COARSE RICCI CURVATURE WITH APPLICATIONS TO MANIFOLD LEARNING

ANTONIO G. ACHE AND MICAH W. WARREN

ABSTRACT. We consider the framework used by Bakry and Emery in their work on logarithmic Sobolev inequalities to define a notion of coarse Ricci curvature on smooth metric measure spaces alternative to the notion proposed by Y. Ollivier. We discuss applications of our construction to the manifold learning problem, specifically to the statistical problem of estimating the Ricci curvature of a submanifold of $\mathbb{R}^N$ from a point cloud assuming that the sample has smooth density with definite lower bounds. More generally we are able to approximate a 1-parameter family of Ricci curvatures that include the Bakry-Emery Ricci curvature.

CONTENTS

1. Introduction 2
   1.1. Background and Motivation 2
   1.2. Carré Du Champ and Bochner Formula 5
   1.3. Iterated Carré du Champ and Coarse Ricci Curvature 7
   1.4. Statement of Results 10
   1.5. Final Remarks 14
   1.6. Organization of the Paper 15
   1.7. Acknowledgements 15

2. Bias Error Estimates 16
   2.1. Bias for Submanifold of Euclidean Space 16
   2.2. Proof of Proposition 2.1 16
   2.3. Bias for Smooth Metric Measure Space with a Density 19

3. Convergence of Coarse Ricci to Actual Ricci on Smooth Manifolds 20

4. Empirical Processes and Convergence 21
   4.1. Estimators of the Carré Du Champ and the Iterated Carré Du Champ in the uniform case 22
   4.2. Glivenko-Cantelli Classes 23
   4.3. Ambient Covering Numbers 25
   4.4. Sample Version of the Carré du Champ 28
   4.5. Subexponential Decay and Almost Sure Convergence 32
   4.6. Proof of Theorem 1.12 34

5. Almost Sure Convergence in the non-Uniform Case 38

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-1161498.
1. INTRODUCTION

In [BN08], Belkin and Niyogi show that the graph Laplacian of a point cloud of data samples taken from a submanifold in Euclidean space converges to the Laplace-Beltrami operator on the underlying manifold. Our goal in this paper is to demonstrate that this process can be continued to approximate Ricci curvature as well. This answers a question of Singer and Wu [SW12, pg. 1103], allowing one to approximate the Hodge Laplacian on 1-forms. The Hodge Laplacian allows one to extract certain topological information, thus we expect our result to have applications to manifold learning.

To this end, we define a notion of coarse Ricci curvature. Coarse Ricci curvature is a quantity defined on pairs of points rather than tangent vectors, thus it can be defined on any metric measure space. We define a family of coarse Ricci curvature operators, which depend on a scale parameter $t$. We show that when taken on a smooth manifold embedded in Euclidean space, these operators converge to the Ricci curvature as $t \to 0$.

The approach is grounded in the Bakry-Emery $\Gamma_2$ calculus. In particular, we appeal to the Bochner formula

$$\Gamma_2(f, f) = \text{Ric}(\nabla f, \nabla f) + \|D^2 f\|^2.$$  

(1.1)

Iterating the approximate Laplacian operators in [BN08] one can construct an approximate $\Gamma_2$ operator, and test this operator on a set of “linear” functions. This defines a coarse Ricci curvature on any two points from a submanifold. This approach recovers the $\text{Ric}_\infty$ tensor and can be modified to recover the standard Ricci curvature as well, provided the volume density is smooth.

We show two types of results. First, the coarse Ricci operators converge as $t \to 0$ when taken on a fixed smooth submanifold. Secondly, there exists an explicit choice of scales $t_n \to 0$ such that the quantities converge almost surely when computed from a set of $n$ points sampled from a smooth probability distribution on the manifold.

1.1. Background and Motivation. The motivation for the paper stems from both the theory of Ricci lower bounds on metric measure spaces and the theory of manifold learning.

1.1.1. Ricci Curvature Lower Bounds on Metric Measure Spaces. One of our main sources of intuition for understanding Ricci curvature is the problem of reinterpreting lower bounds on the Ricci curvature in such a way that it becomes stable under Gromov-Hausdorff limits and thus defining “weak” notion of lower bounds on Ricci curvature. This theory has undergone significant development over the last 10-15 years. A major achievement is the work of Lott-Villani [LV09] and Sturm [Stu06a, Stu06b], which defines Ricci curvature lower bounds on metric measure spaces. These
definitions work quite beautifully provided the metric space is also a length space, but fail to be useful on discrete spaces. The underlying calculus for this theory lies in optimal transport: Given a metric measure space \((X, d, \mu)\) one can consider the space \(\mathcal{B}\) of all Borel probability measures with the 2-Wasserstein distance, which we will denote by \(W_2\), and is given by

\[
W_2(\mu_1, \mu_2) = \inf_{\gamma \in \Pi} \int_{X \times X} d(x, y) \, d\gamma(x, y),
\]

where \(\Pi\) is the set of all probability measures in \(\mathcal{B}\) whose marginals are the measures \(\mu_1\) and \(\mu_2\), i.e. if \(P_i : X \times X \to X\) for \(i = 1, 2\) are the projections onto the first and second factors respectively, then \(\mu_i = (P_i)_* \gamma\) (the push-forward of \(\gamma\) by \(P_i\)). One of the crucial ideas in the work of Lott-Villani and Sturm is to use tools from optimal transport to define a notion of convexity with respect to 2-Wasserstein geodesics (i.e., geodesics with respect to the Wasserstein distance \(W_2\)) such that lower bounds on the Ricci curvature are equivalent to the geodesic convexity of well-chosen functionals. Further, they show that this convexity property is stable under Gromov-Hausdorff limits. In particular, they consider the functional given by the entropy of a measure \(\mu\), given a background measure \(\nu\), defined by

\[
E(\mu|\nu) = \int \frac{d\mu}{d\nu} \log \left( \frac{d\mu}{d\nu} \right) \, d\nu,
\]

and define a space to have non-negative Ricci curvature if the entropy \((1.3)\) is convex along Wasserstein geodesics. This convexity property along Wasserstein geodesics, or displacement convexity, used by Lott-Villani was in turn inspired by the work of Robert McCann in [McC97]. Before the work of Lott-Villani and Sturm, it was well known that lower bounds on sectional curvature are stable under Gromov-Hausdorff; For more on the theory of Alexandrov spaces, see for example [BBI01].

The work of Lott-Villani and Sturm provides a precise definition of lower bounds on Ricci Curvature on length spaces that a priori may be very rough. More recently, Aaron Naber in [Nab13] used stochastic analysis on manifolds to characterize two-sided bounds on non-smooth spaces. Naber’s approach also relies heavily on the existence of Lipschitz geodesics. On the other hand, it is easy to show (for example, consider the space with two positively measured points unit distance apart) that the Wasserstein space \(W_2\) for discrete spaces does not admit any Lipschitz geodesics. To overcome the lack of Wasserstein geodesics on discrete space, Bonciocat and Sturm [BS09] introduce the notion of an “\(h\)-rough geodesic,” which behaves like a geodesic up to an error \(h\). Then using optimal transport methods, they are able to define a lower bound on Ricci curvature, which depends on the scale \(h\).

The notion of coarse Ricci curvature was developed to characterize Ricci curvature, including lower bounds, on a more general class of spaces. The first definition of coarse Ricci curvature was proposed by Ollivier [Oll09] as a function on pairs of points in a metric measure space. Heuristically, two points should have positive coarse Ricci curvature if geodesics balls near the points are “closer” to each other than the points themselves. One possible precise definition of “closer” is given by \((1.2)\). Comparing
the distance between two points to the distance of normalized unit balls is useful geometric information. It is also natural to consider the heat kernel for some small positive time around each given point. In fact, the intuition becomes blatant in the face of [RS05, Cor. 1.4 (x)], which states that the distance \((L^2)\) decays at an exponential rate given by the Ricci lower bound, as mass spreads out from two points via Brownian motion.

Even earlier, it was shown by Jordan, Kinderlehrer and Otto [JKO98] that the gradient flow of the entropy function \((L^3)\) on \(W_2\)-space agrees with the heat flow on \(L^2\). On nice metric measure spaces, one expects this flow to converge to an invariant measure, the same one from which the entropy was originally defined. Thus in principle, the behavior of the entropy functional \((L^3)\) and the generator of heat flow are fundamentally related. This deep relationship is explored in generality in the recent paper [AGS]. It is natural then to attempt to define Ricci curvature in terms of a Markov process. In fact, Ollivier’s idea was to compare the distance between two points to the distance between the point masses after one step in the Markov process. This is also essentially the idea in Lin-Lu-Yau [LLY11], where a lower bound on Ricci curvature of graphs is defined. In contrast, a \(\Gamma_2\) approach was used by Lin-Yau to define Ricci curvature lower bounds on graphs in [LY10].

Recently, Erbas and Maas [EM12] and Mielke [Mie13] have provided a very natural way to define Ricci curvature for arbitrary Markov chains. This involves creating a new Wasserstein space, called the discrete transportation metric, in which one can implement the idea of Lott-Sturm-Villani, i.e., relating lower bounds on Ricci curvature to geodesic convexity of certain entropy functional. Gigli and Maas [GM13] show that in the limit these transportation metrics converge to the Wasserstein space of the manifold in question as the mesh size goes to zero, at least on the torus. Thus this notion recovers the Ricci curvature in the limit.

Any discussion of lower bounds on Ricci curvature would not be complete without a discussion of isoperimetric inequalities, most notably the log-Sobolev inequality. These inequalities, which hold for positive Ricci lower bounds, were generalized by Lott-Villani. In the coarse Ricci setting, log-Sobolev inequalities have been proved by Ollivier, Lin-Yau and Maas for their respective definitions of Ricci curvature lower bounds. We will formulate a log-Sobolev inequality, which holds at a scale \(t\) as well (see section \(L^3\)). This statement requires a standard Bakry-Emery type condition, which is related, but not precisely, a statement about coarse Ricci curvature.

For further introduction to concepts of coarse Ricci curvature see the survey of Ollivier [Oll10].

1.1.2. The Manifold Learning Problem. Roughly speaking, the manifold learning problem deals with inferring or predicting geometric information from a manifold if one is only given a point cloud on the manifold, i.e., a sample of points drawn from the manifold at random according to a certain distribution, without any further information. It is clear then that the manifold learning problem is highly relevant to machine learning and to theory of pattern recognition. An example of an object related to the geometry of an embedded submanifold \(\Sigma\) of Euclidean space that one can “learn”
or estimate from a point cloud is the \emph{rough Laplacian} or Laplace-Beltrami operator. Given an embedding \( F : \Sigma \rightarrow \mathbb{R}^N \) consider its induced metric \( g \). By the rough Laplacian of \( g \) we mean the operator defined on functions by \( \Delta_g f = g^{ij} \nabla_i \nabla_j f \) where \( \nabla \) is the Levi-Civita connection of \( g \). Belkin and Niyogi showed in [BN08] that given a uniformly distributed point cloud on \( \Sigma \) there is a 1-parameter family of operators \( L_t \), which converge to the Laplace-Beltrami operator \( \Delta_g \) on the submanifold. More precisely, the construction of the operators \( L_t \) is based on an approximation of the heat kernel of \( \Delta_g \), and in particular the parameter \( t \) can be interpreted as a choice of scale. In order to learn the rough Laplacian \( \Delta_g \) 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.) one can choose a scale \( t_n \) in such a way that the operators \( L_{t_n} \) converge almost surely to the rough Laplacian \( \Delta_g \). This step follows essentially from applying a quantitative version of the \emph{law of large numbers}. Thus one can almost surely learn spectral properties of a manifold. While in [BN08] it is assumed that the sample is uniform, it was proved by Coifman and Lafon in [CL06] that if one assumes more generally that the distribution of the data points has a smooth, strictly positive density in \( \Sigma \), then it is possible to normalize the operators \( L_t \) in [BN08] to recover the rough Laplacian. More generally, the results in [CL06] and [SW12] 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 \( \rho \) is a smooth function. Since then, Singer and Wu have developed methods for learning the rough Laplacian of an embedded submanifold on 1-forms using \emph{Vector Diffusion Maps} (VDM) (see for example [SW12]). The relationship of Ricci curvature to the Hodge Laplacian on 1-forms is given by the Weitzenböck formula.

In this paper we consider the problem of learning the Ricci curvature of an embedded submanifold \( \Sigma \) of \( \mathbb{R}^N \) at a point from a point cloud. The idea is to construct a notion of coarse Ricci curvature that will serve as a sample estimator of the actual Ricci curvature of the embedded submanifold \( \Sigma \). In order to explain our results we provide more background in the next section.

1.2. Carré Du Champ and Bochner Formula. Bakry and Emery [BE85] introduced the notion of \emph{Carré du Champ} which we now recall. Let \( P_t \) be a 1-parameter family of operators of the form

\[
P_t f(x) = \int_M f(y) p_t(x, dy),
\]

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.

\[
P_{t+s} = P_t \circ P_s.
\]

\[
P_0 = \text{Id}.
\]
In $\mathbb{R}^n$, an example of $P_t$ is the Brownian motion, defined by the density
\begin{equation}
 p_t(x, dy) = \frac{1}{(2\pi t)^{n/2}} e^{-\frac{|x-y|^2}{2t}} dy, t \geq 0.
\end{equation}

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
\begin{equation}
 Lf = \lim_{t \to 0} t^{-1}(P_tf - f),
\end{equation}

We consider a bilinear form which has been introduced in potential theory by J.P. Roth [Rot74] and by Kunita in probability theory [Kun69] and measures the failure of $L$ from satisfying the Leibnitz rule. This bilinear form is defined as
\begin{equation}
 \Gamma(L, u, v) = \frac{1}{2} \left( L(uv) - L(u)v - uL(v) \right).
\end{equation}

When $L$ is the rough Laplacian with respect to the metric $g$, then
\begin{equation}
 \Gamma(\Delta_g, u, v) = \langle \nabla u, \nabla v \rangle.
\end{equation}

We will also consider the iterated Carré du Champ introduced by Bakry and Emery denoted by $\Gamma_2$ and defined by
\begin{equation}
 \Gamma_2(L, u, v) = \frac{1}{2} \left( L(\Gamma(L, u, v)) - \Gamma(L, Lu, v) - \Gamma(L, u, Lv) \right).
\end{equation}

**Notation 1.1.** When considering the operators (1.9) and (1.11) we will use the slightly cumbersome three-parameter notation, as the main results will be stated in terms of a family of operators $\{L_t\}$.

Note that if we restrict our attention to the case $L = \Delta_g$ the Bochner formula yields
\begin{equation}
 \Gamma_2(\Delta_g, u, v) = \frac{1}{2} \Delta \langle \nabla u, \nabla v \rangle - \frac{1}{2} \langle \nabla \Delta_g u, \nabla v \rangle - \frac{1}{2} \langle \nabla u, \nabla \Delta_g v \rangle
\end{equation}
\begin{equation}
 = \text{Ric}(\nabla u, \nabla v) + \langle \text{Hess}_u, \text{Hess}_v \rangle.
\end{equation}

We observe immediately that if $\nabla u = e_i$ and $\text{Hess}_u = 0$ that one can recover the Ricci tensor via
\begin{equation}
 \Gamma_2(\Delta_g, u, u) = \text{Ric}(e_i, e_i).
\end{equation}

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 $\Gamma_2$. With this in mind, they define a curvature-dimension condition for an operator $L$ on a space $X$ 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
\begin{equation}
 \Gamma_2(L, f, f) \geq \frac{1}{N}(Lf)^2 + k\Gamma(L, f, f)
\end{equation}
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 particular, when considering a smooth metric measure space $(M^n, g, e^{-\rho}d\text{vol})$ one has the natural diffusion operator

\begin{equation}
\Delta_{\rho} u = \Delta u - \langle \nabla \rho, \nabla u \rangle,
\end{equation}

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

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

and moreover, they showed the equivalence between the $CD(k, N)$ condition \((1.15)\) and the bound $\text{Ric}_N \geq k$.

The main goal of this paper is to show that one can recover both the Ric and $\text{Ric}_\infty$ for embedded smooth metric measure spaces, by approximating the operators $\Delta_g$ and $\Delta_{\rho}$. We can recover the Ricci curvature of a submanifold of $\mathbb{R}^d$ from (1.14) via the following observation: For any tangent vector $\xi$ at a point $x$, the function

\begin{equation}
f_{\xi}(z) = \langle \xi, z \rangle
\end{equation}

has vanishing Hessian at $x$ when restricted to the submanifold, and has gradient

\begin{equation}
\nabla f_{\xi}(z)|_{z=x} = \xi,
\end{equation}

thus (1.14) yields that

\begin{equation}
\Gamma_2(\Delta_g, f_{\xi}, f_{\xi})(x) = \text{Ric}(\xi, \xi).
\end{equation}

1.3. Iterated Carré du Champ and Coarse Ricci Curvature. In this section we provide a definition of coarse Ricci curvature on general metric measure spaces, using a family of operators which are intended to approximate a Laplace operator on a 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 at a given scale $t$. 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.

As noted above, to recover the Ricci curvature on a manifold, we need to evaluate $\Gamma_2$ on functions with vanishing Hessian at a point. In this section, we do not assume that the space $X$ is Euclidean or has a tangent space. On a general metric measure space, where a tangent space may not be available, the test functions should distinguish different directions. Since we do not have tangent vectors, we replace the notion of tangent vectors with the notion of direction pointing from one point towards another. The coarse Ricci curvature will then be defined on pairs of points. For submanifolds
in Euclidean space, the obvious choice is the linear function whose gradient is the vector which points from a point \( x \) to a point \( y \). To define the coarse Ricci curvature on a general space \( X \) we need to generalize this idea to some sort of locally almost linear function. We propose the following function, which serves our purposes. Given \( x, y \in X \) define

\[
 f_{x,y}(z) = \frac{1}{2} \frac{d^2(y, z) - d^2(x, z) - d^2(x, y)}{d(x, y)}.
\]

One can check that, in the Euclidean case, this is simply the linear function with unit gradient pointing from \( x \) to \( y \) and that, for \( x \) very near \( y \) on a Riemannian manifold, this is (up to high order) the coordinate function in normal coordinates at \( x \) determined by the normal coordinates of \( y \). This leads us to the following definition of coarse Ricci curvature.

**Definition 1.2.** Given an operator \( L \) we define the coarse Ricci curvature for \( L \) as

\[
 \text{Ric}_L(x, y) = \Gamma_2(L, f_{x,y}, f_{x,y})(x).
\]

To see this is consistent with the classical notions, note that this defines a coarse Ricci curvature on a Riemannian manifold as

\[
 \text{Ric}_{\Delta_g}(x, y)(x) = \Gamma_2(\Delta_g, f_{x,y}, f_{x,y})(x).
\]

In Proposition 1.8 we will see how (1.23) recovers the Ricci curvature on the manifold.

**Remark 1.3.** \( \text{Ric}_L(x, y) \) need not be symmetric.

1.3.1. **Approximations of the Laplacian, Carré du Champ and its iterate.** We now construct operators which can be thought of as approximations of the Laplacian on metric measure spaces. This construction is reminiscent of the approximation constructed by Belkin-Niyogi in [BN08] and more generally Coifman-Lafon in [CL06].

Consider a metric measure space \((X, d, \mu)\) with the Borel \(\sigma\)-algebra such that \(\mu(X) < \infty\). Given \( t > 0 \), let \( \theta_t \) be given by

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

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

\[
 L_tf(x) = \frac{2}{t \theta_t(x)} \int_X (f(y) - f(x)) e^{-\frac{d^2(x, y)}{2t}} d\mu(y).
\]

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

\[
 \Gamma(L_t, f, h) = \frac{1}{2} (L_t(fh) - (L_tf)h - f(L_th)),
\]

which simplifies to

\[
 \Gamma(L_t, f, h)(x) = \frac{1}{t \theta_t(x)} \int_X e^{-\frac{d^2(x, y)}{2t}} (f(y) - f(x))(h(y) - h(x)) d\mu.
\]
In a similar fashion we define the iterated Carré du Champ of \( L_t \) to be
\[
\Gamma_2(L_t, f, h) = \frac{1}{2} (L_t(\Gamma(L_t, f, h)) - \Gamma(L_t, L_t f, h) - \Gamma(L_t, f, L_t h)).
\]

**Remark 1.4.** This definition of \( L_t \) differs from Belkin-Niyogi operator in that we normalize by \( \theta_t(x) \) instead of \((2\pi t)^{d/2}\) for an assumed manifold dimension \( d \).

1.3.2. **Empirical Coarse Ricci Curvature 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
\[
\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\xi_i}
\]
(1.29)
where \( \delta_{\xi_i} \) is the atomic point measure at the point \( \xi_i \) (also called \( \delta \)-mass). For any function \( f : \rightarrow \mathbb{R} \) we will use the notation
\[
\mu_n f = \int_X f(y) d\mu_n(y) = \frac{1}{n} \sum_{j=1}^{n} f(\xi_j).
\]
(1.30)

**Notation 1.5.** We will use the “hat” notation (for example \( \hat{L}_t \)) to distinguish those operators, measures, or \( t \)-densities which have been constructed from a sample of finite points.

To be more precise, we define the operator \( \hat{L}_t \) as
\[
\hat{L}_t f(x) = \frac{2}{\theta_t(x)} \int_X (f(y) - f(x)) e^{-\frac{d^2(y,x)}{2t}} d\mu_n(y),
\]
(1.31)
where
\[
\hat{\theta}_t(x) = \int_X e^{-\frac{d^2(y,x)}{2t}} d\mu_n(y) = \frac{1}{n} \sum_{j=1}^{n} e^{-\frac{d(\xi_j,x)^2}{2t}},
\]
(1.32)
and of course
\[
\int_X (f(y) - f(x)) e^{-\frac{d^2(y,x)}{2t}} d\mu_n(y) = \frac{1}{n} \sum_{j=1}^{n} e^{-\frac{d(\xi_j,x)^2}{2t}} (f(\xi_j) - f(x)).
\]
(1.33)

The sample version of Carré du Champ will be the bilinear form \( \Gamma(\hat{L}_t, f, h) \) which from (1.27) takes the form
\[
\Gamma(\hat{L}_t, f, h)(x) = \frac{1}{\theta_t(x)} \frac{1}{n} \sum_{j=1}^{n} e^{-\frac{d^2(\xi_j,x)}{2t}} (f(\xi_j) - f(x))(h(\xi_j) - h(x)),
\]
(1.34)
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
\[
\Gamma_2(\hat{L}_t, f, h) = \frac{1}{2} \left( \hat{L}_t(\Gamma(\hat{L}_t)f, h) - \Gamma(\hat{L}_t, \hat{L}_t f, h) - \Gamma(\hat{L}_t, f, \hat{L}_t h) \right).
\]
(1.35)
Our notion of empirical coarse Ricci curvature is simply

**Definition 1.6.** Let be a sample of points of $X$. The empirical Ricci curvature of $(X,d,\mu)$ at a scale $t$ with respect to a sample $\{\xi_1, \ldots, \xi_n\}$ at $(x,y) \in X \times X$, which will be denoted by $\hat{\text{Ric}}(x,y)$ is

$$\hat{\text{Ric}}(x,y) = \Gamma_2(\hat{L}_t, f_{x,y}, f_{x,y}).$$

(1.36)

**Remark 1.7.** Observe that (1.36) does not make use of the measure $\mu$. We will discuss the role played by the measure when we discuss applications to Manifold Learning in Section 1.4.2.

We are now in position to state our main results.

1.4. **Statement of Results.**

1.4.1. *Convergence of Coarse Ricci Curvature on Smooth Manifolds.* We start by showing that $\text{Ric}_{\triangle g}$ has the following expected property of coarse Ricci curvature on smooth manifolds.

**Proposition 1.8.** Suppose that $M$ is a smooth Riemannian manifold. Let $V \in T_xM$ with $g(V,V) = 1$. Then

$$\text{Ric}(V,V) = \lim_{\lambda \to 0} \text{Ric}_{\triangle g}(x,\exp_x(\lambda V)).$$

(1.37)

The proof of Proposition 1.8 will be given in Section 3 and follows from a classical result of Synge in [Syn31], which provides an expansion for the square of the geodesic distance at a point in normal coordinates.

1.4.2. *Applications to Manifold Learning.* We now show how our notions of coarse Ricci curvature and empirical coarse Ricci curvature at a scale have applications to the Manifold Learning Problem. For the rest of subsection 1.4.2 we will consider a closed embedded submanifold $\Sigma$ of $\mathbb{R}^N$, and the metric measure space will be $(\Sigma, \|\cdot\|, d\text{vol})$, where

- $\|\cdot\|$ is the distance function in the ambient space $\mathbb{R}^N$,
- $d\text{vol}_\Sigma$ is the volume element corresponding to the metric $g$ induced by the embedding of $\Sigma$ 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 $d\text{vol}_\Sigma$.
- 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 $\Sigma$ 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 with intrinsic geodesic distance. As pointed out by Belkin-Niyogi [BN08, Lemma 4.3], the chordal and intrinsic distance functions on a submanifold differ to 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 are able to show the following.

**Theorem 1.9.** Let $\Sigma^d \subset \mathbb{R}^N$ be a closed embedded submanifold, let $g$ be the Riemannian metric induced by the embedding, and let $(\Sigma, \| \cdot \|, d\text{vol}_\Sigma)$ be the metric measure space defined with respect to the ambient distance. Then there exists a constant $C_1$ depending on the geometry of $\Sigma$ and the function $f$ such that

$$\sup_{x \in \Sigma} | \Gamma_2(\Delta_g, f, f)(x) - \Gamma_2(L_t, f, f)(x) | \leq C_1(\Sigma, D^5f) t^{1/2}. \quad (1.38)$$

Theorem 1.9 will follow from Corollary 2.2 which is proved in Section 2.

**Corollary 1.10.** With the hypotheses of Theorem 1.9 we have

$$\text{Ric}_{\Delta_g}(x, y) = \lim_{t \to 0} \Gamma_2(L_t, f_x, y, f_y)(x). \quad (1.39)$$

**Theorem 1.11.** Let $\Sigma^d \subset \mathbb{R}^N$ be a closed embedded submanifold, and let $g$ be the metric induced by the embedding. Let $\gamma(s)$ be a unit speed geodesic in $\Sigma$ such that $\gamma(0) = x$. There exists constants $C_2, C_3$ depending on the geometry of $\Sigma$ such that

$$| \text{Ric}(\gamma'(0), \gamma'(0)) - \text{Ric}_{L_t}(x, \gamma(s)) | \leq C_2 t^{1/2} + C_3 s. \quad (1.40)$$

In principle, Corollary 1.10 combined with Proposition 1.8 give us a way to approximate the Ricci curvature.

Now we address the problem of choosing a scale depending on the size of the data and the dimension of the submanifold $\Sigma$, such that the sequence of empirical Ricci curvatures corresponding to the size of the data converge almost surely to the actual Ricci curvature of $\Sigma$ at a point. In order to simplify the presentation of our results, we start by stating the simplest possible case, which correspond to a uniformly distributed i.i.d. sample $\{\xi_1, \ldots, \xi_n\}$. We will be able to relax this assumption to i.i.d. samples whose distributions have a smooth everywhere positive density.

**Theorem 1.12.** Consider the metric measure space $(\Sigma, \| \cdot \|, d\text{vol}_\Sigma)$ 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 $\Sigma$. For $\sigma > 0$, let

$$t_n = n^{-\frac{1}{3d+1+\frac{1}{\sigma}}} \quad (1.41)$$

Then,

$$\sup_{\xi \in \Sigma} | \hat{\Gamma}_2(L_{t_n}, f, f)(\xi) - \Gamma_2(L_{t_n}, f, f)(\xi) | \xrightarrow{a.s.} 0. \quad (1.42)$$

The proof of Theorem 1.12 even in the uniform case, requires using ideas from the theory of empirical processes. We will provide the necessary background in Section 4. As pointed out in Section 1.1.2, 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 1.12 is proved in Section 4.

Corollary 1.13. Let $\Sigma^d \subset \mathbb{R}^N$ be an embedded submanifold and consider the metric measure space $(\Sigma, \| \cdot \|, \text{dvol}_2)$. Suppose that we have an i.i.d. uniformly distributed sample $\xi_1, \ldots, \xi_n$ drawn from $\Sigma$. Let

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

for any $\sigma > 0$. Then

$$\sup_{x \in \Sigma} \left| \hat{\Gamma}_2(L_{t_n}, f_{x,y}, f_{x,y})(x) - \text{Ric}_{\Delta g}(x, y) \right| \xrightarrow{a.s.} 0.$$ (1.44)

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

Remark 1.14. The convergence is better if $t_n$ is chosen to go to zero slower than in (4.82). In particular, if one replaces $d$ with an upper bound on $d$, then Theorem 1.12 and Corollary 1.13 still hold.

We now state our results for non-uniformly distributed samples and show how we can recover more general objects than the Ricci Curvature, for example the Bakry-Emery tensor if we sample adequately.

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

$$\Delta_\rho u = \Delta_g u - \nabla \rho \cdot \nabla u.$$ (1.45)

In [CL06], the authors consider a family of operators $L_\alpha^t$ which converge to $\Delta(1-\alpha)\rho$. Note that a standard computation (cf [Vil09, Page 384]) gives

$$\Gamma_2(\Delta_2(1-\alpha)\rho, f, f) = \frac{1}{2} \Delta_g \| \nabla f \|^2_g - \langle \nabla \rho, \nabla \Delta_g f \rangle_g + 2(1-\alpha)\nabla^2_g \rho(\nabla f, \nabla f).$$ (1.46)

We adapt [CL06] to our setting: Recall that

$$\theta_t(x) = \int_X e^{-\frac{d^2(x,y)}{2t}} d\mu(y),$$ (1.47)

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

$$\theta_t,\alpha(x) = \int_X e^{-\frac{|x-y|^2}{2t}} \frac{1}{[\theta_t(y)]^\alpha} d\mu(y).$$ (1.48)
We can define the operator
\begin{equation}
L_t^\alpha f(x) = \frac{2}{t \theta_{t,\alpha}(x)} \int e^{-\frac{(y-x)^2}{4t}} \frac{1}{[\theta_t(y)]^{\alpha}} (f(y) - f(x)) \, d\mu(y)
\end{equation}
and again obtain bilinear forms $\Gamma(L_t^\alpha, f, f)$ and $\Gamma_2(L_t^\alpha, f, f)$. For the rest of the section we will consider the metric measure space $(\Sigma, \| \cdot \|, e^{-\rho(d\mu)}$ where $\Sigma^d \subset \mathbb{R}^N$ is an embedded submanifold, $\| \cdot \|$ is the ambient distance and $\rho$ is a smooth function in $\Sigma$. We again take all the operators $L_t, \Gamma_t(L_t, \cdot, \cdot)$ and $\Gamma_2(L_t, \cdot, \cdot)$ and their sample counterparts $\hat{L}_t, \Gamma_t(\hat{L}_t, \cdot, \cdot)$ and $\Gamma_2(\hat{L}_t, \cdot, \cdot)$ with respect to the data of $(\Sigma, \| \cdot \|, e^{-\rho(d\mu)})$.

**Theorem 1.15.** Let $\Sigma^d \subset \mathbb{R}^N$ be an embedded submanifold and consider the smooth metric measure space $(\Sigma, \| \cdot \|, e^{-\rho(d\mu)})$. Let $f \in C^5(\Sigma)$ such that $\|f\|_{C^5} \leq M$. There exists $C_4 = C_4(\Sigma, M, \rho)$ such that
\begin{equation}
\sup_{\xi \in \Sigma} |\Gamma_2(L_1^t, f, f)(\xi) - \Gamma_2(\triangle_{2(1-\alpha)^\rho}, f, f)(\xi)| \leq C_4 t^{1/2}.
\end{equation}

**Corollary 1.16.** With the hypotheses of Theorem 1.15, if in addition we assume that $\|D^2f\| (\xi) = 0$ and choose $\alpha = 1/2$ we have for $C_5 = C_5(\Sigma, M, \rho)$
\begin{equation}
|\text{Ric}_\infty(\nabla f, \nabla f)(\xi) - \Gamma_2(L_1^{1/2}, f, f)(\xi)| \leq C_5 t^{1/2}.
\end{equation}

**Corollary 1.17.** With the hypotheses of Theorem 1.15, if in addition we assume that $\|D^2f\| (\xi) = 0$ and choose $\alpha = 1$, we have for $C_6 = C_6(\Sigma, M, \rho)$
\begin{equation}
|\text{Ric}(\nabla f, \nabla f)(\xi) - \Gamma_2(L_1^t, f, f)(\xi)| \leq C_6 t^{1/2}.
\end{equation}

In particular, if the density is positive and smooth enough, we can still recover the Ricci tensor.

Finally, adapting (1.31) to the modified operators we have

**Theorem 1.18.** Let $\Sigma \subset \mathbb{R}^N$ be a closed embedded submanifold of $\mathbb{R}^N$ and consider the metric measure space $(\Sigma, \| \cdot \|, e^{-\rho(d\mu)})$. Let $\xi_1, \ldots, \xi_n, \ldots$ be an i.i.d. sample whose distribution has density precisely $e^{-\rho(d\mu)}$, and let $[\hat{\Gamma}_t]_2$ be the sample version of $[\Gamma_t]_2$. Choosing $t_n = n^{-\gamma}$ with $\gamma = \frac{1}{4d + 4 + \sigma}$, where $\sigma$ is a positive number, we have
\begin{equation}
\sup_{\xi \in \Sigma} \left| \hat{\Gamma}_2(L_{t_n}^\alpha, f, f)(\xi) - \Gamma_2(L_{t_n}^\alpha, f, f)(\xi) \right| \xrightarrow{a.s.} 0.
\end{equation}

The proof of Theorem 1.18 will be given in Section (5).

**Corollary 1.19.** Let $\Sigma^d \subset \mathbb{R}^N$ be an embedded submanifold and consider the metric measure space $(\Sigma, \| \cdot \|, e^{-\rho(d\mu)})$. Suppose that we have an i.i.d. sample $\{\xi_1, \ldots, \xi_n\}$ whose distribution has density $e^{-\rho(d\mu)}$. Letting
\begin{equation}
t_n = n^{-\frac{1}{4d + 4 + \sigma}},
\end{equation}

for $\sigma > 0$, it follows that $\Gamma_2(L_{t_n}^\alpha, f_{x,y}, f_{x,y})(x)$ converges almost surely to the coarse Ricci curvature
\begin{equation}
Ric_{\Delta_2(1-\alpha)}(x,y).
\end{equation}

1.5. **Final Remarks.** At this point we can make comparisons to other definitions of coarse Ricci curvature. Lin and Yau [LY10], following Chung and Yau [CY96], consider curvature dimension lower bounds of the form (1.15), in the particular case where $L$ is the graph Laplacian using the standard distance function on graphs. This class of metric spaces is quite restricted - in their setting they are able to show that every locally finite graph satisfies a $CD(2, -1)$ condition. The approach by Lin, Lu and Yau in [LLY11] follows the ideas of Ollivier [Oll09], using an optimal transport distance like (1.2), using distance instead of squared distance. With this metric, they compare mass distributions after short diffusion times, and take the limit as the diffusion time approaches zero. They are able to show a Bonnet-Myers type theorem which holds for graphs with positive Ricci curvature. While the Bonnet-Myers result holds in the classical Riemannian setting, it fails for general notions of Ricci curvature, for example, when the Ricci curvature is derived from the standard OrnsteinUhlenbeck process. Ollivier’s definition [Oll09, Definition 3] is very general, and also uses the $W_1$ Wasserstein metric. Ollivier uses $\varepsilon$-geodesics to obtain local-to-global results. Note that the notion of $\varepsilon$-geodesics is stronger than $h$-rough geodesics seen in [BS09].

For emphasis, we summarize that one of the main advantages of our approach is that it uses only integration, and does not require any minimization. In particular, we do not use optimal transport or any notion of geodesics. Further, our approach lends itself to making explicit bounds on the convergence given points randomly selected according to a distribution.

From the theory of Gross and Bakry-Emery, one can show that whenever a heat group together with the invariant measure satisfies standard properties (self adjointness, ergodicity, Leibnitz and product rules) a Bakry-Emery condition of the form
\begin{equation}
\Gamma_2(f,f) \geq K \Gamma(f,f)
\end{equation}
implies a log-Sobolev inequality of the form
\begin{equation}
\int f \log f \, d\mu \leq \frac{1}{2K} \int \frac{\Gamma(f,f)}{f} \, d\mu
\end{equation}
for all $f > 0f$ with $\int f \, d\mu = 1$. In particular, these conditions do hold for any of the approximate operators $L_t$. Thus the log-Sobolev inequality appropriate to our situation takes the following formulation. We omit the proof.

**Theorem 1.20.** Suppose that $\Sigma$ is an embedded submanifold of Euclidean space and $d\mu$ is a measure on $\Sigma$. Suppose that
\begin{equation}
\Gamma_2(L_t, f, f) \geq K \Gamma(L_t, f, f)
\end{equation}
for all $f \in C^4(\Sigma)$. Then for all $f > 0$ with $\int f d\mu = 1$, we have

\begin{equation}
\int \Sigma f \log f d\mu \leq \frac{1}{2Kt} \int \int \frac{1}{\theta_t(x)} e^{-\frac{||x-y||^2}{4t}} \frac{1}{f(x)} (f(y) - f(x))^2 d\mu(y) d\mu(x).
\end{equation}

Note that the condition (1.58) is a priori stronger than a coarse Ricci lower bound of the form

\begin{equation}
\text{Ric}_{L_t}(x, y) = \Gamma_2(L_t, f_{x,y}, f_{x,y}) \geq K
\end{equation}

as the latter only tests the condition (1.58) on a subset of functions and at certain points. While the conditions (1.58) (1.60) are equivalent in the infinitesimal case, it is unclear to us if the weaker condition (1.60) implies any meaningful isoperimetric inequalities in the general case.

Our results show that one can give a definition of coarse 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 [FMN13] 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 estimate the dimension of a submanifold from a sample of points.

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 Palais-Palais-Karcher in [KPP14], specifically on methods for generating point clouds on implicit surfaces using Monte Carlo simulation and the Cauchy-Crofton formula.

1.6. Organization of the Paper. In section 2 we prove Theorems 1.9 and 1.15. In Section 3 we show that on smooth manifolds our definition of coarse Ricci curvature at a scale converges to the actual Ricci curvature. In Section 4 we provide the necessary background on the theory of empirical processes for proving Theorems 1.12 and its corollaries. Finally, in Section 5 we prove Theorem 1.18.

1.7. 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 his attention. The second author would like to thank Jan Maas for useful conversation, and Matthew Kahle for stoking his interest in the topic.
2. Bias Error Estimates

2.1. Bias for Submanifold of Euclidean Space. In this section we prove Theorem 1.9. The theorem will follow from Proposition 2.1 and Corollary 2.2 below. For simplicity we will assume that \((\Sigma, d\text{vol}_\Sigma)\) has unit volume. Recall the definitions (1.21), (1.25), (1.26), (1.27) and (1.28).

Proposition 2.1. Suppose that \(\Sigma^d\) is a closed, embedded, unit volume submanifold of \(\mathbb{R}^N\). For any \(x \in \Sigma\) and for any functions \(f, h \in C^5(\Sigma)\) we have

\[
(2.1) \quad \frac{(2\pi t)^{d/2}}{\theta_t(x)} = 1 + tG_1(x) + t^{3/2}R_1(x),
\]

\[
(2.2) \quad \Gamma_t(f, h)(x) = (\nabla f(x), \nabla h(x)) + t^{1/2}G_2(x, J^2(f)(x), J^2(h)(x))
\]

\[
(2.3) \quad + tG_3(x, J^3(f)(x), J^3(h)(x)) + t^{3/2}R_2(x, J^4(f)(x), J^4(h)(x)),
\]

\[
(2.4) \quad L_t f(x) = \Delta_g f(x) + t^{1/2}G_4(x, J^3(f)(x)) + tG_5(x, J^4(f)(x)) + t^{3/2}R_3(x, J^5(f)(x)),
\]

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 bounded in terms of its arguments.

Corollary 2.2. We have the following expansions

\[
(2.5) \quad L_t(\Gamma_t(f, f))(x) = \Delta_g \|\nabla f(x)\|^2_g + t^{1/2}R_4(x, J^5(f)(x)),
\]

\[
(2.6) \quad \Gamma_t(L_t f, f)(x) = (\nabla \Delta_g f(x), \nabla f(x)) + t^{1/2}R(x, J^5(f)(x)),
\]

2.2. Proof of Proposition 2.1. Our first goal is to fix a local structure which we will use to define the quantities \(G_i\) and \(R_i\) that appear in Proposition 2.1. Choose a point \(x \in \Sigma\), and an identification of tangent plane \(T_x \Sigma\) with \(\mathbb{R}^n\). Locally we may make a smooth choice of ordered orthonormal frame nearby points \(\Sigma\) so that at each point \(y\) there now is a fixed identification of the tangent plane. At each nearby point \(y \in \Sigma\), we can represent \(\Sigma\) as the graph of a function \(U_y\) over the tangent plane \(T_y \Sigma\). Each \(U_y\) will satisfy

\[
(2.7) \quad U_y(0) = 0,
\]

\[
(2.8) \quad DU_y(0) = 0.
\]

By our choice of identification, the functions \(U_y\) are well defined and for \(y\) near \(x\) and \(z \in T_y \Sigma\) near 0, the function \((y, z) \mapsto U_y(z)\) has the same regularity as \(\Sigma\). Fixing a point \(x\), consider a function \(f\) on \(\Sigma\). The function \(f\) is locally well defined as a function over the tangent plane, i.e.

\[
(2.9) \quad f(y) = f(y, U_x(y)) \quad \text{for} \quad y \in T_x \Sigma.
\]

With the above identification we obtain coordinates on the tangent plane at \(x\), and we may take derivatives of \(f\) in this new coordinate system to define the \(m\)-jet of \(f\).
at the point $x$ by

\[(2.10)\quad J^m f(x) = (f(x), Df(x), \ldots D^m f(x)).\]

More concretely, all derivatives in \((2.10)\) are taken with respect to the variable $y$ in \((2.9)\). Since $\Sigma$ is compact, there exists $\tau_0 > 0$ such that for every $y \in \Sigma$ we have

1. The function $U_y$ is defined and smooth on $B_{\tau_0}(0) \subset T_y \Sigma$,
2. $B_{\mathbb{R}^N, \tau_0}(y) \cap \Sigma$ is contained in the graph of $U_y$ over the ball $B_{\tau_0}(0) \subset T_y \Sigma$ where $B_{\mathbb{R}^N, \tau_0}(y)$ is the ball in $\mathbb{R}^N$ centered at $y$ with respect to the ambient distance.

We will use the following notation: given $y \in \Sigma$ and $\tau_0 > 0$ as above, we let

\[(2.11)\quad \Sigma_{y, \tau_0} = \Sigma \cap \{(z, U_y(z)), y \in B_x(0) \subset T_y \Sigma\},\]

in other words, $\Sigma_{y, \tau_0}$ is the part of $\Sigma$ contained in the graph of $U_y$ on $B_{\tau_0}(0) \subset T_y \Sigma$. Observe that with this notation, the statement in (2) above simply says that

\[(2.12)\quad B_{\mathbb{R}^N, \tau_0}(y) \cap \Sigma \subset \Sigma_{y, \tau_0}.\]

Observe that for any $f \in L^\infty(\Sigma)$ we have

\[(2.13)\quad \int_{\Sigma} f(y) e^{-\frac{\|x-y\|^2}{2t}} d\mu(y) = \int_{\Sigma_{x, \tau_0}} f(y) e^{-\frac{\|x-y\|^2}{2t}} d\mu(y)
+ \int_{\Sigma \setminus \Sigma_{x, \tau_0}} f(y) e^{-\frac{\|x-y\|^2}{2t}} d\mu(y),\]

and by (2)

\[(2.14)\quad \int_{\Sigma \setminus \Sigma_{x, \tau_0}(x)} f(y) e^{-\frac{\|x-y\|^2}{2t}} d\mu(y) \leq \|f\|_{L^\infty(\Sigma)} e^{-\frac{\|x\|^2}{2t}},\]

Note also, that for any polynomial $p(z)$, there is a constant $C$ such that on

\[(2.15)\quad \int_{\mathbb{R}^d \setminus B_{\tau_0}/\sqrt{t}} e^{-\frac{\|z\|^2}{2t}} p(z) dz \leq C(p) e^{-\frac{\|z\|^2}{2t}}.\]

The volume form over $T_x \Sigma$ will be

\[(2.16)\quad \mu_x(z) dz = \sqrt{\det(\delta_{ij} + \langle D_i U_x(z), D_j U_x(z)\rangle)}dz.\]

If in \((1.24)\) we choose our distance to be the ambient distance $\| \cdot \|$ in $\mathbb{R}^N$ and the measure $\mu$ to be the volume measure in $\Sigma$, the density $\theta_t(x)$ takes the form

\[(2.17)\quad \theta_t(x) = \int_{\Sigma} e^{-\frac{\|x-y\|^2}{2t}} d\mu(y).\]

In the following, we will use $T_k f(x)(y)$ to denote the $k$-th order term in the Taylor expansion of $f$ at $x$, in the variable $y$. 


We now prove (2.1). Observe that
\[
\theta_t(x) - \int_{\Sigma \setminus \Sigma_{x, r_0}} e^{-\frac{\|x-y\|^2}{2t}} d\mu(y).
\]
\[
= \int_{B_{r_0}} e^{-\frac{\|z\|^2 + \|U_x(z)\|^2}{2t}} \mu_x(z) dz
\]
\[
= t^{n/2} \int_{B_{r_0}/\sqrt{t}} e^{-\frac{\|w\|^2}{2}} e^{-\frac{\|U_x(\sqrt{t}w)\|^2}{2t}} \mu_x(\sqrt{t}w) dw
\]
\[
= t^{n/2} \int_{B_{r_0}/\sqrt{t}} e^{-\frac{\|w\|^2}{2}} e^{-\frac{\|U_x(\sqrt{t}w)\|^2}{2t}} \mu_x(\sqrt{t}w) dw
\]
\[
- t^{n/2} \int_{B_{r_0}/\sqrt{t}} e^{-\frac{\|w\|^2}{2}} e^{-\frac{\|U_x(\sqrt{t}w)\|^2}{2t}} \mu_x(\sqrt{t}w) dw
\]

Now considering (2.17), (2.8) we have
\[
(2.19) \quad \mu_x(\sqrt{t}w) = 1 + tT_2 \mu_x(0)(w) + t^{3/2} R_2 \mu_x(0, w)
\]
\[
(2.20) \quad e^{-\frac{\|U_x(\sqrt{t}w)\|^2}{2t}} = 1 + tT_4 \left[ e^{-\frac{\|U_x(\sqrt{t})\|^2}{2t}}(0)(w) + t^{3/2} R_4 \left[ e^{-\frac{\|U_x(\sqrt{t})\|^2}{2t}}(0) \right] \right] (0, w).
\]

Expanding, collecting lower order terms, integrating and absorbing the exponentially decaying terms into \(t^{3/2} R_1(x, z)\) using (2.15) and (2.16) yields (2.1).

Next we prove (2.4). First compute
\[
\int_{\Sigma} (f(y) - f(x)) e^{-\frac{\|x-y\|^2}{2t}} d\mu(y).
\]
\[
\int_{B_{r_0}} (f(y) - f(x)) e^{-\frac{\|x-y\|^2}{2t}} \mu_x(y) dy = t^{d/2} \int_{B_{r_0}/\sqrt{t}} e^{-\frac{\|w\|^2}{2}} \left( f(\sqrt{t}w) - f(0) \right) \mu_x(\sqrt{t}w) dw
\]

Now,
\[
(2.21) \quad f(\sqrt{t}z) - f(0) = \sqrt{t}T_1 f(z) + tT_2 f(z) + t^{3/2} T_3 f(z) + t^2 T_4 f(z) + t^{5/2} R_4(z)
\]
and also recall (2.19), (2.20). Note that if \(A\) is a symmetric \(d \times d\) matrix we have the identity
\[
(2.22) \quad \int_{\mathbb{R}^d} z^T A z dz = (2\pi)^{d/2} \text{tr}(A),
\]
where \(\text{tr}\) denotes Trace. From this it follows that
\[
(2.23) \quad \int_{\mathbb{R}^d} e^{-\frac{\|z\|^2}{2}} T_2 f(0)(z) dz = (2\pi)^{d/2} \text{tr}(T_2 f(0)).
\]
Again, expanding, collecting lower order terms, integrating odd and even terms, and absorbing the exponentially decaying terms via (2.15) and (2.16) yields
\[
t^{-d/2} \int_{\Sigma} (f(y) - f(x)) e^{-\frac{\|x-y\|^2}{2t}} d\mu(y) = t (2\pi)^{d/2} \Delta g f + t^{3/2} G_4(J^3 f(x))
\]
\[
+ t^2 G_5(J^4 f(x)) + t^{5/2} R_3(J^5 f(x)).
\]
Combining with (2.1) yields (2.4). A very similar calculation yields (2.2).

2.2.1. **Proof of Corollary 2.2** Directly from (2.2)

(2.24) \[ L_t(\Gamma_t(f, f))(x) = L_t \left\{ \|\nabla f\|^2(x) + t^{1/2}G_2(x, J^2(f)(x)) + tG_3(x, J^3(f)(x)) \right\} \]

This last term can be bounded directly by the definition of \( L_t \):

\[ t^{3/2} |L_t R(x)| = t^{3/2} \left| \frac{2}{t} \right| \int (R(y) - R(x)) e^{-\frac{\|x - y\|^2}{2t}} d\mu(y) \right| \leq \left| \frac{2}{t} \right| \int e^{-\frac{\|x - y\|^2}{2t}} d\mu(y) \right| \]

\[ = 4t^{1/2} \|R\|_{L^\infty}. \]

The first three terms are differentiable, so can be dealt with directly by (2.4), giving an expression involving \( J^5(f)(x) \). The estimate (2.6) follows from a similar argument. The result follows by combining the above lemmata for the first term, and then directly bounding the second term.

2.3. **Bias for Smooth Metric Measure Space with a Density.** The bias estimate Theorem 1.9 for a metric measure space with density will follow from the following proposition whose proof is very similar to that of Proposition 2.1. Recall definitions (1.48) etc.

**Proposition 2.3.** *Let \( f \in C^5 \). We have the following expansions*

(2.26) \[ L_t^\alpha f(x) = \Delta_g f(x) + (1 - \alpha) \langle \nabla f(x), \nabla \rho(x) \rangle_g + t^{1/2}G_1(x, J^3(f)) + t^{3/2}R_1(x, J^5(f)) \]

(2.27) \[ \Gamma_t^\alpha(f, h)(x) = \langle \nabla f, \nabla h \rangle_g + t^{1/2}G_2(x, J^2(f), J^2(h)) + t^{3/2}R_2(x, J^4(f), J^4(h)). \]

**Proof.** Following the proof of Proposition 2.1 we have the following expansions

(2.28) \[ \theta_t(x) = (2\pi t)^{d/2} e^{-\rho(x)} \left( 1 + tG_1(x, \rho) + t^{3/2}R_1(x, \rho) \right), \]

(2.29) \[ \theta_{t, \alpha}(x) = (2\pi t)^{(1-\alpha)d/2} e^{(\alpha-1)\rho(x)} \left( 1 + tG_2(x, \rho)R_2(x, \rho) \right), \]

as \( t \to 0 \). Also, taking coordinates on the tangent plane of \( \Sigma \) at the point \( x \) and identifying \( x \) with 0 we have the expansion

(2.30) \[ \frac{d\mu_x(z)}{\theta_t(z)^\alpha} = e^{(\alpha-1)\rho(0)} \left( 1 + (1 - \alpha)\langle D\rho(0), z \rangle + O(\|z\|^2) \right) dz, \]

which holds in a small neighborhood of 0. The rest of the proposition follows from a straightforward computation.
Corollary 2.4. We have the following expansions

\[
L_t^\alpha (\Gamma^\alpha_t(f, h))(x) = \Delta \langle \nabla f, \nabla h \rangle_g(x) + (1 - \alpha) \langle \nabla \rho, \nabla \langle \nabla f, \nabla h \rangle_g \rangle_g(x) + t^{1/2} R_3(x, J^5(f), J^5(h), J^5(\rho)),
\]

\[
\Gamma^\alpha_t(L_t^\alpha f, h)(x) = \langle \nabla \Delta_g f, \nabla h \rangle_g(x) + (1 - \alpha) \langle \nabla f, \nabla \langle \nabla h \rangle_g \rangle_g(x) + t^{1/2} R_4(x, J^5(f), J^5(h), J^5(\rho)).
\]

as \( t \to 0 \).

From Corollary 2.4 we obtain Theorem 1.15.

3. Convergence of Coarse Ricci to Actual Ricci on Smooth Manifolds

We start by proving Proposition 1.8

Proof of Proposition 1.8. Take normal coordinates and without loss of generality assume that \( V = \partial_1 \). Then in these coordinates

\[
f_{x, \text{exp}_x(\lambda V)}(z) = \frac{d^2(\lambda e_1, z) - \|z\|^2 - \lambda^2}{\lambda} = \langle e_1, z \rangle + G(\lambda, z)
\]

where \( G(\lambda, z) \) is a smooth function which vanishes to fourth order, by a classical result of Synge [Syn31, Section 5]. Applying the Bochner formula and taking the limit gives the result. \( \square \)

We now prove Theorem 1.11

Proof of Theorem 1.11. Our goal is to show that

\[
|Ric(\gamma'(0), \gamma'(0)) - Ric_{L_t}(x, \gamma(s))| \leq C_1 t^{1/2} + C_2 s.
\]

First, note that letting

\[
f_s = f_{x, \gamma(s)}(\cdot) = \left\langle \frac{\gamma(s) - x}{|\gamma(s) - x|}, \cdot \right\rangle
\]

and

\[
f_0 = \langle \gamma'(0), \cdot \rangle
\]

we have

\[
|Ric_{L_t}(x, \gamma(s)) - Ric(\gamma'(0), \gamma'(0))| = |\Gamma_2(L_t, f_s, f_s)(x) - Ric(\gamma'(0), \gamma'(0))|
\]

\[
= |\Gamma_2(L_t, f_s, f_s)(x) - \Gamma_2(\Delta_g, f_s, f_s)(x) + \Gamma_2(\Delta_g, f_0, f_0)(x) - \Gamma_2(\Delta_g, f_0, f_0)(x) + \Gamma_2(\Delta_g, f_0, f_0)(x) - Ric(\gamma'(0), \gamma'(0))|
\]

\[
\leq R(J^5f_s)t^{1/2} + C_7(\Sigma)s.
\]

Here we have used the following the facts:
First,

\[ |\Gamma_2(L_t, f_s, f_s)(x) - \Gamma_2(\triangle_g, f_s, f_s)(x)| \leq R(x, J^5 f_s) t^{1/2} \]

by Corollary 2.2.

Second, a straightforward computation yields that for any two functions \( f_0, f_s \)

\[
|\Gamma_2(\triangle_g, f_s, f_s)(x) - \Gamma_2(\triangle_g, f_0, f_0)(x)| \leq (\|f_0\|_{C^2} + \|f_s\|_{C^2}) \|f_s - f_0\|_{C^2}
\]

+ \( \|f_0\|_{C^3} \|f_s - f_0\|_{C^1} + \|f_s\|_{C^1} \|f_s - f_0\|_{C^3} \).

Since the functions \( f_0, f_s \) are ambient linear functions restricted to a submanifold, the higher derivatives are well-controlled. The derivatives of the difference are controlled as follows

\[
\|f_s - f_0\|_{C^3} = \left\| \left\langle \frac{\gamma(s) - x}{\|\gamma(s) - x\|} - \gamma'(0), \cdot \right\rangle \right\|_{C^3} \\
\leq \left\| \frac{\gamma(s) - x}{\|\gamma(s) - x\|} - \gamma'(0) \right\| \|z\|_{C^3}
\]

where \( \|z\|_{C^3} \) is the norm of the derivatives of the coordinate functions, which is also controlled by the geometry of \( \Sigma \). Certainly, for any unit speed curve with bounded curvature we have

\[
\left\| \frac{\gamma(s) - x}{\|\gamma(s) - x\|} - \gamma'(0) \right\| \leq \kappa s.
\]

The curvature of any geodesic inside \( \Sigma \) is controlled by the geometry of \( \Sigma \). Finally, because

\[ \nabla f_0 = \gamma'(0) \]

we have by the Bochner formula

\[
\Gamma_2(\triangle_g, f_0, f_0)(x) - Ric(\gamma'(0), \gamma'(0)) = \|D^2 f_0\|.
\]

Using the tangent plane as coordinates at a point, it is easy to compute that the Hessian of any coordinate function vanishes at the origin. The vector \( \gamma'(0) \) is in the tangent space, so we conclude that (3.7) vanishes.

\[ \square \]

4. EMPirical PROCesSES AND CONVERGENCE

The goal of this section is to prove Theorem 1.12. 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 [vdVW96]. See also [SW13] 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 $\mu$ is the volume measure $d\text{vol}_\Sigma$. Recall that our formal definition of the Carré du Champ of $L_t$ with respect to the uniform distribution is given by

\begin{equation}
\Gamma(L_t, f, h) = \frac{1}{t} \frac{1}{\theta_t(x)} \left( \int \int e^{-\frac{||x-y||^2}{2t}} (f(y) - f(x)) (h(y) - h(x)) d\mu(y) \right).
\end{equation}

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

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

and recall that we defined the $t$-Laplace operator by

\begin{equation}
L_t f(x) = \frac{2}{t} \frac{1}{\theta_t(x)} \int e^{-\frac{||x-y||^2}{2t}} (f(y) - f(x)) d\mu(y),
\end{equation}

and its sample version is

\begin{equation}
\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}} (f(\xi_j) - f(x)).
\end{equation}

Recall that the iterated Carré du Champ is

\begin{equation}
\Gamma_2(L_t, f, h) = \frac{1}{2} (L_t \Gamma(L_t, f, h) - \Gamma(L_t, L_t f, f) - \Gamma(L_t, f, L_t h)).
\end{equation}

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

\begin{equation}
\Gamma_2(L_t, f, f) = \frac{1}{2} (L_t (\Gamma_t(f, f)) - 2 \Gamma_t(L_t f, f)).
\end{equation}

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

\begin{equation}
\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_k||^2}{2t}} \frac{||x-\xi_k||^2}{2t} (f(\xi_j) - f(\xi_k))^2
\end{equation}

\begin{equation}
- \frac{1}{t^2 n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} \frac{1}{\theta_t^2(x)} e^{-\frac{||x-\xi_j||^2}{2t}} \frac{||x-\xi_k||^2}{2t} (f(\xi_k) - f(x))^2
\end{equation}

\begin{equation}
- \frac{2}{t^2 n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} \frac{1}{\theta_t(x) \theta_t(\xi_j)} e^{-\frac{||x-\xi_j||^2}{2t}} \frac{||x-\xi_k||^2}{2t} (f(\xi_k) - f(\xi_j)) (f(\xi_j) - f(x))
\end{equation}

\begin{equation}
+ \frac{2}{t^2 n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} \frac{1}{\theta_t^2(\xi_j)} e^{-\frac{||x-\xi_j||^2}{2t}} \frac{||x-\xi_k||^2}{2t} (f(\xi_k) - f(x)) (f(\xi_j) - f(x)).
\end{equation}

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 $\xi_1, \ldots, \xi_n$ be i.i.d. random variables on the probability space $(\Sigma, \mathcal{B}, \mu)$ where $\mathcal{B}$ is the Borel $\sigma$-algebra of $\Sigma$, and let $f : \Sigma \to [-K, K]$ be a Borel measurable function with $K > 0$. Then for the corresponding empirical measure $\mu_n$ and any $\varepsilon > 0$ we have

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

Observe, however, that (4.7)-(4.10) is a non-linear expression which will involve non-trivial interactions between the data points $\xi_1, \ldots, \xi_n$. This non-trivial interaction between the points $\xi_1, \ldots, \xi_n$ 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

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

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

(a) $f \in L^1(\mu)$ for any $f \in \mathcal{F}$,

(b) For any i.i.d. sample $\xi_1, \ldots, \xi_n$ drawn from $\Sigma$ whose distribution is $\mu$ we have uniform convergence in probability in the sense that for any $\varepsilon > 0$

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

**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 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 $\Sigma$ and totally bounded in $L^\infty(\Sigma)$. Given $\delta > 0$ we let $\mathcal{N}(\mathcal{F}, \delta)$ be the $L^\infty$-covering number of $\mathcal{F}$, i.e.,

\begin{equation}
\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}\}.
\end{equation}

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

\begin{equation}
Pr^* \left\{ \sup_{f \in \mathcal{F}} |\mu_n f - \mu f| \geq \varepsilon \right\} \leq 2\mathcal{N} \left( \mathcal{F}, \frac{\varepsilon}{4} \right) e^{-\frac{\varepsilon^2 n}{8M^2}}.
\end{equation}

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 $\delta$ 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

\begin{equation}
|\mu_n f - \mu f| \leq 2\delta + |\mu_n \phi - \mu \phi|,
\end{equation}

and clearly

\begin{equation}
\sup_{f \in \mathcal{F}} |\mu_n f - \mu f| \leq 2\delta + \max_{\phi \in \mathcal{G}} |\mu_n \phi - \mu \phi|.
\end{equation}

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

\begin{equation}
Pr^* \left\{ \sup_{f \in \mathcal{F}} |\mu_n f - \mu f| \geq \varepsilon \right\} \leq Pr \left\{ \max_{\phi \in \mathcal{G}} |\mu_n \phi - \mu \phi| \geq \varepsilon/2 \right\},
\end{equation}

and by Hoeffding’s inequality we have

\begin{equation}
Pr \left\{ \max_{\phi \in \mathcal{G}} |\mu_n \phi - \mu \phi| \geq \frac{\varepsilon}{2} \right\} \leq 2\mathcal{N} \left( \mathcal{F}, \frac{\varepsilon}{4} \right) e^{-\frac{\varepsilon^2 n}{8M^2}},
\end{equation}

which implies the lemma.

We will frequently use the following classes of functions

\begin{equation}
\mathcal{F}_{f,h}^t = \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\},
\end{equation}

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

where $f, h$ in (4.20) are fixed functions. We will also use $\mathcal{F}_f^t$ to denote the class $\mathcal{F}_{f,f}^t$. Observe that for the classes in (4.20) and (4.21) one has

\begin{equation}
M_{\mathcal{F}_{f,h}^t} = \sup_{\phi \in \mathcal{F}_{f,h}^t} \|\phi\|_{L^\infty(\Sigma)} \leq \left( \frac{2}{e} \right) t^{1/2} \|f\|_{\text{Lip}} \|h\|_{\text{Lip}},
\end{equation}

\begin{equation}
M_{\mathcal{G}^t} = \sup_{\psi \in \mathcal{G}^t} \|\psi\|_{L^\infty(\Sigma)} = t^{1/2}.
\end{equation}
4.3. **Ambient Covering Numbers.** In this subsection we show that the computation of covering numbers of the classes of functions \( F_{t,f,h} \) and \( G_t \) introduced in (4.20), (4.21) 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* \( \Sigma \) defined by

\[
A(\Sigma, \delta) = \inf \{ m : \exists A \text{ with } |A| = m \text{ and } \Sigma \subset \bigcup_{a \in A} B_{\mathbb{R}^N, \delta}(a) \}.
\]

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

\[
\left| t^{1/2} e^{-\frac{\| \xi - \tilde{\xi} \|^2}{2t}} - t^{1/2} e^{-\frac{\| \zeta - \tilde{\xi} \|^2}{2t}} \right| \leq e^{-1/2} \| \xi - \tilde{\xi} \|.
\]

In particular

\[
N(G_t, \varepsilon) \leq A(\Sigma, e^{1/2} \varepsilon).
\]

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

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

and therefore

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

and therefore we have the Lipschitz estimate

\[
|\psi_t(\xi, \zeta) - \psi_t(\tilde{\xi}, \zeta)| \leq e^{-1/2} \| \xi - \tilde{\xi} \|.
\]

\( \square \)

**Corollary 4.7.** Fix a function \( 0 \neq h \in L^\infty(\Sigma) \) with \( \| h \|_{L^\infty(\Sigma)} \leq C \) and consider the class of functions

\[
H^t_{h} = \{ \psi_t(\zeta, \cdot) h(\cdot) : \zeta \in \Sigma \}.
\]

Then for every \( \varepsilon > 0 \) we have

\[
N(H^t_{h}, \varepsilon) \leq A \left( \Sigma, e^{1/2} \varepsilon \right).
\]

**Proof.** We use Lemma 4.6 to obtain the estimate

\[
\left\| \psi_t(\zeta, \cdot) h(\cdot) - \psi_t(\zeta', \cdot) h(\cdot) \right\|_{L^\infty(\Sigma)} \leq C \left\| \psi_t(\zeta, \cdot) - \psi_t(\zeta', \cdot) \right\|_{L^\infty(\Sigma)} \leq C e^{-1/2} \| \zeta - \zeta' \|,
\]

from which the corollary follows. \( \square \)

For Lemma 4.8 we will use the following notation:
For a Lipschitz function $f$ defined on the ambient space $\mathbb{R}^N$ we will write $\|f\|_{\text{Lip}}$ to denote the Lipschitz norm of $f$ with respect to the ambient distance $\| \cdot \|$, i.e.,

\[ \|f\|_{\text{Lip}} = \inf_{x,y \in \mathbb{R}^N, x \neq y} \left\{ \frac{|f(x) - f(y)|}{\|x - y\|} \right\}. \tag{4.34} \]

For a function $f \in C^k(\mathbb{R}^N)$ we will use $\|f\|_{C^k}$ to denote the following

\[ \|f\|_{C^k} = \|f\|_{L^\infty(\Sigma)} + \sum_{j=1}^k \|D_j f\|_{L^\infty(\Sigma)}. \tag{4.35} \]

In particular, when in (4.35) we write $\|D_j f\|_{C^k}$ we mean

\[ \|D_j f\|_{L^\infty(\Sigma)} = \sup_{x \in \Sigma} \|D_j f(x)\|, \tag{4.36} \]

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

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

\[ \sup_{\zeta \in \mathbb{R}^d} |\phi_t(\xi, \zeta) - \phi_t(\xi', \zeta)| \leq C(f,h)\|\xi - \xi'\|, \tag{4.37} \]

where

\[ C(f,h) = C_0 (\|f\|_{\text{Lip}} \|h\|_{\text{Lip}} + \|f\|_{C^1} \|h\|_{\text{Lip}} + \|h\|_{C^1} \|f\|_{\text{Lip}}) \tag{4.38} \]

and $C_0$ is a universal constant. Thus

\[ \mathcal{N}(F_{f,h}^t, \delta) \leq \mathcal{A} \left( \Sigma, \frac{\delta}{C(f,h)} \right). \tag{4.39} \]

**Proof.** Let $\phi_t(\xi, \zeta) \in F_{f,h}^t$, then we have

\[ D_\xi \phi_t(\xi, \zeta) = t^{-1/2} \left( \frac{\xi - \zeta}{\sqrt{2t}} - \frac{\| \xi - \zeta \|^2}{4t} \right) (f(\xi) - f(\zeta)) (h(\xi) - h(\zeta)) \tag{4.40} \]

\[ + t^{-1/2} e^{-\|\xi - \zeta\|^2/4t} D_\xi f(\xi) (h(\xi) - f(\zeta)) \tag{4.41} \]

\[ + t^{-1/2} e^{-\|\xi - \zeta\|^2/4t} D_\xi h(\xi) (f(\xi) - f(\zeta)), \tag{4.42} \]

and then

\[ \|D_\xi \phi_t(\xi, \zeta)\| \leq \|f\|_{\text{Lip}} \|h\|_{\text{Lip}} \left( \frac{\| \xi - \zeta \|^3}{t^{3/2}} \right) e^{-\|\xi - \zeta\|^2/2t} \tag{4.43} \]

\[ + \frac{\| \xi - \zeta \|}{t^{1/2}} e^{-\|\xi - \zeta\|^2/2t} \|f\|_{C^1} \|h\|_{\text{Lip}} \tag{4.44} \]

\[ + \frac{\| \xi - \zeta \|}{t^{1/2}} e^{-\|\xi - \zeta\|^2/2t} \|h\|_{C^1} \|f\|_{\text{Lip}} \tag{4.45} \]

\[ \leq C_0 (\|f\|_{\text{Lip}} \|h\|_{\text{Lip}} + \|f\|_{C^1} \|h\|_{\text{Lip}} + \|f\|_{\text{Lip}} \|h\|_{C^1}). \tag{4.46} \]
where
\begin{equation}
C_0 = \max \left( \sup_{\rho > 0} \{ \rho^3 e^{-\rho^2/2} \}, \sup_{\rho > 0} \{ \rho e^{-\rho/2} \} \right).
\end{equation}

It follows that for any $\xi \in \Sigma$ we have the Lipschitz estimate
\begin{equation}
|\phi_t(\xi, \zeta) - \phi_t(\xi, \zeta')| \leq C(f, h)\|\xi - \xi'\|.
\end{equation}

The following can be obtained in a similar fashion.

**Lemma 4.9.** Let $f \in C^1$ and let $v_t$ be given by
\[ v_t(\xi, \zeta) = e^{-\frac{\|\xi - \zeta\|^2}{2t}}(f(\xi) - f(\zeta)). \]
We then have the following estimate
\begin{equation}
|v_t(\xi, \zeta) - v_t(\xi', \zeta)| \leq \left( \frac{2t}{e} \|f\|_{\text{Lip}} + \|f\|_{C^1} \right) \|\xi - \xi'\|.
\end{equation}

**Proof.** Follows from
\begin{equation}
|D_x v_t(\xi, \zeta)| \leq \frac{\|\xi - \zeta\|^2}{t} e^{-\frac{\|\xi - \zeta\|^2}{2t}} \|f(\xi) - f(\zeta)\|
+ \frac{e^{-\frac{\|\xi - \zeta\|^2}{2t}}}{t} \|D_\xi f(\xi)\|,
\end{equation}
and
\begin{equation}
\sup_{\rho > 0} \rho^2 e^{-\rho^2} = \frac{2}{e}.
\end{equation}

Lemmas 4.6 and 4.8 say that we can relate covering numbers of the classes $F_{f,h}^t, G^t$ to ambient covering numbers of the submanifold $\Sigma$. In order to estimate $A(\Sigma, \delta)$ we need to introduce the notion of *reach* of an embedded submanifold of $\mathbb{R}^N$. For every $\varepsilon > 0$ we can consider the $\varepsilon$ neighborhood of $\Sigma$
\begin{equation}
\Sigma_\varepsilon = \{ x \in \mathbb{R}^N : d(x, \Sigma) < \varepsilon \},
\end{equation}
where $d(\cdot, \Sigma)$ measures the distance from points in $\mathbb{R}^N$ to $\Sigma$ with respect to the ambient norm $\| \cdot \|$. If $\Sigma$ 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
\begin{enumerate}
\item $\varphi$ is smooth,
\item $\varphi(x)$ is the unique point in $\Sigma$ such that $\|\varphi(x) - x\| = d(x, \Sigma)$ for all $x \in \Sigma_\varepsilon$.
\item $x - \varphi(x) \in T_{\varphi(x)}^\perp \Sigma$,
\item $\varphi(y + z) \equiv \varphi(y)$ for all $y \in \Sigma$ and $z \in (T_y \Sigma)^\perp$ with $\|z\| < \varepsilon$,
\item For any vector $V \in \mathbb{R}^N$, $DV \varphi(x) = p_{\varphi(x)}(V)$, where $p_{\varphi(x)}(V)$ is the orthogonal projection of $V$ onto $T_{\varphi(x)} \Sigma$.
\end{enumerate}
See for example [Sim96, Theorem 1]. The map $\varphi$ is called *nearest point projection* onto $\Sigma$. 
Definition 4.10. Let $\Sigma$ be an embedded submanifold of $\mathbb{R}^N$. The reach of $\Sigma$ is the number
\[ \tau = \sup_{\varepsilon > 0} \{ \text{There exists a nearest point projection in } \Sigma_{\varepsilon} \}. \]

We have the following result

Theorem 4.11 ([FMN13] Corollary 6). Suppose that $\Sigma$ 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
\[ U(r) = V \left( \frac{1}{\tau^d} + r^d \right), \]
then for any $\varepsilon > 0$ there is an $\varepsilon$ net of $\Sigma$ 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,
\[ A(\Sigma, \varepsilon) \leq C_d U(\varepsilon^{-1}). \]

Corollary 4.12. We have the following bounds
(a) \[ N(F^t_{f,h}, \varepsilon) \leq C_d V \left( \frac{1}{\tau^d} + \left( \frac{C(f,h)}{\varepsilon} \right)^d \right) \]
(b) \[ N(G^t, \varepsilon) \leq C_d V \left( \frac{1}{\tau^d} + \left( \frac{1}{e^{1/2} \varepsilon} \right)^d \right), \]
(c) \[ N(H^t_{h}, \varepsilon) \leq C_d V \left( \frac{1}{\tau^d} + \left( \frac{C}{e^{1/2} \varepsilon} \right)^d \right). \]

where $C$ is an upper bound for $\|h\|_{L^\infty(\Sigma)}$.

4.4. Sample Version of the Carré du Champ. In this section we are still assuming that the distribution of the sample $\xi_1, \ldots, \xi_n$ in $\Sigma$ is uniform, i.e., $d\mu = d\text{vol}_{\Sigma}$. 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 therefore there exists $t_0 > 0$ such that for $0 < t < t_0$ we have
\[ 2(2\pi)^{d/2} \geq \theta_t \geq \frac{1}{2}(2\pi)^{d/2}. \]

If we let
\[ \lambda_0 = \frac{(2\pi)^{d/2}}{4}, \]
we have for $0 < t < t_0$ the inequality
\[ \theta_t(x) \geq 2t^{d/2} \lambda_0, \]
which will be a convenient normalization for us in the sequel. The main lemma in this section is the following.

**Lemma 4.13.** Suppose $\mathcal{F}$ is a class of functions totally bounded in $L^\infty(\Sigma)$ of the form

$$f(x, \cdot) \text{ for } x \in \Sigma$$

for a fixed $f \in L^\infty(\Sigma \times \Sigma)$ and

$$M = \|f\|_{L^\infty(\Sigma \times \Sigma)}.$$

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

$$\varepsilon t^{d+1/2} \lambda_0 < M.$$ \hfill (4.62)

Then

$$\Pr^* \left\{ \sup\limits_{x \in \Sigma} \left| t^{-1/2} \frac{\mu_n f(x, \cdot)}{\theta_t(x)} - t^{-1/2} \frac{\mu f(x, \cdot)}{\theta_t(x)} \right| \geq \varepsilon \right\} \leq 2 \mathcal{N} \left( \mathcal{G}', \frac{\varepsilon t^{d+1} \lambda_0^2}{4M} \right) \exp \left( -\frac{\varepsilon^2 n^t 2d \lambda_0^4}{8M^2} \right)$$ \hfill (4.63)

$$+ 2 \mathcal{N} \left( \mathcal{F}, \frac{\varepsilon t^{d+1} \lambda_0^2}{4} \right) \exp \left( -\frac{\varepsilon^2 \lambda_0^2 t^{2d+1}}{8M^2} \right).$$ \hfill (4.64)

Before proving Lemma 4.13 we will prove the following elementary lemma.

**Lemma 4.14.** Let $\xi, \zeta$ be positive random variables. For any $\varepsilon > 0$ we have

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

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

$$\frac{1}{\zeta} - \frac{1}{\xi} = \frac{\xi - \zeta}{\xi \zeta} = \frac{\xi - \zeta}{\zeta \xi - \xi^2 + \xi^2}$$

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

$$= \frac{|\zeta - \xi|}{-\xi |\zeta - \xi| + \xi^2},$$

and from

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

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 4.13. It is easy to prove from Lemma 4.14 and (4.61) that for any $\delta > 0$ we have

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

Let us write
\[
t^{-1/2} \left( \frac{\mu_n f(x, \cdot)}{\hat{\theta}_t(x)} - \frac{\mu f(x, \cdot)}{\theta_t(x)} \right) = t^{-1/2} \mu_n f(x, \cdot) \left( \frac{1}{\hat{\theta}_t(x)} - \frac{1}{\theta_t(x)} \right)
+ t^{-1/2} \frac{1}{\theta_t(x)} \left[ \mu_n f(x, \cdot) - \mu f(x, \cdot) \right].
\]

Thus
\[
Pr\left\{ t^{-1/2} \sup_{x \in \Sigma} \left| \frac{\mu_n f(x, \cdot)}{\hat{\theta}_t(x)} - \frac{\mu f(x, \cdot)}{\theta_t(x)} \right| \geq \varepsilon \right\}
\leq Pr\left\{ \sup_{x \in \Sigma} \left| \frac{1}{\hat{\theta}_t(x)} - \frac{1}{\theta_t(x)} \right| \geq \frac{\varepsilon t^{1/2}}{2M} \right\}
+ Pr\left\{ \sup_{x \in \Sigma} \left| \mu_n f(x, \cdot) - \mu f(x, \cdot) \right| \geq \varepsilon \lambda_0 t^{(d+1)/2} \right\}.
\]

Analyzing the first term and using (4.62) and (4.66)-(4.67) leads us to the inequality
\[
Pr\left\{ \sup_{x \in \Sigma} \left| \frac{1}{\hat{\theta}_t(x)} - \frac{1}{\theta_t(x)} \right| \geq \frac{\varepsilon t^{1/2}}{2M} \right\}
\leq Pr\left\{ \sup_{x \in \Sigma} \left| \theta_t(x) - \hat{\theta}_t(x) \right| \geq \frac{2\varepsilon t^{d+1/2} \lambda_0^2}{M + \varepsilon t^{d+1/2} \lambda_0} \right\}
\leq Pr\left\{ \sup_{x \in \Sigma} \left| \theta_t(x) - \hat{\theta}_t(x) \right| \geq \frac{t^{d+1/2} \lambda_0^2}{M} \right\}
\leq 2N \left( G, \frac{\varepsilon t^{d+1/2} \lambda_0^2}{4M} \right) e^{-\frac{\varepsilon^2 t^{2d} \lambda_0^4}{8Mt^2}},
\]
where we have applied Lemma 4.5 to the class $G^t$ (in particular we have used (4.23)). Finally,

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

\section*{In view of Lemma 4.13 we introduce the following notation}

\noindent \textbf{Definition 4.15.} Given a class of functions $\mathcal{F}$ as in the statement of Lemma 4.13 and positive numbers $t, \varepsilon, M$ we define for compactness of notation the following function

\begin{equation}
Q_t(\mathcal{F}, \varepsilon, M, n) = 2N \left( G^t, \varepsilon t^{d+1} \lambda_0^2 \right) \exp \left( -\frac{\varepsilon^2 \lambda_0^2 t^{2d+1}}{8M^2} \right) \exp \left( -\frac{\varepsilon^2 \lambda_0^2 t^{2d+1}}{8M^2} \right),
\end{equation}

\begin{equation}
+ 2N \left( \mathcal{F}, \varepsilon t^{d+1} \lambda_0^2 \right) \exp \left( -\frac{\varepsilon^2 \lambda_0^2 t^{2d+1}}{8M^2} \right) \exp \left( -\frac{\varepsilon^2 \lambda_0^2 t^{2d+1}}{8M^2} \right).
\end{equation}

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

\noindent \textbf{Corollary 4.16.} Letting $K = \left( \frac{2}{c} \right) \|f\|_{\text{Lip}} \|h\|_{\text{Lip}}$ we have

\begin{equation}
\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}_{f,h}^t, \varepsilon, K t^{1/2}, n).
\end{equation}

\textit{Proof.} Recall that

\begin{equation}
\hat{\Gamma}_t(L_t, f, h)(x) = \frac{1}{t \theta_t} \sum_{j=1}^n e^{-\frac{||x-\xi_j||^2}{2t}} (f(\xi_j) - f(x))(h(\xi_j) - h(x))
\end{equation}

\begin{equation}
= \frac{t^{-1/2}}{\theta_t(x)} \sum_{j=1}^n \phi_t(x, \xi_j) = t^{-1/2} \frac{\mu_n(\phi_t(x, \cdot))}{\theta_t(x)},
\end{equation}

where $\phi_t$ is given by the definition of the class $\mathcal{F}_{f,h}^t$ in (4.20) and therefore we can apply Lemma 4.13 to the class $\mathcal{F}_{f,h}^t$ and use the bound

\begin{equation}
\sup_{\psi \in \mathcal{F}_{f,h}^t} \|\psi\|_{L^\infty(\Sigma)} \leq t^{1/2} K.
\end{equation}

which follows from (4.22).

\qed
If we now consider the $t$-Laplacean $L_t$ 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

\begin{equation}
\hat{L}_t h(x) - L_t h(x) = \frac{2}{\theta_t(x) t n} \sum_{j=1}^n e^{-\frac{\|\xi_j-x\|^2}{2t}} (h(\xi_j) - h(x))
\end{equation}

(4.75)

\begin{equation}
= \frac{2}{t \theta_t(x)} \int_{\Sigma} e^{-\frac{\|\xi-x\|^2}{2t}} (h(\xi) - h(x)) d\mu(\xi)
\end{equation}

(4.76)

\begin{equation}
= \frac{2}{\theta_t(x) t n} \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)
\end{equation}

(4.77)

\begin{equation}
= \frac{2\mu_n \psi_t(x,\cdot) h(\cdot)}{t^{3/2} \theta_t(x)} - \frac{2\mu \psi_t(x,\cdot) h(\cdot)}{t^{3/2} \theta_t(x)}.
\end{equation}

(4.78)

Observe that

\begin{equation}
\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}{\theta_t} - \frac{\mu \eta}{\theta_t} \right| \geq t\varepsilon \right\}.
\end{equation}

(4.79)

We have obtained

**Corollary 4.17.** Fix a function $h \in L^\infty(\Sigma)$. If we set $M = \|h\|_{L^\infty}$ we have

\begin{equation}
\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^{1/2} M, n).
\end{equation}

(4.80)

**Proof.** The proof follows from combining Lemma 4.13 with (4.79) and the fact that

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

(4.81)

\[ \square \]

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 purpose of introducing some notation. This notation will be used in later sections.

**Theorem 4.18.** Consider the metric measure space $(\Sigma, \|\cdot\|, d\text{vol}_\Sigma)$ where $\Sigma^d \subset \mathbb{R}^N$ is a smooth closed embedded submanifold. Suppose that we have an uniformly distributed i.i.d. sample $\{\xi_1, \ldots, \xi_n\}$ of points from $\Sigma$. For $\sigma > 0$, let

\begin{equation}
t_n = n^{-\frac{1}{2d+\sigma}}.
\end{equation}

(4.82)

Then we have for fixed $f, h \in \text{Lip}(\Sigma)$

\[ \sup_{\xi \in \Sigma} \left| \hat{\Gamma}(L_{t_n}, f, h)(\xi) - \Gamma(L_{t_n}, f, h)(\xi) \right| \xrightarrow{\text{a.s.}} 0. \]
Proof. For a fixed \( n \) and \( \varepsilon \), we have that
\[
\Pr \left\{ \sup_{x \in \Sigma} \left| \hat{\Gamma}(L_{t_n}, f, h)(x) - \Gamma(L_{t_n}, f, h)(x) \right| \geq \varepsilon \right\} \leq Q_{t_n}(F_{f,h}^{t_n}, \varepsilon, K_{t_n}^{1/2}, n).
\]
Now plugging in the expression for \( t_n \) we observe a bound of the form
\[
(4.83) \quad Q_{t_n}(F_{f,h}^{t_n}, \varepsilon, K_{t_n}^{1/2}) \leq p \left( n^{\frac{1}{2(2d+\sigma)}} \frac{1}{\varepsilon} \right) \exp \left( -c_1 \varepsilon^2 n^{\sigma/(2d+\sigma)} \right)
\]
where \( p \) is a fixed polynomial bound and \( c_1 > 0 \) is a constant. Thus with \( \varepsilon \) fixed, we have
\[
\sum_{n=1}^{\infty} Q_{t_n}(F_{f,h}^{t_n}, \varepsilon, K_{t_n}^{1/2}, n) < \infty.
\]
Applying the Borel-Cantelli Lemma gives the almost sure convergence. \( \square \)

Now we see that for almost sure convergence, one requires a bound of the form \((4.83)\). For this reason, we introduce notation for use in the sequel: consider a function
\[
Q : \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{N} \to \mathbb{R}^+.
\]
We say that \( Q(t, \varepsilon, n) \in O_{BC}(\beta) \) if
\[
\sum_{n=1}^{\infty} Q(n^{-\frac{1}{\beta+\sigma}}, \varepsilon, n) < \infty
\]
for all \( \sigma, \varepsilon > 0 \). Clearly the definition gives
\[
O_{BC}(\beta) \subset O_{BC}(\beta')
\]
for \( \beta' > \beta \). We also observe that
\[
O_{BC}(\beta) + O_{BC}(\beta') \in O_{BC}(\max \{ \beta, \beta' \})
\]
and that if for a class of functions \( F \), a positive number \( K \) we set \( Q(t, \varepsilon, n) = Q_t(F, \varepsilon, K, n) \) then
\[
Q_t(F^t, \varepsilon, M, n) \in O_{BC}(\beta)
\]
\[
\implies Q_t(F^t, \varepsilon t^\alpha, Kt^\delta, n) \in O_{BC}(\beta + 2\alpha - \min\{2\delta, 1\}).
\]

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

**Corollary 4.19.** For the classes of functions define by \((4.20)\) \((4.21)\) and \((4.30)\) and \( \varepsilon, M > 0 \) fixed we have
\[
Q_t(F, \varepsilon, K, n) \in O_{BC}(2d).
4.6. **Proof of Theorem 1.12** Recall that

\[
\Gamma_2(L_t, f, f) = \frac{1}{2} (L_t (\Gamma_t(f, f)) - 2 \Gamma_t(L_t f, f)),
\]

and that from Remark 4.2

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

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

\[
\hat{L}_t(\hat{\Gamma}_t(f, f)) - L(\Gamma_t(f, f))(x) = \hat{L}_t \left( \hat{\Gamma}_t(f, f) - \Gamma_t(f, f) \right)(x) + \left( \hat{L}_t - L_t \right)(x) \Gamma_t(f, f)
\]

\[
= A_1(x) + A_2(x),
\]

and observe that

\[
|A_1| = \left| \hat{L}_t \left( \hat{\Gamma}_t(f, f) - \Gamma_t(f, f) \right) \right| \leq \frac{4}{t} \sup_{\xi \in \Sigma} \left| \hat{\Gamma}_t(f, f)(\xi) - \Gamma_t(f, f)(\xi) \right|.
\]

In order to estimate \( A_2 = \left( \hat{L}_t - L_t \right)(\Gamma_t(f, f)) \) we note from Proposition 2.1 that \( \Gamma_t(f, f) \) converges to \( |\nabla f|^2 \) in \( L^\infty(\Sigma) \) as \( t \to 0 \) and we can choose \( t_1 \) with \( 0 < t < t_1 < t_0 \) so that

\[
\|\Gamma_t(f, f)\|_{L^\infty(\Sigma)} \leq 2 \|f\|_{L^\text{Lip}}^2,
\]

and the idea will be to use Lemma 4.17. We now estimate the difference

\[
\hat{\Gamma}_t(\hat{L}_t f - L_t f, f)(x) - \Gamma_t(L_t f, f)(x) = \Gamma_t(\hat{L}_t f - L_t f, f)(x) + \left( \hat{\Gamma}_t - \Gamma_t \right)(L_t f, f)(x)
\]

\[
= A_3(x) + A_4(x),
\]

and we note that

\[
|A_3| = \left| \hat{\Gamma}_t(\hat{L}_t f - L_t f, f)(x) \right|
\]

\[
\leq \frac{2}{t} \left( \sup_{\xi \in \Sigma} \left| L_t(f)(\xi) - L_t(f)(\xi) \right| \right) \frac{1}{n \theta_t(x)} \sum_{j=1}^{n} e^{-\frac{|x - \xi_j|^2}{2 \theta_t}} |f(\xi_j) - f(x) |
\]

\[
\leq 2 e^{-1/2 \|f\|_{L^\text{Lip}}}/t^{1/2 \theta_t} \left( \sup_{\xi \in \Sigma} \left| (\hat{L}_t(f)(\xi) - L_t(f)(\xi)) \right| \right),
\]

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

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

\[
\leq \frac{2}{t} \sup_{\xi \in \Sigma} |u(\xi)| \left( \frac{\|v\|_{L^\text{Lip}}}{n \theta_t(x)} \sup_{\rho > 0} \rho e^{-\frac{\rho^2}{2 \theta_t}} \right),
\]

\[
\leq \frac{2}{t} \sup_{\xi \in \Sigma} |u(\xi)| \left( \frac{\|v\|_{L^\text{Lip}}}{n \theta_t(x)} \sup_{\rho > 0} \rho e^{-\frac{\rho^2}{2 \theta_t}} \right),
\]
and

\[(4.98)\quad \sup_{\rho > 0} \rho e^{-\frac{s^2}{2\rho^2}} = t^{1/2}e^{-1/2} \leq t^{1/2}.\]

In view of Lemma (4.13), we will write the bound (4.95) on \(|A_3|\) as

\[(4.99)\quad |A_3| \leq 2e^{-1/2}\|f\|_{\text{Lip}} \left( \sup_{\xi \in \Sigma} \left| (L_t(f)(\xi) - L_t(f)(\xi)) \right| \right)\]

\[(4.100)\quad + \frac{2e^{-1/2}\|f\|_{\text{Lip}}}{t^{1/2}} \left( \frac{1}{\theta_t} - \frac{1}{\theta_t} \right) \left( \sup_{\xi \in \Sigma} \left| (\tilde{L}_t(f)(\xi) - \tilde{L}_t(f)(\xi)) \right| \right).\]

In order to estimate \(|A_4|\), i.e.

\[(4.101)\quad |A_4| = \left| (\tilde{\Gamma}_t - \Gamma_t) (L_t(f, f)) \right|,

we consider the class of functions \(F_{L_t,f,f}^t\) introduced in (4.20) so that we can use Lemma (4.16) however, this implies that we need a Lipschitz estimate on the functions \(L_t f\). Letting \(u_t(x, \xi)\) be as in Lemma (4.9) i.e.

\[(4.102)\quad u_t(x, \xi) = e^{-\frac{\|x-\xi\|^2}{2t}} (f(\xi) - f(x)),\]

we have

\[(4.103)\quad L_t f(x) - L_t f(x') = \frac{\theta_t(x') - \theta_t(x)}{t\theta_t(x)\theta_t(x')} \int_{\Sigma} u_t(x, \xi) d\mu + \frac{1}{t\theta_t(x')} \int_{\Sigma} (u_t(\xi, x) - u_t(\xi, x')) d\mu,

and

\[(4.104)\quad \frac{\theta_t(x') - \theta_t(x)}{t\theta_t(x)\theta_t(x')} \int_{\Sigma} u(\xi, x) d\mu \]

\[(4.105)\quad = \left( \frac{t^{-1/2}}{\theta_t(x')} \int_{\Sigma} |\psi_t(\xi, x) - \psi_t(\xi, x')| d\mu \right) L_t f(x).

and again we can choose \(t_0\) in the discussion surrounding (4.61) so that

\[(4.106)\quad \left| \frac{\theta_t(x') - \theta_t(x)}{t\theta_t(x)\theta_t(x')} \int_{\Sigma} u(\xi, x) d\mu \right| \leq 4 \frac{\text{vol}(\Sigma)}{t^{d+1}} \|f\|_{C^2(\Sigma)} \|x - x'\|.

From Lemma (4.9)

\[(4.107)\quad \left| u_t(x, \xi) - u_t(x', \xi) \right| \leq \left( \frac{2t}{e} \|f\|_{\text{Lip}} + \|f\|_{\text{Lip}} \right) \|x - x'\|

and therefore we can adjust \(t_0\) in (4.61) so that for \(0 < t < t_0\)

\[(4.108)\quad \left| \frac{1}{t\theta_t(x')} \int_{\Sigma} (u_t(\xi, x) - u_t(\xi, x')) d\mu \right| \leq \frac{2\text{vol}(\Sigma)}{t^{d+2}} \left( \frac{2}{e} \|f\|_{\text{Lip}} + \|f\|_{C^1} \right) \|x - x'\|.
We have shown that

\[
\|L_t f\|_{\text{Lip}} \leq \frac{2 \text{vol}(\Sigma)}{t^{3/2}} \left( 2 \|f\|_{C^2} + t^{-1/2} \frac{2}{\epsilon} \|f\|_{\text{Lip}} + t^{-1/2} \|f\|_{C^1} \right).
\]

Let \( A_i^* = \sup_{x \in \Sigma} |A_i(x)| \) for \( i = 1, 2, 3, 4 \). We now have from (4.86) and (4.91)

\[
(4.111) \quad \Pr \left\{ \left| \hat{\Gamma}_t (f, f) - \Gamma_t (f, f) \right| \geq \epsilon \right\} \\
(4.112) \quad \leq \Pr \left\{ A^*_1 \geq \frac{\epsilon}{4} \right\} + \Pr \left\{ A^*_2 \geq \frac{\epsilon}{4} \right\} + \Pr \left\{ A^*_3 \geq \frac{\epsilon}{4} \right\} + \Pr \left\{ A^*_4 \geq \frac{\epsilon}{4} \right\} \\
(4.113) \quad = P_1 + P_2 + P_3 + P_4.
\]

From (4.89) we have

\[
(4.114) \quad P_1 \leq \Pr \left\{ \sup_{\xi \in \Sigma} \left| \Gamma_t (f, f)(\xi) - \Gamma_t (f, f)(\xi) \right| \geq \frac{t \epsilon}{16} \right\} \leq Q_t \left( F_t, \frac{\epsilon}{8}, Kt^{1/2}, n \right),
\]

\[
(4.115) \quad \in O_{BC} (2(d+1)).
\]

For \( A_2 = (\hat{L}_t - L_t) (\Gamma_t (f, f)) \) we apply Corollary 4.17 and the estimate (4.90) to the class \( \mathcal{H}^t_{\Gamma_t (f, f)} \) and obtain

\[
(4.116) \quad P_2 \leq Q_t \left( \mathcal{H}^t_{\Gamma_t (f, f)}, \epsilon t, 2 \|f\|_{\text{Lip}}, n \right) \in O_{BC} (2d + 3).
\]

Using (4.100) we have

\[
(4.117) \quad P_3 \leq \Pr \left\{ \frac{\epsilon}{8} \leq 2 \epsilon^{-1/2} \frac{\|f\|_{\text{Lip}}}{t^{1/2} \theta_t} \left( \sup_{\xi \in \Sigma} \left| (\hat{L}_t (f)(\xi) - L_t (f)(\xi)) \right| \right) \right\} \\
(4.118) \quad + \Pr \left\{ \frac{\epsilon}{8} \leq 2 \epsilon^{-1/2} \frac{\|f\|_{\text{Lip}}}{t^{1/2}} \left( \frac{1}{\theta_t} - \frac{1}{\hat{\theta}_t} \right) \left( \sup_{\xi \in \Sigma} \left| (\hat{L}_t (f)(\xi) - L_t (f)(\xi)) \right| \right) \right\}
\]

and choosing \( t > 0 \) small enough we obtain from (4.61) and Corollary 4.17 the estimate

\[
(4.119) \quad \Pr \left\{ \frac{\epsilon}{8} \leq 2 \epsilon^{-1/2} \frac{\|f\|_{\text{Lip}}}{t^{1/2} \theta_t} \left( \sup_{\xi \in \Sigma} \left| (\hat{L}_t (f)(\xi) - L_t (f)(\xi)) \right| \right) \right\} \\
(4.120) \quad \leq \Pr \left\{ \frac{\epsilon}{8} \frac{\lambda_0 t^{d/2} e^{1/2}}{\|f\|_{\text{Lip}}} \leq \sup_{\xi \in \Sigma} \left| (\hat{L}_t (f)(\xi) - L_t (f)(\xi)) \right| \right\} \\
(4.121) \quad \leq Q_t \left( \mathcal{H}^t_{\Gamma, \epsilon} \frac{\lambda_0 t^{d/2} e^{1/2}}{8 \|f\|_{\text{Lip}}} t^{1/2} \|f\|_{L^\infty}, n \right) \in O_{BC} (3d + 3).
\]

On the other hand, we will make use of the estimate

\[
(4.122) \quad \sup_{\xi \in \Sigma} \left| (\hat{L}_t (f)(\xi) - L_t (f)(\xi)) \right| \leq \frac{4}{t} \|f\|_{L^\infty},
\]
for any \( t > 0 \) to obtain

\[
\text{Pr}\left\{ \frac{\varepsilon}{8} \leq \frac{2t^{-1/2}\|f\|_{\text{Lip}}}{t^{1/2}} \left( \frac{1}{\theta_t} - \frac{1}{\theta_t} \right) \left( \sup_{\xi \in \Sigma} |\hat{L}_t(f)(\xi) - L_t(f)(\xi)| \right) \right\}
\]

(4.123)

\[
\leq \text{Pr}\left\{ \frac{\varepsilon t^{3/2}e^{1/2}}{64\|f\|_{\text{Lip}}} \leq \sup_{\hat{x} \in \Sigma} \left| \frac{1}{\theta_t} - \frac{1}{\theta_t} \right| \right\}
\]

(4.124)

and from (4.66)-(4.67) and Lemma 4.1 we observe that

\[
\text{Pr}\left\{ \frac{\varepsilon t}{8} \leq \frac{2e^{-1/2}\|f\|_{\text{Lip}}}{t^{1/2}} \left( \sup_{\xi \in \Sigma} |\hat{L}_t(f)(\xi) - L_t(f)(\xi)| \right) \right\} \leq \text{Pr}\left\{ \frac{\varepsilon t^3/2e^{1/2}}{64\|f\|_{\text{Lip}}} \leq \sup_{\hat{x} \in \Sigma} \left| \frac{1}{\theta_t} - \frac{1}{\theta_t} \right| \right\}
\]

(4.125)

so that in view of (4.110) we have

\[
\sup_{l \in F_{L, f, f}^t} \|l\|_{\text{Lip}} \leq M_t,
\]

(4.128)

and then

\[
\text{Pr}\left\{ \frac{\varepsilon t^{3/2}e^{1/2}}{64\|f\|_{\text{Lip}}} \leq \sup_{\hat{x} \in \Sigma} \left| \frac{1}{\theta_t} - \frac{1}{\theta_t} \right| \right\} \leq Q_t(F_{L, f, f}, \varepsilon, t^{1/2}M_t, n) = O_{BC}(3d + 1).
\]

(4.129)

Finally, we consider the class \( F_{L, f, f}^t \) and let

\[
M_t = \frac{2\cdot 2\text{vol}(\Sigma)}{e} \cdot \left( \frac{2\|f\|_{C^2} + t^{-1/2}2e\|f\|_{\text{Lip}} + t^{-1/2}\|f\|_{C^1}}{t^{1/2}} \right) \|f\|_{\text{Lip}},
\]

(4.127)

so that in view of (4.110) we have

\[
\sup_{l \in F_{L, f, f}^t} \|l\|_{L^\infty(\Sigma)} \leq M_t,
\]

(4.128)

and then

\[
P_4 \leq Q_t(F_{L, f, f}, \varepsilon, t^{1/2}M_t, n) = O_{BC}(3d + 1).
\]

(4.129)

Using again

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

(4.130)

and from (4.115), (4.116), (4.126) and (4.129) we have

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

(4.131)

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 \( \gamma = \frac{1}{3d + 4 + \sigma} \) and \( \sigma \) is any positive number we have

\[
\sup_{\xi \in \Sigma} \left| \hat{\Gamma}_2(L_{t_n}, f, f) - \Gamma_2(L_{t_n}, f, f) \right| \overset{a.s.}{\longrightarrow} 0.
\]

(4.132)
5. Almost Sure Convergence in the non-Uniform Case

In this section we prove Theorem 1.18 which concerns the case when the random variables in the i.i.d. sample \( \xi_1, \ldots, \xi_n, \ldots \), are not distributed according to \( d\mu = d\text{vol}_\Sigma \), but instead according to a density of the form \( d\mu = e^{-\rho}d\text{vol} \) where \( \rho \) is a \( C^\infty \) function. Since we already explained the use of the theory of empirical processes in Section 4 for proving the convergence of the sample version of the iterated Carré du Champ in the uniform case, in this section we will simply assume familiarity with all the bounds derived in Section 4 and we will write all large deviation estimates in terms of the notation for describing subexponential decay introduced in Section 4.5. In particular, we will not write down any bounds on the covering numbers of totally bounded classes in \( L^\infty(\Sigma) \) that will appear in the proof. We start by considering the following simple lemma. Recall the definition of \( \theta_{t,\alpha}(x) \), in (1.48), noting \( \theta_{t,\alpha}(x) = \theta_{t,\alpha}^0(x) \neq \theta_{t,\alpha}(x) \).

Let
\[
\lambda_* = \frac{1}{4} \inf_{x \in \Sigma} \lim_{t \downarrow 0} t^{-d/2} \theta_t(x),
\]
(5.1)
\[
\lambda^* = \sup_{x \in \Sigma} \lim_{t \downarrow 0} t^{-d/2} \theta_t(x).
\]
(5.2)

Lemma 5.1.

\[
\Pr \left\{ \sup_{x \in \Sigma} |\hat{\theta}^\alpha_t(x) - \theta^\alpha_t(x)| \geq \varepsilon \right\} \leq \Pr \left\{ \sup_{x \in \Sigma} |\hat{\theta}_t(x) - \theta_t(x)| \geq \frac{\varepsilon}{\alpha (\lambda_* t^{d/2})^{\alpha-1}} \right\}
+ \Pr \left\{ \sup_{x \in \Sigma} |\hat{\theta}_t(x) - \theta_t(x)| \geq \lambda_* t^{d/2} \right\}.
\]

Proof. The uniform lower bound in (4.61) depends now on the function \( \rho \), so with the quantities defined above, we can choose \( t_0 > 0 \) such that for \( 0 < t < t_0 \) we have
\[
\lambda_* \leq \frac{\theta_t(x)}{t^{d/2}} \leq \lambda^*.
\]
(5.3)

By the mean value theorem,
\[
|\hat{\theta}^\alpha_t(x) - \theta^\alpha_t(x)| = \alpha \left| \tau \theta_t(x) + (1 - \tau)\hat{\theta}_t(x) \right|^{\alpha-1} |\hat{\theta}_t(x) - \theta_t(x)|,
\]
(5.4)
for some \( 0 \leq \tau \leq 1 \). Thus the event
\[
\left\{ \sup_{x \in \Sigma} |\hat{\theta}^\alpha_t(x) - \theta^\alpha_t(x)| \geq \delta \right\},
\]
is contained in the event (recall \( \alpha \in [0, 1] \))
\[
\left\{ \sup_{x \in \Sigma} |\hat{\theta}_t(x) - \theta_t(x)| \geq \frac{\delta}{\alpha (\lambda_* t^{d/2})^{\alpha-1}} \right\} \cup \left\{ \sup_{x \in \Sigma} |\hat{\theta}_t(x) - \theta_t(x)| \geq \lambda_* t^{d/2} \right\},
\]
and therefore

\[
\Pr \left\{ \sup_{x \in \Sigma} |\theta_t^\alpha(x) - \theta_t^\alpha(x)| \geq \delta \right\} \leq \Pr \left\{ \sup_{x \in \Sigma} |\tilde{\theta}_t(x) - \theta_t(x)| \geq \frac{\delta}{\alpha (\lambda_t t^{d/2})^{\alpha-1}} \right\} \\
+ \Pr \left\{ \sup_{x \in \Sigma} |\tilde{\theta}_t(x) - \theta_t(x)| \geq \lambda_t t^{d/2} \right\}.
\]

\[\square\]

**Lemma 5.2.** For fixed \( \varepsilon > 0 \) we have the estimate

\[
\Pr \left\{ \sup_{x \in \Sigma} |\hat{\theta}_{t,\alpha}(x) - \theta_{t,\alpha}(x)| \geq \varepsilon \right\} \\
\leq \Pr \left\{ \sup_{x \in \Sigma} |\tilde{\theta}_t(x) - \theta_t(x)| \geq \frac{\varepsilon (\lambda_t)^{1+\alpha} t^{\frac{d}{2}(\alpha+1)} 2^{2\alpha}}{2 (1 + \varepsilon \lambda_t t^{d/2})} \right\} \\
+ \Pr \left\{ \sup_{x \in \Sigma} |\tilde{\theta}_t(x) - \theta_t(x)| \geq \lambda_t t^{d/2} \right\}.
\]

**Proof.** Letting \( \sigma_t(x, \xi) = e^{-\|x-\xi\|^2/2t} \) we have

\[
(5.5) \quad \hat{\theta}_{t,\alpha}(x) - \theta_{t,\alpha}(x) = \mu_n \frac{\sigma_t(x, \cdot)}{\theta_t^\alpha} - \mu \frac{\sigma_t(x, \cdot)}{\theta_t^\alpha} = \left( \mu_n \frac{\sigma_t(x, \cdot)}{\theta_t^\alpha} - \mu_n \frac{\sigma_t(x, \cdot)}{\theta_t^\alpha} \right) + \left( \mu_n \frac{\sigma_t(x, \cdot)}{\theta_t^\alpha} - \mu \frac{\sigma_t(x, \cdot)}{\theta_t^\alpha} \right)
\]

so that

\[
(5.7) \quad \Pr \left\{ \sup_{x \in \Sigma} |\hat{\theta}_{t,\alpha}(x) - \theta_{t,\alpha}(x)| \geq \varepsilon \right\} \leq \Pr \left\{ \sup_{x \in \Sigma} \left| \mu_n \frac{\sigma_t(x, \cdot)}{\theta_t^\alpha} - \mu_n \frac{\sigma_t(x, \cdot)}{\theta_t^\alpha} \right| \geq \frac{\varepsilon}{2} \right\} \\
+ \Pr \left\{ \sup_{x \in \Sigma} \left| \mu_n \frac{\sigma_t(x, \cdot)}{\theta_t^\alpha} - \mu_{\frac{\sigma_t(x, \cdot)}{\theta_t^\alpha}} \right| \geq \frac{\varepsilon}{2} \right\}.
\]

Now,

\[
(5.9) \quad \left| \mu_n \frac{\sigma_t(x, \cdot)}{\theta_t^\alpha} - \mu_n \frac{\sigma_t(x, \cdot)}{\theta_t^\alpha} \right| \leq \sup_{x \in \Sigma} \left| \frac{1}{\theta_t^\alpha} - \frac{1}{\theta_t^\alpha} \right|.
\]
From Lemma 5.1 we obtain
\[
\Pr \left\{ \sup_{x \in \Sigma} \left| \mu_n \frac{\sigma_t(x, \cdot)}{\theta_t^\alpha} - \mu_n \frac{\sigma_t(x, \cdot)}{\theta_t^\alpha} \right| \geq \frac{\varepsilon}{2} \right\} \leq \Pr \left\{ \sup_{x \in \Sigma} \left| \frac{1}{\theta_t^\alpha} - \frac{1}{\theta_t^\alpha} \right| \geq \frac{\varepsilon}{2} \right\}
\]
\[
\leq \Pr \left\{ \sup_{x \in \Sigma} \left| \hat{\theta}_t^\alpha(x) - \theta_t^\alpha(x) \right| \geq \frac{\varepsilon}{2} \left( \frac{\theta_t^\alpha}{1 + \frac{\varepsilon}{2} \theta_t^\alpha} \right)^2 \right\}
\]
\[
\leq \Pr \left\{ \sup_{x \in \Sigma} \left| \hat{\theta}_t(x) - \theta_t(x) \right| \geq \frac{\varepsilon}{2} \left( 1 + \varepsilon \lambda_t \frac{1}{d/2} \right)^{\alpha - 1} \right\},
\]
and the result follows. □

For the rest of the section we will use the following notation

- We will write \( a \lesssim b \) to indicate that \( a \leq Cb \) for some positive constant depending on \( d, \alpha \) and \( \rho \).
- We will write \( a \approx b \) to indicate that there exist positive constants \( C_1, C_2 \) with \( C_i = C_i(\alpha, d, \rho) \) for \( i = 1, 2 \), such that \( C_1a \leq b \leq C_2a \).

If we now replace \( \varepsilon > 0 \) in Lemma 5.2 by \( t^d \varepsilon \) for some \( \beta > 0 \) we easily obtain:

**Corollary 5.3.** For \( \beta \in \mathbb{R} \) we have the estimates

(5.10) \[
\Pr \left\{ \sup_{x \in \Sigma} \left| \hat{\theta}_{t,\alpha}(x) - \theta_{t,\alpha}(x) \right| \geq t^\beta \varepsilon \right\} \in O_{BC} (2\beta + (1 + \alpha)d),
\]

(5.11) \[
\Pr \left\{ \sup_{x \in \Sigma} \left| \frac{1}{\theta_t^\alpha(x)} - \frac{1}{\theta_t^\alpha(x)} \right| \geq t^\beta \varepsilon \right\} \in O_{BC}(2\beta + d(1 + \alpha)),
\]

(5.12) \[
\Pr \left\{ \sup_{x \in \Sigma} \left| \frac{1}{\theta_{t,\alpha}(x)} - \frac{1}{\theta_{t,\alpha}(x)} \right| \geq t^\beta \varepsilon \right\} \in O_{BC}(2\beta + d(3 - \alpha)).
\]

**Proof.** The corollary follows from combining Lemma 5.2 with the estimate

(5.13) \[
\theta_{t,\alpha}(x) \approx t^{(1-\alpha)d/2} \text{ as } t \to 0.
\]

□

For the Carré du Champ we have

**Proposition 5.4.** Let \( \varepsilon > 0 \), let \( u, v \neq 0 \) be Lipschitz functions and let \( \Phi_{t,u,v} \) be a function defined by

\[
\Phi_{t,u,v}(x, \xi) = e^{-\frac{\|x-x_\xi\|^2}{2t}} (u(x) - u(\xi))(v(x) - v(\xi)).
\]

We have
(1) The class of functions \( \mathcal{F}_{u,v,\alpha} = \left\{ \frac{\Phi_{t,u,v}(x,\cdot)}{\theta_t^\alpha} : x \in \Sigma \right\} \) is totally bounded in \( L^\infty \) and its covering numbers have polynomial growth, i.e., for some polynomial in two real variables \( p(y,z) \) we have the bound
\[
\mathcal{N}(\mathcal{F}_{u,v,\alpha}, \varepsilon) \leq p(\varepsilon^{-1}, t^{-1}).
\]

(2) There exists constants \( M_1, M_2, M_3 \) with \( M_i = M_i(\alpha, \rho, d) \) for \( i = 1, 2, 3 \), such that
\[
\Pr \left\{ \sup_{x \in \Sigma} \left| \tilde{\Gamma}^\alpha_t(u,v)(x) - \Gamma^\alpha_t(u,v)(x) \right| \geq \varepsilon \right\} \lesssim Q_t \left( \mathcal{F}_{u,v,\alpha}^t, \frac{t^{d/2}\varepsilon}{\|u\|_{\text{Lip}} \|v\|_{\text{Lip}}}, M_1, n \right) + \Pr \left\{ \sup_{x \in \Sigma} \left| \frac{1}{\theta_t^\alpha(x)} - \frac{1}{\tilde{\theta}_t^\alpha(x)} \right| \geq \varepsilon M_2 t^{-\frac{d}{2}(1-\alpha)} \right\} + \Pr \left\{ \sup_{x \in \Sigma} \left| \frac{1}{\theta_t^\alpha(x)} - \frac{1}{\tilde{\theta}_t^\alpha(x)} \right| \geq \varepsilon M_3 t^{-d/2} \right\}.
\]

(3) In particular, for \( \varepsilon > 0 \) fixed and \( \beta \in \mathbb{R} \) we obtain
\[
\Pr \left\{ \sup_{x \in \Sigma} \left| \tilde{\Gamma}^\alpha_t(u,v)(x) - \Gamma^\alpha_t(u,v)(x) \right| \geq \beta \varepsilon \right\} \in O_{BC}(2\beta + 3d + 1).
\]

**Proof.** Part (1) follows from arguments similar to those of Section 4.3. For part (2) we write
\[
(5.14)
\Gamma^\alpha_t(u,v)(x) - \tilde{\Gamma}^\alpha_t(u,v)(x) = \left( \frac{1}{t\theta_t^\alpha(x)} \mu_n \left( \frac{\Phi_{t,u,v}(x,\cdot)}{\theta_t^\alpha} \right) - \frac{1}{t\tilde{\theta}_t^\alpha(x)} \mu_n \left( \frac{\Phi_{t,u,v}(x,\cdot)}{\tilde{\theta}_t^\alpha} \right) \right) + \left( \frac{1}{t\theta_t^\alpha(x)} \mu_n \left( \frac{\Phi_{t,u,v}(x,\cdot)}{\theta_t^\alpha} \right) - \frac{1}{t\tilde{\theta}_t^\alpha(x)} \mu_n \left( \frac{\Phi_{t,u,v}(x,\cdot)}{\tilde{\theta}_t^\alpha} \right) \right) + \left( \frac{1}{t\theta_t^\alpha(x)} \mu_n \left( \frac{\Phi_{t,u,v}(x,\cdot)}{\theta_t^\alpha} \right) - \frac{1}{t\tilde{\theta}_t^\alpha(x)} \mu_n \left( \frac{\Phi_{t,u,v}(x,\cdot)}{\theta_t^\alpha} \right) \right)
\]
\[
(5.15)
= A_1(x) + A_2(x) + A_3(x).
\]

Let \( A_i^* \) be defined by \( A_i^* = \sup_{x \in \Sigma} |A_i(x)| \) for \( i = 1, 2, 3 \). In order to bound \( A_i^* \) we observe that
\[
(5.16)
\left\| \frac{\Phi_{u,v}}{t\theta_t^\alpha} \right\|_{L^\infty} \lesssim \|u\|_{\text{Lip}} \|v\|_{\text{Lip}} t^{-d\alpha/2},
\]
and since \( \theta_t^\alpha \approx t^{(1-\alpha)\frac{d}{2}} \) it follows that for some constant \( M_1 = M_1(\alpha, \rho, d) \)
\[
(5.17)
\Pr \left\{ A_1^* \geq \frac{\varepsilon}{3} \right\} = \Pr \left\{ \sup_{x \in \Sigma} \left| \frac{1}{t\theta_t^\alpha(x)} \mu_n \left( \frac{\Phi_{t,u,v}(x,\cdot)}{\theta_t^\alpha} \right) - \frac{1}{t\tilde{\theta}_t^\alpha(x)} \mu_n \left( \frac{\Phi_{t,u,v}(x,\cdot)}{\tilde{\theta}_t^\alpha} \right) \right| \geq \frac{\varepsilon}{3} \right\} \leq Q_t \left( \mathcal{F}_{\alpha,u,v}^t, \frac{t^{d/2}\varepsilon}{M_1}, M_1 \right).
\]
For $A_2$ we note that
\begin{equation}
|A_2(x)| = \left| \frac{1}{t^\alpha_t(x)} \Phi_{t,u,v}(x, \cdot) - \frac{1}{\theta_t(x)} \Phi_{t,u,v}(x, \cdot) \right|
\end{equation}
\begin{equation}
\leq \left\| \frac{\Phi_{u,v}}{t} \right\|_{L^\infty_{\Sigma}} \left( \sup_{x \in \Sigma} \left| \frac{1}{\theta_t(x)} - \frac{1}{\theta_t^\alpha(x)} \right| \right),
\end{equation}
and since $\left\| \frac{\Phi_{u,v}}{t} \right\|_{L^\infty_{\Sigma}} \lesssim \|u\|_{\text{Lip}} \|v\|_{\text{Lip}}$, and $\|\theta_t\|_{L^\infty} \approx t^\frac{d}{2}(1-\alpha)$ we have from Corollary 5.3 a bound of the form
\begin{equation}
\Pr \left\{ |A_2^*| \geq \frac{\varepsilon}{3} \right\} \leq \Pr \left\{ \sup_{x \in \Sigma} \left| \frac{1}{\theta_t(x)} - \frac{1}{\theta_t^\alpha(x)} \right| \geq \frac{\varepsilon M_2 t^\frac{d}{2}(1-\alpha)}{\|u\|_{\text{Lip}} \|v\|_{\text{Lip}}} \right\},
\end{equation}
for some constant $M_2 = M_2(\alpha, \rho, d)$. Finally, for bounding $A_3$, we can easily obtain uniform bounds (in particular independent of $x$) on the probability of the events
\begin{equation}
\{ \hat{\theta}_t(x) \geq \lambda^* t^{d/2} \},
\end{equation}
\begin{equation}
\{ \hat{\theta}_t(x) < \lambda t^{d/2} \},
\end{equation}
and from (5.18) we obtain a bound of the form
\begin{equation}
\Pr \left\{ |A_3^*| \geq \frac{\varepsilon}{3} \right\} \leq \Pr \left\{ \sup_{x \in \Sigma} \left| \frac{1}{\theta_t(x)} - \frac{1}{\theta_t^\alpha(x)} \right| \geq \frac{t^{d/2} M_3}{\|u\|_{\text{Lip}} \|v\|_{\text{Lip}}} \right\}
\end{equation}
for some $M_3 = M_3(\alpha, \rho, d)$ as needed. Part (3) is a straightforward consequence of part (2).
\hfill \Box

For the operator $L_t^\alpha$ a similar argument yields the following.

**Proposition 5.5.** Let $\varepsilon > 0$, let $u \neq 0$ be a function in $L^\infty(\Sigma)$ and let $A_{t,u}$ be a function defined by

$$A_{t,u}(x, \xi) = e^{-\frac{\|x-\xi\|^2}{2t}} u(\xi).$$

We have
\begin{enumerate}
\item The class of functions $\mathcal{G}_{t,\alpha} = \left\{ \frac{A_{t,u}(x)}{\theta_t^\alpha} : x \in \Sigma \right\}$ is totally bounded in $L^\infty$ and its covering numbers have polynomial growth.
\item There exists constants $M_1, M_2, M_3$ with $M_i = M_i(\alpha, \rho, d)$ for $i = 1, 2, 3$, such that
\end{enumerate}
\begin{equation}
\Pr \left\{ \sup_{x \in \Sigma} \left| \tilde{L}_t^\alpha u(x) - L_t^\alpha u(x) \right| \geq \varepsilon \right\} \lesssim Q_t \left( \mathcal{G}_{\alpha}^t, \frac{t^{d+1} \varepsilon}{\|u\|_{L^\infty}} , M_1, n \right)
\end{equation}
\begin{equation}
+ \Pr \left\{ \sup_{x \in \Sigma} \left| \frac{1}{\theta_t^\alpha(x)} - \frac{1}{\theta_t^\alpha(x)} \right| \geq \frac{t^{d+1} (1-\alpha) M_2 \varepsilon}{\|u\|_{L^\infty}} \right\}
\end{equation}
\begin{equation}
+ \Pr \left\{ \sup_{x \in \Sigma} \left| \frac{1}{\theta_t(x)} - \frac{1}{\theta_t(x)} \right| \geq \frac{t^{d+1} M_3 \varepsilon}{\|u\|_{L^\infty}} \right\},
\end{equation}
where $Q_t$ is as in Theorem 2.4.
(3) In particular, for \( \varepsilon > 0 \) fixed and \( \beta \in \mathbb{R} \) we obtain

\[
\Pr \left\{ \sup_{x \in \Sigma} \left| \hat{L}^\alpha_i u(x) - L^\alpha_i u(x) \right| \geq t^3 \varepsilon \right\} \in O_{BC}(2\beta + 3d + 3).
\]

We now prove Theorem 1.18.

5.1. Proof of Theorem 1.18 Consider the difference as

\[
2 \left( \Gamma_2(L^\alpha_i f, f)(x) - \hat{\Gamma}_2(L^\alpha_i f, f)(x) \right)
\]

\[
= \left( L^\alpha_i (\Gamma^\alpha_i f, f) - \hat{L}^\alpha_i (\Gamma^\alpha_i f, f) \right)(x)
\]

\[
+ 2 \left( \Gamma^\alpha_i (L^\alpha_i f, f) - \hat{\Gamma}^\alpha_i (L^\alpha_i f, f) \right)(x) + 2 \left( \hat{\Gamma}^\alpha_i (L^\alpha_i f, f) - \hat{\hat{\Gamma}}^\alpha_i (L^\alpha_i f, f) \right)(x)
\]

\[
= E_1(x) + E_2(x) + E_3(x) + E_4(x).
\]

We can immediately bound \( E_2 \) and \( E_4 \) in the following way. For \( E_2 \) we clearly have

\[
|E_2| \leq 2 \sup_{x \in \Sigma} \left| \Gamma^\alpha_i (f, f)(x) - \hat{\Gamma}^\alpha_i (f, f)(x) \right|,
\]

and for \( E_4 \)

\[
|E_4| \leq 2t^{-1/2} C_0 \sup_{x \in \Sigma} \left| L^\alpha_i f(x) - \hat{L}^\alpha_i f(x) \right| \frac{1}{\inf_{x \in \Sigma} \theta_{t,\alpha}(x)} \left\| \frac{1}{\theta^\alpha_i (x)} \right\|_{L^\infty},
\]

where \( C_0 = \sup_{\rho > 0} e^{-\rho^2} \rho \). For \( E_1 \) and \( E_3 \) we need \( L^\infty \) estimates on \( \Gamma^\alpha_i \) and Lipschitz estimates on \( L^\alpha_i \). An \( L^\infty \) bound for \( \Gamma^\alpha_i \) for sufficiently small \( t \) is

\[
\left\| \Gamma^\alpha_i (f, f) \right\|_{L^\infty} \leq 2 \left\| f \right\|_{L^\infty}^2,
\]

which again follows from the uniform convergence of \( \Gamma^\alpha_i \) to \( |\nabla f|^2 \) as \( t \to 0 \). In order to derive a Lipschitz estimate for \( L^\alpha_i \) we use the notation in Lemma 4.9 and let \( v_t(x, \xi) = e^{-\frac{t \xi^2}{2}} (f(\xi) - f(x)) \) so that

\[
L^\alpha_i (f, f)(x) - L^\alpha_i (f, f)(x') = \frac{\theta_{t,\alpha}(x') - \theta_{t,\alpha}(x)}{t \theta_{t,\alpha}(x) \theta_{t,\alpha}(x')} \int_{\Sigma} \frac{v_t(x, \xi)}{\theta^\alpha_i (\xi)} d\mu(\xi)
\]

\[
+ \frac{1}{t \theta_{t,\alpha}(x')} \int_{\Sigma} (v_t(x, \xi) - v_t(x', \xi)) \frac{\theta^\alpha_i (\xi)}{\theta_{t,\alpha}(x')} d\mu(\xi)
\]

\[
= \frac{\theta_{t,\alpha}(x) - \theta_{t,\alpha}(x')}{\theta_{t,\alpha}(x')} L^\alpha_i
\]

\[
+ \frac{1}{t \theta_{t,\alpha}(x')} \int_{\Sigma} (v_t(x, \xi) - v_t(x', \xi)) \frac{\theta^\alpha_i (\xi)}{\theta_{t,\alpha}(x')} d\mu(\xi)
\]
and for small $t$ we have
\begin{align*}
\| L^t \|_{L^\infty} &\lesssim \| \Delta f - 2(1 - \alpha) \langle Df, D\rho \rangle \|_{L^\infty}, \\
\frac{\| \theta_{t, \alpha}(x) - \theta_{t, \alpha}(x') \|_{L^\infty}}{\theta_{t, \alpha}(x')} &\lesssim t^{-\frac{d+4}{2}} \text{vol}(\Sigma), \\
\left\| \frac{1}{t \theta_{t, \alpha}(x')} \int_{\Sigma} \left( \frac{1}{\theta^\alpha(\xi)} \left( \frac{v_t(x, \xi) - v_t(x', \xi)}{\theta^\alpha(\xi)} \right) \right) \mu(\xi) \right\|_{L^\infty} &\lesssim \text{vol}(\Sigma) t^{-\frac{d+2}{2}} \left( \frac{2t}{\varepsilon} \| f \|_{\text{Lip}} + \| f \|_{C^1} \right) \| x - x' \|,
\end{align*}
where the inequality in (5.41)-(5.42) follows from Lemma 4.9. We conclude that
\begin{align*}
\| L^t f \|_{\text{Lip}} &\lesssim t^{-\frac{d}{2}} C(f, \rho, \Sigma),
\end{align*}
where $C(f, \rho, \Sigma)$ depends on the $C^2$ norm of $f$, the Lipschitz norm of $\rho$ and the volume of $\Sigma$. As in the uniform case, we let $E^*_i = \sup_{x \in \Sigma} | E_i(x) |$ and we consider the estimate
\begin{align*}
\text{Pr}\left\{ \Gamma_2(L^t(f, f)(\xi) - \hat{\Gamma}_2(L^t(f, f)(\xi)) \geq \varepsilon \right\} &\leq \text{Pr}\left\{ E^*_1 \geq \frac{\varepsilon}{4} \right\} + \text{Pr}\left\{ E^*_2 \geq \frac{\varepsilon}{4} \right\} + \text{Pr}\left\{ E^*_3 \geq \frac{\varepsilon}{4} \right\} + \text{Pr}\left\{ E^*_4 \geq \frac{\varepsilon}{4} \right\} \\
&= P_1 + P_2 + P_3 + P_4.
\end{align*}
From Proposition 5.5 and (5.34) it follows immediately that
\begin{align*}
P_1 &= O_{BC}(3d + 3),
\end{align*}
as $t \to 0$. From (5.32) and Proposition 5.4 we obtain
\begin{align*}
P_2 &= O_{BC}(3d + 3),
\end{align*}
as $t \to 0$. From (5.43) and Proposition 5.4 we have
\begin{align*}
P_3 &= O(4d + 3),
\end{align*}
as $t \to 0$. Finally, in order to bound $P_4$ we consider the bound for $E_4$ in (5.33) and observe that from the convergence of $\hat{\theta}_{t, \alpha}$ and $\hat{\theta}_{t}$ to $\theta_{t, \alpha}$ and $\theta_{t}$ respectively we have in probability a bound of the form
\begin{align*}
\inf_{x \in \Sigma} \hat{\theta}_{t, \alpha}(x) \left\| \frac{1}{\hat{\theta}^\alpha_t} \right\|_{L^\infty} &\approx t^{-d/2},
\end{align*}
so that from Proposition 5.5 we see that
\begin{align*}
P_4 &= O_{BC}(4d + 4)
\end{align*}
as $t \to 0$. Since $P_4$ is least decaying term in (5.46) we obtain
\begin{align*}
\text{Pr}\left\{ \sup_{\xi \in \Sigma} | \Gamma_2(L^t(f, f)(\xi) - \hat{\Gamma}_2(L^t(f, f)(\xi)) \geq \varepsilon \right\} &\in O_{BC}(4d + 4).
\end{align*}
With the choice of scale $t_n = n^{-\gamma}$ with $0 < \gamma < \frac{1}{4d+4}$ we see from the Borel-Cantelli lemma that

$$\sup_{\xi \in \Sigma} \left| \Gamma_2(L_t^\alpha(f,f))^{\alpha} - \tilde{\Gamma}_2(L_t^\alpha(f,f))^{\alpha} \right| \quad \text{a.s.} \to 0.$$  

(5.53)  

(5.54)

**REFERENCES**

[AGS] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré, *Bakry-Émery curvature-dimension condition and riemannian ricci curvature bound.*

[BBI01] Dmitri Burago, Yuri Burago, and Sergei Ivanov, *A course in metric geometry*, Graduate Studies in Mathematics, vol. 33, American Mathematical Society, Providence, RI, 2001. MR 1835418 (2002e:53053)

[BÉ85] D. Bakry and Michel Émery, *Diffusions hypercontractives*, Séminaire de probabilités, XIX, 1983/84, Lecture Notes in Math., vol. 1123, Springer, Berlin, 1985, pp. 177–206. MR 889476 (88j:60131)

[BN08] Mikhail Belkin and Partha Niyogi, *Towards a theoretical foundation for Laplacian-based manifold methods*, J. Comput. System Sci. 74 (2008), no. 8, 1289–1308. MR 2460286 (2009m:68207)

[BS09] Anca-Iuliana Bonciocat and Karl-Theodor Sturm, *Mass transportation and rough curvature bounds for discrete spaces*, J. Funct. Anal. 256 (2009), no. 9, 2944–2966. MR 2502429 (2010i:53066)

[CL06] Ronald R. Coifman and Stéphane Lafon, *Diffusion maps*, Appl. Comput. Harmon. Anal. 21 (2006), no. 1, 5–30. MR 2238665 (2008a:60210)

[CY96] F. R. K. Chung and S.-T. Yau, *Logarithmic Harnack inequalities*, Math. Res. Lett. 3 (1996), no. 6, 793–812. MR 1426537 (97k:58182)

[EM12] Matthias Erbar and Jan Maas, *Ricci curvature of finite Markov chains via convexity of the entropy*, Arch. Ration. Mech. Anal. 206 (2012), no. 3, 997–1038. MR 2989449

[FMMN13] Charles Fefferman, Sanjoy Mitter, and Hariharan Narayanan, *Testing the manifold hypothesis*, arXiv:1310.0425, 2013.

[GM13] Nicola Gigli and Jan Maas, *Gromov-Hausdorff convergence of discrete transportation metrics*, SIAM J. Math. Anal. 45 (2013), no. 2, 879–899. MR 3045651

[JKO98] Richard Jordan, David Kinderlehrer, and Felix Otto, *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal. 29 (1998), no. 1–17. MR 1617171 (2000b:35258)

[KPP14] Andreas Karcher, Richard Palais, and Bob Palais, *Point clouds: distributing points uniformly on a surface*, preprint, 2014.

[Kun69] Hiroshi Kunita, *Absolute continuity of Markov processes and generators*, Nagoya Math. J. 36 (1969), 1–26. MR 0250387 (40 #3626)

[LLY11] Yong Lin, Linyuan Lu, and Shing-Tung Yau, *Ricci curvature of graphs*, Tohoku Math. J. (2) 63 (2011), no. 4, 605–627. MR 2872958

[LV09] John Lott and Cédric Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. of Math. (2) 169 (2009), no. 3, 903–991. MR 2480619 (2010i:53068)

[LY10] Yong Lin and Shing-Tung Yau, *Ricci curvature and eigenvalue estimate on locally finite graphs*, Math. Res. Lett. 17 (2010), no. 2, 343–356. MR 2644381 (2011e:05068)

[McC97] Robert J. McCann, *A convexity principle for interacting gases*, Adv. Math. 128 (1997), no. 1, 153–179. MR 1451422 (98e:82003)

[Mie13] Alexander Mielke, *Geodesic convexity of the relative entropy in reversible Markov chains*, Calc. Var. Partial Differential Equations 48 (2013), no. 1-2, 1–31. MR 3090532
[Nab13] Aaron Naber, *Characterization of bounded ricci curvature on smooth and nonsmooth settings*, arXiv:1306.6512, 2013.

[Oll09] Yann Ollivier, *Ricci curvature of Markov chains on metric spaces*, J. Funct. Anal. 256 (2009), no. 3, 810–864. MR 2484937 (2010j:58081)

[Oll10] ———, *A survey of Ricci curvature for metric spaces and Markov chains*, Probabilistic approach to geometry, Adv. Stud. Pure Math., vol. 57, Math. Soc. Japan, Tokyo, 2010, pp. 343–381. MR 2648269 (2011d:58087)

[Rot74] 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 (1974), 1103–1106. MR 0350047 (50 #2540)

[Sim96] 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. MR 1399562 (98c:58042)

[Stu06a] Karl-Theodor Sturm, *On the geometry of metric measure spaces. I*, Acta Math. 196 (2006), no. 1, 65–131. MR 2237206 (2007k:53051a)

[Stu06b] ———, *On the geometry of metric measure spaces. II*, Acta Math. 196 (2006), no. 1, 133–177. MR 2237207 (2007k:53051b)

[SW12] A. Singer and H.-T. Wu, *Vector diffusion maps and the connection Laplacian*, Comm. Pure Appl. Math. 65 (2012), no. 8, 1067–1144. MR 2928092

[SW13] Amit Singer and Hau-Tieng Wu, *Spectral convergence of the connection laplacian from random samples*, arXiv:1306.1587, 2013.

[Syn31] J. L. Synge, *A Characteristic Function in Riemannian Space and its Application to the Solution of Geodesic Triangles*, Proc. London Math. Soc. S2-32 (1931), no. 1, 241. MR 1575991

[vdVW96] 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. MR 1385671 (97g:60035)

[Vil09] Cédric Villani, *Optimal transport*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338, Springer-Verlag, Berlin, 2009, Old and new. MR 2459454 (2010f:49001)

[vRS05] Max-K. von Renesse and Karl-Theodor Sturm, *Transport inequalities, gradient estimates, entropy, and Ricci curvature*, Comm. Pure Appl. Math. 58 (2005), no. 7, 923–940. MR 2142879 (2006j:53048)