(x) = \sum_{i=1}^N (f(x) - f(x_i)) \mathbb{1}_{x_i \sim x},$$

and therefore by Bernstein's inequality (see Van der Vaart and Wellner, 1996, Part. 2), using $|f(x) - f(x_i)| \mathbb{1}_{x \sim x_i} \lesssim \|f\|_{\mathcal{C}^\beta(\mathcal{M})} h_n$,

$$\mathbb{P}_0(|h_n^2 \mu_x \mathcal{L}f(x) - N h_n^2 P_0(B_{\mathbb{R}^D}(x, h_n)) T_{h_n} f(x)| > u) \leq \exp\left(-c \frac{u^2}{N h_n^{d+2} + h_n u}\right)$$

and

$$\begin{aligned} \mathbb{P}_0(|\mu_x - N P_0(B_{\mathbb{R}^D}(x, h_n))| > u/h_n^2) &\leq \exp\left(-c \frac{u^2/h_n^4}{N h_n^d + u/h_n^2}\right) \\ &\leq \exp\left(-c \frac{u^2}{N h_n^{d+4} + u h_n^2}\right). \end{aligned}$$

Hence

$$\begin{aligned} \mathbb{P}_0(|h_n^2 \mu_x \mathcal{L}f(x) - h_n^2 \mu_x T_{h_n} f(x)| > 2u) &\leq \mathbb{P}_0(|h_n^2 \mu_x \mathcal{L}f(x) - N h_n^2 P_0(B_{\mathbb{R}^D}(x, h_n)) T_{h_n} f(x)| > u) \\ &\quad + \mathbb{P}_0(|N h_n^2 P_0(B_{\mathbb{R}^D}(x, h_n)) T_{h_n} f(x) - h_n^2 \mu_x T_{h_n} f(x)| > u) \\ &\leq \exp\left(-c \frac{u^2}{N h_n^{d+2} + h_n u}\right) + \mathbb{P}_0(|N P_0(B_{\mathbb{R}^D}(x, h_n)) - \mu_x| > u/h_n^2 \|f\|_{\mathcal{C}^\beta(\mathcal{M})}) \\ &\leq \exp\left(-c \frac{u^2}{N h_n^{d+2} + h_n u}\right) + \exp\left(-c \frac{u^2}{N h_n^{d+4} + u h_n^2}\right) \end{aligned}$$

Therefore, using the fact that $\mu_x \lesssim N h_n^d$ on $V_N(c)$

$$\begin{aligned} \mathbb{P}_0(|\mathcal{L}f(x) - T_{h_n} f(x)| > u) &\leq \mathbb{P}_0(V_N(c)^c) + \mathbb{P}_0(|h_n^2 \mu_x \mathcal{L}f(x) - h_n^2 \mu_x T_{h_n} f(x)| > c N h_n^{d+2} u) \\ &\leq \mathbb{P}_0(V_N(c)^c) + \exp\left(-c \frac{(N h_n^{d+2} u)^2}{N h_n^{d+2} + N h_n^{d+3} u}\right) + \exp\left(-c \frac{(N h_n^{d+2} u)^2}{N h_n^{d+4} + N h_n^{d+4} u}\right) \\ &= \mathbb{P}_0(V_N(c)^c) + \exp\left(-c \frac{N^2 h_n^{2d+4} u^2}{N h_n^{d+2} + N h_n^{d+3} u}\right) + \exp\left(-c \frac{N^2 h_n^{2d+4} u^2}{N h_n^{d+4} + N h_n^{d+4} u}\right) \\ &= \mathbb{P}_0(V_N(c)^c) + \exp\left(-c \frac{N h_n^{d+2} u^2}{1 + h_n u}\right) + \exp\left(-c \frac{N h_n^d u^2}{1 + u}\right) \end{aligned}$$

For any $0 < h_n \leq 1, u > 0$ we have $\frac{h_n^2}{1+h_n u} \leq \frac{1}{1+u}$, therefore

$$\mathbb{P}_0 (|\mathcal{L}f(x) - T_{h_n}f(x)| > u) \leq \mathbb{P}_0 (V_N(c)^c) + 2 \exp \left(-c \frac{N h_n^{d+2} u^2}{1 + h_n u} \right)$$

By a union bound this gives

$$\mathbb{P}_0 \left(\|\mathcal{L}f - T_{h_n}f\|_{L^\infty(\nu)} > u \right) \leq N \mathbb{P}_0 (V_N(c)^c) + 2N \exp \left(-c \frac{N h_n^{d+2} u^2}{1 + h_n u} \right)$$

We have also $\mathbb{P}_0 (V_N(c)^c) \leq e^{-cN h_n^d}$, so that for any $H > 0$, taking $u = M_0 \left(\frac{\ln N}{N} \right)^{\frac{1}{2}} h_n^{-(1+d/2)}$ with M_0 large enough (that depends on H) we obtain $h_n u = o(1)$ by assumption 2 and

$$\mathbb{P}_0 \left(\|\mathcal{L}f - T_{h_n}f\|_{L^\infty(\nu)} > M_0 \left(\frac{\ln N}{N} \right)^{\frac{1}{2}} h_n^{-(1+d/2)} \right) \leq N^{-H}$$

■

C.3 Proof of Theorem 5

Recall that $\hat{f} = \sum_{j=1}^J \langle u_j | Y \rangle_{L^2(\nu)} u_j$ so that

$$\hat{f} - f_0 = \sum_{j=1}^J \langle u_j | f_0 \rangle_{L^2(\nu)} u_j - f_0 + \sum_{j=1}^J \langle u_j | \underline{\varepsilon} \rangle_{L^2(\nu)} u_j, \quad \underline{\varepsilon} \sim \mathcal{N}(0, \sigma^2 I_n)$$

where I_n is the identity matrix in \mathbb{R}^n . Therefore

$$\|\hat{f} - f_0\|_{L^2(\nu)} \leq \left\| \sum_{j=1}^J \langle u_j | f_0 \rangle_{L^2(\nu)} u_j - f_0 \right\|_{L^2(\nu)} + \left\| \sum_{j=1}^J \langle u_j | \underline{\varepsilon} \rangle_{L^2(\nu)} u_j \right\|_{L^2(\nu)}.$$

We have by orthonormality of (u_1, \dots, u_{J_n})

$$\left\| \sum_{j=1}^J \langle u_j | \underline{\varepsilon} \rangle_{L^2(\nu)} u_j \right\|_{L^2(\nu)}^2 = \sum_{j=1}^{J_n} \langle u_j | \underline{\varepsilon} \rangle_{L^2(\nu)}^2$$

and also that

$$\mathbb{E}_0 \left[\langle u_j | \underline{\varepsilon} \rangle_{L^2(\nu)}^2 \middle| x_{1:n} \right] = \mathbb{E}_0 \left[\left(u_j^T \text{Diag}(\nu) \underline{\varepsilon} \right)^2 \middle| x_{1:n} \right] = \sigma^2 u_j^T \text{Diag}(\nu^2) u_j = \sigma^2 \sum_{i=1}^n u_j(x_i)^2 \nu_{x_i}^2$$

On A_N defined in lemma 21, with $N = n$ and $h_- = h_n$ which satisfies the assumption 2 we have $\nu_{x_i} \lesssim \frac{1}{n}$ so that

$$\mathbb{E}_0 \left[\langle u_j | \underline{\varepsilon} \rangle_{L^2(\nu)}^2 \middle| x_{1:n} \right] \lesssim \frac{1}{n} \|u_j\|_{L^2(\nu)}^2 = \frac{1}{n}.$$

This implies that

$$\sum_{i=1}^{J_n} \langle u_j | \varepsilon \rangle_{L^2(\nu)}^2 \lesssim \frac{1}{n} \sum_{j=1}^{J_n} Z_j(x_{1:n})^2, \quad \text{where} \quad Z_j(x_{1:n}) = \frac{\langle u_j | \varepsilon \rangle_{L^2(\nu)}}{\mathbb{E}_0 \left[\langle u_j | \varepsilon \rangle_{L^2(\nu)}^2 \middle| x_{1:n} \right]^{1/2}}$$

In particular conditionally on $x_{1:n}$, $Z_j(x_{1:n}) \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$. Hence as in Green et al. (2021, Appendix. C1), by a result of Laurent and Massart (2000) we get for some $c > 0$

$$\begin{aligned} \mathbb{P}_0 \left(\left\| \sum_{j=1}^{J_n} \langle u_j | \varepsilon \rangle_{L^2(\nu)} u_j \right\|_{L^2(\nu)}^2 > c \frac{J_n}{n} \right) &\leq \mathbb{P}_0(A_N^c) + \mathbb{E}_0 \left[\mathbb{1}_{A_N} \mathbb{P}_0 \left(\left\| \sum_{j=1}^{J_n} \langle u_j | \varepsilon \rangle_{L^2(\nu)} u_j \right\|_{L^2(\nu)}^2 > c \frac{J_n}{n} \middle| x_{1:n} \right) \right] \\ &\leq \mathbb{P}_0(A_N^c) + \exp(-J_n) \end{aligned} \quad (13)$$

Moreover

$$\|f_{0J_n} - f_0\|_{L^2(\nu)}^2 \leq \|p_{J_n}(e^{-t\mathcal{L}}f_t) - f_0\|_{L^2(\nu)}^2$$

where $p_{J_n}(e^{-t\mathcal{L}}f_t)$ is defined in Theorem 13. Therefore, using Theorem 13, we deduce that choosing t accordingly,

$$\|f_{0J_n} - f_0\|_{L^2(\nu)}^2 \leq \|p_{J_n}(e^{-t\mathcal{L}}f_t) - f_0\|_{L^2(\nu)}^2 \leq \tilde{\varepsilon}_n(J_n)$$

Finally combining this with (13), we obtain that some $C > 0$

$$\mathbb{P}_0 \left(\|\hat{f} - f_0\|_{L^2(\nu)} > C\varepsilon_n \right) \rightarrow 0$$

where

$$\varepsilon_n = \sqrt{\frac{J_n}{n}} + (\ln N)^{\lceil \beta/2 \rceil} \max \left(1, \left(\frac{J_n^{-2/d} \ln N}{h_n^2} \right)^{\lceil \beta/2 \rceil} \left(h_n^\beta + \mathbb{1}_{\beta > 1} \left(\frac{\ln N}{N h_n^d} \right)^{1/2} h_n \right) \right)$$

In order to get a result in expectation, simply notice that

$$\mathbb{E}_0 \left[\|\hat{f} - f_0\|_n^q \right] \leq c\varepsilon_n^q + \mathbb{E}_0 \left[\|\hat{f} - f_0\|_n^q \mathbb{1}_{\|\hat{f} - f_0\|_n > c\varepsilon_n} \right]$$

Since by definition

$$\|\hat{f} - f_0\|_n = \min_{f \in \text{span}(u_1, \dots, u_J)} \|f - f_0\|_n \underbrace{\leq}_{f=0} \|f_0\|_n \leq \|f_0\|_{L^\infty(\mathcal{M})}$$

This implies

$$\mathbb{E}_0 \left[\|\hat{f} - f_0\|_n^q \right] \leq c\varepsilon_n^q + \|f_0\|_{L^\infty(\mathcal{M})}^q \mathbb{P}_0 \left[\|\hat{f} - f_0\|_n > c\varepsilon_n \right]$$

By the first part of the proof, for any $H > 0$ we have

$$\mathbb{P}_0 \left[\|\hat{f} - f_0\|_n > c\varepsilon_n \right] \lesssim n^{-H} + e^{-J_n} + \mathbb{P}_0(A_N^c)$$

Since h_n satisfies assumption 2 and $J_n \geq \ln^\kappa n$, $\kappa > d \geq 1$, for $H > 0$ large enough we get

$$\mathbb{P}_0 \left[\|\hat{f} - f_0\|_n > c\varepsilon_n \right] \lesssim \varepsilon_n^q$$

which concludes the proof. ■

Appendix D. Proofs of section 3.2

We first recall an adaptation to our setting of the general prior mass and testing approach of Ghosal, Ghosh, and van der Vaart (2000); Ghosal and van der Vaart (2007) that we use to prove posterior contraction rates with respect to the empirical L^2 distance $\|\cdot\|_n$. Note that we apply the approach on a high probability event A_N (which does not change the result), and since in the Gaussian regression setting the Kullback-Leibler divergence and variation are both proportional to the (squared) empirical L^2 norm $\|\cdot\|_n^2$, the prior mass condition on the Kullback-Leibler neighbourhood is merely a condition on the prior mass of the $\|\cdot\|_n$ -balls centered at f_0 (see Lemma 2.7 in Ghosal and van der Vaart (2017) as well as Chapter 8 for more details on the prior mass and testing approach).

Theorem 37. Contraction rate theorem

Assume that $\varepsilon_n \rightarrow 0$, $n\varepsilon_n^2 \rightarrow \infty$ and A_N be an event of the form $A_N = \{x_{1:N} \in \Omega_N\}$ for some $\Omega \subset (\mathbb{R}^D)^N$ satisfying $\mathbb{P}_0(A_N) \rightarrow 1$. If there exists fixed constants $C_1 > 1 + \frac{1}{2\sigma^2}$, $C_2 > 4\sqrt{1 + 2\sigma^2}$, $C_3 > 0$ and $\mathcal{F}_n \subset \mathbb{R}^V$ such that, on A_N

- $\Pi[\|f - f_0\|_n \leq \varepsilon_n] \geq \exp(-n\varepsilon_n^2)$
- $\Pi[\mathcal{F}_n^c] \leq e^{-C_1 n\varepsilon_n^2}$
- $\ln N\left(\frac{C_2 \varepsilon_n}{2}, \mathcal{F}_n, \|\cdot\|_n\right) \leq C_3 n\varepsilon_n^2$

where $N(\varepsilon, \mathcal{F}, \|\cdot\|_n)$ denotes the covering number of the set \mathcal{F} at scale ε with respect to the $\|\cdot\|_n$ -metric, then for $M > 0$ large enough

$$\mathbb{E}_0[\Pi[\|f - f_0\|_n > M\varepsilon_n | \mathbb{X}^n]] \rightarrow 0$$

D.1 Proof of Theorem 3

As shown by Theorem 37, the proof is based on first bounding from below the prior mass of neighbourhoods of f_0 , which is done in Lemma 38 below and then to control the entropy by a sieve sequence, which is done in Lemma 39. Recall that

$$\varepsilon_n(J_n, h_n) = \sqrt{\frac{J_n \ln N}{n}} + (\ln N)^{\lceil \beta/2 \rceil} \max\left(1, \left(\frac{J_n^{-2/d} \ln N}{h_n^2}\right)^{\lceil \beta/2 \rceil}\right) \left(h_n^\beta + \mathbb{1}_{\beta > 1} \left(\frac{\ln N}{N h_n^d}\right)^{1/2} h_n\right)$$

with $J_n \geq \ln^\kappa N$, $\kappa > d$ and h_n satisfying assumption 2.

With $z = z_n = z_1 (n\varepsilon_n^2)^{1/b_2}$ with $z_1 > 0$ and $\mathcal{F}_n = \mathcal{F}_{z_n, J_n}$. Its $L^\infty(\nu)$ -metric entropy is then, according to lemma 39, bounded by

$$\forall \zeta > 0, \quad \ln N\left(\zeta \varepsilon_n, \mathcal{F}_n, \|\cdot\|_{L^\infty(\nu)}\right) \lesssim J_n \ln N \lesssim n\varepsilon_n^2.$$

Moreover by the choice of z_n we have, for some $c > 0$

$$\Pi[\mathcal{F}_n^c | J_n, h_n] \leq J_n e^{-b_1 z_n^{b_2}} \leq e^{-c z_1^{b_2} n\varepsilon_n^2}$$

By taking $z_1 > 0$ large enough we conclude the proof of Theorem 3 by applying Lemma 38 together with Theorem 37. ■

D.2 Proof of Theorem 6

The proof is based on that of Theorem 3, but showing that J and h can be chosen in a data dependent way. To do that notice that by the non adaptive case, for any $H > 0$ there exists a constant $C > 0$ such that

$$\mathbb{P}_0 \left(\Pi \left[\|f - f_0\|_{L^\infty(\nu)} > C\varepsilon_n(J_n, h_n) | J_n, h_n \right] < \exp(-nC^2\varepsilon_n^2) \right) \leq N^{-H}$$

where $\varepsilon_n(J_n, h_n)$ is given by (5). If for some $J_0, h_0, h_1 > 0$

$$J_n = J_0 (\ln N)^{\frac{2d(1+2\tau/d)\lceil\beta/2\rceil}{2\beta+d}} n^{\frac{d}{2\beta+d}}, \quad h_n \in \left[\frac{h_0 J_n^{-1/d}}{2 \ln^{\tau/d} N}, \frac{h_1 J_n^{-1/d}}{\ln^{\tau/d} N} \right]$$

then

$$\varepsilon_n(J_n, h_n) \lesssim n^{-\frac{\beta}{2\beta+d}} (\ln N)^{\frac{(2\tau+d)\lceil\beta/2\rceil - \beta(\tau+2\beta/d)}{2\beta+d}}$$

Since

$$\Pi \left[\|f - f_0\|_{L^\infty(\nu)} \leq C\varepsilon_n \right] \geq \pi_J(J_n) \pi_h(h_n | J_n) \Pi \left[\|f - f_0\|_{L^\infty(\nu)} > C\varepsilon_n | J_n, h_n \right]$$

By assumption we have for some a_1, b_1, b_2

$$\pi_J(J_n) \geq e^{-a_1 J_n L_{J_n}} \geq e^{-a_1 J_n \ln N} \text{ and } \pi_h(h_n | J_n) \geq b_1 e^{-b_2 h_n^{-d}}$$

Therefore, since $\varepsilon_n \geq (nh_n^d)^{-1/2} + \sqrt{\frac{J_n \ln N}{n}}$, for some $C' > 0$ we have

$$\mathbb{P}_0 \left(\Pi \left[\|f - f_0\|_{L^\infty(\nu)} \leq C'\varepsilon_n \right] < e^{-nC'^2\varepsilon_n^2} \right) \leq N^{-H}$$

To verify the entropy condition, consider for some $k > 0$ the set $\mathcal{F}_n = \bigcup_{\substack{J \leq kJ_n \\ h \in \mathcal{H}_J}} \mathcal{F}_{z_n, J, h}$, where $\mathcal{F}_{z, J, h}$ is defined in lemma 39 and $z_n = u(n\varepsilon_n^2)^{1/b_2}$ with u, k large enough. Then

$$\Pi[\mathcal{F}_n^c] \leq \Pi_J(J > kJ_n) + \sum_{\substack{J \leq kJ_n \\ h \in \mathcal{H}_J}} J e^{-b_1 z_n^{b_2}} \leq e^{-CJ_n} + (kJ_n)^2 e^{-b_1 u^{b_2} n \varepsilon_n^2} \max_J \#\mathcal{H}_J \leq e^{-Cn\varepsilon_n^2}$$

for any $C > 0$ by choosing k, u large enough, since $\#\mathcal{H}_J \leq K_1 \exp(K_2 J \ln N) \leq K_1 \exp(K'_2 n \varepsilon_n^2)$ for some $K_1, K_2, K'_2 > 0$. Finally for all $\zeta > 0$

$$N(\zeta\varepsilon_n, \mathcal{F}_n, \|\cdot\|_{L^\infty(\nu)}) \lesssim n\varepsilon_n^2$$

which concludes the proof. ■

D.3 Lemmas 38 and 39

Lemma 38. *Prior thickness*

Let $J_n \in \{1, \dots, N\}$, $J_n \geq \ln^\kappa N$, $\kappa > d$ and h_n satisfying assumption 2. Consider

the prior defined by (3). Then for any $H > 0$ there exists $c > 0$ such that with the rate $\epsilon_n(J_n, h_n)$ as defined in (5), We have

$$\mathbb{P}_0 \left(\Pi \left[\|f - f_0\|_{L^\infty(\nu)} \leq c\epsilon_n(J_n, h_n) | J_n, h_n \right] < \exp(-n\epsilon_n(J_n, h_n)^2) \right) \leq N^{-H}.$$

In particular, for any $\tau > d/2$ and

$$h_n = n^{-\frac{1}{2\beta+d}} (\ln n)^{-\frac{1-\tau-2(1+2\tau/d)\lceil\beta/2\rceil}{2\beta+d}}, J_n = \frac{h_n^{-d}}{\ln^\tau N}$$

then

$$\epsilon_n(J_n, h_n) \asymp (\ln n)^{\frac{(2\tau+d)\lceil\beta/2\rceil+(1-\tau)\beta}{2\beta+d}} n^{-\frac{\beta}{2\beta+d}}.$$

Proof Throughout the proof we write $\epsilon_n = \epsilon_n(J_n, h_n)$. We use theorem 13 and distinguish the cases : for some $k > 0$ large enough, with $t = t_n = k\lambda_{J_n}^{-1} \ln N$

1. If $\beta \leq 2$ we have

$$\begin{aligned} \forall f : V \rightarrow \mathbb{R}, \|f - f_0\|_{L^\infty(\nu)} &\leq \|f - p_J(e^{-t\mathcal{L}}f_0)\|_{L^\infty(\nu)} + \|f_0 - p_J(e^{-t\mathcal{L}}f_0)\|_{L^\infty(\nu)} \\ &\lesssim \|f - p_J(e^{-t\mathcal{L}}f_0)\|_{L^\infty(\nu)} + O_{P_0} \left(h_n^\beta (\lambda_{J_n}^{-1} h_n^{-2} \ln N) \right). \end{aligned}$$

Hence for all $c > 0$, there exists $c' > 0$, such that

$$\|f - p_J(e^{-t\mathcal{L}}f_0)\|_{L^\infty(\nu)} \leq c' h_n^\beta \lambda_{J_n}^{-1} h_n^{-2} \ln N \implies \|f - f_0\|_{L^\infty(\nu)} \leq \epsilon_n.$$

Moreover if $f = \sum_{j=1}^J z_j u_j$, using proposition 20 we get

$$\|f - p_J(e^{-t\mathcal{L}}f_0)\|_{L^\infty(\nu)} = \left\| \sum_{j=1}^J \left(z_j - \langle u_j | e^{-t\mathcal{L}} f_0 \rangle_{L^2(\nu)} \right) u_j \right\|_{L^\infty(\nu)} \leq N \max_{1 \leq j \leq N} \left| z_j - \langle u_j | e^{-t\mathcal{L}} f_0 \rangle_{L^2(\nu)} \right|.$$

Hence,

$$\begin{aligned} &\Pi \left[\|f - p_{J_n}(e^{-t\mathcal{L}}f_0)\|_{L^\infty(\nu)} \leq c' h_n^\beta \lambda_{J_n}^{-1} h_n^{-2} \ln N \mid J_n, h_n \right] \\ &\geq \Pi \left[\max_{1 \leq j \leq N} \left| z_j - \langle u_j | e^{-t\mathcal{L}} f_0 \rangle_{L^2(\nu)} \right| \leq \frac{c' h_n^\beta \lambda_{J_n}^{-1} h_n^{-2} \ln N}{N} \right] \\ &\geq \prod_{j=1}^{J_n} \Pi \left[\left| z_j - \langle u_j | e^{-t\mathcal{L}} f_0 \rangle_{L^2(\nu)} \right| \leq \frac{c' h_n^\beta \lambda_{J_n}^{-1} h_n^{-2} \ln N}{N} \right] \\ &\geq \prod_{j=1}^{J_n} \Pi \left[\left| z_j - \langle u_j | e^{-t\mathcal{L}} f_0 \rangle_{L^2(\nu)} \right| \leq \frac{c' h_n^\beta \ln N}{N} \right] \\ &\geq \left(\frac{c' h_n^\beta \ln N}{N} \inf_{[-K, K]} \Psi \right)^{J_n} \text{vol}(B_{\mathbb{R}^{J_n}}(0, 1)), \end{aligned}$$

where we have used $\lambda_{J_n} \leq 2h_n^{-2}$, which holds because $\|\mathcal{L}\|_{\mathcal{L}(L^\infty(\nu))} \leq 2$, and where $K > 0$ is chosen to be larger than

$$\max_{1 \leq j \leq J_n} \left| \langle u_j | e^{-t\mathcal{L}} f_0 \rangle_{L^2(\nu)} \right| + \frac{c'h_n^\beta \ln N}{N} \lesssim 1 + \underbrace{\|e^{-t\mathcal{L}}\|_{\mathcal{L}(L^\infty(\nu))}}_{=1} \|f_0\|_{L^\infty(\mathcal{M})} < +\infty.$$

Hence

$$-\ln \Pi \left[\|f - p_J(e^{-t\mathcal{L}} f_0)\|_{L^\infty(\nu)} \leq c'h_n^\beta \lambda_{J_n}^{-1} h_n^{-2} \ln N \mid J_n, h_n \right] \lesssim J_n \ln N.$$

Now $\varepsilon_n \geq \sqrt{\frac{J_n \ln N}{n}} \iff J_n \ln N \leq n\varepsilon_n^2$, therefore

$$\begin{aligned} \mathbb{P}_0 \left(\Pi \left[\|f - p_J(e^{-t\mathcal{L}} f_0)\|_{L^\infty(\nu)} \leq \varepsilon_n \mid J_n, h_n \right] < \exp(-n\varepsilon_n^2) \right) \\ \leq \mathbb{P}_0 \left(\|f_0 - p_{J_n}(e^{-t\mathcal{L}} f_0)\|_{L^\infty(\nu)} > \varepsilon_n/2 \right) \\ \leq n^{-H} \end{aligned}$$

for any $H > 0$.

2. The case $\beta > 2$ is similar, the only difference being that we use the bound

$$\left| \langle u_j | e^{-t\mathcal{L}} f_t \rangle_{L^2(\nu)} \right| \leq \left\| \sum_{l=0}^k \frac{(t\mathcal{L})^l e^{-t\mathcal{L}}}{l!} \right\|_{\mathcal{L}(L^2(\nu))} \|f_0\|_{L^2(\nu)} \leq \|f_0\|_{L^\infty(\mathcal{M})} \sup_{s>0} \sum_{l=0}^k \frac{s^l e^{-s}}{l!} < +\infty$$

■

Lemma 39. *For some $z \geq 1$ and $J \in \{1, \dots, N\}$ let*

$$\mathcal{F}_{z,J,h} = \left\{ \sum_{j=1}^J a_j u_j(h) : |a_j| \leq z \right\} \subset \mathbb{R}^V, \quad z > 0$$

Then

$$\Pi[\mathcal{F}_{z,J,h}^c | J, h] \leq J e^{-b_1 z^{b_2}}$$

and, there exists $\varepsilon_1, C_1 > 0$ such that for any $\varepsilon \leq \varepsilon_1$

$$N \left(\varepsilon, \mathcal{F}_{z,J,h}, \|\cdot\|_{L^\infty(\nu)} \right) \leq \left(\frac{C_1 z N}{\varepsilon} \right)^{2J}$$

Proof Proof of lemma 39

We drop the subscript h for ease of notation. Using the representation $f = \sum_{j=1}^J Z_j u_j$, the upper bound

$$\Pi[\mathcal{F}_{z,J}^c | J] \leq \sum_{j=1}^J \Pi[|Z_j| > z | J] \leq J \Psi([-z, z]^c) \leq J e^{-b_1 z^{b_2}}$$

follows by the assumption on the tails of Ψ . Moreover by proposition 20, if \mathcal{E} is an $\frac{\varepsilon}{JN}$ -net of $[-z, z]^J$ then $\left\{ \sum_{j=1}^J f_j u_j : f \in \mathcal{E} \right\}$ is an ε -net of $\mathcal{F}_{z,J}$ (in $L^\infty(\nu)$). Since we can always take

$$\#\mathcal{E} \leq \left(\frac{C_1 z J N}{\varepsilon} \right)^J \leq \left(\frac{C_1 z N^2}{\varepsilon} \right)^J \leq \left(\frac{C_1^{1/2} z^{1/2} N}{\varepsilon^{1/2}} \right)^{2J} \leq \left(\frac{C_1^{1/2} c N}{\varepsilon} \right)^{2J}$$

for $\varepsilon \leq \varepsilon_1$ and some $C_1, \varepsilon_1 > 0$ this gives the result by changing $C_1^{1/2}$ into a new C_1 . \blacksquare

References

- Eddie Aamari and Clément Levrard. Nonasymptotic rates for manifold, tangent space and curvature estimation. *The Annals of Statistics*, 47(1):177–204, February 2019. ISSN 0090-5364, 2168-8966. doi: 10.1214/18-AOS1685.
- Eddie Aamari, Clément Berenfeld, and Clément Levrard. Optimal reach estimation and metric learning. *The Annals of Statistics*, 51(3):1086–1108, June 2023. ISSN 0090-5364, 2168-8966. doi: 10.1214/23-AOS2281.
- Nir Ailon and Bernard Chazelle. The Fast Johnson–Lindenstrauss Transform and Approximate Nearest Neighbors. *SIAM Journal on Computing*, 39(1):302–322, January 2009. ISSN 0097-5397. doi: 10.1137/060673096.
- Martin T. Barlow. *Random Walks and Heat Kernels on Graphs*. London Mathematical Society Lecture Note Series. Cambridge University Press, Cambridge, 2017.
- Mikhail Belkin and Partha Niyogi. Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. In *Advances in Neural Information Processing Systems*, volume 14. MIT Press, 2001.
- Mikhail Belkin and Partha Niyogi. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Computation*, 15(6):1373–1396, June 2003. ISSN 0899-7667. doi: 10.1162/08997660321780317.
- Clément Berenfeld and Marc Hoffmann. Density estimation on an unknown submanifold. *Electronic Journal of Statistics*, 15(1):2179–2223, January 2021. ISSN 1935-7524, 1935-7524. doi: 10.1214/21-EJS1826.
- Mira Bernstein, Vin de Silva, John C Langford, and Joshua B Tenenbaum. Graph approximations to geodesics on embedded manifolds.
- Jeff Calder and Nicolas Garcia Trillos. Improved spectral convergence rates for graph Laplacians on epsilon-graphs and k-NN graphs, June 2020.
- Ismaël Castillo, Gérard Kerkycharian, and Dominique Picard. Thomas Bayes’ walk on manifolds. *Probability Theory and Related Fields*, 158(3):665–710, April 2014. ISSN 1432-2064. doi: 10.1007/s00440-013-0493-0.

- Frédéric Chazal and Bertrand Michel. An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists. *Frontiers in Artificial Intelligence*, 4, September 2021. ISSN 2624-8212. doi: 10.3389/frai.2021.667963.
- Frédéric Chazal, Vin de Silva, Marc Glisse, and Oudot Steve. *The Structure and Stability of Persistence Modules*. Springer, 2016.
- Ronald R. Coifman and Stéphane Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, July 2006. ISSN 1063-5203. doi: 10.1016/j.acha.2006.04.006.
- Thierry Coulhon and Adam Sikora. Gaussian heat kernel upper bounds via the Phragmén-Lindelöf theorem. *Proceedings of the London Mathematical Society. Third Series*, 96(2): 507–544, 2008. ISSN 0024-6115. doi: 10.1112/plms/pdm050.
- Thierry Coulhon, Gerard Kerkycharian, and Pencho Petrushev. Heat kernel generated frames in the setting of Dirichlet spaces, June 2012.
- Francesco Denti, Diego Doimo, Alessandro Laio, and Antonietta Mira. The generalized ratios intrinsic dimension estimator. *Scientific Reports*, 12(1):20005, November 2022. ISSN 2045-2322. doi: 10.1038/s41598-022-20991-1.
- Vincent Divol and Frédéric Chazal. The density of expected persistence diagrams and its kernel based estimation. *Journal of Computational Geometry*, 10(2):127–153, 2019. ISSN 1920-180X. doi: 10.20382/jocg.v10i2a7.
- Vincent Divol and Théo Lacombe. Understanding the topology and the geometry of the space of persistence diagrams via optimal partial transport. *Journal of Applied and Computational Topology*, 5(1):1–53, March 2021. ISSN 2367-1734. doi: 10.1007/s41468-020-00061-z.
- David B. Dunson, Hau-Tieng Wu, and Nan Wu. Spectral Convergence of Graph Laplacian and Heat Kernel Reconstruction in L^∞ from Random Samples, June 2021.
- David B. Dunson, Hau-Tieng Wu, and Nan Wu. Graph based Gaussian processes on restricted domains. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 84(2):414–439, 2022. ISSN 1467-9868. doi: 10.1111/rssb.12486.
- Bernardo Fichera, Viacheslav Borovitskiy, Andreas Krause, and Aude Billard. Implicit Manifold Gaussian Process Regression, February 2024.
- Nicolás García Trillos, Moritz Gerlach, Matthias Hein, and Dejan Slepčev. Error Estimates for Spectral Convergence of the Graph Laplacian on Random Geometric Graphs Toward the Laplace–Beltrami Operator. *Foundations of Computational Mathematics*, 20(4):827–887, August 2020. ISSN 1615-3383. doi: 10.1007/s10208-019-09436-w.
- Christopher Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman. Minimax Manifold Estimation. *Journal of Machine Learning Research*, 13(43):1263–1291, 2012. ISSN 1533-7928.

- S. Ghosal and A. van der Vaart. Convergence rates of posterior distributions for non iid observations. *aos*, 35(1):192–223, 2007.
- Subhashis Ghosal and Aad van der Vaart. *Fundamentals of Nonparametric Bayesian Inference*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2017. ISBN 978-0-521-87826-5. doi: 10.1017/9781139029834.
- Subhashis Ghosal, Jayanta K. Ghosh, and Aad W. van der Vaart. Convergence rates of posterior distributions. *The Annals of Statistics*, 28(2):500–531, April 2000. ISSN 0090-5364, 2168-8966. doi: 10.1214/aos/1016218228.
- Franziska Göbel and Gilles Blanchard. Volume Doubling Condition and a Local Poincaré Inequality on Unweighted Random Geometric Graphs, March 2020.
- Alden Green, Sivaraman Balakrishnan, and Ryan Tibshirani. Minimax Optimal Regression over Sobolev Spaces via Laplacian Regularization on Neighborhood Graphs. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, pages 2602–2610. PMLR, March 2021.
- Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. ISSN 01621459, 1537274X.
- William B. Johnson, Joram Lindenstrauss, and Gideon Schechtman. Extensions of lipschitz maps into Banach spaces. *Israel Journal of Mathematics*, 54(2):129–138, June 1986. ISSN 1565-8511. doi: 10.1007/BF02764938.
- Jisu Kim, Alessandro Rinaldo, and Larry Wasserman. Minimax Rates for Estimating the Dimension of a Manifold. *Journal of Computational Geometry*, pages 42–95 Pages, February 2019. doi: 10.20382/jocg.v10i1a3.
- B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28(5):1302–1338, October 2000. ISSN 0090-5364, 2168-8966. doi: 10.1214/aos/1015957395.
- Jialu Liu and Jiawei Han. Spectral Clustering. In *Data Clustering*. Chapman and Hall/CRC, 2014. ISBN 978-1-315-37351-5.
- David Loiseaux, Mathieu Carrière, and Andrew J. Blumberg. A Framework for Fast and Stable Representations of Multiparameter Persistent Homology Decompositions, June 2023.
- David Loiseaux, Luis Scoccola, Mathieu Carrière, Magnus Bakke Botnan, and Steve Oudot. Stable Vectorization of Multiparameter Persistent Homology using Signed Barcodes as Measures, February 2024.
- Zacharie Naulet and Éric Barat. Some Aspects of Symmetric Gamma Process Mixtures. *Bayesian Analysis*, 13(3):703–720, September 2018. ISSN 1936-0975, 1931-6690. doi: 10.1214/17-BA1058.

- J. R. Norris. *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1997. ISBN 978-0-521-63396-3. doi: 10.1017/CBO9780511810633.
- Karl Pearson. LIII. On lines and planes of closest fit to systems of points in space. November 1901. doi: 10.1080/14786440109462720.
- Mathew Penrose. *Random Geometric Graphs*. Oxford University Press, May 2003. ISBN 978-0-19-850626-3. doi: 10.1093/acprof:oso/9780198506263.001.0001.
- Paul Rosa, Slava Borovitskiy, Alexander Terenin, and Judith Rousseau. Posterior Contraction Rates for Matérn Gaussian Processes on Riemannian Manifolds. *Advances in Neural Information Processing Systems*, 36:34087–34121, December 2023.
- Judith Rousseau. On the frequentist properties of bayesian nonparametric methods. *Annual Review of Statistics and Its Application*, 3(Volume 3, 2016):211–231, 2016. ISSN 2326-831X. doi: 10.1146/annurev-statistics-041715-033523.
- Sam T. Roweis and Lawrence K. Saul. Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science*, 290(5500):2323–2326, December 2000. doi: 10.1126/science.290.5500.2323.
- Daniel Sanz-Alonso and Ruiyi Yang. Unlabeled Data Help in Graph-Based Semi-Supervised Learning: A Bayesian Nonparametrics Perspective. *Journal of Machine Learning Research*, 23(97):1–28, 2022. ISSN 1533-7928.
- Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation*, 10(5):1299–1319, July 1998. ISSN 0899-7667. doi: 10.1162/089976698300017467.
- Zhaoyang Shi, Krishnakumar Balasubramanian, and Wolfgang Polonik. Adaptive and non-adaptive minimax rates for weighted Laplacian-eigenmap based nonparametric regression, October 2023.
- Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*, 290(5500):2319–2323, December 2000. doi: 10.1126/science.290.5500.2319.
- Aad van der Vaart and Harry van Zanten. Information Rates of Nonparametric Gaussian Process Methods. *Journal of Machine Learning Research*, 12(60):2095–2119, 2011. ISSN 1533-7928.
- Aad Van der Vaart and Jon Wellner. *Weak Convergence and Empirical Processes*. Springer, 1996.
- Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, December 2007. ISSN 1573-1375. doi: 10.1007/s11222-007-9033-z.
- Caroline L. Wormell and Sebastian Reich. Spectral Convergence of Diffusion Maps: Improved Error Bounds and an Alternative Normalization. *SIAM Journal on Numerical Analysis*, 59(3):1687–1734, January 2021. ISSN 0036-1429. doi: 10.1137/20M1344093.