RICCI CURVATURE AND THE MANIFOLD LEARNING PROBLEM

ANTONIO G. ACHE AND MICAH W. WARREN

Abstract. Consider a sample of n points taken i.i.d from a submanifold Σ of Euclidean space. We show that there is a way to estimate the Ricci curvature of Σ with respect to the induced metric from the sample. Our method is grounded in the notions of *Carré du Champ* for diffusion semi-groups, the theory of empirical processes and local Principal Component Analysis.

Contents

The first author was partially supported by a postdoctoral fellowship of the National Science Foundation, award No. DMS-1204742.

The second author was partially supported by NSF Grant DMS-1438359.

1. Introduction

In this paper we are concerned with the structure of sets of large data in high dimensions. Even though we deal with sets of points in $\mathbb{R}^N$ for N large, a common assumption when studying large data sets is that the points lie on or in the vicinity of an embedded low dimensional submanifold Σ^d of $\mathbb{R}^N$. This assumption is oftentimes called *the manifold assumption* and the study of geometric and topological properties of sets satisfying the manifold assumption is what we call nowadays *manifold learning*. The interest in the structure of large data sets comes from the need of organizing information arising from many different sources, for example, images, signals, genomes and other outcomes. Even though there has been significant progress in the manifold learning problem in the last decade, a fundamental question remains unanswered: construct an algorithm for *learning* or effectively estimating the curvature of a manifold that is being approximated by a point cloud. In this paper we lay the theoretical foundation for estimating the Ricci curvature of an embedded submanifold of $\mathbb{R}^N$ if one only knows a point cloud approximating the submanifold. In particular, combining with the recent work of Singer and Wu [25] and PCA (Principal Component Analysis), we offer a construction which takes a point cloud and generates the geometric information as follows.

- **Input:** A sequence of points $\xi_1, ..., \xi_n \subset \mathbb{R}^N$ and a bandwidth parameter t.

- **Output:** For each point ξ_i, an approximate basis for a tangent space at ξ_i, and Ricci curvature matrix approximating the Ricci curvature with respect to this basis.

  If points are sampled randomly from a smooth submanifold, then for n large and t small, there will be an orthogonal connection matrix O_{ξ_i,ξ_j} approximating the connection between the tangent spaces of nearby points ξ_i, ξ_j, (see [25]).

There is a choice of kernel and and PCA cutoff parameter, which may affect the output, but will not affect the limit when the points are sampled from a smooth submanifold as $n \to \infty$.

Some advantages of our method are the following:

- It generates a weighted Ricci curvature which takes into account the underlying probability density.
- We do not need to approximate the derivatives of the underlying metric.
- It generates an approximate Ricci curvature even without assuming that the manifold has a constant dimension
- Our method allows one to study the convergence of the sample version of Ricci curvature to its actual value based on extrinsic information like the *reach* of the submanifold (see Definition 4.10).
- When the connection forms O_{ξ_i,ξ_j} are available, it allows one to approximate the Hodge Laplacian on 1-forms.

The following is our main result: For a more precise statement, see Theorem B in section 2.1.1.

Theorem A (Approximation of the Ricci Curvature)**.** *Consider the metric measure space* $(\Sigma, \|\cdot\|, d\mathrm{vol}_\Sigma)$ *where* $\Sigma^d \subset \mathbb{R}^N$ *is a smooth closed embedded submanifold. Suppose that we have a uniformly distributed i.i.d. sample* $\{\xi_1, \ldots, \xi_n\}$ *of points from* Σ*. For* $x \in \Sigma$*, one can define a sequence of* d*-tuples of orthogonal vectors, and* $d \times d$ *Ricci matrices* $\hat{R}_{i,j}$ *representing the Ricci curvature on these vectors, such at if* $\eta \in T_x\Sigma$*, then*

$$\left|\hat{R}_{i,j}\eta^i\eta^j - \mathrm{Ric}_x(\eta,\eta)\right| \xrightarrow{\text{a.s}} 0,$$

where η^i *are the components of the vector* η *projected onto the* d*-tuple of vectors approximating the tangent plane.*

We will see that the sequence of tangent spaces can be be constructed using a method known as *local Principal Component Analysis* (local PCA). Since this construction is a crucial part of the article [25], we will devote Section 5 to a fairly detailed explanation of the construction of vectors using PCA. Construction of the Ricci curvature operators is described in section 2.

We will also state, (but not prove) a corresponding theorem that applies when the sample of points is distributed i.i.d withrespect to a volume form other than the given Riemannian volume form. See Theorem D.

Our method is based on purely on distance, not on combinatorics. In short, we iterate an approximate, kernel-generated Laplacian to obtain a Bakry-Emery operator. To obtain a "coarse Ricci curvature" this is applied to a function which should be approximately linear at a point. To obtain a Ricci curvature on the PCA basis, we apply this operator to linear functions with differentials determined by the PCA.

Our notion does not assume or attempt to create a triangulation of the manifold. There have been a number of works recently that study the Forman-Ricci curvature [13, 19, 27, 36]. Forman-Ricci curvature [11] (like ours) is based on the Bakry-Emery approach, but requires a graph structure so is inherently combinatorial. Indeed, our main results do not contradict results regarding failures of curvature convergence for *triangulated* approximations to a manifold [17, pg. vii].

Our interest in the Ricci curvature is not arbitrary, but is motivated by very concrete problems in applied mathematics. One example of these problems comes from topological data analysis : estimating the spectrum of the Hodge Laplacian on 1-forms with respect to the induced metric of an embedded submanifold Σ of $\mathbb{R}^N$. The spectrum of the Hodge Laplacian is important in the study of topological properties of Σ. Singer and Wu's analysis based on Vector Diffusion Maps [25] represents significant progress in the estimation of the spectrum of the so called *connection* or *rough* Laplacian on 1-forms. Although since the publication of the seminal paper by Carlsson [7] there has been an explosion of interest in topological data analysis, cf. [8], there does not yet seem to be an effective way to estimate the Hodge Laplacian of an embedded manifold. We remark that an effective algorithm for learning the Ricci curvature of an embedded submanifold could in principle provide us with a method for estimating the Hodge Laplacian on 1-forms in view of the *Weitzenböck* formula. More precisely, given a metric g on Σ and a 1-form $X \in \Omega^1(\Sigma^d)$ we know that

$$\Delta_g X = -\Delta_H X + \mathrm{Ric}(X),$$

where $\mathrm{Ric}(X)$ is the Ricci endomorphism applied to X, i.e., $\mathrm{Ric}(X) = g^{jk}\mathrm{Ric}_{ij}X_k$. In terms of linear algebra, if we have taken n points from a d dimensional manifold, with n sufficiently large, PCA will give us a basis for the tangent space, in which case the space of 1-forms can be described as a vector in $\mathbb{R}^{nd}$ - that is, for each point in $\mathbb{R}^n$ we choose a d-vector. The Hodge Laplacian then becomes a map

$$\Delta_H : \mathbb{R}^{nd} \to \mathbb{R}^{nd}$$

which one can analyze. In the current paper, we do not attempt to prove spectral convergence, as one can see from the [26] that this is expected to be somewhat involved.

Another motivation for the problem of learning the Ricci curvature of a submanifold is to measure the robustness of networks, in particular in biology [35, 34, 29, 19]. In [29], a connection has been established between Ricci curvature of graphs and the robustness of cancer networks. Moreover, it has been suggested that robustness of cancer networks is associated to a certain "entropy" and that the Ricci curvature of a graph is closely related to such entropy. The notion of Ricci curvature used in [29] is Ollivier's coarse Ricci curvature. The recent paper [19] compares different notions of Ricci curvature when applied to biological networks. Again we note that most of the methods are based on combinatorial approaches.

As we will see, our method is based on the fact that it is possible to estimate the *rough Laplacian* of the induced metric of an embedded submanifold of $\mathbb{R}^N$. Given an embedded submanifold Σ^d of $\mathbb{R}^N$, and an embedding $F : \Sigma^d \to \mathbb{R}^N$, the *metric induced by* F is given in coordinates by $g_{ij} = \langle D_i F, D_j F\rangle$ where $\langle \cdot, \cdot\rangle$ is the Euclidean inner product in $\mathbb{R}^N$ and D means differentiation with respect to coordinates on Σ^d that are determine by the way Σ^d is embedded into $\mathbb{R}^N$. By rough Laplacian of g we mean the operator defined on functions by $\Delta_g f = g^{ij}\nabla_i\nabla_j f$ where ∇ is the Levi-Civita connection of g and f is a smooth function defined on Σ. Belkin and Niyogi showed in [4] that given a uniformly distributed point cloud on Σ there is a 1-parameter family of operators L_t, which converge to the Laplace-Beltrami operator Δ_g on the submanifold. More precisely, the construction of the operators L_t is based on an approximation of the heat kernel of Δ_g, and in particular the bandwidth parameter t can be interpreted as a choice of scale. In order to learn the rough Laplacian Δ_g from a point cloud it is necessary to write a sample version of the operators L_t. Then, supposing we have n data points that are independent and identically distributed (abbreviated by i.i.d.) and whose common distribution is the uniform distribution on Σ^d with respect its induced metric, one can choose a bandwidth parameter t_n in such a way that the operators L_{t_n} applied to a smooth function f on Σ converge almost surely to the rough Laplacian Δ_g. This step follows essentially from applying a quantitative version of the *law of large numbers*. Thus one can almost surely learn spectral properties of a manifold. While in [4] it is assumed that the sample is uniform, it was proved by Coifman and Lafon in [9] that if one assumes more generally that the distribution of the data points has a smooth, strictly positive density in Σ, then it is possible to normalize the operators L_t in [4] to recover the rough Laplacian. More generally, the results in [9] and [25] show that it is possible to recover a whole family of operators that include the Fokker-Planck operator and the weighted Laplacian $\Delta_\rho f = \Delta f - \langle \nabla \rho, \nabla f \rangle$ associated to the smooth metric measure space $(M, g, e^{-\rho} d\text{vol})$, where ρ is a smooth function. As the Bakry-Emery Ricci tensor can be obtained by iterating Δ_ρ, the Bakry-Emery Ricci tensor can be approximated by iterating approximations of Δ_ρ. Following [4], Singer and Wu have recently developed methods for learning the rough Laplacian of an embedded submanifold on 1-forms using *Vector Diffusion Maps* (VDM) (see for example [25]).

It is the goal of this paper, together with [2] and [1], to demonstrate that the above discussed approximation of the rough Laplacian can be continued to approximate Ricci curvature as well. In fact, our approximation method is based on writing sample counterparts of the Ricci curvature. More generally, we will show that it is possible to define sample counterparts of more general objects, for example of the notions of *Carré du Champ* and *iterated Carré du Champ* associated to a diffusion semi-group. Our idea for estimating the Carré du Champ (and ultimately the Ricci curvature) from a sample is closely related to the results in [2, 1]. For example, in [1] we define a family of coarse Ricci curvatures which depend on a scale parameter t, and show that when taken on a smooth embedded submanifold on Euclidean space, these recover the Ricci curvature as $t \to 0^+$. We will show that as long as we sample points adequately from the submanifold Σ, it is possible to choose a scale t_n depending only on the size of the data set (equal to n) to obtain almost sure convergence to the actual Ricci curvature of the submanifold at a given point. We will summarize the results in [2, 1] relevant to the present article in Section 3 (for example Theorem 3.1 and Proposition 3.2).

Our results show that one can give a definition of a sample version of Ricci curvature at a scale on general metric measure spaces that converges to the actual Ricci curvature on smooth Riemannian manifolds. Moreover, our definition of empirical coarse Ricci curvature at a scale can be thought of as an extension of Ricci curvature to a class of discrete metric spaces, namely those obtained from sampling points from a smooth closed embedded submanifold of $\mathbb{R}^N$. Note however, that in order to obtain convergence of the empirical coarse Ricci curvature at a scale to the actual Ricci curvature we need to assume that there is a manifold which fits the distribution of the data. Recently, Fefferman-Mitter-Narayanan in [10] have developed an algorithm for testing the hypothesis that there exists a manifold which fits the distribution of a sample, however, a problem that remains open is how to best estimate the dimension of a submanifold from a sample of points. See section 5.3 for more discussion.

In another vein, there is much current interest in a converse problem: The development of algorithms for generating point clouds on manifolds or even on surfaces. Recently, there has been progress in this direction by Karcher-Palais-Palais in [12], specifically on methods for generating point clouds on implicit surfaces using Monte Carlo simulation and the Cauchy-Crofton formula.

1.1. **Computational issues.** A naive dense implementation of our algorithm is undoubtably slow, and becomes unfeasible for large numbers of points. Indeed, inspecting the formulas in Section 4.1, we see that to compute the value of the coarse Ricci curvature at a point (x, y) requires on order of n^3 computations. This reduces to $O(n^2)$ if we are in the presence of a constant density based on assuming also that one knows the dimension of the underlying manifold . Note that $O(n^2)$ is the order of naive dense computation of an optimal transport problem via linear programming. Thus computing the full coarse Ricci object should require $O(n^5)$. Of course, even storing the distance matrix itself can become prohibitive for large enough n. Nonetheless, for smaller sets (say of size $n < 500$) we were able construct a "coarse Ricci flow" based on our construction that was successful in clustering some simple data sets [28]. On the other hand, it appears that improvements can be made with very little loss in accuracy, by using only nearest neighbors or truncating the kernel, obtaining a sparse distance function and sparse coarse Ricci object.

1.2. **Organization of the paper.** This paper is devoted to proving Theorems B and C, which are the ingredients for Theorem A. In the section 2 we review the relevant background. In Section 3 we summarize some of the results in [1]. The core of the paper will be Section 4 devoted to the prove of Theorem C. In Section 5 we review the construction of local PCA in [25] and show how can we combine this construction with Theorem C to prove Theorem B. We also include a discussion of dimension reduction and estimation. For the reader's convenience we provide a table of notations at the end of section 2.

1.3. **Acknowledgements.** The authors would like to thank Amit Singer, Hau-Tieng Wu and Charles Fefferman for constant encouragement. The first author would like to express gratitude to Adolfo Quiroz for very useful conversations on the topic of empirical processes, and to Richard Palais for bringing his work to the attention of both authors. The second author would like to thank Jan Maas for useful conversations, and Matthew Kahle for stoking his interest in the topic. We would also like to thank Bartek Siudeja for some code implementations of the algorithm. Finally, the authors are grateful to the anonymous referee for providing many comments and suggestions that helped improving the presentation of the paper.

2. Background and Definitions

In this section we recall (cf. [2]) how Ricci curvature on general metric spaces can be constructed with an operator, in particular the infinitesimal generator of a diffusion semi-group. When the space is a metric measure space, we use a family of operators which are intended to approximate a Laplace operator on the space at scale t. As this definition holds on metric measure spaces constructed from sampling points from a manifold, we can define an *empirical or sample version* of the Ricci curvature, given a bandwidth parameter t. As mentioned above, this last construction will have an application to the manifold learning problem, namely it will serve to predict the

Ricci curvature of an embedded submanifold of $\mathbb{R}^N$ if one only has a point cloud on the manifold and the distribution of the sample has a smooth positive density.

2.0.1. *Carré du champ.* We now recall how Bakry and Emery [3] related the notion of *Carré du Champ* to Ricci curvature. Let P_t be a 1-parameter family of operators of the form

$$P_t f(x) = \int_M f(y) p_t(x, dy), \tag{2.1}$$

where f is a bounded measurable function defined on M and $p_t(x, dy)$ is a non-negative kernel. We assume that P_t satisfies the semi-group property, i.e.

$$P_{t+s} = P_t \circ P_s. \tag{2.2}$$

$$P_0 = \mathrm{Id}. \tag{2.3}$$

In $\mathbb{R}^n$, an example of P_t is the *heat semi-group*, defined by the density

$$p_t(x, dy) = \frac{1}{(2\pi t)^{n/2}} e^{-\frac{|x-y|^2}{2t}} dy, t \geq 0. \tag{2.4}$$

If now P_t is a diffusion semi-group defined on (M, g), we let L be the infinitesimal generator of P_t, which is densely defined in L^2 by

$$Lf = \lim_{t \to 0} t^{-1}(P_t f - f). \tag{2.5}$$

We consider a bilinear form which has been introduced in potential theory by J.P. Roth [20] and by Kunita in probability theory [14] and measures the failure of L from satisfying the Leibnitz rule. This bilinear form is defined as

$$\Gamma(L, u, v) = \frac{1}{2}\left(L(uv) - L(u)v - uL(v)\right). \tag{2.6}$$

The expression given by (2.6) is called *Carré du Champ.* When L is the rough Laplacian with respect to the metric g, then

$$\Gamma(\Delta_g, u, v) = \langle \nabla u, \nabla v \rangle_g.$$

We will also consider the *iterated Carré du Champ* introduced by Bakry and Emery denoted by Γ_2 and defined by

$$\Gamma_2(L, u, v) = \frac{1}{2}\left(L(\Gamma(L, u, v)) - \Gamma(L, Lu, v) - \Gamma(L, u, Lv)\right). \tag{2.7}$$

Note that if we restrict our attention to the case $L = \Delta_g$ the Bochner formula yields

$$\begin{aligned} \Gamma_2(\Delta_g, u, v) &= \frac{1}{2}\Delta\langle \nabla u, \nabla v\rangle_g - \frac{1}{2}\langle \nabla \Delta_g u, \nabla v\rangle_g - \frac{1}{2}\langle \nabla u, \nabla \Delta_g v\rangle_g \\ &= \mathrm{Ric}(\nabla u, \nabla v) + \langle \nabla_g^2 u, \nabla_g^2 v\rangle_g. \end{aligned} \tag{2.8}$$

We observe immediately that if $\nabla u = e_i$ and $\nabla_g^2 u = 0$ one can recover the Ricci tensor via

$$\Gamma_2(\Delta_g, u, u) = \mathrm{Ric}(e_i, e_i). \tag{2.9}$$

A geometric interpretation for the Carré du Champ in (2.6) and its iterate (2.7) is given by their role in formulating the so-called $CD(K, N)$ *curvature condition* due to Bakry and Emery. The fundamental observation of Bakry and Emery is that the properties of Ricci curvature lower bounds can be observed and exploited by using the bilinear form Γ_2. With this in mind, they define a curvature-dimension condition for an operator L on a smooth metric measure space $(X, g, d\nu)$ as follows. If there exist measurable functions $k : X \to \mathbb{R}$ and $N : X \to [1, \infty]$ such that for every f on a set of functions dense in $L^2(X, d\nu)$ the inequality

$$\Gamma_2(L, f, f) \geq \frac{1}{N}(Lf)^2 + k\Gamma(L, f, f) \tag{2.10}$$

holds, then the space X together with the operator L satisfies the $CD(k, N)$ *condition*, where k stands for curvature and N for dimension. In [3], it is shown that when considering a smooth metric measure space $(M^n, g, e^{-\rho}d\mathrm{vol})$ one has a natural diffusion operator given by

$$\Delta_\rho u = \Delta u - \langle \nabla \rho, \nabla u \rangle, \tag{2.11}$$

corresponding to the variation of the Dirichlet energy with respect to the measure $e^{-\rho}d\mathrm{vol}$. By studying the properties of Δ_ρ, Bakry and Emery arrive at the following dimension and weight dependent definition of the Ricci tensor:

$$\mathrm{Ric}_N = \begin{cases} \mathrm{Ric} + \mathrm{Hess}_\rho & \text{if } N = \infty, \\ \mathrm{Ric} + \mathrm{Hess}_\rho - \frac{1}{N-n}(d\rho \otimes d\rho) & \text{if } n < N < \infty, \\ \mathrm{Ric} + \mathrm{Hess}_\rho - \infty(d\rho \otimes d\rho) & \text{if } N = n, \\ -\infty & \text{if } N < n, \end{cases} \tag{2.12}$$

and moreover, they showed the equivalence between the $CD(k, N)$ condition (2.10) and the bound $\mathrm{Ric}_N \geq k$. The $CD(k, N)$ condition that we stated above is also related to the *displacement convexity* condition used by Lott and Villani in [16] to study the stability of lower bounds on Ricci curvature under limits in the Gromov-Hausdorff sense. For the problem of estimating Ricci curvature from a point cloud, the Bochner formula (2.8) and (2.9) give a direct connection between the Carré du Champ, its iterate and Ricci curvature.

2.0.2. *Approximations of the Laplacian, Carré du Champ and its iterate.* Following [4] and [9], we recall how to construct operators which can be thought of as approximations of the Laplacian on metric measure spaces. Consider a metric measure space (X, d, μ) with a Borel σ-algebra such that $\mu(X) < \infty$. Given $t > 0$, let θ_t be given by

$$\theta_t(x) = \int_X e^{-\frac{d^2(x,y)}{2t}} d\mu(y). \tag{2.13}$$

We define a 1-parameter family of operators L_t as follows: given a function f on X let

$$L_t f(x) = \frac{2}{t\theta_t(x)} \int_X (f(y) - f(x))\, e^{-\frac{d^2(x,y)}{2t}} d\mu(y). \tag{2.14}$$

With respect to this L_t one can define a Carré du Champ on appropriately integrable functions f, h by

$$\Gamma(L_t, f, h) = \frac{1}{2} \left(L_t(fh) - (L_t f)h - f(L_t h) \right), \tag{2.15}$$

which simplifies to

$$\Gamma(L_t, f, h)(x) = \frac{1}{t\theta_t(x)} \int_X e^{-\frac{d^2(x,y)}{2t}} (f(y) - f(x))(h(y) - h(x)) d\mu(y). \tag{2.16}$$

In a similar fashion we define the iterated Carré du Champ of L_t to be

$$\Gamma_2(L_t, f, h) = \frac{1}{2} \left(L_t(\Gamma(L_t, f, h)) - \Gamma(L_t, L_t f, h) - \Gamma(L_t, f, L_t h) \right). \tag{2.17}$$

Remark 2.1. Note that Belkin and Niyogi [4, pg 1295, eq (6)] normalize by a factor $(4\pi t)^{d/2}$, which requires knowledge of the dimension. Our definition of L_t (2.14) differs from Belkin-Niyogi operator in that we normalize by $\theta_t(x)$ instead. This has a cost in that convergence may be slower, but has the advantage of being dimensionless, allowing our general discussion to fit into the framework of spaces with lower Ricci curvature bound, for example, the disjoint union of two manifolds of different dimensions or a sequence of manifolds which may be collapsing.

2.0.3. *Empirical Carré du Champ at a given scale.* We can also define empirical versions of $L_t, \Gamma(L_t, \cdot, \cdot)$ and $\Gamma_2(L_t, \cdot, \cdot)$. On a space which consists of n points $\{\xi_1, ..., \xi_n\}$ sampled from a manifold, it is natural to consider the *empirical measure* defined by

$$\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\xi_i} \tag{2.18}$$

where δ_{ξ_i} is the atomic point measure at the point ξ_i (also called δ-mass). For any function $f : X \to \mathbb{R}$ we will use the notation

$$\mu_n f = \int_X f(y) d\mu_n(y) = \frac{1}{n} \sum_{j=1}^{n} f(\xi_j).$$

Notation 2.2. We will use the "hat" notation (for example $\hat{L}_t$) to distinguish those operators, measures, or t-densities that have been constructed from a sample of finite points.

To be more precise, we define the operator $\hat{L}_t$ as

$$\hat{L}_t f(x) = \frac{2}{t\hat{\theta}_t(x)} \int_X (f(y) - f(x))\, e^{-\frac{d^2(x,y)}{2t}} d\mu_n(y), \tag{2.19}$$

where

$$\hat{\theta}_t(x) = \int_X e^{-\frac{d^2(x,y)}{2t}} d\mu_n(y) = \frac{1}{n}\sum_{j=1}^{n} e^{-\frac{d(\xi_j,x)^2}{2t}}, \tag{2.20}$$

and of course

$$\int_X (f(y)-f(x))\, e^{-\frac{d^2(x,y)}{2t}} d\mu_n(y) = \frac{1}{n}\sum_{j=1}^{n} e^{-\frac{d(\xi_j,x)^2}{2t}} (f(\xi_j)-f(x)). \tag{2.21}$$

The *sample version of Carré du Champ* will be the bilinear form $\Gamma(\hat{L}_t, f, h)$ which from (2.16) takes the form

$$\Gamma(\hat{L}_t, f, h)(x) = \frac{1}{t\hat{\theta}_t(x)}\frac{1}{n}\sum_{j=1}^{n} e^{-\frac{d^2(\xi_j,x)}{2t}} (f(\xi_j)-f(x))(h(\xi_j)-h(x)). \tag{2.22}$$

We denote the iterated Carré du Champ corresponding to $\hat{L}_t$ by $\Gamma_2(\hat{L}_t, f, h)$, and by this we mean

$$\Gamma_2(\hat{L}_t, f, h) = \frac{1}{2}\left(\hat{L}_t(\Gamma(\hat{L}_t f, h)) - \Gamma(\hat{L}_t, \hat{L}_t f, h) - \Gamma(\hat{L}_t, f, \hat{L}_t h)\right). \tag{2.23}$$

2.1. **Statement of Results.**

2.1.1. *Applications to Manifold Learning.* We now show how our notion of empirical Carré du Champ at a given scale has applications to the Manifold Learning Problem. For the rest of subsection 2.1.1 we will consider a closed, smooth, embedded submanifold Σ of $\mathbb{R}^N$, and the metric measure space will be $(\Sigma, \|\cdot\|, d\mathrm{vol})$, where

- $\|\cdot\|$ is the distance function in the ambient space $\mathbb{R}^N$,
- $d\mathrm{vol}_\Sigma$ is the volume element corresponding to the metric g induced by the embedding of Σ into $\mathbb{R}^N$.

In addition we will adopt the following conventions

- All operators L_t, $\Gamma(L_t, \cdot, \cdot)$ and $\Gamma_2(L_t, \cdot, \cdot)$ will be taken with respect to the distance $\|\cdot\|$ and the measure $d\mathrm{vol}_\Sigma$.
- All sample versions $\hat{L}_t$, $\Gamma(\hat{L}_t, \cdot, \cdot)$ and $\Gamma_2(\hat{L}_t, \cdot, \cdot)$ are taken with respect to the ambient distance $\|\cdot\|$.

The choice of the above metric measure space is consistent with the setting of manifold learning in which no assumption on the geometry of the submanifold Σ is made, in particular, we have no a priori knowledge of the geodesic distance and therefore we can only hope to use the chordal distance as a reasonable approximation for the geodesic distance. We will show that while our construction at a scale t involves only information from the ambient space, the limit as t tends to 0 will recover the Ricci curvature of the submanifold. As pointed out by Belkin-Niyogi [4, Lemma 4.3], the chordal and intrinsic distance squared functions on a smooth submanifold disagree first at fourth order near a point , so while much of the analysis is done on submanifolds, the intrinsic geometry will be recovered in the limit.

We now address the problem of choosing a bandwidth parameter depending on the size of the data and the dimension of the submanifold Σ, such that the sequence of empirical Ricci curvatures corresponding to the size of the data converge almost surely to the actual Ricci curvature of Σ at a point. In order to simplify the presentation of our results, we start by stating the simplest possible case, which corresponds to a uniformly distributed i.i.d.sample $\{\xi_1, \ldots, \xi_n\}$. The more general case of distributions with strictly positive density with respect to the Lebesgue measure can be explored using the same methods presented here, but the proof is quite lengthy.

Theorem B (Approximation of the Ricci Curvature)**.** *Consider the metric measure space* $(\Sigma, \|\cdot\|, d\mathrm{vol}_\Sigma)$ *where* $\Sigma^d \subset \mathbb{R}^N$ *is a smooth closed embedded submanifold. Suppose that we have a uniformly distributed i.i.d. sample* $\{\xi_1, \ldots, \xi_n\}$ *of points from* Σ*. For* $\sigma > 0$*, let*

$$t_n = n^{-\frac{1}{3d+3+\sigma}}. \tag{2.24}$$

For $x \in \Sigma$ *there exists a sequence of* d*-tuples of orthogonal vectors, and* $d \times d$ *Ricci matrices* $\hat{R}_{i,j}$ *representing the Ricci curvature on these vectors, such at if* $\eta \in T_x\Sigma$*, then*

$$\left| \hat{R}_{i,j}\eta^i\eta^j - \mathrm{Ric}_x(\eta, \eta) \right| \xrightarrow{\text{a.s}} 0, \tag{2.25}$$

where η^i *are the components of the vector* η *projected onto the* d*-tuple of vectors approximating the tangent plane.*

See Section 5 for more on the PCA construction. Even though the approximation method used to obtain Theorem B is inspired by the notion of *Coarse Ricci* curvature introduced by the authors in [2, 1], Theorem B relies heavily on a precise estimation of the tangent space at a point by means of local PCA, as opposed to the approximation proposed in [2, 1] based on the construction of an "auxiliary tangent space" by taking segments within the point cloud.

Remark 2.3. The convergence is more certain (but perhaps slower) if t_n is chosen to go to zero slower than in (2.24), i.e., slower than

$$t_n = n^{-\frac{1}{3d+3+\sigma+\tau}},$$

where τ is any positive number. In particular, if one replaces d with an upper bound on d, then Theorems B, C, and Corollary A still hold. We also remark that we do not attempt compute the rate which one minimizes the mean-squared error of the estimation - that is we do not attempt to compute or justify a version of *Silverman's rule of thumb* (see [23]). It could be that there is bandwidth parameter which gives a faster but less sure (i.e. with lower probability) convergence rate.

Besides local PCA, the proof of Theorem B relies heavily on the following theorem for the iterated Carré du Champ: Theorem B in turn follows from an approximation result for the iterated Carré du Champ:

Theorem C. *Consider the metric measure space $(\Sigma, \|\cdot\|, d\mathrm{vol}_\Sigma)$ where $\Sigma^d \subset \mathbb{R}^N$ is a smooth closed embedded submanifold. For $\sigma > 0$, let*

$$t_n = n^{-\frac{1}{3d+3+\sigma}}. \tag{2.26}$$

(a) *If $f \in C^\infty(\mathbb{R}^N)$, then*

$$\sup_{\xi\in\Sigma}\left|\hat{\Gamma}_2(\hat{L}_{t_n}, f, f)(\xi) - \Gamma_2(L_{t_n}, f, f)(\xi)\right| \xrightarrow{a.s.} 0.$$

(b) *If $\mathscr{L}_M$ is the class of linear functions*

$$\mathscr{L}_M = \{f_\eta|_\Sigma : f_\eta(x) = \langle \eta, x\rangle, \|\eta\| \le M, x \in \mathbb{R}^N\},$$

where $\|\cdot\|$ is the ambient distance in $\mathbb{R}^N$ and $f_\eta|_\Sigma$ is the restriction of the linear function $f_\eta : \mathbb{R}^N \to \mathbb{R}$ given by $f_\eta(x) = \langle \eta, x\rangle$ to Σ, then

$$\sup_{f\in\mathscr{L}_M}\sup_{\xi\in\Sigma}\left|\hat{\Gamma}_2(\hat{L}_{t_n}, f, f)(\xi) - \Gamma_2(L_{t_n}, f, f)(\xi)\right| \xrightarrow{a.s.} 0.$$

The proof of Theorem C requires using ideas from the theory of *empirical processes* for which we will provide the necessary background in Section 4. As pointed out in the introduction, since we are interested in recovering an object from its sample version, we are forced to consider a law of large numbers in order to obtain convergence in probability or almost surely. The problem is that the sample version of $\Gamma_2(L_t, \cdot, \cdot)$ involves a high correlation between the data points, destroying independence and any hope of applying large number results directly. The idea then is to reduce the convergence of the sample version of $\Gamma_2(L_t, \cdot, \cdot)$ to the application of a uniform law of large numbers to certain classes of functions. Theorem C is proved in Section 4. In section 5 we will prove that Theorem C indeed implies Theorem B. This will require results from [1].

2.1.2. *Smooth Metric Measure Spaces and non-Uniformly Distributed Samples.* Consider a smooth metric measure space $(M, g, e^{-\rho}d\mathrm{vol})$ and let Δ_ρ be the operator

$$\triangle_\rho u = \triangle_g u - \langle \nabla\rho, \nabla u\rangle_g.$$

In [9], the authors consider a family of operators L_t^α which converge to $\triangle_{2(1-\alpha)\rho}$. Note that a standard computation (cf [32, Page 384]) gives

$$\Gamma_2(\triangle_{2(1-\alpha)\rho}, f, f) = \frac{1}{2}\Delta_g\|\nabla f\|_g^2 - \langle\nabla\rho, \nabla\Delta_g f\rangle_g + 2(1-\alpha)\nabla_g^2\rho(\nabla f, \nabla f).$$

We adapt [9] to our setting: Recall that

$$\theta_t(x) = \int_X e^{-\frac{d^2(x,y)}{2t}}d\mu(y),$$

and define, for $\alpha \in [0, 1]$

$$\theta_{t,\alpha}(x) = \int_X e^{-\frac{d^2(x,y)}{2t}}\frac{1}{[\theta_t(y)]^\alpha}d\mu(y). \tag{2.27}$$

We can define the operator

$$L_t^\alpha f(x) = \frac{2}{t}\frac{1}{\theta_{t,\alpha}(x)}\int_X e^{-\frac{d^2(x,y)}{2t}}\frac{1}{[\theta_t(y)]^\alpha}\left(f(y)-f(x)\right)d\mu(y), \tag{2.28}$$

and again obtain bilinear forms $\Gamma(L_t^\alpha, f, f)$ and $\Gamma_2(L_\alpha^t, f, f)$. For the rest of the section we will consider the metric measure space $(\Sigma, \|\cdot\|, e^{-\rho}d\mathrm{vol}_\Sigma)$ where $\Sigma^d \subset \mathbb{R}^N$ is an embedded submanifold, $\|\cdot\|$ is the ambient distance and ρ is a smooth function in Σ. We again take all the operators $L_t^\alpha, \Gamma_t(L_t^\alpha, \cdot, \cdot)$ and $\Gamma_2(L_t^\alpha, \cdot, \cdot)$ and their sample counterparts $\hat{L}_t^\alpha, \Gamma_t(\hat{L}_t^\alpha, \cdot, \cdot)$ and $\Gamma_2(\hat{L}_t^\alpha, \cdot, \cdot)$ with respect to the data of $(\Sigma, \|\cdot\|, e^{-\rho}d\mathrm{vol}_\Sigma)$.

Based on estimates in [2] and calculations similar to the proof of Theorem C, we can prove

Theorem D (Non-uniform case)**.** *Consider the metric space $(\Sigma, \|\cdot\|)$ where $\Sigma^d \subset \mathbb{R}^N$ is a smooth closed embedded submanifold. Suppose that we have an i.i.d. sample $\{\xi_1, \ldots, \xi_n\}$ of points from Σ whose common distribution has density $e^{-\rho}$. For $\sigma > 0$, let*

$$t_n = n^{-\frac{1}{4d+4+\sigma}}.$$

Then, for any $x \in \Sigma$ and any $\eta \in T_x\Sigma$ there exists a sequence of functions f_n constructed from the data such that

$$\left|\hat{\Gamma}_2^\alpha(\hat{L}_{t_n}^\alpha, f_n, f_n)(\xi) - \mathrm{Ric}_{x,\alpha}(\eta,\eta)\right| \xrightarrow{a.s.} 0,$$

and where

$$\mathrm{Ric}_\alpha = \mathrm{Ric} + 2(1-\alpha)\nabla_g^2\rho. \tag{2.29}$$

We omit the proof of Theorem D. Heuristically, the order of decay of t_n is not hard to determine, but the proof is exceeding long and tedious.

2.2. **Notation and table of definitions.** Below we give a table of object that frequently appear.

Table 1. Summary of Notations used in the paper

Notation	Description	Where Found
Σ^d	Embedded submanifold of $\mathbb{R}^N$	
N, d	Dimensions of ambient space and submanifold	
n	Number of sample points	
t	Bandwidth parameter	
ξ_i	point sampled from Σ^d	
$\Gamma(L, u, v)$	First Carré du champ	(2.6)
$\Gamma_2(L, u, v)$	Iterated Carré du champ	(2.7)
L_t	Finite scaled approximate Laplacian	(2.14)
Δ_g	Laplace-Beltrami operator on a manifold	
$\theta_t(x)$	Scale-t density at x	(2.13)
μ_n	measure determined by sampling n-points	(2.18)
$\hat{L}_t$	Empirical Laplace at scale t	(2.19)
$\Gamma(\hat{L}_t, f, h)$	Sampled Carré du Champ	(2.22)
$\Gamma_2(\hat{L}_t, f, h)$	Sampled iterated Carré du Champ	(2.23)
$\\| \cdot \\|$	distance function on $\mathbb{R}^N$	
t_n	Scale parametric chosen based on n	(4.101)
$F_{x,y}(z)$	Approximate signed distance function from x to y	(3.6)
$f_{x,y}(z)$	Scaled approximate signed distance function	(3.7)
μ	Probability distribution on Σ^d	
$\mathcal{N}(\mathcal{F}, \delta)$	Covering number of a set of functions, for radius δ	(4.13)
$\\|f\\|_{\text{A-t-Lip}}$	Almost t Lipschitz norm	(4.20)
$\mathcal{F}^t_{f,h}, \mathcal{G}^t, \mathcal{H}^t_h$	Classes of functions	(4.21)(4.22), Lemma 4.15
τ	The reach of Σ	Definition 4.10
$\mathcal{A}(X, \delta)$	Ambient covering number of a set $X \subset \mathbb{R}^N$	(4.26)
$U_{V,\tau}(r)$	Covering bound function	Theorem 4.11
$Q_t(\mathcal{F}, \varepsilon, C, n)$		(4.91)
$\mathcal{Q}(t, \varepsilon, n)$		(4.103)
$O_{BC}(\beta)$		(4.104)
$C^*(\Sigma)$, $C^*_1(\Sigma)$	Constants depending on Σ	Convention 4.14

We also give a table for the different Ricci curvature operators

TABLE 2. Notation used for different Ricci Curvature operators

Notation	Extended Description	Where Found
$\mathrm{Ric}_L(x,y)$	Coarse Ricci operator determined by operator L and evaluated at x and y. This is roughly quadratic in $d(x,y)$.	Definition 3.3
$\mathrm{RIC}_L(x,y)$	Life-sized Ricci operator determined by operator L and evaluated at x and y. This need not vanish as $x \to y$.	Definition 3.4
$\mathrm{Ric}_{\Delta_g}(x,y)$	Coarse Ricci operator determined by Laplace operator Δ_g and evaluated at x and y. This is computable when Δ_g is available.	Definition 3.3, applied to Δ_g
$\hat{R}_{i,j}$	Ricci matrices defined w.r.t PCA basis for $T_x\Sigma^d$	(5.2)
$\mathrm{Ric}_x(\cdot,\cdot)$	Classical Ricci 2-form on $T_x\Sigma^d$	
Ric_α	Weighted Ricci curvature with parameter α and density $e^{-\rho}$	(2.29)

3. Summary of previous results. Proof of Theorem B

We now recall the following result proved in [1]:

Theorem 3.1 (See [1]). *Let $\Sigma^d \subset \mathbb{R}^N$ be a closed embedded submanifold, let g be the Riemannian metric induced by the embedding, and let $(\Sigma, \|\cdot\|, d\mathrm{vol}_\Sigma)$ be the metric measure space defined with respect to the ambient distance. Given any $f \in \mathscr{C}_M \subset C^5(\Sigma)$ where $\mathscr{C}_M$ is the class of functions*

$$\mathscr{C}_M = \left\{ f : \|f\|_{C^5(\Sigma)} \leq M \right\},$$

there exists a constant C_1 depending on the geometry of Σ and the function f such that

$$\sup_{x\in\Sigma} |\Gamma_2(\Delta_g, f, f)(x) - \Gamma_2(L_t, f, f)(x)| < C_1(\Sigma, M)t^{1/2}.$$

A fundamental step for proving Theorem 3.1 is the following proposition shown in [1]: For simplicity we will assume that $(\Sigma, d\mathrm{vol}_\Sigma)$ has unit volume. Recall the definitions (2.13), (2.14), (2.15), (2.16) and (2.17).

Proposition 3.2 (See [1]). *Suppose that Σ^d is a closed, embedded, unit volume submanifold of $\mathbb{R}^N$. Let also g be the metric induced by the embedding of Σ^d into $\mathbb{R}^N$.*

For any x in Σ and for any functions f, h in $C^5(\Sigma)$ we have

$$\frac{(2\pi t)^{d/2}}{\theta_t(x)} = 1 + tG_1(x) + t^{3/2}R_1(x), \tag{3.1}$$

$$\Gamma(L_t, f, h)(x) = \langle \nabla_\Sigma f(x), \nabla_\Sigma h(x)\rangle_g + t^{1/2}G_2(x, J^2(f)(x), J^2(h)(x)) \tag{3.2}$$

$$+ tG_3(x, J^3(f)(x), J^3(h)(x)) + t^{3/2}R_2(x, J^4(f)(x), J^4(h)(x)), \tag{3.3}$$

$$L_t f(x) = \Delta_g f(x) + t^{1/2}G_4(x, J^3(f)(x)) + tG_5(x, J^4 f(x)) + t^{3/2}R_3(x, J^5 f(x)), \tag{3.4}$$

and

$$\Gamma_2(L_t, f, f)(x) = \Gamma_2(\Delta_g, f, f)(x) + t^{1/2}R_5(x, J^5 f(x)) \tag{3.5}$$

where each G_i is a locally defined function, which is smooth in its arguments, and $J^k(u)$ is a locally defined k-jet of the function u. Also, each R_i is a locally defined function of x which is uniformly bounded in terms of its arguments.

We will show in Section 5 that Propositions 3.1 and 3.2 are needed to prove Theorem B.

3.1. **Life-Sized Coarse Ricci Curvature.** As mentioned above, in [2, 1], the authors have formulated a notion of Coarse Ricci Curvature alternative to Ollivier's Coarse Ricci curvature. The purpose of this section is to formulate the results of this paper in terms of the notions developed in [2, 1]. In particular, we show how Ricci curvature can be approximated using test functions different to the linear functions in Theorem B.

Heuristically, the Bochner formula (2.8) defines the Ricci curvature on the gradient of a function ∇u, up to an error term determined by the Hessian of u. In other words, if u is a function with small Hessian at a point, then the Bochner formula provides a good approximation for the Ricci curvature in the direction ∇u. For this reason, given a pair of points x and y, we attempt to construct a "linear" function that has gradient "pointing" from x to y. For any $x, y, x \neq y \in X$, we define

$$F_{x,y}(z) = \frac{1}{d(x,y)}\frac{1}{2}\left(d^2(x,y) - d^2(y,z) + d^2(z,x)\right). \tag{3.6}$$

Notice that in Euclidean space, this simplifies to

$$F_{x,y}(z) = \frac{1}{\|y-x\|}\left(x^2 + \langle z, y-x\rangle\right) = \frac{1}{\|y-x\|}\left(x^2 + z\cdot(y-x)\right),$$

which has gradient a constant unit vector

$$\nabla F_{x,y} = \frac{(y-x)}{\|y-x\|}.$$

We also define for $x, y \in X$, the following function

$$f_{x,y}(z) = \frac{1}{2}\left(d^2(x,y) - d^2(y,z) + d^2(z,x)\right). \tag{3.7}$$

In Euclidean space, this has gradient

$$\nabla f_{x,y} = (y - x).$$

The function F is chosen so that the gradient will not vanish as the points x, y approach each other, whereas the function f is chosen so that a quadratic form on the gradient $\nabla f_{x,y}$ will be the same order as the distance-squared function $d^2(x, y)$ as x, y approach each other. Both have natural interpretations, as we will see below. Note that our definitions (3.7, 3.6) are designed as an approximation on general metric spaces, when computing coarse Ricci curvature. When computing Ricci curvature on a known tangent vector in Euclidean space, such a function is unneccessary. In the combinatorial setting, one can make a special choice which minimizes the contribution of the Hessian in the Bochner formula, see [13, pg. 659].

Definition 3.3. *Given an operator L we define the coarse Ricci curvature for L as*

$$\mathrm{Ric}_L(x, y) = \Gamma_2(L, f_{x,y}, f_{x,y}).$$

In principle, the functions in (3.6) and (3.7) serve as a substitute of the linear functions in Theorem B (see also Section 5). Notice that the quantity in (3.3) is the same order as distance squared. To obtain a quantity that does not vanish near the diagonal, we use (3.6):

Definition 3.4. *Given an operator L we define the life-sized coarse Ricci curvature for L as*

$$\mathrm{RIC}_L(x, y) = \Gamma_2(L, F_{x,y}, F_{x,y}).$$

From this, one can define notions of *empirical coarse Ricci curvature*, by taking the sample versions of (3.3) and (3.4)
Inspired by Theorems B and C, the results at the end of Section 4 will lead easily to the following.

Corollary A. *Let $\Sigma^d \subset \mathbb{R}^N$ be an embedded submanifold and consider the metric measure space $(\Sigma, \|\cdot\|, \mathrm{dvol}_\Sigma)$. Suppose that we have an i.i.d. uniformly distributed sample $\xi_1, \ldots, \xi_n$ drawn from Σ. Let*

$$t_n = n^{-\frac{1}{3d+3+\sigma}}, \tag{3.8}$$

for any $\sigma > 0$. Then

$$\sup_{x \in \Sigma} \left| \hat{\Gamma}_2(L_{t_n}, F_{x,y}, F_{x,y})(x) - \mathrm{RIC}_{\Delta_g}(x, y) \right| \xrightarrow{a.s.} 0.$$

In other words, there is a choice of scale depending on the size of the data and the dimension of the submanifold for which the corresponding empirical life-sized coarse Ricci curvatures converge almost surely to the life-sized coarse Ricci curvature.

The proof is given at the end of Section 4. Another result proved in [1] is

Corollary 3.5. *With the hypotheses of Theorem 3.1 we have*

$$\mathrm{Ric}_{\Delta_g}(x, y) = \lim_{t \to 0} \Gamma_2(L_t, f_{x,y}, f_{x,y})(x).$$

We note that the relation between the coarse Ricci curvature and the Ricci curvature is as follows.

Proposition 3.6 (See [1])**.** *Suppose that M is a smooth Riemannian manifold. Let $V \in T_xM$ with $g(V,V) = 1$. Then*

$$\mathrm{Ric}(V,V) = \lim_{\lambda \to 0} \mathrm{RIC}_{\triangle_g}(x, \exp_x(\lambda V)).$$

4. Empirical Processes and Convergence. Proof of Theorem C

The goal of this section is to prove Theorem C. This will be done using tools from the theory of empirical processes in order to establish uniform laws of large numbers in a sense that we will explain in Sections 4.2 through 4.6. For a standard reference in the theory of empirical processes, see [30]. See also [26] for further applications of the theory of empirical processes to the recovery of diffusion operators from a sample.

4.1. **Estimators of the Carré Du Champ and the Iterated Carré Du Champ in the uniform case.** Let us assume that the measure μ is the volume measure $d\mathrm{vol}_\Sigma$. Recall that our formal definition of the Carré du Champ of L_t with respect to the uniform distribution is given by

$$\Gamma(L_t, f, h) = \frac{1}{t}\frac{1}{\theta_t(x)}\left(\int_\Sigma e^{-\frac{\|x-y\|^2}{2t}}\left(f(y) - f(x)\right)\left(h(y) - h(x)\right) d\mu(y)\right). \tag{4.1}$$

It is clear from (4.1) that a sample estimator of the Carré Du Champ at a point x is given by

$$\hat{\Gamma}(L_t, f, f)(x) = \frac{1}{t}\frac{1}{\hat{\theta}_t(x)}\left(\frac{1}{n}\sum_{j=1} e^{-\frac{\|x-\xi_j\|^2}{2t}}\left(f(\xi_j) - f(x)\right)\left(h(\xi_j) - h(x)\right)\right), \tag{4.2}$$

and recall that we defined the t-Laplace operator by

$$L_t f(x) = \frac{2}{t}\frac{1}{\theta_t(x)}\int_\Sigma e^{-\frac{\|x-y\|^2}{2t}}\left(f(y) - f(x)\right) d\mu(y), \tag{4.3}$$

and its sample version is

$$\hat{L}_t f(x) = \frac{2}{t}\frac{1}{\hat{\theta}_t(x)}\frac{1}{n}\sum_{j=1}^{n} e^{-\frac{\|x-\xi_j\|^2}{2t}}\left(f(\xi_j) - f(x)\right). \tag{4.4}$$

Recall that the iterated Carré du Champ is

$$\Gamma_2(L_t, f, h) = \frac{1}{2}\left(L_t\Gamma(L_t, f, h) - \Gamma(L_t, L_t f, f) - \Gamma(L_t, f, L_t h)\right). \tag{4.5}$$

For simplicity, we will evaluate $\Gamma_2(L_t, \cdot, \cdot)$ at a pair (f, f) instead of (f, h) and by symmetry it is clear that we obtain

$$\Gamma_2(L_t, f, f) = \frac{1}{2}\left(L_t(\Gamma_t(f, f)) - 2\Gamma_t(L_t f, f)\right). \tag{4.6}$$

Combining the sample versions of Γ_t and L_t we obtain a sample version for $\Gamma_2(L_t, f, f)$

$$\hat{\Gamma}_2(L_t,f,f)(x)=\frac{1}{t^2}\frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\frac{1}{\hat{\theta}_t(\xi_k)\hat{\theta}_t(x)}e^{-\frac{\|x-\xi_j\|^2}{2t}-\frac{\|\xi_j-\xi_k\|^2}{2t}}(f(\xi_j)-f(\xi_k))^2 \tag{4.7}$$

$$-\frac{1}{t^2n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\frac{1}{\hat{\theta}_t^2(x)}e^{-\frac{\|x-\xi_j\|^2}{2t}-\frac{\|x-\xi_k\|^2}{2t}}(f(\xi_k)-f(x))^2 \tag{4.8}$$

$$-\frac{2}{t^2n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\frac{1}{\hat{\theta}_t(x)\hat{\theta}_t(\xi_j)}e^{-\frac{\|x-\xi_j\|^2}{2t}-\frac{\|\xi_j-\xi_k\|^2}{2t}}(f(\xi_k)-f(\xi_j))(f(\xi_j)-f(x)) \tag{4.9}$$

$$+\frac{2}{t^2n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\frac{1}{\hat{\theta}_t^2(x)}e^{-\frac{\|x-\xi_j\|^2}{2t}-\frac{\|\xi_j-\xi_k\|^2}{2t}}(f(\xi_k)-f(x))(f(\xi_j)-f(x)). \tag{4.10}$$

In principle, the convergence analysis for (4.7)-(4.10) can be done using the following standard result in large deviation theory.

Lemma 4.1 (Hoeffding's Lemma)**.** *Let $\xi_1,\dots,\xi_n$ be i.i.d. random variables on the probability space $(\Sigma,\mathcal{B},\mu)$ where $\mathcal{B}$ is the Borel σ-algebra of Σ, and let $f:\Sigma\to[-K,K]$ be a Borel measurable function with $K>0$. Then for the corresponding empirical measure μ_n and any $\varepsilon>0$ we have*

$$\Pr\{|\mu_n f-\mu f|\geq\varepsilon\}\leq 2e^{-\frac{\varepsilon^2 n}{2K^2}}.$$

Observe, however, that (4.7)-(4.10) is a non-linear expression which will involve non-trivial interactions between the data points $\xi_1,\dots,\xi_n$. This non-trivial interaction between the points $\xi_1,\dots,\xi_n$ will produce a loss of independence and we will *not* be able to apply Hoeffding's Lemma directly to (4.7)-(4.10). In order to address this difficulty we will establish several uniform laws of large numbers which will provide us with a large deviation estimate for (4.7)-(4.10).

Remark 4.2. *We will not use directly the expression* (4.7)-(4.10)*, instead we will write* (4.7)-(4.10) *schematically in the form*

$$\hat{\Gamma}_2(L_t(f,f)(x)=\frac{1}{2}\left(\hat{L}_t\left(\hat{\Gamma}(f,f)\right)(x)-2\hat{\Gamma}_t(\hat{L}_tf,f)(x)\right), \tag{4.11}$$

which is clearly equivalent to (4.7)-(4.10)*.*

4.2. **Glivenko-Cantelli Classes.** A *Glivenko-Cantelli* class of functions is essentially a class of functions for which a uniform law of large numbers is satisfied.

Definition 4.3. *Let μ be a fixed probability distribution defined on Σ. A class $\mathcal{F}$ of functions of the form $f:\Sigma\to\mathbb{R}$ is* Glivenko-Cantelli *if*

(a) *$f\in L^1(d\mu)$ for any $f\in\mathcal{F}$,*
(b) *For any i.i.d. sample $\xi_1,\dots,\xi_n$ drawn from Σ whose distribution is μ we have uniform convergence in probability in the sense that for any $\varepsilon>0$*

$$\lim_{n\to\infty}\Pr{}^*\left\{\sup_{f\in\mathcal{F}}|\mu_n f-\mu f|>\varepsilon\right\}=0. \tag{4.12}$$

Remark 4.4. Note that in general we have to consider outer probabilities $\Pr^*$ instead of $\Pr$ because the class $\mathcal{F}$ may not be countable and the supremum $\sup_{f\in\mathcal{F}}|\mu_n f-\mu f|$ may not be measurable. On the other hand, if the class $\mathcal{F}$ is separable in $L^\infty(\Sigma)$, then we can replace $\Pr^*$ by $\Pr$. While all of the classes that we will encounter in this paper will be separable in $L^\infty(\Sigma)$, we use $\Pr^*$ when we deal with a general class.

Let $\mathcal{F}$ be a class of functions defined on Σ and totally bounded in $L^\infty(\Sigma)$. Given $\delta>0$ we let $\mathcal{N}(\mathcal{F},\delta)$ be the L^∞ δ-covering number of $\mathcal{F}$, i.e.,

$$\mathcal{N}(\mathcal{F},\delta)=\inf\{m:\mathcal{F}\text{ is covered by } m \text{ balls of radius } \delta \text{ in the } L^\infty \text{ norm}\}. \tag{4.13}$$

Lemma 4.5. *Let $\mathcal{F}$ be an equicontinuous class of functions in $L^\infty(\Sigma)$ that satisfies $\sup_{f\in\mathcal{F}}\{\|f\|_{L^\infty(\Sigma)}\}\le M<\infty$ for some $M>0$. Then for any distribution μ which is absolutely continuous with respect to $d\mathrm{vol}_\Sigma$, the class $\mathcal{F}$ is μ-Glivenko-Cantelli. Moreover, if $\xi_1,\dots,\xi_n$ is an i.i.d. sample drawn from Σ with distribution μ we have*

$$\Pr^*\left\{\sup_{f\in\mathcal{F}}|\mu_n f-\mu f|\ge\varepsilon\right\}\le 2\mathcal{N}\left(\mathcal{F},\frac{\varepsilon}{4}\right)e^{-\frac{\varepsilon^2 n}{8M^2}}.$$

Proof. By equicontinuity of $\mathcal{F}$, it follows from the Arzelà-Ascoli theorem that $\mathcal{F}$ is precompact in the $L^\infty(\Sigma)$ norm and hence totally bounded in $L^\infty(\Sigma)$. In particular for every $\delta>0$, the number $\mathcal{N}(\mathcal{F},\delta)$ is finite. Let $\mathcal{G}$ be a finite class such that the union of all balls with center in $\mathcal{G}$ and radius δ covers $\mathcal{F}$ and $|\mathcal{G}|=\mathcal{N}(\mathcal{F},\delta)$. For any $f\in\mathcal{F}$ there exists $\phi\in\mathcal{G}$ such that $\|f-\phi\|_{L^\infty(\Sigma)}<\delta$ and we obtain

$$|\mu_n f-\mu f|\le 2\delta+|\mu_n\phi-\mu\phi|, \tag{4.14}$$

and clearly

$$\sup_{f\in\mathcal{F}}|\mu_n f-\mu f|\le 2\delta+\max_{\phi\in\mathcal{G}}|\mu_n\phi-\mu\phi|. \tag{4.15}$$

Fixing $\varepsilon>0$ and choosing $\delta=\varepsilon/4$ we observe that

$$\Pr^*\left\{\sup_{f\in\mathcal{F}}|\mu_n f-\mu f|\ge\varepsilon\right\}\le\Pr\left\{\max_{\phi\in\mathcal{G}}|\mu_n\phi-\mu\phi|\ge\frac{\varepsilon}{2}\right\}, \tag{4.16}$$

and by Hoeffding's inequality we have

$$\Pr\left\{\max_{\phi\in\mathcal{G}}|\mu_n\phi-\mu\phi|\ge\frac{\varepsilon}{2}\right\}\le 2\mathcal{N}\left(\mathcal{F},\frac{\varepsilon}{4}\right)e^{-\frac{\varepsilon^2 n}{8M^2}}, \tag{4.17}$$

which implies the lemma. □

We now define the space $\mathrm{Lip}(\mathbb{R}^N)$ of functions with bounded Lipschitz semi-norm in $\mathbb{R}^N$. To be clear, we will be using the following norms and semi-norms:

$$\|f\|_{\mathrm{Lip}}=\sup_{x,y\in\mathbb{R}^N,x\neq y}\left\{\frac{|f(x)-f(y)|}{\|x-y\|}\right\}, \tag{4.18}$$

where f is a Lipschitz function defined on all of $\mathbb{R}^N$.

$$\|f\|_{C^k(\Sigma)} = \|f\|_{L^\infty(\Sigma)} + \sum_{j=1}^{k} \|D_j f\|_{L^\infty(\Sigma)}. \tag{4.19}$$

where

$$\|D_j f\|_{L^\infty(\Sigma)} = \sup_{x\in\Sigma} \|D_j f(x)\|,$$

i.e., the norm $\|D_j f(x)\|$ is the norm in the ambient space $\mathbb{R}^N$, and

(4.20)

$$\|f\|_{\text{A-t-Lip}} = \inf\left\{A + B : |f(x) - f(y)| \le A\|x-y\| + Bt^{1/2} \text{ for all } x, y \in \mathbb{R}^N\right\}$$

this t-almost Lipschitz norm being weaker than the Lipschitz norm. We will frequently use the following classes of functions

$$\mathcal{F}^t_{f,h} = \left\{\phi_t(\xi,\zeta) = t^{-1/2} e^{-\frac{\|\xi-\zeta\|^2}{2t}} (f(\xi) - f(\zeta))(h(\xi) - h(\zeta)) : \xi \in \Sigma\right\}, \tag{4.21}$$

$$\mathcal{G}^t = \left\{\psi_t(\xi,\zeta) = t^{1/2} e^{-\frac{\|\xi-\zeta\|^2}{2t}} : \xi \in \Sigma\right\}, \tag{4.22}$$

where f, h in (4.21) are fixed functions defined on all of $\mathbb{R}^N$. Given a class of functions $\mathscr{S}$ we use $\mathscr{M}_{\mathscr{S}}$ to denote

$$\mathscr{M}_{\mathscr{S}} = \sup_{f\in\mathscr{S}} \{\|f\|_{L^\infty(\Sigma)}\}.$$

With this notation we easily find that

$$\mathscr{M}_{\mathcal{F}^t_{f,h}} = \sup_{\phi\in\mathcal{F}^t} \{\|\phi\|_{L^\infty(\Sigma)}\} \le \left(\frac{2}{e}\right) t^{1/2} \|f\|_{\text{Lip}} \|h\|_{\text{Lip}}, \tag{4.23}$$

$$\mathscr{M}_{\mathcal{F}^t_{f,h}} \le t^{1/2} \|f\|_{\text{A-t-Lip}} \|h\|_{\text{Lip}}, \tag{4.24}$$

$$\mathscr{M}_{\mathcal{G}^t} = \sup_{\psi\in\mathcal{G}^t} \{\|\psi\|_{L^\infty(\Sigma)}\} = t^{1/2}. \tag{4.25}$$

4.3. **Ambient Covering Numbers.** In this subsection we show that the computation of covering numbers of the classes of functions $\mathcal{F}^t_{f,h}$ and $\mathcal{G}^t$ introduced in (4.21), (4.22) reduces to the computation of covering numbers of submanifolds of $\mathbb{R}^N$. For this purpose we will need the notion of *ambient covering number of* a totally bounded set X of $\mathbb{R}^N$ defined by

$$\mathcal{A}(X,\delta) = \inf\{m : \exists A \text{ with } |A| = m \text{ and } X \subset \bigcup_{a\in A} B_{\mathbb{R}^N,\delta}(a)\}. \tag{4.26}$$

where the ball $B_{\mathbb{R}^N,\delta}(a)$ is taken with respect to the ambient distance $\|\cdot\|$.

Lemma 4.6. *For any $\xi, \tilde\xi, \zeta \in \mathbb{R}^N$ we have*

$$\left| t^{1/2} e^{-\frac{\|\zeta-\xi\|^2}{2t}} - t^{1/2} e^{-\frac{\|\zeta-\tilde\xi\|^2}{2t}} \right| \le e^{-1/2} \|\xi - \tilde\xi\|. \tag{4.27}$$

In particular

$$\mathcal{N}(\mathcal{G}^t, \varepsilon) \leq \mathcal{A}(\Sigma, e^{1/2}\varepsilon). \tag{4.28}$$

Proof. Observe that the function $\psi_t(\xi, \zeta) = t^{1/2} e^{-\frac{\|\zeta-\xi\|^2}{2t}}$ satisfies

$$D_\xi \psi_t(\xi, \zeta) = t^{1/2} \frac{(\zeta - \xi)}{t} e^{-\frac{\|\xi-\zeta\|^2}{2t}}, \tag{4.29}$$

and therefore

$$\sup_{\xi,\zeta\in\mathbb{R}^N} \|D_\xi \psi_t(\xi, \zeta)\| \leq \sqrt{2} \sup_{\rho>0} \left\{ \rho e^{-\rho^2} \right\} = e^{-\frac{1}{2}}, \tag{4.30}$$

and therefore we have the Lipschitz estimate

$$|\psi_t(\xi, \zeta) - \psi_t(\tilde{\xi}, \zeta)| \leq e^{-1/2} \|\xi - \tilde{\xi}\|. \tag{4.31}$$

□

Corollary 4.7. *Fix a function $h \in L^\infty(\Sigma)$ with $h \not\equiv 0$ and such that $\|h\|_{L^\infty(\Sigma)} \leq C$. Consider the class of functions*

$$\mathcal{H}_h^t = \left\{ \psi_t(\zeta, \cdot) h(\cdot) : \zeta \in \Sigma \right\}. \tag{4.32}$$

Then for every $\varepsilon > 0$ we have

$$\mathcal{N}(\mathcal{H}_h^t, \varepsilon) \leq \mathcal{A}\left(\Sigma, \frac{e^{1/2}}{C} \varepsilon \right). \tag{4.33}$$

Proof. We use Lemma 4.6 to obtain the estimate

$$\left\| \psi_t(\zeta, \cdot) h(\cdot) - \psi_t(\zeta', \cdot) h(\cdot) \right\|_{L^\infty(\Sigma)} \leq C \|\psi_t(\zeta, \cdot) - \psi_t(\zeta', \cdot)\|_{L^\infty(\Sigma)} \tag{4.34}$$

$$\leq C e^{-1/2} \|\zeta - \zeta'\|, \tag{4.35}$$

from which the corollary follows. □

Lemma 4.8. *For any $\phi_t(\xi, \cdot), \phi_t(\xi', \cdot) \in \mathcal{F}_{f,h}^t$ we have*

$$\sup_{\zeta\in\mathbb{R}^d} |\phi_t(\xi, \zeta) - \phi_t(\xi', \zeta)| \leq C(f, h) \|\xi - \xi'\|,$$

where

$$C(f, h) = C_0 \|f\|_{\mathrm{Lip}} \|h\|_{\mathrm{Lip}}, \tag{4.36}$$

and C_0 is a universal constant. Thus

$$\mathcal{N}(\mathcal{F}_{f,h}^t, \delta) \leq \mathcal{A}\left(\Sigma, \frac{\delta}{C(f, h)} \right). \tag{4.37}$$

Proof. Let $\phi_t(\xi,\cdot) \in \mathcal{F}^t_{f,h}$. Fixing ζ and differentiating in ξ we have

$$D_\xi \phi_t(\xi,\zeta) = t^{-1/2}\frac{(\zeta-\xi)}{2t}e^{-\frac{\|\xi-\zeta\|^2}{2t}}(f(\xi)-f(\zeta))(h(\xi)-h(\zeta)) \tag{4.38}$$

$$+\, t^{-1/2}e^{-\frac{\|\xi-\zeta\|^2}{2t}}D_\xi f(\xi)(h(\xi)-f(\zeta)) \tag{4.39}$$

$$+\, t^{-1/2}e^{-\frac{\|\xi-\zeta\|^2}{2t}}D_\xi h(\xi)(f(\xi)-f(\zeta)), \tag{4.40}$$

and then

$$\|D_\xi \phi_t(\xi,\zeta)\| \le \|f\|_{\mathrm{Lip}}\|h\|_{\mathrm{Lip}}\frac{\|\xi-\zeta\|^3}{t^{3/2}}e^{-\frac{\|\xi-\zeta\|^2}{2t}} \tag{4.41}$$

$$+\, \frac{\|\xi-\zeta\|}{t^{1/2}}e^{-\frac{\|\xi-\zeta\|^2}{2t}}\|f\|_{\mathrm{Lip}}\|h\|_{\mathrm{Lip}} \tag{4.42}$$

$$+\, \frac{\|\xi-\zeta\|}{t^{1/2}}e^{-\frac{\|\xi-\zeta\|^2}{2t}}\|h\|_{\mathrm{Lip}}\|f\|_{\mathrm{Lip}} \tag{4.43}$$

$$\le C_0\|f\|_{\mathrm{Lip}}\|h\|_{\mathrm{Lip}}, \tag{4.44}$$

where

$$C_0 = 3\max\left(\sup_{\rho>0}\{\rho^3 e^{-\rho^2/2}\}, \sup_{\rho>0}\{\rho e^{-\rho/2}\}\right).$$

Note that all of the above estimates hold if we differentiate with respect to points in $\mathbb{R}^N$ as opposed to points restricted to Σ (this is because we are differentiating with respect to coordinates in the ambient space). It follows that for any $\zeta \in \Sigma$ and any $\xi, \xi' \in \Sigma$ we have the Lipschitz estimate

$$|\phi_t(\xi,\zeta) - \phi_t(\xi',\zeta)| \le C(f,h)\|\xi-\xi'\|.$$

□

The following can be obtained in a similar fashion.

Lemma 4.9. *Let $f \in C^1(\mathbb{R}^N)$ and let υ_t be given by*

$$\upsilon_t(\xi,\zeta) = e^{-\frac{\|\xi-\zeta\|^2}{2t}}(f(\xi)-f(\zeta)).$$

We then have the following estimate

$$\left|\upsilon_t(\xi,\zeta) - \upsilon_t(\xi',\zeta)\right| \le \left(\frac{2t}{e}+1\right)\|f\|_{\mathrm{Lip}}\|\xi-\xi'\|. \tag{4.45}$$

Proof. As before,

$$|D_\xi \upsilon_t(\xi,\zeta)| \le \frac{\|\xi-\zeta\|}{t}e^{-\frac{\|\xi-\zeta\|^2}{2t}}|f(\xi)-f(\zeta)| + e^{-\frac{\|\xi-\zeta\|^2}{2t}}\|D_\xi f(\xi)\| \tag{4.46}$$

$$\le \frac{\|\xi-\zeta\|^2}{t}e^{-\frac{\|\xi-\zeta\|^2}{2t}}\|f\|_{\mathrm{Lip}} + \sup_{\xi\in\Sigma}\|D_\xi f(\xi)\|, \tag{4.47}$$

and

$$\sup_{\rho>0}\rho^2 e^{-\rho^2/2} = \frac{2}{e}. \tag{4.48}$$

□

Lemmas 4.6 and 4.8 say that we can relate covering numbers of the classes $\mathcal{F}^t_{f,h}, \mathcal{G}^t$ to ambient covering numbers of the submanifold Σ. In order to estimate $\mathcal{A}(\Sigma, \delta)$ we need to introduce the notion of *reach* of an embedded submanifold of $\mathbb{R}^N$. Recall that for every $\varepsilon > 0$ we can consider the ε neighborhood of Σ

$$\Sigma_\varepsilon = \{x \in \mathbb{R}^N : d(x, \Sigma) < \varepsilon\}, \tag{4.49}$$

where $d(\cdot, \Sigma)$ measures the distance from points in $\mathbb{R}^N$ to Σ with respect to the ambient norm $\|\cdot\|$. If Σ is a smooth, embedded submanifold of $\mathbb{R}^N$, for $\varepsilon > 0$ sufficiently small we can define a smooth map $\varphi : \Sigma_\varepsilon \to \Sigma$ such that

(1) φ is smooth,
(2) $\varphi(x)$ is the unique point in Σ such that $\|\varphi(x) - x\| = d(x, \Sigma)$ for all $x \in \Sigma_\varepsilon$.
(3) $x - \varphi(x) \in T^\perp_{\varphi(x)}\Sigma$,
(4) $\varphi(y + z) \equiv \varphi(y)$ for all $y \in \Sigma$ and $z \in (T_y\Sigma)^\perp$ with $\|z\| < \varepsilon$,
(5) For any vector $V \in \mathbb{R}^N$, $D_V\varphi(x) = p_{\varphi(x)}(V)$, where $p_{\varphi(x)}(V)$ is the orthogonal projection of V onto $T_{\varphi(x)}\Sigma$.

See for example [24, Theorem 1]. The map φ is called *nearest point projection* onto Σ.

Definition 4.10. *Let Σ be an embedded submanifold of $\mathbb{R}^N$. The reach of Σ is the number*

$$\tau = \sup\{\varepsilon > 0 : \textit{There exists a nearest point projection in } \Sigma_\varepsilon\}. \tag{4.50}$$

We quote the following result

Theorem 4.11 ([10] Corollary 6)**.** *Suppose that Σ is a d-dimensional embedded submanifold of $\mathbb{R}^N$ with volume V and reach $\tau > 0$ and let $U : \mathbb{R}_+ \to \mathbb{R}_+$ be the function*

$$U_{V,\tau}(r) = \left(\frac{1}{\tau^d} + r^d\right) V, \tag{4.51}$$

then for any $\varepsilon > 0$ there is an ε-net of Σ with respect to the ambient distance $\|\cdot\|$ of no more than $C_d U(\varepsilon^{-1})$ points where C_d is a dimensional constant. In particular,

$$\mathcal{A}(\Sigma, \varepsilon) \le C_d U(\varepsilon^{-1}).$$

Corollary 4.12. *We have the following bounds*

(a)

$$\mathcal{N}(\mathcal{F}^t_{f,h}, \varepsilon) \le C_d V \left(\frac{1}{\tau^d} + \left(\frac{C(f,h)}{\varepsilon}\right)^d\right), \tag{4.52}$$

(b)

$$\mathcal{N}(\mathcal{G}^t, \varepsilon) \le C_d V \left(\frac{1}{\tau^d} + \left(\frac{1}{e^{1/2}\varepsilon}\right)^d\right), \tag{4.53}$$

(c)

$$\mathcal{N}(\mathcal{H}_h^t, \varepsilon) \leq C_d V \left(\frac{1}{\tau^d} + \left(\frac{C}{e^{1/2}\varepsilon} \right)^d \right), \tag{4.54}$$

where $C(f,h)$ *in* (4.52) *is given by* (4.36) *and* C *in* (4.54) *is an upper bound for* $\|h\|_{L^\infty(\Sigma)}$.

Remark 4.13. From (4.52) we can obtain effective bounds for $\mathcal{N}(\mathcal{F}_{f,h}^t, \varepsilon)$ in the sense that these bounds depend only on bounds for $\|f\|_{\mathrm{Lip}}, \|h\|_{\mathrm{Lip}}$.

We will adopt the following convention

Convention 4.14. From now on, we will use the following convention

- We will use $C^*(\Sigma)$ to denote the number
$$C^*(\Sigma) = \sup_{\xi \in \Sigma} \{\|\xi\|\}.$$
- We will use $C_1^*(\Sigma)$ to denote a constant that depends on the embedding coordinates of Σ and on $L^\infty(\Sigma)$ bounds on II_Σ (second fundamental form of Σ) and its derivatives. This constant may vary from line to line.
- From now on, we will use C_d, α_d and β_d to denote dimensional constants that may change from line to line.

We now derive covering bounds with respect to the ambient distance for the class

$$\mathscr{L}_M = \{f_\eta|_\Sigma : f_\eta(x) = \langle \eta, x \rangle, \|\eta\| \leq M, x \in \mathbb{R}^N\},$$

where $\|\cdot\|$ is the ambient distance in $\mathbb{R}^N$ and $f_\eta|_\Sigma$ is the restriction of the linear function $f_\eta : \mathbb{R}^N \to \mathbb{R}$ given by $f_\eta(x) = \langle \eta, x \rangle$ to Σ, which was also introduced in the statement of Theorem C. Note that for functions u, u' in $\mathscr{L}_M$ we can write $u(\xi) = \langle \eta, \xi \rangle$ and $u'(\xi) = \langle \eta', \xi \rangle$ where $\|\eta\|, \|\eta'\| \leq M$ and therefore

$$|u(\xi) - u'(\xi)| \leq C^*(\Sigma)\|\eta - \eta'\|.$$

We conclude that

$$\mathcal{N}(\mathscr{L}_M, \delta) \leq \mathcal{A}\left(B_M(0), \frac{\delta}{C^*(\Sigma)} \right).$$

On the other hand, it is well known that for any $\varepsilon > 0$

$$\mathcal{A}(B_M(0), \varepsilon) \leq \left(\frac{2M + \varepsilon}{\varepsilon} \right)^N,$$

see for example [18, Chapter 4]. It follows that we have bounds on covering numbers of the form

$$\mathcal{N}(\mathscr{L}_M, \delta) \leq \left(\frac{2MC^*(\Sigma) + \delta}{\delta} \right)^N. \tag{4.55}$$

We now prove uniform estimates on the covering number for the collection of classes $\mathcal{F}^t_{f,h}$ and $\mathcal{H}^t_h$ for f, h in the class $\mathscr{L}_M$ defined in the statement of Theorem C.

Lemma 4.15. *Consider the classes of functions*

(4.56)
$$\mathcal{F}^t = \left\{ \phi_t(f,h,\xi)(\zeta) = t^{-1/2} e^{-\frac{\|\xi-\zeta\|^2}{2t}} (f(\xi) - f(\zeta))(h(\xi) - h(\zeta)) : \xi \in \Sigma, f, h \in \mathscr{L}_M \right\},$$

(4.57)
$$\mathcal{H}^t = \left\{ \omega_t(h,\xi)(\zeta) = t^{1/2} e^{-\frac{\|\xi-\zeta\|^2}{2t}} h(\zeta) : \xi \in \Sigma, h \in \mathscr{L}_M \right\},$$

(4.58)
$$\mathcal{F}^t_* = \left\{ \phi^*_t(f,h,\xi)(\zeta) = t^{-1/2} e^{-\frac{\|\xi-\zeta\|^2}{2t}} (L_t f(\xi) - L_t f(\zeta))(h(\xi) - h(\zeta)) : \xi \in \Sigma, f, h \in \mathscr{L}_M \right\},$$

(4.59)
$$\mathcal{H}^t_* = \left\{ \omega^*_t(f,h,\xi)(\zeta) = t^{1/2} e^{-\frac{\|\xi-\zeta\|^2}{2t}} \Gamma_t(f,h)(\zeta) : f, h \in \mathscr{L}_M, \xi \in \Sigma \right\}.$$

Then

$$\mathcal{N}\left(\mathcal{F}^t, \varepsilon\right) \le C_d V \left(\frac{1}{\tau^d} + \left(\frac{3C_0 M^2}{\varepsilon} \right)^d \right) \left(\frac{12M^2 e^{-1/2} C^*(\Sigma) + \varepsilon}{\varepsilon} \right)^{2N}, \tag{4.60}$$

$$\mathcal{N}(\mathcal{H}^t, \varepsilon) \le C_d V \left(\frac{1}{\tau^d} + \left(\frac{2C^*(\Sigma) e^{-1/2} M}{\varepsilon} \right)^d \right) \left(\frac{4M t^{1/2} C^*(\Sigma) + \varepsilon}{\varepsilon} \right)^N, \tag{4.61}$$

where V is the volume of Σ, C_0 is the constant in Lemma 4.8 and $C^*(\Sigma)$ is as in Convention 4.14. In addition, assuming that $t > 0$ is sufficiently small, we have bounds of the form

$$\mathcal{N}(\mathcal{H}^t_*, \varepsilon) \le C_d U \left(\frac{\varepsilon}{3M^2 \beta_d} \right) \left(\frac{12\alpha_d C^*(\Sigma) M + \varepsilon}{\varepsilon} \right)^{2N}, \tag{4.62}$$

and

(4.63)

$$\mathcal{N}(\mathcal{F}^t_*, \varepsilon) \le$$
$$\le C_d \left(\frac{12M^2 e^{-1/2} C^*(\Sigma)}{t\varepsilon} + 1 \right)^N \left(\frac{12M^2 e^{-1/2} C^*_1(\Sigma) C^*(\Sigma)}{\varepsilon} + 1 \right)^N U \left(\frac{\varepsilon}{3M^2 C_0 C^*_1(\Sigma)} \right).$$

where $C^*_1(\Sigma)$ is as in convention 4.14. In all of the above inequalities, $U(r)$ is as in (4.51).

Proof. Without loss of generality let us assume that Σ has unit volume with respect to the induced metric. In order to prove (4.60), we estimate the difference

$$\left| \phi_t(f,h,\xi)(\zeta) - \phi_t(\tilde{f}, \tilde{h}, \xi')(\zeta) \right|,$$

for arbitrary $f, h, \tilde{f}, \tilde{h} \in \mathscr{L}_M$ and $\xi, \xi' \in \Sigma$. Note that

$$\begin{aligned}\left|\phi_t(f,h,\xi)(\zeta) - \phi_t(\tilde{f},\tilde{h},\xi')(\zeta)\right| \leq& \left|\phi_t(f,h,\xi)(\zeta) - \phi_t(\tilde{f},h,\xi)(\zeta)\right| \\ &+ \left|\phi_t(\tilde{f},h,\xi)(\zeta) - \phi_t(\tilde{f},\tilde{h},\xi)(\zeta)\right| \\ &+ \left|\phi_t(\tilde{f},\tilde{h},\xi)(\zeta) - \phi_t(\tilde{f},\tilde{h},\xi')(\zeta)\right|.\end{aligned}$$

Note that

$$\begin{aligned}\left|\phi_t(f,h,\xi)(\zeta) - \phi_t(\tilde{f},h,\xi)(\zeta)\right| &= t^{-1/2}e^{-\frac{\|\xi-\zeta\|^2}{2t}}\left|h(\xi)-h(\zeta)\right|\cdot\left|f(\xi)-\tilde{f}(\xi)\right| \\ &+ t^{-1/2}e^{-\frac{\|\xi-\zeta\|^2}{2t}}\left|h(\xi)-h(\zeta)\right|\cdot\left|f(\zeta)-\tilde{f}(\zeta)\right| \\ &\leq 2e^{-1/2}M\|f-\tilde{f}\|_{L^\infty(\Sigma)}.\end{aligned}$$

Similarly,

$$\left|\phi_t(\tilde{f},h,\xi)(\zeta) - \phi_t(\tilde{f},\tilde{h},\xi)(\zeta)\right| \leq 2e^{-1/2}M\|h-\tilde{h}\|_{L^\infty(\Sigma)}.$$

Finally, from Lemma 4.8 we obtain

$$\left|\phi_t(\tilde{f},\tilde{h},\xi)(\zeta) - \phi_t(\tilde{f},\tilde{h},\xi')(\zeta)\right| \leq C(\tilde{f},\tilde{h})\|\xi-\xi'\| \leq C_0M^2\|\xi-\xi'\|.$$

If we now turn to the computation of $L^\infty(\Sigma)$ covering numbers, we obtain the bound

$$\mathcal{N}(\mathcal{F}^t,\varepsilon) \leq \mathcal{N}\left(\mathscr{L}_M, \frac{\varepsilon}{6Me^{-1/2}}\right)^2 \cdot \sup_{f,h\in\mathscr{L}_M}\mathcal{N}\left(\mathcal{F}^t_{f,h}, \frac{\varepsilon}{3}\right), \tag{4.64}$$

and observe from (4.37),(4.52) and (4.55) that for any $\delta > 0$ we have the bounds

$$\mathcal{N}\left(\mathscr{L}_M,\delta\right) \leq \left(\frac{2MC^*(\Sigma)+\delta}{\delta}\right)^N, \tag{4.65}$$

$$\sup_{f,h\in\mathscr{L}_M}\mathcal{N}\left(\mathcal{F}^t_{f,h},\delta\right) \leq \mathcal{A}\left(\Sigma, \frac{\delta}{C_0M^2}\right). \tag{4.66}$$

In view of Theorem 4.11 and (4.64)

$$\mathcal{N}\left(\mathcal{F}^t,\varepsilon\right) \leq \left(\frac{12M^2e^{-1/2}C^*(\Sigma)+\varepsilon}{\varepsilon}\right)^{2N}\cdot C_d\left(\frac{1}{\tau^d} + \left(\frac{3C_0M^2}{\varepsilon}\right)^d\right),$$

where C_d is a dimensional constant and τ is the reach of Σ. In order to prove (4.61), we use the estimate

$$\begin{aligned}\left|\omega_t(\zeta,x)h(x) - \omega_t(\zeta',x)\tilde{h}(x)\right| &\leq \left|\left(\omega_t(\zeta,x)-\omega_t(\zeta',x)\right)h(x)\right| + \omega_t(\zeta',x)\left|h(x)-\tilde{h}(x)\right| \\ &\leq C^*(\Sigma)Me^{-1/2}\|\zeta-\zeta'\| + t^{1/2}\|h-\tilde{h}\|_{L^\infty(\Sigma)}.\end{aligned} \tag{4.67}$$

In order to obtain (4.67) we have used inequality (4.31), $\|\psi_t\|_{L^\infty(\Sigma\times\Sigma)} = t^{1/2}$ and $\sup_{h\in\mathscr{L}_M} \|h\|_{L^\infty(\Sigma)} \leq MC^*(\Sigma)$. It follows that

$$\begin{aligned}\mathcal{N}\left(\mathcal{H}^t,\varepsilon\right) &\leq \mathcal{A}\left(\Sigma,\frac{\varepsilon}{2C^*(\Sigma)e^{-1/2}M}\right)\cdot\mathcal{N}\left(\mathscr{L}_M,\frac{\varepsilon}{2t^{1/2}}\right)\\ &\leq \mathcal{A}\left(\Sigma,\frac{\varepsilon}{2C^*(\Sigma)e^{-1/2}M}\right)\cdot\mathcal{A}\left(B_M(0),\frac{\varepsilon}{2t^{1/2}C^*(\Sigma)}\right)\\ &\leq C_d\left(\frac{1}{\tau^d}+\left(\frac{2C^*(\Sigma)e^{-1/2}M}{\varepsilon}\right)^d\right)\left(\frac{4Mt^{1/2}C^*(\Sigma)+\varepsilon}{\varepsilon}\right)^N,\end{aligned}$$

as claimed. For the proof of (4.62) we observe that

$$\begin{aligned}\left|\omega_t^*(f,h,\xi)(\zeta)-\omega_t^*(\tilde{f},\tilde{h},\xi')(\zeta)\right| \leq &\left|\omega_t^*(f,h,\xi)(\zeta)-\omega_t^*(\tilde{f},h,\xi)(\zeta)\right|\\ &+\left|\omega_t^*(\tilde{f},h,\xi,\zeta)-\omega_t^*(\tilde{f},\tilde{h},\xi,\zeta)\right|\\ &+\left|\omega_t^*(\tilde{f},\tilde{h},\xi)(\zeta)-\omega_t^*(\tilde{f},\tilde{h},\xi')(\zeta)\right|.\end{aligned}$$

Observe that

(4.68)

$$\left|\omega_t^*(f,h,\xi)(\zeta)-\omega_t^*(\tilde{f},h,\xi)(\zeta)\right| \leq \frac{2\|f-\tilde{f}\|_{L^\infty(\Sigma)}e^{-\frac{\|\xi-\zeta\|^2}{2t}}}{t^{1/2}\theta_t(\zeta)}\int_\Sigma e^{-\frac{\|\eta-\zeta\|^2}{2t}}\left|h(\eta)-h(\zeta)\right|d\mu(\eta).$$

Note also that since $t^{-d/2}\theta_t(x)$ converges uniformly to $(2\pi)^{d/2}$ on Σ as $t\to 0^+$, there exists a number $t_0>0$ depending only on the embedding coordinates of Σ such that for $0<t<t_0$ one has the estimate

$$\begin{aligned}\int_\Sigma e^{-\frac{\|\eta-\zeta\|^2}{2t}}\left|h(\eta)-h(\zeta)\right|d\mu(\eta) &= \int_\Sigma e^{-\frac{\|\eta-\zeta\|^2}{4t}}e^{-\frac{\|\eta-\zeta\|^2}{4t}}\left|h(\eta)-h(\zeta)\right|d\mu(\eta)\\ &\leq \|h\|_{\mathrm{Lip}}\int_\Sigma e^{-\frac{\|\eta-\zeta\|^2}{4t}}e^{-\frac{\|\eta-\zeta\|^2}{4t}}\|\eta-\zeta\|\,d\mu(\eta)\\ &\leq \sqrt{2t}e^{-1/2}\|h\|_{\mathrm{Lip}}\int_\Sigma e^{-\frac{\|\eta-\zeta\|^2}{4t}}d\mu(\eta)\\ &\leq M\sqrt{2t}2^{\frac{d+2}{2}}(2\pi)^{d/2}t^{d/2}.\end{aligned}$$

For the above inequality we have used that

$$\lim_{t\to 0^+} t^{-d/2}\int_\Sigma e^{-\frac{\|\eta-x\|^2}{4t}}d\mu(\eta) = 2^{d/2}(2\pi)^{d/2}$$

uniformly on Σ. It follows that

$$\begin{aligned}\left|\omega_t^*(f,h,\xi)(\zeta)-\omega_t^*(\tilde{f},h,\xi)(\zeta)\right| &\leq \alpha_d\|f-\tilde{f}\|_{L^\infty(\Sigma)}M\\ &= \alpha_d M\|f-\tilde{f}\|_{L^\infty(\Sigma)},\end{aligned} \tag{4.69}$$

for some dimensional constant α_d. Analogously, for $0 < t < t_0$ we have

$$\left|\omega_t^*(\tilde{f}, h, \xi, \zeta) - \omega_t^*(\tilde{f}, \tilde{h}, \xi, \zeta)\right| \leq \alpha_d M \|h - \tilde{h}\|_{L^\infty(\Sigma)}. \tag{4.70}$$

We also have

$$\left|\omega_t^*(\tilde{f}, \tilde{h}, \xi)(\zeta) - \omega_t^*(\tilde{f}, \tilde{h}, \xi')(\zeta)\right| \tag{4.71}$$

$$\leq \frac{e^{-1/2}\|\xi - \xi'\|}{t\theta_t(\xi)} \int_\Sigma e^{-\frac{\|\zeta-\eta\|^2}{2t}} \left|\tilde{f}(\eta) - \tilde{f}(\zeta)\right| \left|\tilde{h}(\eta) - \tilde{h}(\zeta)\right| d\mu(\eta) \tag{4.72}$$

$$\leq \beta_d e^{-1/2} M^2 \|\xi - \xi'\|. \tag{4.73}$$

Combining (4.69), (4.70) and (4.73) with an argument similar to the one used to prove (4.60) we can easily prove (4.62). For the proof of (4.63), we start with the inequality

$$\begin{aligned}\left|\phi_t^*(f, h, \xi)(\zeta) - \phi_t^*(\tilde{f}, \tilde{h}, \xi')(\zeta)\right| \leq & \left|\phi_t^*(f, h, \xi)(\zeta) - \phi_t^*(\tilde{f}, h, \xi)(\zeta)\right| \\ & + \left|\phi_t^*(\tilde{f}, h, \xi, \zeta) - \phi_t^*(\tilde{f}, \tilde{h}, \xi, \zeta)\right| \\ & + \left|\phi_t^*(\tilde{f}, \tilde{h}, \xi)(\zeta) - \phi_t^*(\tilde{f}, \tilde{h}, \xi')(\zeta)\right|,\end{aligned}$$

where $f, h, \tilde{f}, \tilde{h}$ are in $\mathscr{L}_M$ and ξ, ξ' are in Σ. A simple estimate shows that

$$\left|\phi_t^*(f, h, \xi)(\zeta) - \phi_t^*(\tilde{f}, h, \xi)(\zeta)\right| \leq \frac{2Me^{-1/2}\|f - \tilde{f}\|_{L^\infty(\Sigma)}}{t}. \tag{4.74}$$

Similarly

$$\left|\phi_t^*(\tilde{f}, h, \xi)(\zeta) - \phi_t^*(\tilde{f}, \tilde{h}, \xi)(\zeta)\right| \leq 2e^{-1/2}\|h - \tilde{h}\|_{L^\infty(\Sigma)} \Lambda(L_t \tilde{f}), \tag{4.75}$$

where[1]

$$\Lambda(L_t\tilde{f}) = \sup_{x, x' \in \Sigma} \left\{ \frac{\left|L_t\tilde{f}(x) - L_t\tilde{f}(x')\right|}{\|x - x'\|} \right\}.$$

From Proposition 3.2 if follows that there exists $T_0 > 0$ depending only on the embedding coordinates of Σ such that for $0 < t < T_0$ we have the expansion

$$L_t\tilde{f}(x) = \Delta_\Sigma \tilde{f}(x) + t^{1/2} R(x),$$

where $\Delta_\Sigma \tilde{f}$ is the Laplacian of $\tilde{f}$ with respect to the induced metric and where $R(x)$ is computed in terms of local integrals of the 3-jet of $\tilde{f}(x)$ as pointed out in (3.4). Observe now that for $\tilde{f}$ in $\mathscr{L}_M$ we can write $\tilde{f}(x) = \langle \eta, x \rangle$ for some η with $\|\eta\| \leq M$, and therefore

$$\nabla^2_\Sigma \tilde{f}(x) = -\eta^\perp \mathrm{II}_\Sigma(x),$$

[1] Note that the quotient $\Lambda(L_t\tilde{f})$ is not exactly the Lipschitz semi-norm that we defined in (4.18) because the supremum used in the definition of $\Lambda(L_t\tilde{f})$ is taken over points in Σ as opposed to (4.18) which involves points in all of $\mathbb{R}^N$.

where $\eta^{\perp}$ is the normal component of η with respect to $T_x\Sigma$ and $\mathrm{II}_{\Sigma}(x)$ is the second fundamental form of Σ at x. In particular, (3.4) shows that

$$L_t\tilde{f}(x) = -\eta^{\perp}H_{\Sigma}(x) + t^{1/2}R(x),$$

where $H_{\Sigma}(x)$ is the mean curvature of Σ at x and $R(x)$ is computed in terms of II_{Σ} and its derivatives. We conclude that we have a bound of the form

$$\left|\Lambda(L_t\tilde{f})\right| \leq M \cdot C_1^*(\Sigma), \tag{4.76}$$

and inserting (4.76) into (4.75) we obtain

$$\left|\phi_t^*(\tilde{f}, h, \xi)(\zeta) - \phi_t^*(\tilde{f}, \tilde{h}, \xi)(\zeta)\right| \leq 2e^{-1/2}MC_1^*(\Sigma)\|h - \tilde{h}\|_{L^\infty(\Sigma)}. \tag{4.77}$$

In a similar fashion, for $0 < t < T_0$ we also have

$$\left|\phi_t^*(\tilde{f}, \tilde{h}, \xi)(\zeta) - \phi_t^*(\tilde{f}, \tilde{h}, \xi^{'})(\zeta)\right| \leq C_0M^2C_1^*(\Sigma)\|\xi - \xi^{'}\|. \tag{4.78}$$

Combining (4.74),(4.77) and (4.78) we conclude that for t with $0 < t < T_0$ we have the bound

$$\begin{aligned}\left|\phi_t^*(f, h, \xi)(\zeta) - \phi_t^*(\tilde{f}, \tilde{h}, \xi^{'})(\zeta)\right| \leq{}& \frac{2Me^{-1/2}}{t}\|f - \tilde{f}\|_{L^\infty(\Sigma)} \\ &+ 2Me^{-1/2}C_1^*(\Sigma)\|h - \tilde{h}\|_{L^\infty(\Sigma)} \\ &+ M^2C_0C_1^*(\Sigma)\|\xi - \xi^{'}\|,\end{aligned}$$

for arbitrary $f, h, \tilde{f}, \tilde{h}$ in $\mathscr{L}_M$ and $\xi, \xi^{'}$ in Σ. We obtain in turn that for some dimensional constant C_d and for $t > 0$ sufficiently small we have the inequality

$$\begin{aligned}\mathcal{N}(\mathcal{F}_*^t, \varepsilon) &\leq \mathcal{N}\left(\mathscr{L}_M, \frac{t\varepsilon}{6Me^{-1/2}}\right) \cdot \mathcal{N}\left(\mathscr{L}_M, \frac{\varepsilon}{6Me^{-1/2}C_1^*(\Sigma)}\right) \cdot \mathcal{A}\left(\Sigma, \frac{\varepsilon}{3M^2C_0C_1^*(\Sigma)}\right) \\ &\leq C_d\left(\frac{12M^2e^{-1/2}C^*(\Sigma)}{t\varepsilon} + 1\right)^N \left(\frac{12M^2e^{-1/2}C_1^*(\Sigma)C^*(\Sigma)}{\varepsilon} + 1\right)^N U\left(\frac{\varepsilon}{3M^2C_0C_1^*(\Sigma)}\right).\end{aligned}$$

This proves inequality (4.63). □

The following lemma is easily proved using the tools developed in Lemma 4.15.

Lemma 4.16. *The classes* (4.56),(4.57),(4.58) *and* (4.59) *satisfy the following bounds*

$$\mathscr{M}_{\mathcal{F}^t} = \sup_{u \in \mathcal{F}^t}\|u\|_{L^\infty(\Sigma)} \leq 2t^{1/2}e^{-1}M^2, \tag{4.79}$$

$$\mathscr{M}_{\mathcal{H}^t} = \sup_{v \in \mathcal{H}^t}\|v\|_{L^\infty(\Sigma)} \leq t^{1/2}MC^*(\Sigma), \tag{4.80}$$

In addition, for $t > 0$ sufficiently small we also have

$$\mathscr{M}_{\mathcal{H}_*^t} = \sup_{v \in \mathcal{H}_*^t}\|v\|_{L^\infty(\Sigma)} \leq \beta_d M^2t^{1/2}, \tag{4.81}$$

and

$$\mathscr{M}_{\mathcal{F}_*^t} = \sup_{v \in \mathcal{F}_*^t}\|v\|_{L^\infty(\Sigma)} \leq C_1^*(\Sigma)M^2t^{1/2}. \tag{4.82}$$

For the above inequalities, $C^*(\Sigma), C_1^*(\Sigma)$ are as in Convention 4.14 and β_d is a dimensional constant as in Lemma 4.15.

4.4. **Sample Version of the Carré du Champ.** In this section we are still assuming that the distribution of the sample $\xi_1, \dots, \xi_n$ in Σ is *uniform*, i.e., $d\mu = d\mathrm{vol}_\Sigma$ and Σ has unit volume with respect to the induced metric. In this case we know that $\lim_{t\to 0^+} t^{-d/2}\theta_t = (2\pi)^{d/2}$ uniformly in $L^\infty(\Sigma)$ and moreover there exists $t_0 > 0$ depending only on the embedding coordinates of Σ such that for $0 < t < t_0$ we have

$$2(2\pi)^{d/2} \geq t^{-d/2}\theta_t \geq \frac{1}{2}(2\pi)^{d/2}.$$

If we let

$$\lambda_0 = \frac{(2\pi)^{d/2}}{4}, \tag{4.83}$$

we have for $0 < t < t_0$ the inequality

$$\theta_t(x) \geq 2t^{d/2}\lambda_0, \tag{4.84}$$

which will be a convenient normalization for us in the sequel. The main lemma in this section is the following.

Lemma 4.17. *Let $\mathcal{J}$ be given by a class of functions of the form*

$$\varphi(f, x)(\cdot) = f(x, \cdot)$$

for $x \in \Sigma$ and $f \in \mathcal{F}$ where $\mathcal{F}$ is a totally bounded class of functions in $L^\infty(\Sigma \times \Sigma)$. Let

$$C = \sup_{f\in\mathcal{F}} \|f\|_{L^\infty(\Sigma\times\Sigma)}.$$

Suppose also $0 < t < t_0$ and ε is small enough so that

$$\varepsilon t^{d+1/2}\lambda_0 < C. \tag{4.85}$$

Then $\mathcal{J}$ is totally bounded in $L^\infty(\Sigma)$ and

$$\mathrm{Pr}^*\left\{\sup_{f\in\mathcal{F}}\sup_{x\in\Sigma}\left| t^{-1/2}\frac{\mu_n f(x,\cdot)}{\hat{\theta}_t(x)} - t^{-1/2}\frac{\mu f(x,\cdot)}{\theta_t(x)}\right| \geq \varepsilon\right\} \tag{4.86}$$

$$\leq 2\mathcal{N}\left(\mathcal{G}^t, \frac{\varepsilon t^{d+1}\lambda_0^2}{4C}\right)\exp\left(-\frac{\varepsilon^2 t^{2d+1}\lambda_0^4}{8C^2}n\right) \tag{4.87}$$

$$2\mathcal{N}\left(\mathcal{F}, \frac{\varepsilon\lambda_0 t^{(d+1)/2}}{4}\right)\exp\left(-\frac{\varepsilon^2\lambda_0^2 t^{d+1} n}{8C^2}\right). \tag{4.88}$$

Before proving Lemma 4.17 we will prove the following elementary lemma.

Lemma 4.18. *Let ξ, ζ be positive random variables. For any $\varepsilon > 0$ we have*

$$\Pr\left\{\left|\frac{1}{\xi} - \frac{1}{\zeta}\right| \geq \varepsilon\right\} \leq \Pr\left\{|\zeta - \xi| \geq \frac{\varepsilon\xi^2}{1 + \varepsilon\xi}\right\}.$$

Proof. Assume $0 < \zeta < \xi$

$$\begin{aligned}\frac{1}{\zeta} - \frac{1}{\xi} = \frac{\xi - \zeta}{\xi\zeta} &= \frac{\xi - \zeta}{\xi\zeta - \xi^2 + \xi^2} \\ &= \frac{\zeta - \xi}{\xi\left(\zeta - \xi\right) + \xi^2} \\ &= \frac{|\zeta - \xi|}{-\xi|\zeta - \xi| + \xi^2},\end{aligned}$$

and from

$$\frac{|\zeta - \xi|}{-\xi\left|\zeta - \xi\right| + \xi^2} \leq \varepsilon$$

we have

$$|\zeta - \xi| \leq \frac{\varepsilon\xi^2}{1 + \varepsilon\xi}.$$

For the case $0 < \xi < \zeta$ we have

$$\frac{1}{\xi} - \frac{1}{\zeta} = \frac{\zeta - \xi}{\xi(\zeta - \xi) + \xi^2} = \frac{|\xi - \zeta|}{\xi|\xi - \zeta| + \xi^2}.$$

and $\frac{|\xi - \zeta|}{\xi|\xi - \zeta| + \xi^2} \geq \varepsilon$ implies

$$(1 + \varepsilon\xi)|\xi - \zeta| \geq (1 - \varepsilon\xi)|\xi - \zeta| \geq \varepsilon\xi^2.$$

□

Proof of Lemma 4.17. It is easy to prove from Lemma 4.18 and (4.84) that for any $\delta > 0$ we have

$$\Pr\left\{\sup_{x\in\Sigma}\left|\frac{1}{\hat{\theta}_t(x)} - \frac{1}{\theta_t(x)}\right| \geq \delta\right\} \tag{4.89}$$

$$\leq \Pr\left\{\sup_{x\in\Sigma}\left|\theta_t(x) - \hat{\theta}_t(x)\right| \geq \frac{4\delta t^d\lambda_0^2}{1 + 2\delta t^{d/2}\lambda_0}\right\}. \tag{4.90}$$

Let us write

$$\begin{aligned}t^{-1/2}\left(\frac{\mu_n f(x,\cdot)}{\hat{\theta}_t(x)} - \frac{\mu f(x,\cdot)}{\theta_t(x)}\right) &= t^{-1/2}\mu_n f(x,\cdot)\left(\frac{1}{\hat{\theta}_t(x)} - \frac{1}{\theta_t(x)}\right) \\ &\quad + t^{-1/2}\frac{1}{\theta_t(x)}\left[\mu_n f(x,\cdot) - \mu f(x,\cdot)\right].\end{aligned}$$

Thus

$$\begin{aligned}
&\Pr\left\{t^{-1/2}\sup_{f\in\mathcal{F}}\sup_{x\in\Sigma}\left|\frac{\mu_n f(x,\cdot)}{\hat{\theta}_t(x)}-\frac{\mu f(x,\cdot)}{\theta_t(x)}\right|\geq\varepsilon\right\}\\
&\leq\Pr\left\{\sup_{x\in\Sigma}\left|\frac{1}{\hat{\theta}_t(x)}-\frac{1}{\theta_t(x)}\right|\geq\frac{\varepsilon}{2}\frac{t^{1/2}}{C}\right\}\\
&+\Pr\left\{\sup_{f\in\mathcal{F}}\sup_{x\in\Sigma}|\mu_n f(x,\cdot)-\mu f(x,\cdot)|\geq\varepsilon\lambda_0 t^{(d+1)/2}\right\}.
\end{aligned}$$

Analyzing the first term using (4.85) and (4.89)-(4.90) leads us to the inequality

$$\begin{aligned}
&\Pr\left\{\sup_{x\in\Sigma}\left|\frac{1}{\hat{\theta}_t(x)}-\frac{1}{\theta_t(x)}\right|\geq\frac{\varepsilon}{2}\frac{t^{1/2}}{C}\right\}\\
&\leq\Pr\left\{\sup_{x\in\Sigma}\left|\theta_t(x)-\hat{\theta}_t(x)\right|\geq\frac{2\varepsilon t^{d+1/2}\lambda_0^2}{C+\varepsilon t^{d+1/2}\lambda_0}\right\}\\
&\leq\Pr\left\{\sup_{x\in\Sigma}\left|\theta_t(x)-\hat{\theta}_t(x)\right|\geq\frac{\varepsilon t^{d+1/2}\lambda_0^2}{C}\right\}\\
&\leq 2\mathcal{N}\left(\mathcal{G}^t,\frac{\varepsilon t^{d+1}\lambda_0^2}{4C}\right)\exp\left(-\frac{\varepsilon^2 t^{2d+1}\lambda_0^4}{8C^2}n\right),
\end{aligned}$$

where we have applied Lemma 4.5 to the class $\mathcal{G}^t$ (in particular we have used (4.25)). Finally,

$$\begin{aligned}
&\Pr\left\{\sup_{f\in\mathcal{F}}\sup_{x\in\Sigma}|\mu_n f(x,\cdot)-\mu f(x,\cdot)|\geq\varepsilon\lambda_0 t^{(d+1)/2}\right\}\\
&\leq 2\mathcal{N}\left(\mathcal{F},\frac{\varepsilon\lambda_0 t^{(d+1)/2}}{4}\right)\exp\left(-\frac{\varepsilon^2\lambda_0^2 t^{(d+1)}n}{8C^2}\right).
\end{aligned}$$

□

In view of Lemma 4.17, we introduce the following notation

Definition 4.19. *Given a class of functions $\mathcal{F}$ as in the statement of Lemma 4.17 and positive numbers t,ε,M we define for compactness of notation the following function*

$$Q_t(\mathcal{F},\varepsilon,C,n)=2\mathcal{N}\left(\mathcal{G}^t,\frac{\varepsilon t^{d+1}\lambda_0^2}{4t^{1/2}C}\right)\exp\left(-\frac{\varepsilon^2 t^{2d+1}\lambda_0^4}{8\left(t^{1/2}C\right)^2}n\right) \tag{4.91}$$

$$+\,2\mathcal{N}\left(\mathcal{F},\frac{\varepsilon\lambda_0 t^{(d+1)/2}}{4}\right)\exp\left(-\frac{\varepsilon^2\lambda_0^2 t^{d+1}n}{8\left(t^{1/2}C\right)^2}\right). \tag{4.92}$$

Remark 4.20. Because most of the function classes that we will encounter in our analysis satisfy bounds of the form (4.23), we include the $t^{1/2}$ factor in the expression for C.

Corollary 4.21. *If $\mathcal{F}$ is any of the classes of functions defined by* (4.21), (4.22), (4.32), (4.56),(4.57), (4.58), (4.58) *we have*

$$\Pr{}^{*}\left\{\sup_{x\in\Sigma}\left|t^{-1/2}\frac{\mu_n f(x,\cdot)}{\hat{\theta}_t(x)}-t^{-1/2}\frac{\mu f(x,\cdot)}{\theta_t(x)}\right|\geq\varepsilon\right\}\leq Q_t(\mathcal{F},\varepsilon,C,n). \tag{4.93}$$

For (4.21)*, the constant C depends on Σ, $\|f\|_{\mathrm{Lip}}$ and $\|h\|_{\mathrm{Lip}}$. For* (4.22)*, C depends on Σ. For* (4.32)*, C depends on $\|h\|_{\infty}$ and Σ. For* (4.56), (4.57), (4.58) *and* (4.59)*, C depends on Σ, M and dimensional constants.*

As a corollary we obtain the rate of convergence in probability of the sample Carré du Champ to its expected value.

Corollary 4.22. *We have the following bounds*

(a) *Fixing f and h and letting*

$$K=K\left(f,h\right)=\min\left\{\left(\frac{2}{e}\right)\|f\|_{\mathrm{Lip}}\|h\|_{\mathrm{Lip}},\|f\|_{A\text{-}t\text{-}\mathrm{Lip}}\|h\|_{\mathrm{Lip}}\right\},$$

we have

$$\Pr\left\{\sup_{x\in\Sigma}\left|\hat{\Gamma}_t(f,h)(x)-\Gamma_t(f,h)(x)\right|\geq\varepsilon\right\}\leq Q_t(\mathcal{F}^t_{f,h},\varepsilon,K,n).$$

(b) *For the class $\mathscr{L}_M$ we have*

$$\Pr\left\{\sup_{f,h\in\mathscr{L}_M}\sup_{x\in\Sigma}\left|\hat{\Gamma}_t(f,h)(x)-\Gamma_t(f,h)(x)\right|\geq\varepsilon\right\}\leq Q_t(\mathcal{F}^t,\varepsilon,K,n),$$

where $K=\frac{2M^2}{e}$.

(c) *In addition, for $t>0$ sufficiently small we have*

$$\Pr\left\{\sup_{f,h\in\mathscr{L}_M}\sup_{x\in\Sigma}\left|\hat{\Gamma}_t(L_tf,h)(x)-\Gamma_t(L_tf,h)(x)\right|\geq\varepsilon\right\}\leq Q_t(\mathcal{F}^t_{*},\varepsilon,K,n),$$

where $K=C_1^*(\Sigma)M^2$, and $C_1^*(\Sigma)$ is as in Convention 4.14.

Proof. Let us only prove (a) since the proofs of (b) and (c) are very similar. Recall that

$$\hat{\Gamma}_t(L_t,f,h)(x)=\frac{1}{t}\frac{1}{\hat{\theta}_t}\sum_{j=1}^{n}e^{-\frac{\|x-\xi_j\|^2}{2t}}(f(\xi_j)-f(x))(h(\xi_j)-h(x)) \tag{4.94}$$

$$=\frac{t^{-1/2}}{\hat{\theta}_t(x)}\sum_{j=1}^{n}\phi_t(x,\xi_j)=t^{-1/2}\frac{\mu_n(\phi_t(x,\cdot))}{\hat{\theta}_t(x)}, \tag{4.95}$$

where ϕ_t is given by the definition of the class $\mathcal{F}^t_{f,h}$ in (4.21) and therefore we can apply Lemma 4.17 to the class $\mathcal{F}^t_{f,h}$ and use the bound

$$\sup_{\phi\in\mathcal{F}^t_{f,h}}\|\phi\|_{L^\infty(\Sigma)}\leq t^{1/2}K.$$

which follows from (4.23). $\square$

If we now consider the t-Laplacian L_t and its sample version $\hat{L}_t$, we see that the deviation of $\hat{L}_t$ from L_t on a function $h \in L^\infty(\Sigma)$ with $\|h\|_{L^\infty(\Sigma)} \leq M$ simplifies to

$$\hat{L}_t h(x) - L_t h(x) = \frac{2}{t\hat{\theta}_t(x)n} \sum_{j=1}^{n} e^{-\frac{\|\xi_j - x\|^2}{2t}} (h(\xi_j) - h(x)) \tag{4.96}$$

$$- \frac{2}{t\theta_t(x)} \int_\Sigma e^{-\frac{\|\xi - x\|^2}{2t}} (h(\xi) - h(x)) d\mu(\xi) \tag{4.97}$$

$$= \frac{2}{\hat{\theta}_t(x)tn} \sum_{j=1}^{n} e^{-\frac{\|\xi_j - x\|^2}{2t}} h(\xi_j) - \frac{2}{t\theta_t(x)} \int_\Sigma e^{-\frac{\|\xi - x\|^2}{2t}} h(\xi) d\mu(\xi) \tag{4.98}$$

$$= \frac{2\mu_n \psi_t(x,\cdot)h(\cdot)}{t^{3/2}\hat{\theta}_t(x)} - \frac{2\mu\psi_t(x,\cdot)h(\cdot)}{t^{3/2}\theta_t(x)}. \tag{4.99}$$

Observe that

$$\Pr\left\{\sup_{x\in\Sigma} \left|\hat{L}_t h(x) - L_t h(x)\right| \geq \varepsilon\right\} \leq \Pr\left\{t^{-1/2} \sup_{\eta \in \mathcal{H}_h^t} \left|\frac{\mu_n \eta}{\hat{\theta}_t} - \frac{\mu\eta}{\theta_t}\right| \geq t\varepsilon\right\}. \tag{4.100}$$

We have obtained

Corollary 4.23. *We have the following bounds*

(a) *Fix a function $h \in L^\infty(\Sigma)$. If we set $C = \|h\|_{L^\infty}$ we have*

$$\Pr\left\{\sup_{x\in\Sigma} \left|\hat{L}_t h(x) - L_t h(x)\right| \geq \varepsilon\right\} \leq Q_t(\mathcal{H}_h^t, \varepsilon t, C, n).$$

(b) *For the class $\mathscr{L}_M$ we have the estimate*

$$\Pr\left\{\sup_{h\in\mathscr{L}_M} \sup_{x\in\Sigma} \left|\hat{L}_t h(x) - L_t h(x)\right| \geq \varepsilon\right\} \leq Q_t(\mathcal{H}^t, \varepsilon t, C, n),$$

where $C = MC^*(\Sigma)$.

(c) *In addition, we have for every $t > 0$ the estimate*

$$\Pr\left\{\sup_{f,h\in\mathscr{L}_M} \sup_{x\in\Sigma} \left|\hat{L}_t\left(\Gamma_t(f,h)\right)(x) - L_t\left(\Gamma_t(f,h)\right)(x)\right| \geq \varepsilon\right\} \leq Q_t(\mathcal{H}_*^t, \varepsilon t, C, n),$$

where $C = \beta_d M^2$, where $\beta_d > 0$ is a dimensional constant.

Proof. The proof of (a) follows from combining Lemma 4.17 with (4.100) and the fact that

$$\sup_{\eta\in\mathcal{H}_h^t} \|\eta\|_{L^\infty(\Sigma)} \leq t^{1/2}\|h\|_{L^\infty(\Sigma)} = t^{1/2}C.$$

The proofs of (b) and (c) are similar to the proof of (a) but make use of Lemmas (4.15) and (4.16).

□

4.5. **Subexponential Decay and Almost Sure Convergence.** The goal of this subsection is to demonstrate that the decay rate for the quantities $Q_t(\mathcal{F}, \varepsilon, M, n)$ implies the almost sure convergence. We illustrate the Borel-Cantelli type proof in this section for the purpose of introducing some notation. This notation will be used in later sections.

Theorem 4.24. *Consider the metric measure space* $(\Sigma, \|\cdot\|, d\mathrm{vol}_\Sigma)$ *where* $\Sigma^d \subset \mathbb{R}^N$ *is a smooth closed embedded submanifold. Suppose that we have a uniformly distributed i.i.d. sample* $\{\xi_1, \ldots, \xi_n\}$ *of points from* Σ*. For* $\sigma > 0$*, let*

$$t_n = n^{-\frac{1}{2d+\sigma}}. \tag{4.101}$$

Then

(a) *For fixed* $f, h \in \mathrm{Lip}(\Sigma)$ *we have*
$$\sup_{\xi \in \Sigma} \left| \hat{\Gamma}(L_{t_n}, f, h)(\xi) - \Gamma(L_{t_n}, f, h)(\xi) \right| \xrightarrow{a.s.} 0$$
as $n \to \infty$.

(b) *For the class* $\mathscr{L}_M$ *we have*
$$\sup_{f,h \in \mathscr{L}_M} \sup_{\xi \in \Sigma} \left| \hat{\Gamma}(L_{t_n}, f, h)(\xi) - \Gamma(L_{t_n}, f, h)(\xi) \right| \xrightarrow{a.s.} 0$$
as $n \to \infty$.

(c) *In addition we have*
$$\sup_{f,h \in \mathscr{L}_M} \sup_{\xi \in \Sigma} \left| \hat{\Gamma}(L_{t_n}, L_{t_n} f, h)(\xi) - \Gamma(L_{t_n}, L_{t_n} f, h)(\xi) \right| \xrightarrow{a.s.} 0$$
as $n \to \infty$.

Proof. For part (a), if we fix n and ε, we have

$$\Pr\left\{ \sup_{x \in \Sigma} \left| \hat{\Gamma}(L_{t_n}, f, h)(x) - \Gamma(L_{t_n}, f, h)(x) \right| \geq \varepsilon \right\} \leq Q_{t_n}(\mathcal{F}^{t_n}_{f,h}, \varepsilon, K, n)$$

where $K = K(f, h)$ as in Corollary 4.22. Now plugging in the expression for t_n we observe a bound of the form

$$Q_{t_n}(\mathcal{F}^{t_n}_{f,h}, \varepsilon, K) \leq p\left(n^{\frac{1}{2(2d+\sigma)}}, \frac{1}{\varepsilon} \right) \exp\left(-c_1 \varepsilon^2 n^{\sigma/(2d+\sigma)} \right) \tag{4.102}$$

where p is a fixed polynomial bound and $c_1 > 0$ is a constant. Thus with ε fixed, we have

$$\sum_{n=1}^{\infty} Q_{t_n}(\mathcal{F}^{t_n}_{f,h}, \varepsilon, K, n) < \infty.$$

Applying the Borel-Cantelli Lemma gives the almost sure convergence. The proof of part (b) is analogous, since we have the estimate

$$\Pr\left\{ \sup_{f,h \in \mathscr{L}_M} \sup_{x \in \Sigma} \left| \hat{\Gamma}(L_{t_n}, f, h)(x) - \Gamma(L_{t_n}, f, h)(x) \right| \geq \varepsilon \right\} \leq Q_{t_n}(\mathcal{F}^{t_n}, \varepsilon, K, n),$$

with $K = \frac{2}{e}M^2$. Observe that $Q_{t_n}(\mathcal{F}^{t_n}, \varepsilon, K, n)$ satisfies a bound of the form (4.102), but this time the coefficients of the polynomial p depend on Σ, M and N as seen in Lemma 4.15. Part (c) follows from the estimate

$$\Pr\left\{\sup_{f,h\in\mathscr{L}_M}\sup_{x\in\Sigma}\left|\hat{\Gamma}(L_t, L_t f, h)(x) - \Gamma(L_t, L_t f, h)(x)\right| \geq \varepsilon\right\} \leq Q_t(\mathcal{F}^t_*, \varepsilon, K, n),$$

where $K = M^2 C_1^*(\Sigma)$. □

Now we see that for almost sure convergence, one requires a bound of the form (4.102). For this reason, we introduce notation for use in the sequel: consider a function

$$\mathcal{Q} : \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{N} \to \mathbb{R}^+. \tag{4.103}$$

We say that $\mathcal{Q}(t, \varepsilon, n) \in O_{BC}(\beta)$ if

$$\sum_{n=1}^{\infty} \mathcal{Q}(n^{-\frac{1}{\beta+\sigma}}, \varepsilon, n) < \infty \tag{4.104}$$

for all $\sigma, \varepsilon > 0$. Clearly the definition gives

$$O_{BC}(\beta) \subset O_{BC}(\beta')$$

for $\beta' > \beta$. We also observe that

$$O_{BC}(\beta) + O_{BC}(\beta') \in O_{BC}\left(\max\{\beta, \beta'\}\right).$$

Lemma 4.25. *For the classes of functions defined by* (4.21), (4.22), (4.32), (4.56), (4.57), (4.58), (4.59), *for fixed* $\varepsilon, M > 0$ *and for* $t > 0$ *small depending on the embedding coordinates of* Σ *we have*

$$Q_t(\mathcal{F}, \varepsilon, M, n) \in O_{BC}(2d).$$

Proof. Plugging in

$$t = n^{-\frac{1}{\beta+\sigma}}$$

the dominant exponential factor in (4.91) becomes

$$\exp\left(-\frac{\varepsilon^2\lambda_0^4}{8C} n^{\frac{\beta+\sigma-2d}{\beta+\sigma}}\right).$$

Clearly, for

$$\beta + \sigma - 2d > 0$$

we have

$$\sum_{n=1}^{\infty} \mathcal{Q}(n^{-\frac{1}{\beta+\sigma}}, \varepsilon, n) < \infty.$$

□

We record the following, which is clear from the definitions and Corollary 4.12.

Corollary 4.26. *Given any of the classes above, let*

$$\mathcal{Q}(t,\varepsilon,n) = Q_t(\mathcal{F}, \varepsilon t^{\alpha}, K t^{\delta}, n).$$

where δ is a real number. Then

$$\mathcal{Q}(t,\varepsilon,n) \in O_{BC}(2d + 2\alpha - 2\delta).$$

4.6. **Proof of Theorem C.** We will only carry out the proof of part (b). The proof of part (a) is analogous. Recall that

$$\Gamma_2(L_t, f, f) = \frac{1}{2}\left(L_t\left(\Gamma_t(f,f)\right) - 2\Gamma_t(L_t f, f)\right),$$

and that from Remark 4.2

$$\hat{\Gamma}_2(L_t, f, f) = \frac{1}{2}\left(\hat{L}_t\left(\hat{\Gamma}_t(f,f)\right) - 2\hat{\Gamma}_t(\hat{L}_t f, f)\right),$$

so we will start by estimating the difference $\hat{L}_t\left(\hat{\Gamma}_t(f,f)\right) - L_t\left(\Gamma_t(f,f)\right)$ which we write as

$$\left[\hat{L}_t(\hat{\Gamma}_t(f,f)) - L(\Gamma_t(f,f))\right](x) = \hat{L}_t\left(\hat{\Gamma}_t(f,f) - \Gamma_t(f,f)\right)(x) \tag{4.105}$$

$$+ \left(\hat{L}_t - L_t\right)\Gamma_t(f,f)(x) \tag{4.106}$$

$$= A_1(x) + A_2(x), \tag{4.107}$$

and observe that

$$\|A_1\|_{L^\infty(\Sigma)} = \sup_{x\in\Sigma}\left|\hat{L}_t\left(\hat{\Gamma}_t(f,f)(x) - \Gamma_t(f,f)(x)\right)\right| \le \frac{4}{t}\sup_{f\in\mathscr{L}_M}\sup_{\xi\in\Sigma}\left|\hat{\Gamma}_t(f,f)(\xi) - \Gamma_t(f,f)(\xi)\right|. \tag{4.108}$$

In order to estimate $A_2 = \left(\hat{L}_t - L_t\right)\left(\Gamma_t(f,f)\right)$ we make use of Corollary 4.23. We now estimate the difference

$$\hat{\Gamma}_t(\hat{L}_t f, f)(x) - \Gamma_t(L_t f, f)(x) = \hat{\Gamma}_t(\hat{L}_t f - L_t f, f)(x) + \left(\hat{\Gamma}_t - \Gamma_t\right)(L_t f, f)(x) \tag{4.109}$$

$$= A_3(x) + A_4(x), \tag{4.110}$$

and we note

$$\|A_3\|_{L^\infty(\Sigma)} = \sup_{x\in\Sigma}\left|\hat{\Gamma}_t(\hat{L}_t f - L_t f, f)(x)\right| \tag{4.111}$$

$$\le \sup_{x\in\Sigma}\frac{2}{t}\left(\sup_{\xi\in\Sigma}\left|\hat{L}_t(f)(\xi) - L_t(f)(\xi)\right|\right)\frac{1}{n\hat{\theta}_t(x)}\sum_{j=1}^{n} e^{-\frac{\|\xi_j - x\|^2}{2t}}|f(\xi_j) - f(x)| \tag{4.112}$$

$$\le 2\frac{e^{-1/2}\|f\|_{\mathrm{Lip}}}{t^{1/2}}\sup_{x\in\Sigma}\frac{1}{\hat{\theta}_t(x)}\left(\sup_{\xi\in\Sigma}\left|(\hat{L}_t(f)(\xi) - L_t(f)(\xi))\right|\right)$$

$$\le 2\frac{e^{-1/2}M}{t^{1/2}}\sup_{x\in\Sigma}\frac{1}{\hat{\theta}_t(x)}\left(\sup_{f\in\mathscr{L}_M}\sup_{\xi\in\Sigma}\left|(\hat{L}_t(f)(\xi) - L_t(f)(\xi))\right|\right) \tag{4.113}$$

where we have used the fact that for functions u, v

$$\left|\hat{\Gamma}_t(u,v)(x)\right| \leq \frac{2}{t}\sup_{\xi\in\Sigma}|u(\xi)|\left(\frac{1}{n\hat{\theta}_t(x)}\sum_{j=1}^{n} e^{-\frac{\|x-\xi_j\|^2}{2t}}|v(x)-v(\xi_j)|\right) \tag{4.114}$$

$$\leq \frac{2}{t}\sup_{\xi\in\Sigma}|u(\xi)|\left(\frac{\|v\|_{\mathrm{Lip}}}{n\hat{\theta}_t(x)}\sup_{\rho>0}\rho e^{-\frac{\rho^2}{2t}}\right), \tag{4.115}$$

and

$$\sup_{\rho>0}\rho e^{-\frac{\rho^2}{2t}} = t^{1/2}e^{-1/2} \leq t^{1/2}.$$

In view of Lemma (4.17), we will write the bound (4.113) on $\|A_3\|_{L^\infty(\Sigma)}$ as

$$\|A_3\|_{L^\infty(\Sigma)} \leq 2\frac{e^{-1/2}M}{t^{1/2}\theta_t}\sup_{x\in\Sigma}\frac{1}{\theta_t(x)}\left(\sup_{f\in\mathscr{L}_M}\sup_{\xi\in\Sigma}\left|(\hat{L}_t(f)(\xi)-L_t(f)(\xi))\right|\right) \tag{4.116}$$

$$+\frac{2e^{-1/2}M}{t^{1/2}}\sup_{x\in\Sigma}\left|\frac{1}{\hat{\theta}_t(x)}-\frac{1}{\theta_t(x)}\right|\left(\sup_{f\in\mathscr{L}_M}\sup_{\xi\in\Sigma}\left|(\hat{L}_t(f)(\xi)-L_t(f)(\xi))\right|\right). \tag{4.117}$$

In order to estimate $\|A_4\|_{L^\infty(\Sigma)}$, i.e.

$$\|A_4\|_{L^\infty(\Sigma)} = \sup_{x\in\Sigma}\left|\left(\hat{\Gamma}_t-\Gamma_t\right)(L_t(f,f))(x)\right|,$$

we will make use of Corollary 4.22.

Observe now that the terms A_1, A_2, A_3, A_4 above are random variables with expressions of the form $A_i = A_i(f,x)$, let $A_i^* = \sup_{f\in\mathscr{L}_M}\sup_{x\in\Sigma}|A_i(f,x)|$ for $i = 1,2,3,4$. We now have from (4.105) and (4.109)

$$\Pr\left\{\sup_{f\in\mathscr{L}_M}\sup_{x\in\Sigma}\left|\hat{\Gamma}_2(L_t,f,f)-\Gamma_2(L_t,f,f)\right|(x)\geq\varepsilon\right\} \tag{4.118}$$

$$\leq \Pr\left\{A_1^*\geq\frac{\varepsilon}{4}\right\}+\Pr\left\{A_2^*\geq\frac{\varepsilon}{4}\right\}+\Pr\left\{A_3^*\geq\frac{\varepsilon}{4}\right\}+\Pr\left\{A_4^*\geq\frac{\varepsilon}{4}\right\} \tag{4.119}$$

$$= P_1+P_2+P_3+P_4. \tag{4.120}$$

From (4.108) and Corollary 4.22 we have

$$P_1 \leq \Pr\left\{\sup_{f\in\mathscr{L}_M}\sup_{\xi\in\Sigma}\left|\hat{\Gamma}_t(f,f)(\xi)-\Gamma_t(f,f)(\xi)\right|\geq\frac{t\varepsilon}{16}\right\}\leq Q_t\left(\mathcal{F}_f^t,\frac{t\varepsilon}{16},2e^{-1}M^2,n\right), \tag{4.121}$$

$$\in O_{BC}\left(2d+2\right). \tag{4.122}$$

The statement about the convergence order follows from Corollary 4.26.

For $A_2 = \left(\hat{L}_t - L_t\right)(\Gamma_t(f,f))$ we apply part (c) of Corollary 4.23 together with (4.81) and obtain

$$P_2 \le Q_t\left(\mathcal{H}_*^t, \varepsilon t, \beta_d M^2, n\right) \in O_{BC}\left(2d+2\right). \tag{4.123}$$

Using (4.117) we have

$$P_3 \le \Pr\left\{\frac{\varepsilon}{8} \le 2\frac{e^{-1/2}M}{t^{1/2}\theta_t}\sup_{x\in\Sigma}\frac{1}{\theta_t(x)}\left(\sup_{f\in\mathscr{L}_M}\sup_{\xi\in\Sigma}\left|(\hat{L}_t(f)(\xi) - L_t(f)(\xi))\right|\right)\right\} \tag{4.124}$$

$$+\Pr\left\{\frac{\varepsilon}{8} \le \frac{2e^{-1/2}M}{t^{1/2}}\sup_{x\in\Sigma}\left|\frac{1}{\hat{\theta}_t(x)} - \frac{1}{\theta_t(x)}\right|\left(\sup_{f\in\mathscr{L}_M}\sup_{\xi\in\Sigma}\left|(\hat{L}_t(f)(\xi) - L_t(f)(\xi))\right|\right)\right\} \tag{4.125}$$

and choosing $t > 0$ small enough we obtain from (4.84) and Corollary 4.23 the estimate

$$\Pr\left\{\frac{\varepsilon}{8} \le 2\frac{e^{-1/2}M}{t^{1/2}\theta_t}\left(\sup_{f\in\mathscr{L}_M}\sup_{\xi\in\Sigma}\left|(\hat{L}_t(f)(\xi) - L_t(f)(\xi))\right|\right)\right\} \tag{4.126}$$

$$\le \Pr\left\{\frac{\varepsilon}{8}\frac{\lambda_0 t^{\frac{d+2}{2}}e^{1/2}}{M} \le \sup_{f\in\mathscr{L}_M}\sup_{\xi\in\Sigma}\left|(\hat{L}_t(f)(\xi) - L_t(f)(\xi))\right|\right\} \tag{4.127}$$

$$\le Q_t\left(\mathcal{H}^t, \frac{\varepsilon}{8}\frac{\lambda_0 t^{\frac{d+1}{2}}e^{1/2}}{M}t, C, n\right) \in O_{BC}(3d+3), \tag{4.128}$$

where $C = C^*(\Sigma)M \ge \sup_{f\in\mathscr{L}_M}\|f\|_{L^\infty(\Sigma)}$. Here the increased order is $t^{\frac{d+1}{2}}$ and t which is $2d + d + 1 + 2$. On the other hand, we will make use of the estimate

$$\sup_{f\in\mathscr{L}_M}\sup_{\xi\in\Sigma}\left|(\hat{L}_t(f)(\xi) - L_t(f)(\xi))\right| \le \frac{4}{t}\sup_{f\in\mathscr{L}_M}\|f\|_{L^\infty} \le \frac{4C^*(\Sigma)M}{t} = \frac{4C}{t},$$

for any $t > 0$, to obtain

$$\Pr\left\{\frac{\varepsilon}{8} \le \frac{2e^{-1/2}M}{t^{1/2}}\sup_{x\in\Sigma}\left|\frac{1}{\hat{\theta}_t(x)} - \frac{1}{\theta_t(x)}\right|\left(\sup_{f\in\mathscr{L}_M}\sup_{\xi\in\Sigma}\left|(\hat{L}_t(f)(\xi) - L_t(f)(\xi))\right|\right)\right\} \tag{4.129}$$

$$\le \Pr\left\{\frac{\varepsilon t^{3/2}e^{1/2}}{64M} \le \sup_{x\in\Sigma}\left|\frac{1}{\hat{\theta}_t(x)} - \frac{1}{\theta_t(x)}\right|\right\} \tag{4.130}$$

and from (4.89)-(4.90) and Lemma 4.1 we observe that

$$\Pr\left\{\frac{\varepsilon}{8} \le \frac{2e^{-1/2}M}{t^{1/2}}\sup_{x\in\Sigma}\left|\frac{1}{\hat{\theta}_t} - \frac{1}{\theta_t}\right|\left(\sup_{f\in\mathscr{L}_M}\sup_{\xi\in\Sigma}\left|(\hat{L}_t(f)(\xi) - L_t(f)(\xi))\right|\right)\right\} \in O_{BC}(d+3).$$

We then conclude that

$$P_3 \in O_{BC}(3d+3). \tag{4.131}$$

Finally, in order to estimate P_4, we use part (c) of Lemma 4.22, namely

$$\Pr\left\{\sup_{f\in\mathscr{L}_M}\sup_{x\in\Sigma}\left|\hat{\Gamma}_t(L_tf,f)(x)-\Gamma_t(L_tf,f)(x)\right|\geq\varepsilon\right\}\leq Q_t(\mathcal{F}_*^t,\varepsilon,K,n),$$

where $K=C_1^*(\Sigma)M^2$ (from (4.82)) and therefore

$$P_4=\Pr\left\{A_4^*\geq\frac{\varepsilon}{4}\right\}\leq Q_t(\mathcal{F}_*^t,\frac{\varepsilon}{4},K,n)\in O_{BC}(2d). \tag{4.132}$$

Using again

$$\Pr\left\{\left|\hat{\Gamma}_2(L_t,f,f)-\Gamma_2(L_t,f,f)\right|\geq\varepsilon\right\}\leq P_1+P_2+P_3+P_4,$$

and from (4.122), (4.123), (4.131) and (4.132) we have

$$\Pr\left\{\left|\hat{\Gamma}_2(L_t,f,f)-\Gamma_2(L_t,f,f)\right|\geq\varepsilon\right\}\in O_{BC}(3d+3), \tag{4.133}$$

and it follows from the Borel-Cantelli argument given in Section 4.5 that for any sequence of the form $t_n=n^{-\gamma}$ where $\gamma=\frac{1}{3d+3+\sigma}$ and σ is any positive number we have

$$\sup_{\xi\in\Sigma}\left|\hat{\Gamma}_2(L_{t_n},f,f)-\Gamma_2(L_{t_n},f,f)\right|\xrightarrow{\text{a.s.}}0.$$

This proves Theorem B.

Now with the convergence for each fixed function f we can prove Corollary A (see section 3).

Proof of Corollary A. We work on a compact smooth submanifold of Euclidean space. With the ambient distance function, there is no cut locus, and the set of functions given by (recall (3.6))

$$\mathcal{R}=\{F_{x,y}:(x,y)\in\Sigma\times\Sigma\},$$

is uniformly bounded in C^5. The map

$$\Sigma\times\Sigma\to C^5(\Sigma)$$
$$(x,y)\to F_{x,y}$$

is a Lipschitz map. It follows that we can take a finite δ-net $\mathcal{G}$ with respect to the L^∞ topology, and the net size will grow at worst polynomially. That is for a given

$$f\in\mathcal{R}$$

there exists

$$f^*\in\mathcal{G}$$

such that

$$\|f-f^*\|_{L^\infty}<\delta.$$

The constant δ will be chosen below. We want to estimate the probability

$$P=\Pr\left\{\sup_{f\in\mathcal{R}}\sup_{\xi\in\Sigma}\left|\hat{\Gamma}_2(L_t,f,f)\,(\xi)-\Gamma_2(\Delta_g,f,f)\,(\xi)\right|\geq\varepsilon\right\}.$$

We have

$$\begin{aligned}\hat{\Gamma}_2(L_t,f,f)\,(\xi)-\Gamma_2(\Delta_g,f,f)\,(\xi) &= \hat{\Gamma}_2(L_t,f,f)\,(\xi)-\hat{\Gamma}_2(L_t,f^*,f^*)\,(\xi)\\ &+\hat{\Gamma}_2(L_t,f^*,f^*)\,(\xi)-\Gamma_2(L_t,f^*,f^*)\,(\xi)\\ &+\Gamma_2(L_t,f^*,f^*)\,(\xi)-\Gamma_2(L_t,f,f)\,(\xi)\\ &+\Gamma_2(L_t,f,f)\,(\xi)-\Gamma_2(\Delta_g,f,f)\,(\xi)\,.\end{aligned}$$

Thus,

$$\begin{aligned}P &\leq \Pr\left\{\sup_{f\in\mathcal{R}}\sup_{\xi\in\Sigma}\left|\hat{\Gamma}_2(L_t,f,f)\,(\xi)-\hat{\Gamma}_2(L_t,f^*,f^*)\right|\geq\frac{\varepsilon}{4}\right\}\\ &+\Pr\left\{\sup_{f\in\mathcal{R}}\sup_{\xi\in\Sigma}\left|\hat{\Gamma}_2(L_t,f^*,f^*)\,(\xi)-\Gamma_2(L_t,f^*,f^*)\,(\xi)\right|\geq\frac{\varepsilon}{4}\right\}\\ &+\Pr\left\{\sup_{f\in\mathcal{R}}\sup_{\xi\in\Sigma}\left|\Gamma_2(L_t,f^*,f^*)\,(\xi)-\Gamma_2(L_t,f,f)\,(\xi)\right|\geq\frac{\varepsilon}{4}\right\}\\ &+\Pr\left\{\sup_{f\in\mathcal{R}}\sup_{\xi\in\Sigma}\left|\Gamma_2(L_t,f,f)\,(\xi)-\Gamma_2(\Delta_g,f,f)\right|\geq\frac{\varepsilon}{4}\right\}\\ &\leq P_1+P_2+P_3+P_4.\end{aligned}$$

First, note that for the bilinear form $\hat{\Gamma}_2$ we have

$$\hat{\Gamma}_2(L_t,f,f)\,(\xi)-\hat{\Gamma}_2(L_t,f^*,f^*)=\hat{\Gamma}_2(L_t,f-f^*,f)\,(\xi)-\hat{\Gamma}_2(L_t,f^*-f,f^*)$$

thus

$$\left|\hat{\Gamma}_2(L_t,f,f)\,(\xi)-\hat{\Gamma}_2(L_t,f^*,f^*)\right|\leq\frac{4}{t^2}\left\|f-f^*\right\|_{L^\infty}\sup_{f\in\mathcal{R}}\left\|f\right\|_{L^\infty}.$$

So we may choose

$$\delta=\frac{\varepsilon}{16C_0}t^{2+\sigma}$$

where

$$C_0=\sup_{F_{x,y}\in\mathcal{R}}\left\|F_{x,y}\right\|_\infty.$$

With this choice $P_1=0$. By the same reasoning, also $P_3=0$.

Next, we have by Theorem 3.1

$$\left|\sup_{\xi\in\Sigma}\Gamma_2(L_t,f,f)\,(\xi)-\Gamma_2(\Delta_g,f,f)\,(\xi)\right|\leq C_5t^{1/2}$$

where

$$C_5=\sup_{F_{x,y}\in\mathcal{R}}\left\|F_{x,y}\right\|_{C^5(\Sigma)}.$$

We conclude that as long as

$$t^{1/2}<\frac{\varepsilon}{4C_5}$$

we have $P_4 = 0$. We are left to show

$$P_2 = \Pr\left\{\sup_{f\in\mathcal{R}}\sup_{\xi\in\Sigma}\left|\hat{\Gamma}_2(L_t, f^*, f^*)(\xi) - \Gamma_2(L_t, f^*, f^*)(\xi)\right| \geq \frac{\varepsilon}{4}\right\} \to 0.$$

But by (4.133) we have, for each individual $f^* \in \mathcal{G}$

$$\Pr\left\{\sup_{\xi\in\Sigma}\left|\hat{\Gamma}_2(L_t, f^*, f^*)(\xi) - \Gamma_2(L_t, f^*, f^*)(\xi)\right| \geq \frac{\varepsilon}{4}\right\} \in O(3d+3),$$

and the size of the set satisfies

$$|\mathcal{G}| \leq \mathcal{N}\left(\mathcal{R}, \frac{\varepsilon}{16C_0}t^{2+\sigma}\right).$$

This in turn is bounded by a polynomial in $\frac{1}{t}$, so we can apply the Borel Cantelli argument and obtain the result. □

5. Local PCA and proof of Theorem B

The goal of this section is to construct a class of test functions that can be inserted in our construction for $\hat{\Gamma}_2(L_{t_n}, \cdot, \cdot)$ to recover the Ricci curvature as stated in Theorem B (in other words, we will explain how to obtain the functions f_n in Theorem B). The key for the construction of these test functions is a method for estimating a basis of the tangent space to Σ^d at a given point $x \in \Sigma$ known as *local PCA* where PCA stands for "Principal Component Analysis". The construction that we are about to describe was developed in [25] Section 2.1 and Appendix B, however, for the reader's convenience we will review the construction without proving any of the theorems shown in [25]. After reviewing the local PCA construction in [25] we will explain how these ideas can be combined with Theorem C to prove Theorem B.

5.1. **Estimating an orthonormal basis of the tangent space to a submanifold at a point.** Let $x \in \Sigma^d$ where Σ^d is again a smooth d-dimensional submanifold of $\mathbb{R}^N$ and let $\xi_1, \xi_2 \ldots, \xi_n$ be data points on Σ^d. As pointed out in the introduction, suppose that we fix an embedding $F : \Sigma^d \to \mathbb{R}^N$ so that we obtain a metric g in Σ induced by F given in local coordinates by

$$g_{ij} = \langle D_iF, D_jF\rangle,$$

where of course $\langle\cdot,\cdot\rangle$ is the inner product of $\mathbb{R}^N$. We assume that the data points $\{\xi_j\}_{j=1}^d$ are uniformly distributed and assuming that we have sufficiently many data points we expect that many of the points $\{\xi_j\}$ will concentrate near x. More precisely, let us fix a positive number $\varepsilon > 0$ and let $B_\varepsilon(x)$ be the geodesic ball of radius ε centered at x with respect to g and let $\{\eta_1, \ldots, \eta_{N_\varepsilon}\}$ be the intersection of $B_\varepsilon(x)$ with the set $\{\xi_1, \ldots, \xi_n\}$. For large n, and an analysis similar to the one presented in the previous section, it is clear that for large n, we can choose ε such that $N_\varepsilon \gg d$ but $N_\varepsilon \ll n$. We now choose a function Φ satisfying

- Φ is supported in $[0,1]$,
- Φ is non increasing in $[0,1]$,
- Φ is C^2 on $[0,1]$.

One common choice for Φ is $\Phi(s) = (1 - s^2)\chi_{[0,1]}(s)$. After choosing Φ, we consider a $N_\varepsilon \times N_\varepsilon$ diagonal matrix D_ε with diagonal entries

$$(D_\varepsilon)_{jj} = \sqrt{\Phi\left(\frac{\|x - \eta_j\|}{\varepsilon}\right)},$$

for $j = 1, \ldots, N_\varepsilon$. At the same time, consider now the $N \times N_\varepsilon$ matrix X_ε whose rows are given by

$$X_\varepsilon = \left[\eta_1 - x, \ldots, \eta_{N_\varepsilon} - x\right],$$

and the $N \times N_\varepsilon$ matrix

$$W_\varepsilon = X_\varepsilon D_\varepsilon.$$

The idea of constructing the matrix W_ε above is to weight the data points so that those points that are closer to x are given preference. Next, the matrix W_ε admits a *singular value decomposition* of the form

$$W_\varepsilon = U_\varepsilon \Lambda_\varepsilon V_\varepsilon^T,$$

where U_ε is a $N \times N_\varepsilon$ matrix whose columns are orthonormal in the Hilbert-Schmidt norm and known as the *left singular vectors* of W_ε, and Λ_ε is a diagonal matrix with non increasing diagonal elements that describe the relative importance of the vectors. Recall that the Hilbert-Schmidt norm is defined for $N \times d$ matrices by

$$\|A\|_{HS} = \sqrt{\operatorname{tr}(A^T A)}.$$

We now consider a $N \times d$ matrix U_ε with orthonormal columns given by taking the first d orthonormal singular vectors of U_ε (each column of U_ε is a singular vector for W_ε). Alternatively, or if d is unknown, we can choose the vectors whose weights in Λ_ε are bigger than a chosen cutoff value, for example $1/2$. On a smooth manifold these will agree for large numbers of points. We will write U_ε as

$$U_\varepsilon = \left[\zeta_{1,\varepsilon}(x), \ldots, \zeta_{d,\varepsilon}(x)\right]. \tag{5.1}$$

The vectors $\zeta_{j,\varepsilon}(x)$ for $j = 1, \ldots, d$ form an orthonormal basis for a d-dimensional subspace of $\mathbb{R}^N$, in fact, this basis will serve as an approximation for a basis to the tangent space $T_x\Sigma$. Strictly speaking, the vectors $\zeta_{j,\varepsilon}$ depend on the point x, and the data points $\xi_1, \ldots, \xi_n$, but for now we will omit that dependence. It is important to keep in mind though that the dependence of $\zeta_{j,\varepsilon}$ on these parameters is highly non-linear and in principle it should create a problem of correlation, however, the empirical theoretic methods that we have discussed above will allow us to deal with this high correlation

Another important part of the construction is the choice of ε, in fact, as the number of data points tends to infinity, ε will tend to zero at a definite rate, more precisely, we will choose ε to be essentially a negative power of n (the number of data points). With these observations in mind, we can state the main theorem that asserts that the columns of the matrix U_ε defined above converge to a basis of the tangent space $T_x\Sigma^d$.

Theorem 5.1 (See Theorem B.1 in [25]). *Let $\{\xi_j\}_{j=1}^n$ be a uniformly distributed sample of data points on the embedded submanifold $\Sigma^d \subset \mathbb{R}^N$ and let us choose ε by $\varepsilon = \varepsilon_n = O(n^{-\frac{6}{d+2}})$. There exists a $N \times d$ matrix $\Theta_\infty(x)$ whose columns are a basis of $F_*\left(T_x\Sigma^d\right)$ and such that with high probability (w.h.p.)*

$$\min_{U \in O(d)} \|U^T U_{\varepsilon_n} - \Theta_\infty(x)\|_{HS} = O(n^{-\frac{3}{d+2}}).$$

Remark 5.2. The estimate with high probability of $\Theta_\infty(x)$ by U_{ε_n} implies that U_{ε_n} converges to $\Theta_\infty(x)$ almost surely.

In the next section we show how to use Theorem 5.1 to prove Theorem B.

5.2. **Proof Of Theorem B.** We start by observing that if one knows a tangent vector η to Σ to a point x, then one can construct a test function such that the iterated Carré du Champ applied to that function is precisely $\mathrm{Ric}_x(\eta, \eta)$.

Proposition 5.3. *Let $F : \Sigma^d \to \mathbb{R}^N$ be an embedding of Σ in $\mathbb{R}^N$ and let $f : \mathbb{R}^N \to \mathbb{R}$ be given by $f(z) = \langle z, \eta\rangle$ where $\eta \in \mathbb{R}^N$ is fixed. If g is the metric induced by the embedding $F : \Sigma \to \mathbb{R}^N$ and if we fix a point $x \in \Sigma$ we obtain*

$$\Gamma_2(\Delta_g, f, f)(x) = \mathrm{Ric}_x(\eta^T, \eta^T) + \left((\eta)^\perp\right)^2 \|\mathrm{II}_\Sigma\|^2_{\Sigma,x},$$

where η^T and $\eta^\perp$ are the orthogonal projections of η onto $F_*\left(T_x\Sigma\right)$ and $\left(F_*\left(T_x\Sigma\right)\right)^\perp$ respectively and II_Σ is the second fundamental form of Σ. In particular, if $\eta \in F_*\left(T_x\Sigma\right)$ we have

$$\Gamma_2(\Delta_g, f, f)(x) = \mathrm{Ric}_x(\eta, \eta).$$

Proof. For simplicity let us only prove the codimension 1 case. By (2.8) we have

$$\Gamma_2(\Delta_g, f, f)(x) = \mathrm{Ric}_x(\nabla_\Sigma f, \nabla_\Sigma f) + \left\|\nabla^2_\Sigma f\right\|^2_{\Sigma,x}.$$

Observe that $(\nabla_\Sigma f)_x$ is obtained by taking the projection of Df (ambient derivative) onto $F_*\left(T_x\Sigma\right)$ and therefore $(\nabla_\Sigma f)_x = \eta^T$. Observe now that we have

$$\nabla^2_\Sigma f = D^2 f - (Df)^\perp \mathrm{II}_\Sigma = -(Df)^\perp \mathrm{II}_\Sigma.$$

The proposition follows. □

Next, we show how the approximate Ricci is constructed from the basis. For each of the vectors $\zeta_j(x)$ determined by the PCA, consider the linear function

$$f_{n,j}(z) = \langle z, \zeta_j\rangle$$

and then define

$$\hat{R}_{i,j} = \hat{\Gamma}_2(L_{t_n}, f_{n,i}, f_{n,j}). \tag{5.2}$$

For any vector (in the tangent space, approximate tangent space, or neither) we can define the approximate Ricci curvature of η by projecting η onto the vectors ζ_j and summing the linear combination as follows. Projecting onto the approximate basis and splitting the vector, let

$$\eta = \eta^A + \eta^\perp = \eta^j \zeta_j + \eta^\perp$$

and define

$$\hat{\mathrm{Ric}}(\eta,\eta) = \hat{R}_{i,j}\eta^i\eta^j.$$

Proof of Theorem B. For a given vector $\eta \in T_xM$ define the function $f(z) = \langle z, \eta\rangle$ as above. Let

$$\eta = \eta^A + \eta^\perp$$

and define $f^A(z) = \langle z, \eta^A\rangle$ and $f^\perp(z) = \langle z, \eta^\perp\rangle$ so that

$$\hat{\mathrm{Ric}}_x(\eta,\eta) = \hat{\Gamma}_2(L_{t_n}, f^A, f^A). \tag{5.3}$$

Now we compute the difference of the actual Ricci and the approximate Ricci, using Proposition 5.3, and (5.3)

$$\mathrm{Ric}_x(\eta,\eta) - \hat{\mathrm{Ric}}_x(\eta,\eta) = \Gamma_2(\Delta_g, f, f) - \hat{\Gamma}_2(L_{t_n}, f^A, f^A)) \tag{5.4}$$

$$= \Gamma_2(\Delta_g, f, f) - \Gamma_2(\Delta_g, f^A, f^A) \tag{5.5}$$

$$+ \Gamma_2(\Delta_g, f^A, f^A) - \Gamma_2(L_{t_n}, f^A, f^A) \tag{5.6}$$

$$+ \Gamma_2(L_{t_n}, f^A, f^A) - \hat{\Gamma}_2(L_{t_n}, f^A, f^A) \tag{5.7}$$

Observe that clearly, all functions f^A are in the class $\mathscr{L}_M$ for some fixed $M > 0$ and therefore

$$\left|\left(\hat{\Gamma}_2(L_{t_n}, f^A, f^A) - \Gamma_2(L_{t_n}, f^A, f^A)\right)\right| \leq \sup_{f\in\mathscr{L}_M}\sup_{\xi\in\Sigma}\left|\hat{\Gamma}_2(L_{t_n}, f, f)(\xi) - \Gamma_2(L_{t_n}, f, f)(\xi)\right|,$$

and from Theorem C, with the given choice of scale t_n, we have that

$$\left|\Gamma_2(L_{t_n}, f^A, f^A) - \hat{\Gamma}_2(L_{t_n}, f^A, f^A)\right| \xrightarrow{\text{a.s.}} 0.$$

Similarly, it follows from Theorem 3.2 that

$$\left|\Gamma_2(\Delta_g, f^A, f^A) - \Gamma_2(L_{t_n}, f^A, f^A)\right| \longrightarrow 0.$$

Now certainly, as the PCA is choosing linear functions that recover the tangent space with high probability in the limit, the linear functions f^A converge with high probability uniformly in all orders to the function f on the manifold. It follows that the term

$$\left|\Gamma_2(\Delta_g, f, f) - \Gamma_2(\Delta_g, f^A, f^A)\right| \xrightarrow{\text{a.s}} 0.$$

Combining the above three limits we have

$$\left|\mathrm{Ric}_x(\eta,\eta) - \hat{\mathrm{Ric}}_x(\eta,\eta)\right| \xrightarrow{\text{a.s}} 0.$$

$\square$

5.3. **Dimension estimation.** The problem of estimating the dimension of the underlying submanifold assuming the manifold hypothesis is a fascinating subject in itself and there is a large number of estimators for the intrinsic dimension in the manifold learning problem that have been proposed in the literature. Some of the early approaches for dimension estimation were based on PCA or also the application of the Vapnik-Chervonenkis classes of sets of separating hyperplanes [31]. These methods lose effectiveness when applied to relatively highly nonlinear problems or in the presence of noise. There is a wide variety of methods that have been proposed in order to overcome these difficulties, and many of them use of techniques in machine learning that are nowadays well known, for example methods based on suitable maximum likelihood estimators applied to distances to the k Nearest Neighbors (kNN). See for example [22, 21] where many of these ideas are discussed in detail. There are also fractal-based methods, and methods based on the concept of ISOMAP, which consists in estimating the distance of points nearby by the geodesic distance whenever possible and, and estimating the distance between points that are far apart by means of the shortest path in the graph that is used to approximate the submanifold. See for example [5], [6] and [33] and the references therein. See also the discussion in [25, page 1073], the references therein and [15].

References

[1] A. Ache and M. Warren. *Approximating coarse Ricci curvature on metric measure spaces with applications to submanifolds of Euclidean space.* arXiv:1505.04171v1, 2015.

[2] Antonio G. Ache and Micah W. Warren. Coarse Ricci curvature as a function on $M \times M$. *Results Math.*, 72(4):1823–1837, 2017.

[3] D. Bakry and Michel Émery. Diffusions hypercontractives. In *Séminaire de probabilités, XIX, 1983/84*, volume 1123 of *Lecture Notes in Math.*, pages 177–206. Springer, Berlin, 1985.

[4] Mikhail Belkin and Partha Niyogi. Towards a theoretical foundation for Laplacian-based manifold methods. *J. Comput. System Sci.*, 74(8):1289–1308, 2008.

[5] J. Bruske and G. Sommer. Intrinsic dimensionality estimation with optimally topology preserving maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:572–575, 1998.

[6] F. Camastra and A. Vinciarelli. Estimating the intrinsic dimension of data with a fractal-based method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:1404–1407, 2002.

[7] Gunnar Carlsson. Topology and data. *Bull. Amer. Math. Soc. (N.S.)*, 46(2):255–308, 2009.

[8] Gunnar Carlsson. Topological pattern recognition for point cloud data. *Acta Numer.*, 23:289–368, 2014.

[9] Ronald R. Coifman and Stéphane Lafon. Diffusion maps. *Appl. Comput. Harmon. Anal.*, 21(1):5–30, 2006.

[10] Charles Fefferman, Sanjoy Mitter, and Hariharan Narayanan. Testing the manifold hypothesis. *J. Amer. Math. Soc.*, 29(4):983–1049, 2016.

[11] Robin Forman. Bochner's method for cell complexes and combinatorial Ricci curvature. *Discrete Comput. Geom.*, 29(3):323–374, 2003.

[12] Andreas Karcher, Richard Palais, and Bob Palais. *Point Clouds: distributing points uniformly on a surface.* preprint, 2014.

[13] Bo'az Klartag, Gady Kozma, Peter Ralli, and Prasad Tetali. Discrete curvature and abelian groups. *Canad. J. Math.*, 68(3):655–674, 2016.

[14] Hiroshi Kunita. Absolute continuity of Markov processes and generators. *Nagoya Math. J.*, 36:1–26, 1969.

[15] Anna V. Little. *Estimating the Intrinsic Dimension of High-Dimensional Data Sets: A Multiscale, Geometric Approach*. ProQuest LLC, Ann Arbor, MI, 2011. Thesis (Ph.D.)–Duke University.

[16] John Lott and Cédric Villani. Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math. (2)*, 169(3):903–991, 2009.

[17] Laurent Najman and Pascal Romon, editors. *Modern approaches to discrete curvature*, volume 2184 of *Lecture Notes in Mathematics*. Springer, Cham, 2017.

[18] David Pollard. *Empirical processes: theory and applications*. NSF-CBMS Regional Conference Series in Probability and Statistics, 2. Institute of Mathematical Statistics, Hayward, CA; American Statistical Association, Alexandria, VA, 1990.

[19] M. Pouryahya, J. Mathews, and A. Tannenbaum. Comparing Three Notions of Discrete Ricci Curvature on Biological Networks. *ArXiv e-prints*, December 2017.

[20] Jean-Pierre Roth. Opérateurs carré du champ et formule de Lévy-Kinchine sur les espaces localement compacts. *C. R. Acad. Sci. Paris Sér. A*, 278:1103–1106, 1974.

[21] A. Rozza, G. Lombardi, C. Ceruti, E. Casiraghi, and P. Campadelli. Novel high intrinsic dimensionality estimators. *Mach. Learn.*, 89(1-2):37–65, 2012.

[22] Alessandro Rozza, Gabriele Lombardi, Marco Rosa, Elena Casiraghi, and Paola Campadelli. IDEA: intrinsic dimension estimation algorithm. In *Image analysis and processing—ICIAP 2011. Part I*, volume 6978 of *Lecture Notes in Comput. Sci.*, pages 433–442. Springer, Heidelberg, 2011.

[23] B. W. Silverman. *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London, 1986.

[24] Leon Simon. *Theorems on regularity and singularity of energy minimizing maps*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 1996. Based on lecture notes by Norbert Hungerbühler.

[25] A. Singer and H.-T. Wu. Vector diffusion maps and the connection Laplacian. *Comm. Pure Appl. Math.*, 65(8):1067–1144, 2012.

[26] Amit Singer and Hau-Tieng Wu. Spectral convergence of the connection Laplacian from random samples. *Inf. Inference*, 6(1):58–123, 2017.

[27] R. P. Sreejith, Karthikeyan Mohanraj, Jürgen Jost, Emil Saucan, and Areejit Samal. Forman curvature for complex networks. *J. Stat. Mech. Theory Exp.*, (6):063206, 26, 2016.

[28] Bartek Suideja and Micah Warren. https://github.com/micah541/ricci, 2015.

[29] Allen Tannenbaum, Chris Sander, Liangjia Zhu, Romeil Sandhu, Ivan Koltsov, Eduard Reznik, Yasin Senbabaoglu, and Tryphon Georgiu. *Graph Curvature and the Robustness of Cancer Networks*. arXiv:1502.04512, 2015.

[30] Aad W. van der Vaart and Jon A. Wellner. *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996. With applications to statistics.

[31] Vladimir N. Vapnik. *Statistical learning theory*. Adaptive and Learning Systems for Signal Processing, Communications, and Control. John Wiley & Sons, Inc., New York, 1998. A Wiley-Interscience Publication.

[32] Cédric Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.

[33] Xiaohui Wang and J. S. Marron. A scale-based approach to finding effective dimensionality in manifold learning. *Electron. J. Stat.*, 2:127–148, 2008.

[34] M. Weber, E. Saucan, and J. Jost. Can one see the shape of a network? *arXiv:1608.07838*, August 2016.

[35] M. Weber, J. Stelzer, E. Saucan, A. Naitsat, G. Lohmann, and J. Jost. Curvature-based Methods for Brain Network Analysis. *arXiv:1707.00180*, July 2017.

[36] Melanie Weber, Emil Saucan, and Jürgen Jost. Characterizing complex networks with forman-ricci curvature and associated geometric flows. *CoRR*, abs/1607.08654, 2016.

E-mail address: tonyache@gmail.com

Department of Mathematics, University of Oregon, Eugene OR 97403
E-mail address: micahw@oregon.edu