APPROXIMATING COARSE RICCI CURVATURE ON METRIC MEASURE SPACES WITH APPLICATIONS TO SUBMANIFOLDS OF EUCLIDEAN SPACE

ANTONIO G. ACHE AND MICAH W. WARREN

Abstract. For a submanifold

$$\Sigma \subset \mathbb{R}^N$$

Belkin and Niyogi showed that one can approximate the Laplacian operator using heat kernels. Using a definition of coarse Ricci curvature derived by iterating Laplacians, we approximate the coarse Ricci curvature of submanifolds Σ in the same way. More generally, on any metric measure space, we are able to approximate a 1-parameter family of coarse Ricci functions that include the coarse Bakry-Emery Ricci curvature.

Contents

1. Introduction 1
2. Coarse Ricci Curvature 2
2.1. Iterated Carré du Champ 2
2.2. Coarse Ricci Curvature 3
2.3. Statement of Results 5
3. Bias Error Estimates 6
3.1. Bias for Submanifold of Euclidean Space 6
3.2. Proof of Proposition 3.1 7
3.3. Bias for Smooth Metric Measure Space with a Density 10
3.4. Convergence of Coarse Ricci to Actual Ricci on Smooth submanifolds 11
References 12

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. (See also [HAvL05].) Our goal in this paper is to demonstrate that this process can be continued to approximate Ricci curvature as well. This is a step towards answering a question of Singer and Wu [SW12, pg. 1103] , in principle allowing one to approximate the Hodge Laplacian on

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

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

1-forms. The Hodge Laplacian allows one to extract certain topological information, thus we expect our result to have applications to manifold learning.

To do this, we need to use a modified notion of coarse Ricci curvature defined in [AW17]. Coarse Ricci curvature is a quantity that is derived from a Laplace-type operator and 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 corresponding smooth Ricci curvature operators as $t \to 0$.

Our goal is to reconstruct the Ricci using the distance function on the ambient space, and approximation of the Laplacian. A problem arises in that the ambient distance squared function manifests an error at fourth order, see [BN08, Lemma 4.3] Because the definition in [AW17] requires five derivatives to recover Ricci tensor, we have to modify this to a quantity that recovers the tensor using only three derivatives.

More specifically, by iterating the approximate Laplacian operators in [BN08] one can construct an approximate Γ_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 Ric_∞ tensor and can be modified to recover the standard Ricci curvature as well, provided the volume density is smooth.

In this paper we accomplish two things. First, following [BN08] we define a coarse Ricci operator at scale t on any metric measure space. Second, we show that these operators converge to the intrinsic coarse Ricci, as $t \to 0$, when taken on a fixed smooth submanifold. In [AW] we show 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.

The motivation for the paper stems from both the theory of Ricci lower bounds on metric measure spaces and the theory of manifold learning. For a more extensive background on the motivation for coarse Ricci and relation to Ricci curvature lower bounds and some other motivating problems, see [AW] and [AW17].

2. Preliminaries and statement of results

2.1. **Iterated Carré du Champ.** Given an operator L we define the Carré du champ as follows.

$$\Gamma(L,u,v) = \frac{1}{2}\left(L(uv) - L(u)v - uL(v)\right). \tag{2.1}$$

We will also consider the *iterated Carré du Champ* introduced by Bakry and Emery [BÉ85] denoted by Γ_2 and defined by

$$\Gamma_2(L,u,v) = \frac{1}{2}\left(L(\Gamma(L,u,v)) - \Gamma(L,Lu,v) - \Gamma(L,u,Lv)\right). \tag{2.2}$$

When L is the rough Laplacian with respect to the metric g, then

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

Notation 2.1. When considering the operators (2.1) and (2.2) we will use the slightly cumbersome three-parameter notation, as the main results will be stated in terms of a family of operators $\{L_t\}$.

2.2. **Coarse Ricci Curvature.** In this section we provide a definition of coarse Ricci curvature on general metric measures spaces, using a family of operators which are intended to approximate a Laplace operator on a space at scale t. The coarse Ricci curvature will then be defined on pairs of points. To obtain a quantity for the operator (2.2), we need a function to evaluate. For submanifolds in Euclidean space, we use a linear function whose gradient is the vector that points from a point x to a point y. On a genereal metric space X, given $x, y \in X$ define

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

Note that in Euclidean space this is

$$f_{x,y}(z) = \langle y - x, z\rangle. \tag{2.3}$$

This leads us to the following definition of coarse Ricci curvature.

Definition 2.2. *Given an operator L we define the coarse Ricci curvature for L as*

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

We recall the main results from [AW17].

Theorem 2.3. *Let*

$$\Delta_\rho v = \Delta_g v - \langle \nabla\rho, \nabla v\rangle_g$$

be the wieghted Laplacian and let

$$\mathrm{Ric}_\infty = \mathrm{Ric} + \nabla_g^2 \rho$$

Then

$$\mathrm{Ric}_\infty(\gamma'(0), \gamma'(0)) = \frac{1}{2}\frac{d^2}{ds^2}\mathrm{Ric}_{\Delta_\rho}(x, \gamma(s)), \tag{2.4}$$

and

$$\mathrm{Ric}_\infty \geq K \tag{2.5}$$

if and only if

$$\mathrm{Ric}_{\Delta_\rho}(x,y) \geq K d^2(x,y).$$

As mentioned in the introduction, the ambient distance squared function osculates the instrinsic distance squared function only to third order on the diagonal along the submanifold. So the above formula could manifest some error terms. To side-step this, we appeal to the Bochner formula, which says

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

We note that if we evaluate Γ_2 on functions with vanishing Hessian at a point, we can recover the Ricci curvature exactly. For submanifolds in Euclidean space, we normalize the functions (2.3) to linear function whose gradient is the *unit* vector that points from a point x to a point y. In particular, given x, y

$$F_{x,y}(z) = \frac{1}{2}\frac{d^2(y,z) - d^2(x,z) - d^2(x,y)}{d(x,y)}.$$

That is

$$F_{x,y}(z) = \langle \frac{y-x}{|y-x|}, z\rangle. \tag{2.6}$$

This leads us to the following definition of *life-sized* coarse Ricci curvature.

Definition 2.4. *Given an operator L we define the life-sized coarse Ricci curvature for L as*

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

As we will see, this also can be used to recover the Ricci curvature, without taking any derivatives.

2.2.1. *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 a slight modification of the approximation constructed by Belkin-Niyogi in [BN08] and more generally Coifman-Lafon in [CL06]. Consider a metric measure space (X, d, μ) with the Borel σ-algebra such that $\mu(X) < \infty$. Given $t > 0$, let θ_t be given by

$$\theta_t(x) = \int_X e^{-\frac{d^2(x,y)}{2t}} d\mu(y). \tag{2.7}$$

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

$$L_t f(x) = \frac{2}{t\theta_t(x)} \int_X (f(y) - f(x))\, e^{-\frac{d^2(x,y)}{2t}} d\mu(y). \tag{2.8}$$

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

$$\Gamma(L_t, f, h) = \frac{1}{2}\left(L_t(fh) - (L_t f)h - f(L_t h)\right), \tag{2.9}$$

which simplifies to

$$\Gamma(L_t, f, h)(x) = \frac{1}{t\theta_t(x)} \int_X e^{-\frac{d^2(x,y)}{2t}} (f(y) - f(x))(h(y) - h(x)) d\mu. \tag{2.10}$$

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

$$\Gamma_2(L_t, f, h) = \frac{1}{2}\left(L_t(\Gamma(L_t, f, h)) - \Gamma(L_t, L_t f, h) - \Gamma(L_t, f, L_t h)\right). \tag{2.11}$$

Remark 2.5. *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.*

2.3. **Statement of Results.** We will consider a closed, smooth, embedded submanifold Σ 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 Σ into $\mathbb{R}^N$.

In addition we will adopt the following conventions

- All operators L_t, $\Gamma(L_t, \cdot, \cdot)$ and $\Gamma_2(L_t, \cdot, \cdot)$ will be taken with respect to the distance $\|\cdot\|$ and the measure $d\text{vol}_\Sigma$.

The choice of the above metric measure space is consistent with the setting of manifold learning in which no assumption on the geometry of the submanifold Σ is made, in particular, we have no a priori knowledge of the geodesic distance and therefore we can only hope to use the chordal distance as a reasonable approximation for the geodesic distance. We will show that while our construction at a scale t involves only information from the ambient space, the limit as t tends to 0 will recover the life-size coarse 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 smooth submanifold differ first at fourth order near a point , so while much of the analysis is done on submanifolds, the intrinsic geometry will be recovered in the limit. We are able to show the following.

Theorem 2.6. *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 Σ and the function f such that*

$$\sup_{x\in\Sigma} |\Gamma_2(\Delta_g, f, f)(x) - \Gamma_2(L_t, f, f)(x)| < C_1(\Sigma, D^5 f) t^{1/2}. \tag{2.12}$$

Theorem 2.6 will follow from Corollary 3.2 which is proved in Section 3.

Corollary 2.7. *With the hypotheses of Theorem 2.6 we have*

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

Theorem 2.6 applies to all functions on the manifold. To obtain the life-size Ricci curvature we apply these to $F_{x,y}$ to obtain the following.

Theorem 2.8. *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 Σ such that $\gamma(0) = x$. There exists constants C_2, C_3 depending on the geometry of Σ such that*

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

This will be proved in section 3.4.

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

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

In [CL06], the authors consider a family of operators L_t^α which converge to $\triangle_{2(1-\alpha)\rho}$. Note that a standard computation (cf [Vil09, Page 384]) gives

$$\Gamma_2(\triangle_{2(1-\alpha)\rho}, f, f) = \frac{1}{2}\Delta_g \|\nabla f\|_g^2 - \langle \nabla\rho, \nabla\Delta_g f\rangle_g + 2(1-\alpha)\nabla_g^2\rho(\nabla f, \nabla f). \tag{2.13}$$

We adapt [CL06] to our setting: Recall that

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

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

$$\theta_{t,\alpha}(x) = \int_X e^{-\frac{d^2(x,y)}{2t}} \frac{1}{[\theta_t(y)]^\alpha} d\mu(y). \tag{2.15}$$

We can define the operator

$$L_t^\alpha f(x) = \frac{2}{t}\frac{1}{\theta_{t,\alpha}(x)} \int e^{-\frac{d^2(x,y)}{2t}} \frac{1}{[\theta_t(y)]^\alpha} \left(f(y) - f(x)\right) d\mu(y) \tag{2.16}$$

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

Theorem 2.9. *Let $\Sigma^d \subset \mathbb{R}^N$ be an embedded submanifold and consider the smooth metric measure space $(\Sigma, \|\cdot\|, e^{-\rho}d\mathrm{vol}_\Sigma)$. Let $f \in C^5(\Sigma)$ such that $\|f\|_{C^5} \leq M$. There exists $C_4 = C_4(\Sigma, M, \rho)$ such that*

$$\sup_{\xi\in\Sigma} \left|\Gamma_2(L_t^\alpha, f, f)(\xi) - \Gamma_2(\triangle_{2(1-\alpha)\rho}, f, f)(\xi)\right| \leq C_4 t^{1/2}. \tag{2.17}$$

In particular, if the density is positive and smooth enough, we can still recover the Ricci curvature, using $\alpha = 1$ in the above expression.

3. Bias Error Estimates

3.1. **Bias for Submanifold of Euclidean Space.** In this section we prove Theorem 2.6. The theorem will follow from Proposition 3.1 and Corollary 3.2 below. For simplicity we will assume that $(\Sigma, d\mathrm{vol}_\Sigma)$ has unit volume. Recall the definitions (2.7), (2.8), (2.9), (2.10) and (2.11).

Proposition 3.1. *Suppose that Σ^d is a closed, embedded, unit volume submanifold of $\mathbb{R}^N$. For any x in Σ and for any functions f, h in $C^5(\Sigma)$ we have*

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

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

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

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

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 3.2. *We have the following expansions*

$$L_t(\Gamma(L_t, f, f))(x) = \Delta_g \|\nabla f(x)\|_g^2 + t^{1/2}R_4(x, J^5(f)(x)), \tag{3.5}$$

$$\Gamma_t(L_t f, f)(x) = \langle \nabla \Delta_g f(x), \nabla f(x)\rangle + t^{1/2}R_5(x, J^5(f)(x)). \tag{3.6}$$

3.2. **Proof of Proposition 3.1.** Our first goal is to fix a local structure which we will use to define the quantities G_i and R_i and $J^k(u)$ that appear in Proposition 3.1. 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 for nearby points in Σ 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 Σ as the graph of a function U_y over the tangent plane $T_y\Sigma$. Each U_y will satisfy

$$U_y(0) = 0, \tag{3.7}$$

$$DU_y(0) = 0. \tag{3.8}$$

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 Σ. Fixing a point x, consider a function f on Σ. The function f is locally well defined as a function over the tangent plane, i.e.

$$f(y) = f(y, U_x(y)) \text{ for } y \in T_x\Sigma. \tag{3.9}$$

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

$$J^m f(x) = (f(x), Df(x), ..D^m f(x)). \tag{3.10}$$

More concretely, all derivatives in (3.10) are taken with respect to the variable y in (3.9). Since Σ 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

$$\Sigma_{y,\tau} = \Sigma \cap \{(z, U_y(z)), y \in B_\tau(0) \subset T_y\Sigma\}, \tag{3.11}$$

in other words, Σ_{y,τ_0} is the part of Σ 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

$$B_{\mathbb{R}^N,\tau_0}(y) \cap \Sigma \subset \Sigma_{y,\tau_0}. \tag{3.12}$$

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

$$\int_\Sigma f(y) e^{-\frac{\|x-y\|^2}{2t}} d\mu(y) = \int_{\Sigma_{x,\tau_0}} f(y) e^{-\frac{\|x-y\|^2}{2t}} d\mu(y) \tag{3.13}$$

$$+ \int_{\Sigma \setminus \Sigma_{x,\tau_0}} f(y) e^{-\frac{\|x-y\|^2}{2t}} d\mu(y), \tag{3.14}$$

and by (2)

$$\int_{\Sigma \setminus \Sigma_{x,\tau_0}(x)} f(y) e^{-\frac{\|x-y\|^2}{2t}} d\mu(y) \leq \|f\|_{L^\infty(\Sigma)} e^{-\frac{\tau_0^2}{2t}}. \tag{3.15}$$

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

$$\left| \int_{\mathbb{R}^d \setminus B_{\tau_0/\sqrt{t}}} e^{-\|z\|^2/2} p(z) dz \right| \leq C(p) e^{-\frac{\tau_0^2}{2t}}. \tag{3.16}$$

The volume form over $T_x\Sigma$ will be

$$\mu_x(z) dz = \sqrt{\det\left(\delta_{ij} + \langle D_i U_x(z), D_j U_x(z) \rangle\right)} dz. \tag{3.17}$$

If in (2.7) we choose our distance to be the ambient distance $\|\cdot\|$ in $\mathbb{R}^N$ and the measure μ to be the volume measure in Σ, the density $\theta_t(x)$ takes the form

$$\theta_t(x) = \int_\Sigma e^{-\frac{\|x-y\|^2}{2t}} d\mu(y). \tag{3.18}$$

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 (3.1). Observe that

$$\begin{aligned}
\theta_t(x) - &\int_{\Sigma\setminus\Sigma_{x,\tau_0}} e^{-\frac{\|x-z\|^2}{2t}} d\mu(z) \\
&= \int_{B_{\tau_0}} e^{-\left(\|z\|^2+\|U_x(z)\|^2\right)/2t}\mu_x(z)dz \\
&= t^{n/2}\int_{B_{\tau_0/\sqrt{t}}} e^{-\|w\|^2/2}e^{-\left\|U_x(\sqrt{t}w)\right\|^2/2t}\mu_x(\sqrt{t}w)dw \\
&= t^{n/2}\int_{\mathbb{R}^d} e^{-\|w\|^2/2}e^{-\left\|U_x(\sqrt{t}w)\right\|^2/2t}\mu_x(\sqrt{t}w)dw \\
&\quad - t^{n/2}\int_{\mathbb{R}^d\setminus B_{\tau_0/\sqrt{t}}} e^{-\|w\|^2/2}e^{-\left\|U_x(\sqrt{t}w)\right\|^2/2t}\mu_x(\sqrt{t}w)dw.
\end{aligned}$$

Now considering (3.17), (3.8) we have

$$\mu_x(\sqrt{t}w) = 1 + tT_2\mu_x(0)(w) + t^{3/2}R_2\mu_x(0,w), \tag{3.19}$$

$$e^{-\frac{\|U_x(\sqrt{t}w)\|^2}{2t}} = 1 + tT_4\left[e^{-\left\|U_x(\sqrt{t}\cdot)\right\|^2/2t}\right](0)(w) + t^{3/2}R_4\left[e^{-\left\|U_x(\sqrt{t}\cdot)\right\|^2/2t}\right](0,w). \tag{3.20}$$

Expanding, collecting lower order terms, integrating and absorbing the exponentially decaying terms into $t^{3/2}R_1(x,z)$ using (3.15) and (3.16) yields (3.1).

Next we prove (3.4). First compute

$$\int_\Sigma (f(y)-f(x))\, e^{-\frac{\|x-y\|^2}{2t}} d\mu(y).$$

$$\int_{B_{\tau_0}} (f(y)-f(x))\, e^{-\frac{\|x-y\|^2}{2t}}\mu_x(y)dy = t^{d/2}\int_{B_{\tau_0/\sqrt{t}}} e^{-\frac{\|z\|^2}{2}}\left(f(\sqrt{t}z)-f(0)\right)\mu_x(\sqrt{t}z)dz$$

Now,

$$f(\sqrt{t}z) - f(0) = \sqrt{t}T_1f(z) + tT_2f(z) + t^{3/2}T_3f(z) + t^2T_4f(z) + t^{5/2}R_4(z) \tag{3.21}$$

and also recall (3.19). (3.20). Note that if A is a symmetric $d\times d$ matrix we have the identity

$$\int_{\mathbb{R}^d} e^{-\frac{\|z\|^2}{2}} z^TAzdz = (2\pi)^{d/2}\mathrm{tr}(A), \tag{3.22}$$

where tr denotes Trace. From this it follows that

$$\int_{\mathbb{R}^d} e^{-\frac{\|z\|^2}{2}} T_2f(0)(z)dz = (2\pi)^{d/2}\mathrm{tr}(T_2f(0)). \tag{3.23}$$

Again, expanding, collecting lower order terms, integrating odd and even terms, and absorbing the exponentially decaying terms via (3.15) and (3.16) yields

$$\begin{aligned}
t^{-d/2}\int_\Sigma (f(y)-f(x))\, e^{-\frac{\|x-y\|^2}{2t}} d\mu(y) &= t\,(2\pi)^{d/2}\,\Delta_g f + t^{3/2}G_4(J^3f(x)) \\
&\quad + t^2G_5(J^4f(x)) + t^{5/2}R_3(J^5f(x)).
\end{aligned}$$

Combining with (3.1) yields (3.4). A very similar calculation yields (3.2).

Proof of Corollary 3.2. Directly from (3.2)

$$L_t(\Gamma(L_t, f, f))(x) = L_t\left\{\|\nabla f\|^2(x) + t^{1/2}G_2(x, J^2(f)(x)) + tG_3(x, J^3(f)(x))\right\} \tag{3.24}$$

$$+ t^{3/2}L_tR_2(x, J^4(f)(x)) \tag{3.25}$$

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

$$\begin{aligned} t^{3/2}\left|L_tR(x)\right| &= t^{3/2}\left|\frac{1}{t}\frac{2}{\theta_t(x)}\int (R(y)-R(x))\, e^{-\frac{\|x-y\|^2}{2t}}d\mu(y)\right| \\ &\leq \left|t^{1/2}\frac{2}{\theta_t(x)}2\left\|R\right\|_{L^\infty}\int e^{-\frac{\|x-y\|^2}{2t}}d\mu(y)\right| \\ &= 4t^{1/2}\left\|R\right\|_{L^\infty}. \end{aligned}$$

The first three terms are differentiable, so can be dealt with directly by (3.4), giving an expression involving $J^5(f)(x)$. The estimate (3.6) follows from a similar argument. The result follows by combining the above lemmata for the first term, and then directly bounding the second term. □

3.3. **Bias for Smooth Metric Measure Space with a Density.** The bias estimate Theorem 2.6 for a metric measure space with density will follow from the following proposition whose proof is very similar to that of Proposition 3.1. Recall definitions (2.15) and (2.16) .

Proposition 3.3. *Let $f \in C^5$. We have the following expansions*

$$\begin{aligned} L_t^\alpha f(x) =& \Delta_g f(x) + (1-\alpha)\langle \nabla f(x), \nabla\rho(x)\rangle_g \\ &+ t^{1/2}G_1(x, J^3(f)) + t^{3/2}R_1(x, J^5(f)) \end{aligned} \tag{3.26}$$

$$\Gamma^\alpha(L_t, f, h)(x) = \langle \nabla f, \nabla h\rangle_g + t^{1/2}G_2(x, J^2(f), J^2(h)) + t^{3/2}R_2(x, J^4(f), J^4(h)). \tag{3.27}$$

Proof. Following the proof of Proposition 3.1 we have the following expansions

$$\theta_t(x) = (2\pi t)^{d/2}e^{-\rho(x)}\left(1 + tG_1(x,\rho) + t^{3/2}R_1(x,\rho)\right), \tag{3.28}$$

$$\theta_{t,\alpha}(x) = (2\pi t)^{(1-\alpha)d/2}e^{(\alpha-1)\rho(x)}\left(1 + tG_2(x,\rho)R_2(x,\rho)\right), \tag{3.29}$$

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

$$\frac{d\mu_x(z)}{\theta_t(z)^\alpha} = e^{(\alpha-1)\rho(0)}\left(1 + (1-\alpha)\langle D\rho(0), z\rangle + O(\|z\|^2)\right)dz, \tag{3.30}$$

which holds in a small neighborhood of 0. The rest of the proposition follows from a straightforward computation. □

Corollary 3.4. *We have the following expansions*

$$\begin{aligned} L_t^\alpha(\Gamma^\alpha(L_t^\alpha, f, h))(x) &= \Delta\langle\nabla f, \nabla h\rangle_g(x) + (1-\alpha)\langle\nabla\rho, \nabla\langle\nabla f, \nabla h\rangle_g\rangle_g(x) \\ &\quad + t^{1/2}R_3(x, J^5(f), J^5(h), J^5(\rho)), \\ \Gamma^\alpha(L_t^\alpha, L_t^\alpha f, h)(x) &= \langle\nabla\Delta_g f, \nabla h\rangle_g + (1-\alpha)\langle\nabla\langle\nabla\rho, \nabla f\rangle_g, \nabla h\rangle_g \\ &\quad + t^{1/2}R_4(x, J^5(f), J^5(h), J^5(\rho)) \end{aligned}$$

as $t \to 0$.

From Corollary 3.4 we obtain Theorem 2.9.

3.4. **Convergence of Coarse Ricci to Actual Ricci on Smooth submanifolds.** We now prove Theorem 2.8:

Proof of Theorem 2.8 . Our goal is to show that

$$|\mathrm{Ric}(\gamma'(0), \gamma'(0)) - \mathrm{RIC}_{L_t}(x, \gamma(s))| \leq C_1 t^{1/2} + C_2 s.$$

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

(3.31)

$$|\mathrm{RIC}_{L_t}(x, \gamma(s)) - \mathrm{Ric}(\gamma'(0), \gamma'(0))| = |\Gamma_2(L_t, f_s, f_s)(x) - \mathrm{Ric}(\gamma'(0), \gamma'(0))|$$

$$= \left| \begin{array}{c} \Gamma_2(L_t, f_s, f_s)(x) - \Gamma_2(\triangle_g, f_s, f_s)(x) \\ +\Gamma_2(\triangle_g, f_s, f_s)(x) - \Gamma_2(\triangle_g, f_0, f_0)(x) \\ +\Gamma_2(\triangle_g, f_0, f_0)(x) - \mathrm{Ric}(\gamma'(0), \gamma'(0)) \end{array} \right| \tag{3.32}$$

$$\leq R(J^5 f_s)t^{1/2} + C_7(\Sigma)s. \tag{3.33}$$

Here we have used the following 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 3.2.

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

(3.34)

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

$$+ \|f_0\|_{C^3}\, \|f_s - f_0\|_{C^1} + \|f_s\|_{C^1}\, \|f_s - f_0\|_{C^3}\,. \tag{3.35}$$

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

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

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

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

The curvature of any geodesic inside Σ is controlled by the geometry of Σ.

Finally, because

$$\nabla f_0 = \gamma'(0)$$

we have by the Bochner formula

$$\Gamma_2(\triangle_g, f_0, f_0)(x) - \mathrm{Ric}(\gamma'(0), \gamma'(0)) = \left\|\nabla^2 f_0\right\|^2. \tag{3.36}$$

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 (3.36) vanishes.

□

References

[AW] A. Ache and M. Warren, *Coarse Ricci curvature and the manifold learning problem*, arXiv:1410.3351.

[AW17] Antonio G. Ache and Micah W. Warren, *Coarse Ricci curvature as a function on $M \times M$*, Results Math. **72** (2017), no. 4, 1823–1837. MR 3735525

[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)

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

[HAvL05] Matthias Hein, Jean-Yves Audibert, and Ulrike von Luxburg, *From graphs to manifolds—weak and strong pointwise consistency of graph Laplacians*, Learning theory, Lecture Notes in Comput. Sci., vol. 3559, Springer, Berlin, 2005, pp. 470–485. MR 2203281 (2006i:68108)

[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

[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)

Mathematics Department, University of Notre Dame, Notre Dame, Indiana, 46556 USA
Email address: aache@nd.edu

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