Available online at www.sciencedirect.com

ScienceDirect

Journal of Computer and System Sciences 74 (2008) 1289–1308

JOURNAL OF COMPUTER AND SYSTEM SCIENCES

www.elsevier.com/locate/jcss

Towards a theoretical foundation for Laplacian-based manifold methods

Mikhail Belkin [a], Partha Niyogi [b,*]

[a] *Ohio State University, Department of Computer Science and Engineering, USA*
[b] *University of Chicago, Departments of Computer Science and Statistics, USA*

Received 15 September 2006; received in revised form 11 May 2007

Available online 30 August 2007

Abstract

In recent years manifold methods have attracted a considerable amount of attention in machine learning. However most algorithms in that class may be termed "manifold-motivated" as they lack any explicit theoretical guarantees. In this paper we take a step towards closing the gap between theory and practice for a class of Laplacian-based manifold methods. These methods utilize the graph Laplacian associated to a data set for a variety of applications in semi-supervised learning, clustering, data representation.

We show that under certain conditions the graph Laplacian of a point cloud of data samples converges to the Laplace–Beltrami operator on the underlying manifold. Theorem 3.1 contains the first result showing convergence of a random graph Laplacian to the manifold Laplacian in the context of machine learning.

Keywords: Laplace–Beltrami operator; Graph Laplacian; Manifold methods

1. Introduction

Manifold methods have become increasingly important and popular in machine learning and have seen numerous recent applications in data analysis including dimensionality reduction, visualization, clustering and classification. The central modeling assumption in all of these methods is that the data resides on or near a low-dimensional submanifold in a higher-dimensional space. It should be noted that such an assumption seems natural for a data-generating source with relatively few degrees of freedom.

However in almost all modeling situations, one does not have access to the underlying manifold but instead approximates it from a point cloud. The most common approximation strategy in these methods it to construct an adjacency graph associated to a point cloud. Most manifold learning algorithms then proceed by exploiting the structure of this graph. The underlying intuition has always been that since the graph is a proxy for the manifold, inference based on the structure of the graph corresponds to the desired inference based on the geometric structure of the manifold. However few theoretical results are available to justify this intuition.

* Corresponding author.
E-mail addresses: mbelkin@cse.ohio-state.edu (M. Belkin), niyogi@cs.uchicago.edu (P. Niyogi).

0022-0000/$ – see front matter © 2007 Published by Elsevier Inc.
doi:10.1016/j.jcss.2007.08.006

An important and popular class of learning methods makes use of the graph Laplacian for various learning applications. It is worth noting that in almost all cases, the graph itself is an empirical object, constructed as it is from sampled data. Therefore any graph-theoretic technique is justifiable only when it can be related to the underlying process generating the data.

In this paper we take the first steps towards a theoretical foundation for manifold-based methods in learning, by showing that under certain conditions the graph Laplacian is directly related to the manifold Laplace–Beltrami operator and converges to it as the amount of data goes to infinity.

1.1. Prior work

Many manifold and graph-motivated learning methods have been recently proposed for various problems and applications, including [3,11,13,25,31] for visualization and data representation, [2,5,9,26,30,33,36,37] for partially supervised classification and [16,20,27,29,32], among others, for spectral clustering. A discussion of various spectral methods and their out-of-sample extensions is given in [6].

The problem of estimating geometric and topological invariants from point cloud data has also recently attracted some attention. Some of the recent work includes estimating geometric invariants of the manifold, such as homology [23,38], geodesic distances [7], and comparing point clouds using Gromov–Hausdorff distance [17].

In this paper we show how the Laplace–Beltrami operator on the manifold can be approximated from point cloud data. More specifically, we show that for the uniform distribution on the manifold, the graph Laplacian converges to the Laplace–Beltrami operator as the number of points increases and the kernel bandwidth is selected appropriately. This convergence is uniform over a class of functions and over points of the manifold. We then show that the same argument also works for an arbitrary probability distribution, when the graph Laplacian converges to a slightly different weighted version of the Laplace–Beltrami operator. We also consider the case of the normalized Laplacian and prove a stronger uniform convergence type result.

This paper grew out of an attempt to analyze the Laplacian Eigenmaps algorithm of [3] and is an extended version of the COLT 2005 paper [4]. Our extensions include a more complete proof of the uniform convergence results for the graph Laplacian as well as a discussion of how the ingredients of the basic proof need to be modified to handle the case of an arbitrary probability distribution. The first results on the convergence of the graph Laplacian in a manifold setting were presented in the PhD thesis [1]. Theorem 3.1 of this paper is a probabilistic version of those results showing the convergence of the empirical graph Laplacian and was first presented in [21].

Those results were subsequently generalized to the case of the arbitrary probability distribution in the 2004 PhD thesis of Stefan Lafon [18] (see also [11]). Further generalization and an empirical convergence result was presented in [15] and an improved rate in [28]. A more subtle analysis of uniform convergence has recently been made in [14]. We also note the closely related work [34], where similar functional objects were studied in a different context. In a non-geometric setting convergence of the graph Laplacian was observed in [8]. We also note [19], where convergence of spectral properties of graph Laplacians, such as their eigenvectors and eigenvalues, was demonstrated in a non-geometric setting for a fixed kernel bandwidth. Connections to geometric invariants of the manifold, such as the Laplace–Beltrami operator, were not considered in that work.

Finally we point out that while the parallel between the geometry of manifolds and the geometry of graphs is well-known in spectral graph theory and in certain areas of differential geometry (see, e.g., [10]) the nature of that parallel is usually not made mathematically precise.

2. Basic objects

We will now introduce the main setting of the paper and the basic objects of study. We will consider a compact smooth manifold $\mathcal{M}$ isometrically embedded in some Euclidean space $\mathbb{R}^N$. This embedding induces a measure corresponding to a volume form μ on the manifold. For example, the volume form for a closed curve, i.e. an embedding of a circle S^1, measures the usual curve length in $\mathbb{R}^N$.

The Laplace–Beltrami operator $\Delta_{\mathcal{M}}$ is a key geometric object associated to a Riemannian manifold. Given $p \in \mathcal{M}$, the tangent space $T_p\mathcal{M}$ can be identified with the affine space of vectors tangent to $\mathcal{M}$ at p. This vector space has the natural inner product induced by the embedding $\mathcal{M} \subset \mathbb{R}^N$.

Given a differentiable function $f:\mathcal{M}\to\mathbb{R}$, let $\nabla_{\mathcal{M}} f$ be the gradient vector field on the manifold. Recall that the gradient vector $\nabla_{\mathcal{M}} f(p)$ points in the direction of the fastest ascent for f at p.

The Laplace–Beltrami operator $\Delta_{\mathcal{M}}$ (on functions) can be defined as divergence[1] of the gradient, $\Delta_{\mathcal{M}} f = -\operatorname{div}(\nabla_{\mathcal{M}} f)$. Alternatively, $\Delta_{\mathcal{M}}$ can be defined as the only operator, such that for any two differentiable functions f, h

$$\int_{\mathcal{M}} h(x)\,\Delta_{\mathcal{M}} f(x)\,d\mu(x) = \int_{\mathcal{M}} \big\langle \nabla_{\mathcal{M}} h(x), \nabla_{\mathcal{M}} f(x)\big\rangle\,d\mu(x) \tag{1}$$

where the inner product is taken in the tangent space of the manifold and μ is the canonical uniform measure. The Laplace–Beltrami operator is one of the classical objects in differential geometry, see, e.g., [24] for a detailed exposition. Recall also that in the Euclidean space, the Laplace–Beltrami operator is the ordinary Laplacian:

$$\Delta f = -\sum_i \frac{\partial^2 f}{\partial x_i^2}.$$

On a k-dimensional manifold $\mathcal{M}$, in a local coordinate system $(x_1,\dots,x_k)$ with a metric tensor g_{ij}, the Laplace–Beltrami operator applied to a function $f(x_1,\dots,x_k)$ is given by

$$\Delta_{\mathcal{M}} f = \frac{1}{\sqrt{\det(g)}}\sum_j \frac{\partial}{\partial x^j}\left(\sqrt{\det(g)}\sum_i g^{ij}\frac{\partial f}{\partial x^i}\right)$$

where g^{ij} are the components of the inverse of the metric tensor G^{-1}.

The eigenfunctions of the Laplace–Beltrami operator form a basis for the space of square integrable functions $L^2(\mathcal{M})$ on the manifold. In the special case, when the manifold is a 1-dimensional circle, the corresponding basis consists of the usual Fourier harmonics $\sin(k\pi x)$ and $\cos(k\pi x)$. Indeed, functions on a circle can be considered as periodic functions on $\mathbb{R}$. The corresponding Laplacian is simply the second derivative $-\frac{\partial^2}{\partial x^2}$. It is easy to see that the Fourier harmonics are the only eigenfunctions of the second derivative in the space of the periodic functions.

If the manifold is taken with a measure ν (given by $d\nu(x) = P(x)\,d\mu(x)$ for some function $P(x)$ and with $d\mu$ being the canonical measure corresponding to the volume form), which is not uniform, a more general notion of the *weighted Laplacian* seems natural. The weighted Laplacian can then be defined as $\Delta_{\mathcal{M},\nu} f(x) = \Delta_P f = \frac{1}{P(x)}\operatorname{div}(P(x)\nabla_{\mathcal{M}} f)$. See [22] for details.

The question addressed in this work is how to reconstruct the Laplace–Beltrami operator on $\mathcal{M}$ given a sample of points from the manifold. Before proceeding, we will need to fix notation and to describe the basic objects under consideration.

Graph Laplacian. Given a sample of n points $x_1,\dots,x_n$ from $\mathcal{M}$, we construct the complete weighted graph associated to that point cloud by taking $x_1,\dots,x_n$ as vertices of the graph and taking the edge weights to be $w_{ij} = e^{-\frac{\|x_i-x_j\|^2}{4t}}$. The corresponding *graph Laplacian* matrix L_n^t is given by

$$\big(L_n^t\big)_{ij} = \begin{cases} -w_{ij}, & \text{if } i \neq j,\\ \sum_k w_{ik}, & \text{if } i = j.\end{cases}$$

We may think of the matrix L_n^t as an operator on functions, defined on the data points:

$$L_n^t f(x_i) = f(x_i)\sum_j e^{-\frac{\|x_i-x_j\|^2}{4t}} - \sum_j f(x_j)e^{-\frac{\|x_i-x_j\|^2}{4t}}.$$

Point cloud Laplace operator. We immediately see that this formulation extends to any function on the ambient space and will denote the corresponding operator by $\mathbf{L}_n^t$:

$$\mathbf{L}_n^t f(x) = f(x)\frac{1}{n}\sum_j e^{-\frac{\|x-x_j\|^2}{4t}} - \frac{1}{n}\sum_j f(x_j)e^{-\frac{\|x-x_j\|^2}{4t}}.$$

[1] We use the minus sign in front of the gradient following a convention that makes the Laplacian a positive operator.

It is clear that $\mathbf{L}_n^t f(x_i) = \frac{1}{n} L_n^t f(x_i)$. We will call the operator $\mathbf{L}_n^t$ the *Laplacian associated to the point cloud* $x_1, \ldots, x_n$.

Functional approximation to the Laplace–Beltrami operator. Given a measure ν on $\mathcal{M}$ we construct the corresponding operator

$$\mathbf{L}^t f(x) = f(x) \int_{\mathcal{M}} e^{-\frac{\|x-y\|^2}{4t}}\, d\nu(y) - \int_{\mathcal{M}} f(y) e^{-\frac{\|x-y\|^2}{4t}}\, d\nu(y).$$

We observe $\mathbf{L}_n^t$ is simply a special form of $\mathbf{L}^t$ corresponding to the Dirac measure supported on $x_1, \ldots, x_n$.

The objects of interest in this paper are the Laplace–Beltrami operator $\Delta_{\mathcal{M}}$, its functional approximation $\mathbf{L}^t$ and its empirical approximation (which is an extension of the graph Laplacian) $\mathbf{L}_n^t$.

3. Main result

Our main contribution is to establish a connection between the graph Laplacian associated to a point cloud and the Laplace–Beltrami operator on the underlying manifold from which the points are drawn.

Consider a compact[2] k-dimensional differentiable manifold $\mathcal{M}$ isometrically embedded in $\mathbb{R}^N$. For our primary result, we will assume that the data is sampled from the uniform probability measure given by the induced metric on $\mathcal{M}$, i.e., the induced metric scaled by the factor $\frac{1}{\mathrm{vol}(\mathcal{M})}$.

Given data points $\mathcal{S}_n = \{x_1, \ldots, x_n\}$ in $\mathbb{R}^N$ sampled i.i.d. from this probability distribution we construct the associated Laplace operator $\mathbf{L}_n^t$. Our main result shows that for a fixed function $f \in C^\infty(\mathcal{M})$ and for a fixed point $p \in \mathcal{M}$, after appropriate scaling the operator $\mathbf{L}_n^t$ converges to the true Laplace–Beltrami operator on the manifold.

Theorem 3.1. *Let data points $x_1, \ldots, x_n$ be sampled from a uniform distribution on a manifold $\mathcal{M} \subset \mathbb{R}^N$. Put $t_n = n^{-\frac{1}{k+2+\alpha}}$, where $\alpha > 0$ and let $f \in C^\infty(\mathcal{M})$. Then the following equality holds*:

$$\lim_{n\to\infty} \frac{1}{t_n (4\pi t_n)^{\frac{k}{2}}} \mathbf{L}_n^{t_n} f(x) = \frac{1}{\mathrm{vol}(\mathcal{M})} \Delta_{\mathcal{M}} f(x)$$

where the limit is taken in probability and $\mathrm{vol}(\mathcal{M})$ *is the volume of the manifold with respect to the canonical measure.*

This theorem is the outcome of joint work with its roots in the results of [1]. It was first communicated in its essentially current form in [21] and appeared with a proof in [4]. The proof of this theorem consists of two parts. The main and more difficult result connects $\mathbf{L}^t$ (with the standard uniform measure) and the Laplace–Beltrami operator $\Delta_{\mathcal{M}}$. We show that when t tends to 0, $\mathbf{L}^t$ (appropriately scaled) converges to $\Delta_{\mathcal{M}}$. The second part uses the basic concentration inequalities of probability theory to show that $\frac{1}{n}\mathbf{L}_n^t$ converges to $\mathbf{L}^t$ when n tends to infinity and the points $x_1, \ldots, x_n$ are sampled from a probability distribution on the manifold.

The proof of this primary theorem contains several key ideas that appear in various generalizations of the result. In the above theorem, we assert the pointwise convergence of $\mathbf{L}_n^t f(p)$ to $\Delta_{\mathcal{M}} f(p)$ for a fixed function f and a fixed point p. Uniformity over all points $p \in \mathcal{M}$ follows almost immediately from the compactness of $\mathcal{M}$. Uniform convergence over a class of functions require a few additional considerations. A version of such a uniform theorem is

Theorem 3.2. *Let data points $x_1, \ldots, x_n$ be sampled from a uniform distribution on a compact manifold $\mathcal{M} \subset \mathbb{R}^N$ and let $\mathcal{F}$ be a family of functions $f \in C^\infty(\mathcal{M})$, with a uniform bound on all derivatives up to order* 3. *Then there exists a sequence of real numbers t_n, $t_n \to 0$, such that in probability*

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F}} \left| \frac{1}{t_n (4\pi t_n)^{\frac{k}{2}}} \mathbf{L}_n^{t_n} f(x) - \frac{1}{\mathrm{vol}(\mathcal{M})} \Delta_{\mathcal{M}} f(x) \right| = 0.$$

[2] It is possible to provide weaker but more technical conditions, which we will not discuss here.

This theorem is proved in Section 6 of this paper.

Another important direction for further generalization is the consideration of an arbitrary probability distribution P on $\mathcal{M}$ according to which points may be sampled. Such a distribution function $P : \mathcal{M} \to \mathbb{R}$ satisfies two properties (i) $P(q) > 0$ for all $q \in \mathcal{M}$ and (ii) integrates to one, i.e., $\int_{\mathcal{M}} P \operatorname{vol}(q) = 1$ where $\operatorname{vol}(q)$ is the volume form at the point q. This direction was systematically pursued by Lafon, Coifman, and others (see [11,18]). In Section 5, we consider this situation and provide an exposition of how the ideas in the proof of Theorem 3.1 need to be modified to cover this more general case. An almost immediate consequence of Theorem 3.1 that is applicable to the case of an arbitrary probability distribution is the following

Theorem 3.3. *Let $\mathcal{M}$ be a compact Riemannian submanifold of $\mathbb{R}^N$ and let $P : \mathcal{M} \to \mathbb{R}$ be a probability distribution function on $\mathcal{M}$ according to which data points $x_1, \ldots, x_n$ are drawn in i.i.d. fashion. Then, for $t_n = n^{-\frac{1}{k+2+\alpha}}$, where $\alpha > 0$, we have*

$$\lim_{n\to\infty} \frac{1}{t_n(4\pi t_n)^{\frac{k}{2}}} \mathbf{L}_n^{t_n} f(x) = P(x)\Delta_{P^2} f(x).$$

We see here that the graph Laplacian converges to a scaled version of the weighted Laplacian $\Delta_{P^2} f$. To remove the scaling factors that occur in the theorem above, one needs to consider the normalized Laplacian. A discussion of this is provided in Section 5 where we show how to construct a proof of convergence of the normalized Laplacian using the tools of this paper. The normalized Laplacian and a generalization to a family of normalized Laplacians was first pointed out in the current manifold learning context in [11,18].

3.1. Laplace operator and the heat equation

The Laplace operator is intimately related to the heat equation that governs the diffusion of heat on a manifold. We will now discuss this relationship and develop some intuitions about the methods used in the proof of Theorem 3.1.

Let us begin by considering the heat flow in Euclidean space, $\mathbb{R}^k$. Recall that the *Laplace operator* in $\mathbb{R}^k$ is defined as

$$\Delta f(x) = -\sum_i \frac{\partial^2 f}{\partial x_i^2}(x).$$

We say that a sufficiently differentiable function $u(x,t)$ satisfies the *heat equation* if

$$\frac{\partial}{\partial t} u(x,t) + \Delta u(x,t) = 0. \tag{2}$$

The heat equation describes diffusion of heat with the initial distribution $u(x,0)$. At any time t, the distribution of heat in $\mathbb{R}^k$ is given by the function $u(x,t)$. The solution to the heat equation is given by a semi-group of heat operators $\mathbf{H}^t$. Given an initial heat distribution $u(x,0) = f(x)$, the heat distribution at time t is given by $u(x,t) = \mathbf{H}^t f(x)$.

It turns out that this operator is given by convolution with the heat kernel, which for $\mathbb{R}^k$ is the usual Gaussian.

$$\mathbf{H}^t f(x) = \int_{\mathbb{R}^k} f(y) H^t(x,y)\, dy,$$

$$H^t(x,y) = (4\pi t)^{-\frac{k}{2}} e^{-\frac{\|x-y\|^2}{4t}}.$$

We summarize this in the following

Theorem 3.4 *(Solution to the heat equation in $\mathbb{R}^k$). Let $f(x)$ be a sufficiently differentiable bounded function. We then have*

$$\mathbf{H}^t f = (4\pi t)^{-\frac{k}{2}} \int_{\mathbb{R}^k} e^{-\frac{\|x-y\|^2}{4t}} f(y)\, dy, \tag{3}$$

$$f(x) = \lim_{t\to 0} \mathbf{H}^t f(x) = (4\pi t)^{-\frac{k}{2}} \int_{\mathbb{R}^k} e^{-\frac{\|x-y\|^2}{4t}} f(y)\, dy. \tag{4}$$

The function $u(x,t) = \mathbf{H}^t f$ satisfies the heat equation

$$\frac{\partial}{\partial t}u(x,t) + \Delta u(x,t) = 0$$

with initial condition $u(x,0) = f(x)$.

The heat equation is the key to approximating the Laplace operator,

$$\Delta f(x) = -\frac{\partial}{\partial t}u(x,t)\bigg|_{t=0} = -\frac{\partial}{\partial t}\mathbf{H}^t f(x)\bigg|_{t=0} = \lim_{t\to 0}\frac{1}{t}\big(f(x) - \mathbf{H}^t f(x)\big). \tag{5}$$

Equation (5) above suggests our scheme for approximating the Laplace operator. Recalling that the heat kernel is the Gaussian which integrates to 1, we observe that

$$\Delta f(x) = \lim_{t\to 0} -\frac{1}{t}\left((4\pi t)^{-\frac{k}{2}}\int\limits_{\mathbb{R}^k} e^{-\frac{\|x-y\|^2}{4t}} f(y)\,dy - f(x)\,(4\pi t)^{-\frac{k}{2}}\int\limits_{\mathbb{R}^k} e^{-\frac{\|x-y\|^2}{4t}}\,dy\right).$$

This quantity can easily be approximated from a point cloud[3] $x_1,\dots,x_n$ by computing the empirical version of the integrals involved:

$$\hat{\Delta} f(x) = \frac{1}{t}\frac{(4\pi t)^{-\frac{k}{2}}}{n}\left(f(x)\sum_i e^{-\frac{\|x_i-x\|^2}{4t}} - \sum_i e^{-\frac{\|x_i-x\|^2}{4t}} f(x_i)\right) = \frac{1}{t(4\pi t)^{\frac{k}{2}}}\mathbf{L}_n^t(f)(x).$$

This intuition can be easily turned into a convergence result for $\mathbb{R}^k$. Extending this analysis to an arbitrary manifold, however, is not as straightforward as it might seem at first blush. The two principal technical issues are the following:

(1) With some very rare exceptions we do not know the exact form of the heat kernel $H^t_{\mathcal{M}}(x,y)$.
(2) Even the asymptotic form of the heat kernel requires knowing the geodesic distance between points in the point cloud. However we can only observe distances in the ambient space $\mathbb{R}^N$.

Remarkably both of these issues can be overcome because of certain missing terms in the asymptotic expansion of various intrinsic objects! This will become clear in the proof that we provide in the next section.

4. Proof of the main results

4.1. Basic differential geometry

Before we proceed further, let us briefly review some basic notions of differential geometry. Assume we have a compact[4] differentiable k-dimensional submanifold of $\mathbb{R}^N$ with the induced Riemannian structure. That means that we have a notion of *length* for curves on $\mathcal{M}$. Given two points $x, y \in \mathcal{M}$ the *geodesic distance* $\operatorname{dist}_{\mathcal{M}}(x,y)$ is the length of the shortest curve connecting x and y. It is clear that $\operatorname{dist}_{\mathcal{M}}(x,y) \geqslant \|x-y\|$.

Given a point $p \in \mathcal{M}$, one can identify the *tangent space* $T_p\mathcal{M}$ with an affine subspace of $\mathbb{R}^N$ passing through p. This space has a natural linear structure with the origin at p. Furthermore it is possible to define the *exponential map* $\exp_p : T_p\mathcal{M} \to \mathcal{M}$. The key property of the exponential map is that it takes lines through the origin in $T_p\mathcal{M}$ to geodesics passing through p. The exponential map is a local diffeomorphism and produces a natural system of coordinates for some neighborhood of p. The Hopf–Rinow theorem (see, e.g., [12]) implies that a compact manifold is *geodesically complete*, i.e. that any geodesic can be extended indefinitely which, in particular, implies that there exists a geodesic connecting any two given points on the manifold.

The Riemannian structure on $\mathcal{M}$ induces a measure (which we will denote by μ) corresponding to the *volume form* (denoted at a point $q \in \mathcal{M}$ as $\operatorname{vol}(q)$). For a compact $\mathcal{M}$ total volume of $\mathcal{M}$ is guaranteed to be finite, which gives rise to the canonical *uniform* probability distribution on $\mathcal{M}$ given by $d\mu(q) = \frac{1}{\operatorname{vol}(\mathcal{M})}\operatorname{vol}(q)$.

[3] We are ignoring the technicalities about the probability distribution for the moment. It is not hard however to show that it is sufficient to restrict the distribution to some open set containing the point x.

[4] We assume compactness to simplify the exposition. A weaker condition will suffice as noted above.

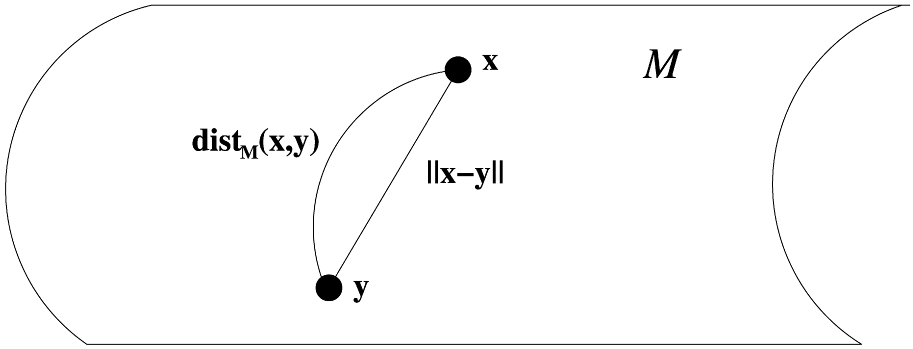

Fig. 1. Geodesic and chordal distance.

Before proceeding with the main proof we state one curious property of geodesics, which will be needed later. It concerns the relationship between $\mathrm{dist}_{\mathcal{M}}(x, y)$ and $\|x - y\|$. The geodesic and chordal distances are shown pictorially in Fig. 1. It is clear that when x and y are close, the difference between these two quantities is small. Interestingly, however, this difference is smaller than one (at least the authors) would expect initially. It turns out (cf., Lemma 4.3) that when the manifold is compact

$$\|x - y\|^2 = \mathrm{dist}^2_{\mathcal{M}}(x, y) - O\big(\mathrm{dist}^4_{\mathcal{M}}(x, y)\big).$$

In other words chordal distance approximates geodesic distance up to order three. This observation and certain consequent properties of the geodesic map make the approximations used in this paper possible.

The Laplace–Beltrami operator $\Delta_{\mathcal{M}}$ is a second order differential operator. The family of diffusion operators $\mathbf{H}^t_{\mathcal{M}}$ satisfies the following properties:

$$\Delta_{\mathcal{M}} \mathbf{H}^t_{\mathcal{M}}(f) = -\frac{\partial}{\partial t} \mathbf{H}^t_{\mathcal{M}}(f) \quad \text{heat equation,}$$

$$\lim_{t \to 0} \mathbf{H}^t_{\mathcal{M}}(f) = f \quad \delta\text{-family property.}$$

In other words, the diffusion of heat on a manifold (from an initial distribution $u(x, 0) = f(x)$) is governed by the heat equation on the manifold. The solution of the heat equation is given by convolution with the *heat kernel*. The corresponding family of integral operators is denoted by $\mathbf{H}^t_{\mathcal{M}}$ (see, e.g., [24]).

Our proof hinges on the fact that in geodesic coordinates the heat kernel can be approximated by a Gaussian (in the ambient Euclidean space) for small values of t and the relationships between distances in the ambient space and geodesics along the submanifold.

4.2. Main proof

We will now proceed with the proof of the main theorem.

First we note that the quantities

$$\int_{\mathcal{M}} e^{-\frac{\|p-x\|^2}{4t}} f(x)\, d\mu(x)$$

and

$$f(p) \int_{\mathcal{M}} e^{-\frac{\|p-x\|^2}{4t}}\, d\mu(x)$$

can be empirically estimated from the point cloud. In particular, consider the operator

$$\mathbf{L}^t f(p) = \frac{1}{t} \frac{1}{(4\pi t)^{\frac{k}{2}}} \int_{\mathcal{M}} e^{-\frac{\|p-x\|^2}{4t}} \big(f(x) - f(p)\big)\, d\mu(x). \tag{6}$$

By the simple law of large numbers, it is clear that for a fixed $t > 0$, a fixed function f and a fixed point $p \in \mathcal{M}$,

$$\lim_{n\to\infty} \frac{1}{t} \frac{1}{(4\pi t)^{\frac{k}{2}}} \frac{1}{n} \mathbf{L}_n^t f(p) = \mathbf{L}^t f(p)$$

where the convergence of the random empirical point cloud Laplacian is in probability.

The bulk of our paper will constitute an analysis of the operator $\mathbf{L}^t$ as $t \to 0$. Then noting that

$$\lim_{n\to\infty} \frac{1}{t} \frac{1}{(4\pi t)^{\frac{k}{2}}} \mathbf{L}_n^t f(p) = \mathbf{L}^t f(p) \quad \text{and} \quad \lim_{t\to\infty} \mathbf{L}^t f(p) = \frac{1}{\mathrm{vol}(\mathcal{M})} \Delta_{\mathcal{M}} f(p)$$

we obtain the main theorem.

The analysis of the limiting behavior of $\mathbf{L}^t$ will be conducted in several steps:

Step 1. Reduction of the integral to a ball on $\mathcal{M}$.

This argument is provided in Section 4.3 where we prove Lemma 4.1 which implies that

$$\lim_{t\to 0} \mathbf{L}^t f(p) = \lim_{t\to 0} \frac{1}{t} \frac{1}{(4\pi t)^{\frac{k}{2}}} \int_B e^{-\frac{\|p-x\|^2}{4t}} \big(f(p) - f(x)\big)\, d\mu(x)$$

where B is an arbitrary open subset of $\mathcal{M}$ containing p.

Step 2. Change of coordinates via the exponential map.

This step is discussed in Section 4.4 where in particular, we derive Eq. (10) that is a suitable asymptotic approximation to $\mathbf{L}^t f$. By changing to exponential coordinates, the integral over the manifold in the definition of $\mathbf{L}^t$ is ultimately reduced to a new integral over a k-dimensional Euclidean space given by Eq. (10).

Step 3. Analysis in Euclidean space $\mathbb{R}^k$.

The last step is an analysis of the integral of Eq. (10) in $\mathbb{R}^k$ and is conducted in Section 4.5. The key Proposition 4.4 proves the limiting behavior of $\mathbf{L}^t$ as $t \to 0$.

This leads to the proof of the main theorem as follows:

Proof of main Theorem 3.1. Recall that the point cloud Laplacian $\mathbf{L}_n^t$ applied to f at p is

$$\mathbf{L}_n^t f(p) = \frac{1}{n} \left(\sum_{i=1}^n e^{-\frac{\|p-x_i\|^2}{4t}} f(p) - \sum_{i=1}^n e^{-\frac{\|p-x_i\|^2}{4t}} f(x_i) \right).$$

We note that $\mathbf{L}_n^t f(p)$ is the empirical average of n independent random variables with the expectation

$$\mathbb{E}\,\mathbf{L}_n^t f(p) = \left(f(p) \int_{\mathcal{M}} e^{-\frac{\|p-y\|^2}{4t}}\, d\mu(y) - \int_{\mathcal{M}} f(y) e^{-\frac{\|p-y\|^2}{4t}}\, d\mu(y) \right). \tag{7}$$

By an application of Hoeffding's inequality (Eq. (7.2)), we have

$$\mathbb{P}\left[\frac{1}{t(4\pi t)^{k/2}} \left| \mathbf{L}_n^t f(p) - \mathbb{E}\,\mathbf{L}_n^t f(p) \right| > \epsilon \right] \leqslant 2e^{-1/2\epsilon^2 nt(4\pi t)^{k/2}}.$$

Choosing t as a function of n by letting $t = t_n = (\frac{1}{n})^{\frac{1}{k+2+\alpha}}$, where $\alpha > 0$, we see that for any fixed $\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{1}{t_n(4\pi t_n)^{k/2}} \left| \mathbf{L}_n^{t_n} f(p) - \mathbb{E}\,\mathbf{L}_n^{t_n} f(p) \right| > \epsilon \right] = 0.$$

Noting that by Proposition 4.4 and Eq. (7)

$$\lim_{n\to\infty} \frac{1}{t_n(4\pi t_n)^{k/2}}\, \mathbf{L}_n^{t_n} = \frac{\Delta_{\mathcal{M}} f(p)}{\mathrm{vol}(\mathcal{M})}$$

we obtain the theorem. $\square$

4.3. Reduction to a ball

In this section, we show that the integral over the entire manifold $\mathcal{M}$ in the operator $\mathbf{L}^t$ can be replaced by an integral over an open subset of $\mathcal{M}$ with no change in the limiting behavior. In other words,

$$\left|\mathbf{L}^t f(p) - \int\limits_B e^{-\frac{\|p-y\|^2}{4t}} \big(f(p) - f(y)\big)\, d\mu(y)\right| = o\big(t^a\big) \tag{8}$$

for any $a > 0$. This is a consequence of the following

Lemma 4.1. *Given an open set $B \subset \mathcal{M}$, $p \in B$, and a function $f \in L^\infty(\mathcal{M})$, for any $a > 0$ as $t \to 0$, we have*

$$\left|\int\limits_B e^{-\frac{\|p-y\|^2}{4t}} f(y)\, d\mu(y) - \int\limits_{\mathcal{M}} e^{-\frac{\|p-y\|^2}{4t}} f(y)\, d\mu(y)\right| = o\big(t^a\big).$$

Proof. Let $d = \inf_{x \notin B} \|p - x\|^2$ and let $M = \mu(\mathcal{M} - B)$ be the measure of the complement to B. Since B is open and p is an interior point of B, $d > 0$. We thus see that

$$\left|\int\limits_B e^{-\frac{\|p-y\|^2}{4t}} f(y)\, d\mu(y) - \int\limits_{\mathcal{M}} e^{-\frac{\|p-y\|^2}{4t}} f(y)\, d\mu(y)\right| \leqslant M \sup_{x \in \mathcal{M}} \big(\big|f(x)\big|\big) e^{-\frac{d^2}{4t}}.$$

As t tends to zero $e^{-\frac{d^2}{4t}}$, decreases at an exponential rate and the claim follows. □

4.4. Exponential change of coordinates

To proceed with our analysis we will need a tool, enabling us to reduce an operator on a manifold to an operator on a Euclidean space. This is done by observing that in a neighborhood of a point p, a Riemannian manifold has a canonical local coordinate system given by the *exponential map* $\exp_p$. Recall that $\exp_p$ is a map : $T\mathcal{M}_p (= \mathbb{R}^k) \to \mathcal{M}$, such that $\exp_p(0) = p$, $\exp_p$ is a local isomorphism and an image of a straight line through the origin in $\mathbb{R}^k$ is a geodesic in $\mathcal{M}$.

In order to reduce the computations to $\mathbb{R}^k$ we will apply the following change of coordinates:

$$y = \exp_p(x)$$

where $y \in \mathcal{M}$. Thus given a function $f : \mathcal{M} \to \mathbb{R}$, we rewrite it in geodesic coordinates (locally around p) by putting

$$\tilde{f}(x) = f\big(\exp_p(x)\big).$$

We will need the following key statement relating the Laplace–Beltrami operator and the Euclidean Laplacian:

Lemma 4.2. *(See, e.g., [24, p. 90].)*

$$\Delta_{\mathcal{M}} f(p) = \Delta_{\mathbb{R}^k} \tilde{f}(0) = -\sum_{i=1}^{k} \frac{\partial^2 \tilde{f}}{\partial x_i^2}(0). \tag{9}$$

Since $\exp_p : T\mathcal{M}_p (= \mathbb{R}^k) \to \mathcal{M}$ is a locally invertible (within the *injectivity radius*) map, we can choose a small open set $\tilde{B} \subset \mathbb{R}^k$, s.t. $\exp_p$ is a diffeomorphism onto its image $B = \exp_p(\tilde{B}) \subset \mathcal{M}$. We will generally let $\tilde{B}$ be a k-dimensional ball of radius ϵ, where ϵ can be chosen appropriately.

By the definition of the exponential map, we see that for any point $y = \exp_p(x)$, we have that $d^2_{\mathcal{M}}(y, p) = \|x\|^2_{\mathbb{R}^k} \geqslant \|y - p\|^2_{\mathbb{R}^N}$ where $p, y \in \mathcal{M} \subset \mathbb{R}^N$ and $x \in T_p = \mathbb{R}^k$. However, crucially for our purposes, it is possible to show that the chordal distance $\|y - p\|_{\mathbb{R}^N}$ and the geodesic distance $\|x\|_{\mathbb{R}^k}$ are related by the following

Lemma 4.3. *For any two points $p, y \in \mathcal{M}$, $y = \exp_p(x)$, the relation between the Euclidean distance and geodesic distance is given by*

$$\|x\|^2_{\mathbb{R}^k} - \|y - p\|^2_{\mathbb{R}^N} = g(x)$$

where $|g(x)| = O(\|x\|^4_{\mathbb{R}^k})$ *uniformly. In other words, there exists a constant* $C > 0$ *such that*

$$0 \leqslant \|x\|^2_{\mathbb{R}^k} - \|y - p\|^2_{\mathbb{R}^N} = g(x) < C\|x\|^4_{\mathbb{R}^k}$$

for all p *and* $y = \exp_p(x)$*. The constant* C *depends upon the embedding of the manifold and bounds on the third derivatives of the embedding coordinates.*

Notation. From now on we will suppress subscripts in the norm notation. For example, we will write $\|x\|$ instead of $\|x\|_{\mathbb{R}^k}$.

Following Eq. (8), we see that an appropriately scaled version of $\mathbf{L}^t f(p)$ can be approximated by the following integral

$$\frac{1}{t}\frac{1}{(4\pi t)^{\frac{k}{2}}}\int\limits_B e^{-\frac{\|y-p\|^2}{4t}}\big(f(p) - f(y)\big)\,d\mu(p)$$

which can be written in exponential coordinates as

$$\frac{1}{\mathrm{vol}(\mathcal{M})}\frac{1}{t}\frac{1}{(4\pi t)^{\frac{k}{2}}}\int\limits_{\tilde{B}} e^{-\frac{\|\exp_p(x)-p\|^2}{4t}}\big(\tilde{f}(0) - \tilde{f}(x)\big)\sqrt{\det(g_{ij})}\,dx$$

where $\det(g_{ij})$ is the determinant of the metric tensor in exponential coordinates. Now we make use of the fact (see [34,35] and references therein) that the metric tensor has an asymptotic expansion in exponential coordinates given by

$$\det(g_{ij}) = 1 - \frac{1}{6}x^T Rx + O\big(\|x\|^3\big)$$

where R is the Ricci curvature tensor. On a smooth, compact manifold $\mathcal{M}$, the elements of R are bounded and therefore, we can write $\det(g_{ij}) = 1 + O(\|x\|^2)$. Combining this with Lemma 4.3, our approximation to $\mathbf{L}^t$ becomes

$$L = \frac{1}{\mathrm{vol}(\mathcal{M})}\frac{1}{t}\frac{1}{(4\pi t)^{\frac{k}{2}}}\int\limits_{\tilde{B}} e^{-\frac{\|x\|^2-g(x)}{4t}}\big(\tilde{f}(0) - \tilde{f}(x)\big)\big(1 + O\big(\|x\|^2\big)\big)\,dx \tag{10}$$

where $\tilde{B}$ is a sufficiently small ball.

4.5. Analysis in $\mathbb{R}^k$

In this manner, by switching to exponential coordinates, we have reduced our task to analysis of an integral in $\mathbb{R}^k$. We begin by considering the Taylor expansion of $\tilde{f}(x)$ about 0 as follows

$$\tilde{f}(x) - \tilde{f}(0) = x \cdot \nabla f + \frac{1}{2}x^T Hx + O\big(\|x\|^3\big)$$

where $\nabla f = (\frac{\partial \tilde{f}}{\partial x_1}, \frac{\partial \tilde{f}}{\partial x_2}, \ldots, \frac{\partial \tilde{f}}{\partial x_k})^T$ is the gradient of $\tilde{f}$ and H is the corresponding Hessian. The remainder term may be bounded in terms of third order derivatives of the function $\tilde{f}$. From now on we will assume that the third order derivatives of f exist and are bounded. In particular, we have that for some constant K and any x

$$\big|\tilde{f}(x) - \tilde{f}(0)\big| \leqslant K\|x\|. \tag{11}$$

We now consider the quantity $e^{-\frac{\|x\|^2-g(x)}{4t}}$ that occurs in the integrand of interest in Eq. (10). We make use of the following approximation:

$$e^{-\frac{\|x\|^2-g(x)}{4t}} = e^{-\frac{\|x\|^2}{4t}}e^{\frac{g(x)}{4t}} = e^{-\frac{\|x\|^2}{4t}}\left(1 + O\left(\frac{1}{4t}g(x)e^{\frac{g(x)}{4t}}\right)\right)$$

where we have used the fact that $e^{\alpha} = 1 + O(\alpha e^{\alpha})$ for $\alpha > 0$.

Armed with these expansions, we can rewrite the integral of Eq. (10) as

$$L = \gamma_t \int_{\tilde{B}} e^{-\frac{\|x\|^2}{4t}} \left(1 + O\left(\frac{g(\|x\|)}{4t} e^{\frac{g(\|x\|)}{4t}}\right)\right) \big(\tilde{f}(0) - \tilde{f}(x)\big)\big(1 + O\big(\|x\|^2\big)\big)\, dx$$

where $\gamma_t = \frac{1}{\mathrm{vol}(\mathcal{M})} \frac{1}{t} \frac{1}{(4\pi t)^{\frac{k}{2}}}$.

To simplify computations we split the integral in three summands

$$L = A_t + B_t + C_t$$

where

$$A_t = -\gamma_t \int_{\tilde{B}} e^{-\frac{\|x\|^2}{4t}} \big(\tilde{f}(x) - \tilde{f}(0)\big)\, dx,$$

$$B_t = -\gamma_t \int_{\tilde{B}} e^{-\frac{\|x\|^2}{4t}} \big(\tilde{f}(x) - \tilde{f}(0)\big) O\big(\|x\|^2\big)\, dx$$

and

$$C_t = -\gamma_t \int_{\tilde{B}} e^{-\frac{\|x\|^2}{4t}} \big(\tilde{f}(x) - \tilde{f}(0)\big)\left(O\left(\frac{g(x)}{4t} e^{\frac{g(x)}{4t}}\right)\big(1 + O\big(\|x\|^2\big)\big)\right) dx.$$

It remains to bound each of the terms A_t, B_t and C_t, which we do in the following proposition

Proposition 4.4. *For A_t, B_t, C_t defined as above, the following hold*

(i) $\lim_{t\to 0} A_t = \frac{1}{\mathrm{vol}(\mathcal{M})} \Delta_{\mathcal{M}} f(p)$,
(ii) $\lim_{t\to 0} B_t = 0$, *and*
(iii) $\lim_{t\to 0} C_t = 0$ *implying that*

$$\lim_{t\to 0} \frac{1}{t(4\pi t)^{k/2}} \mathbf{L}^t f(p) = \frac{1}{\mathrm{vol}(\mathcal{M})} \Delta_M f(p).$$

Proof. We begin by considering A_t. Using the Taylor expansion of $\tilde{f}(x)$, we have

$$A_t = -\gamma_t \int_{\tilde{B}} \left(x \cdot \nabla \tilde{f} + \frac{1}{2} x^T H x + O\big(\|x\|^3\big)\right) e^{-\frac{\|x\|^2}{4t}}\, dx.$$

We now make the following four observations. First, note that

$$\int_{\tilde{B}} x_i e^{-\frac{\|x\|^2}{4t}}\, dx = 0 \tag{12}$$

and for $i \neq j$

$$\int_{\tilde{B}} x_i x_j e^{-\frac{\|x\|^2}{4t}}\, dx = 0. \tag{13}$$

These follow directly from the properties of a zero mean Gaussian distribution with diagonal covariance matrix and the symmetries inherent in such a distribution function. Further, for $i = j$, we see that

$$\lim_{t\to 0} \frac{1}{t} \frac{1}{(4\pi t)^{k/2}} \int_{\tilde{B}} x_i^2 e^{-\frac{\|x\|^2}{4t}} = \mathrm{vol}(\mathcal{M}) \lim_{t\to 0} \gamma_t \int_{\tilde{B}} x_i^2 e^{-\frac{\|x\|^2}{4t}} = 2. \tag{14}$$

Now note that

$$\frac{1}{t}\frac{1}{(4\pi t)^{k/2}}\int\limits_{\tilde{B}} \|x\|^3 e^{-\frac{\|x\|^2}{4t}} = O\big(t^{1/2}\big). \tag{15}$$

These four observations immediately imply that

$$\mathrm{vol}(\mathcal{M}) \lim_{t\to 0} A_t = -\operatorname{tr}(H) = -\sum_{i=1}^{n} \frac{\partial^2 \tilde{f}(0)}{\partial x_i^2} = \Delta_{\mathcal{M}} f(p).$$

We now turn to B_t. Note that

$$|B_t| \leqslant \gamma_t \int\limits_{\tilde{B}} \big|\big(\tilde{f}(x) - \tilde{f}(0)\big)\big| e^{-\frac{\|x\|^2}{4t}} O\big(\|x\|^2\big)\, dx \leqslant \gamma_t \int\limits_{\tilde{B}} e^{-\frac{\|x\|^2}{4t}} O\big(\|x\|^3\big)\, dx$$

where the last inequality follows from Eq. (11). Clearly, B_t tends to 0 as $t \to 0$ from Eq. (15).

Finally, consider C_t. We note that for sufficiently small values of $\|x\|$

$$\|x\|^2 - g(x) \geqslant \frac{\|x\|^2}{2}$$

and so

$$|C_t| \leqslant \gamma_t \int\limits_{\tilde{B}} \big|\big(\tilde{f}(x) - \tilde{f}(0)\big)\big| e^{-\frac{\|x\|^2}{8t}} \big(1 + O\big(\|x\|^2\big)\big)\, dx \leqslant \gamma_t \int\limits_{\tilde{B}} O\bigg(\frac{\|x\|^5}{t}\bigg) e^{-\frac{\|x\|^2}{8t}} \big(1 + O\big(\|x\|^2\big)\big)\, dx.$$

The same argument as above shows that $\lim_{t\to 0} C_t = 0$. □

5. Arbitrary probability distribution

The convergence proof for the case of a uniform probability distribution can easily be extended to cover the case of an arbitrary probability distribution with support on the submanifold of interest. Minimal modification is required as we now demonstrate below.

The setting is as follows: consider $\mathcal{M} \in \mathbb{R}^N$ to be a compact Riemannian submanifold of Euclidean space as before. Let μ be a probability measure with support on $\mathcal{M}$. Under reasonable regularity conditions (essentially, absolutely continuous with respect to the volume measure), μ may be characterized by means of a probability density function $P : \mathcal{M} \to \mathbb{R}^+$ such that

$$\int\limits_{\mathcal{M}} f(y)\, d\mu(y) = \int\limits_{\mathcal{M}} f(y) P(y)\, \mathrm{vol}(y)$$

where $\int_{\mathcal{M}} P(q)\, \mathrm{vol}(q) = 1$ and $\mathrm{vol}(q)$ is the volume form on T_q. Thus, our previous discussion considered the case $P(q) = \frac{1}{\mathrm{vol}(\mathcal{M})}$.

As before, fix a function $f : \mathcal{M} \to \mathbb{R}$ and a point $p \in \mathcal{M}$. Then if $x_1, \ldots, x_n$ are points sampled from $\mathcal{M}$ according to P, the point cloud Laplacian $\mathbf{L}_n^t$ is defined as before as

$$\mathbf{L}_n^t(f)(x) = \frac{1}{n}\bigg\{ f(x) \sum_j e^{-\frac{\|x-x_j\|^2}{4t}} - \sum_j f(x_j) e^{-\frac{\|x-x_j\|^2}{4t}} \bigg\}.$$

By the simple law of large numbers, this (suitably scaled) converges to $\mathbf{L}_P^t$ (where we explicitly denote the dependence on P by the subscript) defined as

$$\mathbf{L}_P^t f(p) = \frac{1}{(4\pi t)^{k/2} t} \int\limits_{\mathcal{M}} e^{-\frac{\|p-x\|^2}{4t}} \big(f(p) - f(x)\big)\, d\mu(x)$$

which by the relation $d\mu(x) = P(x)\,\mathrm{vol}(x)$ can be written as

$$\mathbf{L}_P^t f(p) = \frac{1}{(4\pi t)^{k/2} t} \int_{\mathcal{M}} e^{-\frac{\|p-x\|^2}{4t}} \big(f(p) - f(x)\big) P(x)\,\mathrm{vol}(x).$$

Now consider the function

$$h(x) = \big(f(p) - f(x)\big) P(x).$$

Clearly, $h(p) = 0$ and therefore, we see that

$$\mathbf{L}_P^t f(p) = \mathrm{vol}(\mathcal{M}) \frac{1}{t(4\pi t)^{k/2}} \mathbf{L}^t h(p)$$

where $\mathbf{L}^t$ is the operator associated to the uniform probability distribution. By Proposition 4.4, we see that

$$\lim_{t\to 0} \mathbf{L}_P^t f(p) = \mathrm{vol}(\mathcal{M}) \lim_{t\to 0} \mathbf{L}^t h(p) = \Delta_{\mathcal{M}} h(p).$$

Let us now consider $\Delta_{\mathcal{M}} h(p)$. We see that

$$\Delta_{\mathcal{M}} h(p) = \Delta_{\mathcal{M}}\big(P(x)\big(f(x) - f(p)\big)\big).$$

Using the fact that at the point $p \in \mathcal{M}$, $\Delta(fg) = g\Delta f + f\Delta g + 2\langle \mathrm{grad}\, f, \mathrm{grad}\, g\rangle_{T_p}$, we see that

$$\Delta_{\mathcal{M}} h(p) = P\Delta_{\mathcal{M}} f + 2\langle \mathrm{grad}\, f, \mathrm{grad}\, P\rangle = P\frac{1}{P^2}\mathrm{div}\big(P^2 \,\mathrm{grad}\, f\big) = P\Delta_{P^2} f(p)$$

where $\Delta_{P^2} f$ is the weighted Laplacian defined as

$$\Delta_{P^2} f = \Delta_{\mathcal{M}} f + \frac{2}{P}\langle \mathrm{grad}\, f, \mathrm{grad}\, P\rangle = \frac{1}{P^2}\mathrm{div}\big(P^2 \,\mathrm{grad}\, f\big).$$

For more details on this form of the weighted Laplacian, see [22].

Thus, we can state the following theorem

Theorem 5.1. *Let $\mathcal{M}$ be a compact Riemannian submanifold of $\mathbb{R}^N$ and let $P : \mathcal{M} \to \mathbb{R}$ be a probability distribution function on $\mathcal{M}$ according to which data points $x_1, \ldots, x_n$ are drawn in i.i.d. fashion. Then, for $t_n = n^{-\frac{1}{k+2+\alpha}}$, where $\alpha > 0$, we have*

$$\lim_{n\to\infty} \frac{1}{t_n(4\pi t_n)^{k/2}} \mathbf{L}_n^{t_n} f(x) = P(x)\Delta_{P^2} f(x).$$

5.1. Normalizing the weights

In the previous section, we saw that the empirical point cloud Laplacian converges pointwise to an operator which is the weighted Laplacian scaled (multiplied) by the probability density function at that same point. It turns out that by suitably normalizing the weights of the point cloud Laplacian, the scaling factor may be removed. By a different normalization it is even possible to recover the Laplace Beltrami operator on the manifold thus separating the truly geometric aspects of the manifold from the probability distribution on it.

We now consider the convergence of the Laplacian with normalized weights. As before, we have a smooth, compact, Riemannian manifold $\mathcal{M}$ along with a probability distribution function $P : \mathcal{M} \to \mathbb{R}^+$ on it according to which points $x_1, \ldots, x_n$ are drawn in i.i.d. fashion. We assume that $a \leqslant P(x) \leqslant b$ for all $x \in \mathcal{M}$. Consider a fixed function $f : \mathcal{M} \to \mathbb{R}$ and a point $x \in \mathcal{M}$. The point cloud Laplacian operator may be defined as

$$\mathbf{L}_n^t f(x) = \frac{1}{n}\sum_{i=1}^{n} W(x, x_i)\big(f(x) - f(x_i)\big).$$

If $W(x, x_i) = e^{-\frac{\|x-x_i\|^2}{4t}}$, this corresponds to the operator whose convergence properties have been analyzed in previous sections. The operator with normalized weights corresponds to the choice

$$W(x, x_i) = \frac{1}{t}\frac{G_t(x, x_i)}{\sqrt{\hat{d}_t(x)}\sqrt{\hat{d}_t(x_i)}}$$

where $G_t = \frac{1}{(4\pi t)^{k/2}} e^{-\frac{\|x-x_i\|^2}{4t}}$ is the Gaussian weight. The quantities $\hat{d}_t(x)$ and $\hat{d}_t(x_i)$ are empirical estimates of $d_t(x)$ and $d_t(x_i)$ respectively defined as follows:

$$d_t(x) = \int_{\mathcal{M}} G_t(x, y) P(y) \operatorname{vol}(y)$$

while

$$\hat{d}_t(x_i) = \frac{1}{n-1} \sum_{j \neq i} G_t(x_i, x_j)$$

and

$$\hat{d}_t(x) = \frac{1}{n} \sum_{j \neq i} G_t(x, x_j).$$

We proceed now to outline the argument for the convergence of this operator. We will see how essential ideas for this new convergence argument are already contained in the proof of convergence in the setting with a uniform probability distribution (detailed in Section 4).

5.1.1. Limiting continuous operator

Our first step is to show that for a fixed $t > 0$, the empirical operator converges to the following continuous operator given by

$$\mathbf{L}_P^t f(x) = \frac{1}{t} \int_{\mathcal{M}} \frac{G_t(x, y)}{\sqrt{d_t(x)}\sqrt{d_t(y)}} \big(f(x) - f(y)\big) P(y) \operatorname{vol}(y).$$

This follows from the usual considerations of the law of large numbers applied carefully. To see this, first consider the intermediate operator

$$\mathbf{L}_n^{t\prime} f(x) = \frac{1}{nt} \sum_{i=1}^{n} \frac{G_t(x, x_i)}{\sqrt{d_t(x)}\sqrt{d_t(x_i)}} \big(f(x) - f(x_i)\big).$$

Since $a_t < d_t(x) < b_t$ for all x and $0 < \frac{1}{t} G_t(x, x_i) < \frac{1}{t(4\pi t)^{k/2}}$, the random variable $\frac{G_t(x,x_i)(f(x)-f(x_i))}{t\sqrt{d_t(x)}\sqrt{d_t(x_i)}} \leqslant Q_t$ is bounded and a simple application of Hoeffding's inequality yields an exponential rate at which $\mathbf{L}_n^{t\prime} f(x)$ converges to $\mathbf{L}_P^t f(x)$ in probability as $n \to \infty$. In other words,

$$\mathbb{P}\big[\big|\mathbf{L}_n^{t\prime} f(x) - \mathbf{L}_P^t f(x)\big| > \epsilon\big] \leqslant 2e^{-\frac{\epsilon^2 n}{Q_t}}. \tag{16}$$

Next we note that for each i,

$$\lim_{n\to\infty} \hat{d}_t(x_i) = d_t(x_i)$$

where the convergence is in probability by the simple law of large numbers. Further, since $|G_t| \leqslant \frac{1}{(4\pi t)^{k/2}} = R_t$, the rate is exponential and an application of Hoeffding's inequality yields

$$\mathbb{P}\big[\big|\hat{d}_t(x_i) - d_t(x_i)\big| > \epsilon\big] \leqslant 2e^{-\frac{\epsilon^2(n-1)}{R_t}}.$$

By a union bound (over all i), we have that for all i simultaneously,

$$\mathbb{P}\big[\max_i \big|\hat{d}_t(x_i) - d_t(x_i)\big| > \epsilon\big] \leqslant 2ne^{-\frac{\epsilon^2(n-1)}{R_t}}. \tag{17}$$

It is easy to check that if $\epsilon < \frac{a_t}{2}$ in Eq. (17), with high probability, for all i

$$G_t(x, x_i)\left|\frac{1}{\sqrt{\hat{d}(x)\hat{d}(x_i)}} - \frac{1}{\sqrt{d(x)\,d(x_i)}}\right| = O\left(R_t \frac{\epsilon b_t}{a_t^{3/2}}\right)$$

so that

$$\lim_{n\to\infty} \mathbb{P}\big[\big|\mathbf{L}_n^t f(x) - \mathbf{L}_n^{t\prime} f(x)\big| > \epsilon\big] = 0. \tag{18}$$

Combining Eqs. (16) and (18), we have the result. Rates may be worked out from the above equations and attention to the constants therein.

5.1.2. Analysis of the continuous operator

The rest of the proof revolves around an understanding of the limiting behavior of the underlying continuous operator $\mathbf{L}_P^t$ as $t \to 0$. We begin by noting that by the usual arguments,

$$\lim_{t\to 0} d_t(x) = P(x).$$

By the compactness of $\mathcal{M}$, this convergence holds uniformly and in fact, one can check that

$$d_t(x) = P(x) + O\big(tg(x)\big)$$

where g is a smooth function depending upon higher derivatives of P. Therefore, we can write

$$d_t^{-\frac{1}{2}}(x) = \big(P(x) + O\big(tg(x)\big)\big)^{-\frac{1}{2}} = P(x)^{-\frac{1}{2}} + O(t).$$

Next consider the operator given by

$$\mathbf{L}_P^t f(x) = \frac{1}{t} \int\limits_{\mathcal{M}} \frac{G_t(x,y)}{\sqrt{P(x)}\sqrt{P(y)}} \big(f(x) - f(y)\big) P(y) \operatorname{vol}(y).$$

We first prove that this operator converges pointwise to the weighted Laplacian as $t \to 0$. To see this, first note that by the same analysis as in Lemma 4.1, we have that the above integral can be reduced to the integral over a ball in the limit, i.e.,

$$\begin{aligned}
&\lim_{t\to 0} \frac{1}{t} \int\limits_{B} \frac{G_t(x,y)}{\sqrt{P(x)}\sqrt{P(y)}} \big(f(x) - f(y)\big) P(y) \operatorname{vol}(y) \\
&\quad = \lim_{t\to 0} \frac{1}{t} \int\limits_{\mathcal{M}} \frac{G_t(x,y)}{\sqrt{P(x)}\sqrt{P(y)}} \big(f(x) - f(y)\big) P(y) \operatorname{vol}(y).
\end{aligned} \tag{19}$$

Now, one can follow the logic of Section 4 and write the above expression in exponential coordinates by expanding around T_x and considering the exponential map $\exp_x : T_x \to \mathcal{M}$. Let $\tilde{B}$ be a ball in T_x as usual. Then B may be viewed as the image of this ball under the exponential map. Writing $y = \exp_x(z)$ and recalling that $x = \exp_x(0)$, we have

$$\begin{aligned}
&\frac{1}{t} \int\limits_{B} \frac{G_t(x,y)}{\sqrt{P(x)}\sqrt{P(y)}} \big(f(x) - f(y)\big) P(y) \operatorname{vol}(y) \\
&\quad = \frac{1}{t} \frac{1}{\sqrt{\tilde{P}(0)}} \int\limits_{\tilde{B}} \frac{1}{(4\pi t)^{k/2}} e^{-\frac{\|\exp_x(z) - \exp_x(0)\|^2}{4t}} \sqrt{\tilde{P}(z)} \big(\tilde{f}(z) - \tilde{f}(0)\big) \det g_{ij}\, dz
\end{aligned}$$

where we have used the $\tilde{}$ notation in a similar fashion as in Section 4, i.e., $\tilde{P}(z) = P(\exp_x(z))$ and $\tilde{f}(z) = f(\exp_x(z))$.

Expanding the density P in Taylor series, we write

$$\tilde{P}^{\frac{1}{2}}(z) = \tilde{P}^{\frac{1}{2}}(0) + \frac{1}{2} \tilde{P}^{-\frac{1}{2}}(0) z^T \nabla \tilde{P} + O\big(\|z\|^2\big).$$

Similarly, expanding $\tilde{f}$ in Taylor series, we have

$$\tilde{f}(z) = \tilde{f}(0) + z^T \nabla \tilde{f} + O\big(\|z\|^2\big).$$

Finally, following the discussion in Section 4.5, we note that $e^{-\frac{\|\exp_x(z)-\exp_x(0)\|^2}{4t}}$ is $O(e^{-\frac{\|z\|^2}{4t}} + \frac{\|z\|^4}{t}e^{-\frac{\|z\|^2}{8t}})$ while $\det g_{ij}$ is $O(1+\|z\|^2)$. Thus, we may write the left-hand side of Eq. (19) as

$$\frac{1}{t}\int\limits_B \frac{G_t(x,y)}{\sqrt{P(x)}\sqrt{P(y)}}\big(f(x)-f(y)\big)P(y)\,\mathrm{vol}(y) = \frac{1}{t}\frac{1}{(4\pi t)^{k/2}}\int\limits_{\tilde{B}} A_t(z)B_t(z)C_t(z)\,dz$$

where $A_t(z) = O(e^{-\frac{\|z\|^2}{4t}} + \frac{\|z\|^4}{t}e^{-\frac{\|z\|^2}{8t}})$, $B_t(z) = z^T\nabla\tilde{f} + \frac{1}{2}z^T Hz + O(\|z\|^3)$, and $C_t(z) = \tilde{P}^{-1/2}(0) + \frac{1}{2}z^T\nabla\tilde{P} + O(\|z\|^2)$. Collecting terms in the integrand, letting $t \to 0$, and making use of the observations in Eqs. (12)–(15), we see that

$$\lim_{t\to 0}\mathbf{L}_P^t f(p) = \Delta_{\mathcal{M}} f + \frac{1}{P}\langle \mathrm{grad}\, P, \mathrm{grad}\, f\rangle_{T_p} = \Delta_P f(p).$$

The last step is to note that

$$\frac{1}{t}\int\limits_{\mathcal{M}} G_t(x,y)\left(\frac{1}{\sqrt{d_t(x)d_t(y)}} - \frac{1}{\sqrt{P(x)P(y)}}\right)\big(f(x)-f(y)\big)P(y)\,\mathrm{vol}(y)$$
$$= \frac{1}{t}\int\limits_{\mathcal{M}} G_t(x,y)O(t)\big(f(x)-f(y)\big)P(y)\,\mathrm{vol}(y).$$

By the same arguments with appropriate Taylor expansions, it is easy to check that

$$\lim_{t\to 0}\left|\frac{1}{t}\int\limits_{\mathcal{M}} G_t(x,y)O(t)\big(f(x)-f(y)\big)P(y)\,\mathrm{vol}(y)\right| = \lim_{t\to 0}\int\limits_{\mathcal{M}} G_t(x,y)\big|f(x)-f(y)\big|P(y)\,\mathrm{vol}(y) = 0.$$

Thus we are able to extend our analysis to cover the following theorem:

Theorem 5.2. *Let $\mathcal{M}$ be a compact Riemannian manifold without boundary. Let P be a probability distribution function on $\mathcal{M}$ according to which examples $x_1, \ldots, x_n$ are drawn in i.i.d. fashion. Then assuming that* (i) $0 < a = \inf_{x\in\mathcal{M}} P(x) < b = \sup_{x\in\mathcal{M}} P(x)$ *and* (ii) *P is twice differentiable, we have that there exists a sequence t_n such that $\mathbf{L}_n^{t_n} f(p)$ converges to $\Delta_P f(p)$ where the convergence is in probability.*

6. Uniform convergence

For a fixed function f, let

$$A_f(t) = \frac{1}{t(4\pi t)^{\frac{k}{2}}}\left(\int\limits_{\mathcal{M}} e^{-\frac{\|p-y\|^2}{4t}} f(p)\,d\mu(y) - \int\limits_{\mathcal{M}} e^{-\frac{\|p-y\|^2}{4t}} f(y)\,d\mu(y)\right).$$

Its empirical version from the point cloud is simply

$$\hat{A}_f(t) = \frac{1}{t(4\pi t)^{-\frac{k}{2}}}\frac{1}{n}\sum_{i=1}^{n} e^{-\frac{\|p-y\|^2}{4t}}\big(f(p)-f(x_i)\big) = -\frac{1}{t(4\pi t)^{\frac{k}{2}}}\mathbf{L}_n^t f(p).$$

By the standard law of large numbers, we have that $\hat{A}_f(t)$ converges to $A_f(t)$ in probability. One can easily extend this uniformly over all functions in the following proposition

Proposition 6.1. *Let F be an equicontinuous family of functions with a uniform bound up to the third derivative. Then for each fixed t, we have*

$$\lim_{n\to\infty}\mathbb{P}\Big[\sup_{f\in F}\big|\hat{A}_f(t) - A_f(t)\big| > \epsilon\Big] = 0.$$

Proof. By equicontinuity and the uniform bound, we have by the Arzela–Ascoli theorem that F is a precompact family. Correspondingly, let $F_\gamma \subset F$ be a γ-net in F in the L_∞ topology and let $N(\gamma)$ be the size of this net. This guarantees that for any $f \in F$, there exists $g \in F_\gamma$ such that $\|f - g\|_\infty < \gamma$. By a standard union bound over the finite elements of F_γ, we have

$$\lim_{n\to\infty} \mathbb{P}\Big[\sup_{g\in F_\gamma} \big|\hat{A}_g(t) - A_g(t)\big| > \frac{\epsilon}{2}\Big] = 0.$$

Now for any $f \in F$, we have that

$$\begin{aligned}\big|\hat{A}_f(t) - A_f(t)\big| &\leqslant \big|\hat{A}_f(t) - \hat{A}_g(t) + \hat{A}_g(t) + A_g(t) - A_g(t) - A_f(t)\big| \\ &\leqslant \big|\hat{A}_f(t) - \hat{A}_g(t)\big| + \big|\hat{A}_g(t) - A_g(t)\big| + \big|A_g(t) - A_f(t)\big|.\end{aligned}$$

It is easy to check that for $\gamma = \frac{\epsilon}{4}(4\pi t)^{\frac{k+2}{2}}$, we have

$$\big|\hat{A}_f(t) - A_f(t)\big| < \frac{\epsilon}{2} + \sup_{g\in F_\gamma} \big|\hat{A}_g(t) - A_g(t)\big|.$$

Therefore

$$\mathbb{P}\Big[\sup_{f\in F}\big|\hat{A}_f(t) - A_f(t)\big| > \epsilon\Big] \leqslant \mathbb{P}\Big[\sup_{g\in F_\gamma} \big|\hat{A}_g(t) - A_g(t)\big| > \frac{\epsilon}{2}\Big].$$

Taking limits as n goes to infinity, the result follows. □

Now we note from Proposition 4.4 that for each $f \in F$, we have

$$\lim_{t\to 0}\big(A_f(t) - \Delta_{\mathcal{M}} f(p)\big) = 0.$$

This functional convergence can be extended uniformly over the class F in the following proposition

Proposition 6.2. *Let F be a class of functions uniformly bounded up to three derivatives. Then*

$$\lim_{t\to 0} \sup_{f\in F}\left|\left(A_f(t) - \frac{1}{\mathrm{vol}(\mathcal{M})}\Delta_{\mathcal{M}} f(p)\right)\right| = 0.$$

Proof. The proof is a minor adaptation of that of Proposition 4.4. Following that proof, we note that

$$A_f(t) = \gamma_t \int\limits_{\mathcal{M}} e^{-\frac{\|y-p\|^2}{4t}} \big(f(p) - f(y)\big)\, d\mu(y)$$

where $\gamma_t = \frac{1}{\mathrm{vol}(\mathcal{M})} \frac{1}{t} \frac{1}{(4\pi t)^{\frac{k}{2}}}$.

Using the arguments of Lemma 4.1, and switching to exponential coordinates using $y = \exp_p(x)$, we have

$$A_f(t) = K_t(f) + B_t(f) + C_t(f) + O\big(\|f\|_\infty \gamma_t e^{-\frac{r^2}{4t}}\big)$$

where

$$K_t(f) = \gamma_t \int\limits_{\tilde{B}} e^{-\frac{\|x\|^2}{4t}} \big(\tilde{f}(0) - \tilde{f}(x)\big)\, dx,$$

$$B_t(f) = \gamma_t \int\limits_{\tilde{B}} e^{-\frac{\|x\|^2}{4t}} \big(\tilde{f}(0) - \tilde{f}(x)\big) O\big(\|x\|^2\big)\, dx$$

and

$$C_t(f) = \gamma_t \int\limits_{\tilde{B}} e^{-\frac{\|x\|^2}{4t}} \big(\tilde{f}(0) - \tilde{f}(x)\big)\left(O\left(\frac{g(x)}{4t} e^{\frac{g(x)}{4t}}\right)\big(1 + O\big(\|x\|^2\big)\big)\right) dx$$

where $\tilde{B}$ is a ball in T_p of radius $r > 0$. Examining $K_t(f)$, we see that

$$K_t(f) = \Delta_{\mathcal{M}} f(p)\left(\frac{1}{\mathrm{vol}(\mathcal{M})} - \gamma_t \int\limits_{\tilde{B}^c} e^{-\frac{\|x\|^2}{4t}} \|x\|^2\, dx\right) + \gamma_t \int\limits_{\tilde{B}} e^{-\frac{\|x\|^2}{4t}} O\big(\|x\|^3\big)$$

where the constant in the $O(\|x\|^3)$ term depends on the uniform bounds on the third derivative of the function. Therefore, we have

$$A_f(t) - \frac{1}{\mathrm{vol}(\mathcal{M})}\Delta_{\mathcal{M}} f(p) = B_t(f) + C_t(f) + \left(K_t(f) - \frac{1}{\mathrm{vol}(\mathcal{M})}\Delta_{\mathcal{M}} f(p)\right) + O\big(\|f\|_\infty \gamma_t e^{-\frac{r^2}{4t}}\big).$$

Examining the expressions for $B_t(f)$, $C_t(f)$, and $K_t(f)$, we see that

$$\lim_{t\to 0} \sup_{f\in F} \left| A_f(t) - \frac{1}{\mathrm{vol}(\mathcal{M})}\Delta_{\mathcal{M}} f(p)\right| = 0$$

where we have explicitly used the fact that all F is uniformly bounded up to three derivatives. □

Therefore, from Propositions 6.1 and 6.2, we see that there exists a monotonically decreasing sequence t_n such that $\lim_{n\to\infty} t_n = 0$ for which the following theorem is true

Theorem 6.3. *Let data points $x_1, \dots, x_n$ be sampled from a uniform distribution on a compact manifold $\mathcal{M} \subset \mathbb{R}^N$ and let $\mathcal{F}_C$ be a family of functions $f \in C^\infty(\mathcal{M})$, which is equicontinuous and with a uniform bound on all derivatives up to order* 3. *Then there exists a sequence of real numbers t_n, $t_n \to 0$, such that in probability*

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F}_C} \left| \frac{1}{t(4\pi t_n)^{\frac{k}{2}}} \mathbf{L}_n^{t_n} f(x) - \frac{1}{\mathrm{vol}(\mathcal{M})}\Delta_{\mathcal{M}} f(x)\right| = 0.$$

A similar uniform convergence bound can be shown using the compactness of $\mathcal{M}$ and leads to Theorem 2.

7. Auxiliary results

7.1. Bound on the Gaussian in $\mathbb{R}^k$

Lemma 7.1. *Let $B \in \mathbb{R}^k$ be an open set, such that $y \in B$. Then as $t \to 0$*

$$\int\limits_{\mathbb{R}^k - B} (4\pi t)^{-\frac{k}{2}} e^{-\frac{\|x-y\|^2}{4t}}\, dx = o\left(\frac{1}{t} e^{-\frac{1}{t}}\right).$$

Proof. Without a loss of generality we can assume that $y = 0$. There exists a cube $C_s \subset B$ with side s, centered at the origin. We have

$$0 < \int\limits_{\mathbb{R}^k - B} (4\pi t)^{-\frac{k}{2}} e^{-\frac{\|\mathbf{z}\|^2}{4t}}\, dx < \int\limits_{\mathbb{R}^k - C_s} (4\pi t)^{-\frac{k}{2}} e^{-\frac{\|\mathbf{z}\|^2}{4t}}\, dx.$$

Using the standard substitution $z = \frac{x}{\sqrt{t}}$, we can rewrite the last integral as

$$\int\limits_{\mathbb{R}^k - C_s} (4\pi t)^{-\frac{k}{2}} e^{-\frac{\|x\|^2}{4t}}\, dx = \int\limits_{\mathbb{R}^k - C_{\frac{s}{\sqrt{t}}}} (4\pi)^{-\frac{k}{2}} e^{-\frac{\|z\|^2}{4}}\, dz$$

where $C_{\frac{s}{\sqrt{t}}}$ is a cube with side $\frac{s}{\sqrt{t}}$. The last quantity is the probability that at least one coordinate of a standard multivariate Gaussian is greater than $\frac{s}{\sqrt{t}}$ in absolute value. Using the union bound, we get

$$\int\limits_{\mathbb{R}^k - C_{\frac{s}{\sqrt{t}}}} (4\pi)^{-\frac{k}{2}} e^{-\frac{\|z\|^2}{4}}\, dz \leqslant k(4\pi)^{-\frac{k}{2}} \int\limits_{u \notin [-\frac{s}{\sqrt{t}}, \frac{s}{\sqrt{t}}]} e^{-u^2}\, du = k(4\pi)^{-\frac{k}{2}} \left(\sqrt{\pi} - \sqrt{\pi}\, \mathrm{Erf}\left(\frac{s}{\sqrt{t}}\right)\right).$$

The function $1 - \mathrm{Erf}(x)$ decays faster than e^{-x} and the claim follows. □

7.2. Hoeffding inequality

Theorem 7.2 *(Hoeffding). Let $X_1, \ldots, X_n$ be independent identically distributed random variables, such that $|X_i| \leqslant K$. Then*

$$P\left\{\left|\frac{\sum_i X_i}{n} - \mathbb{E}X_i\right| > \epsilon\right\} < 2\exp\left(-\frac{\epsilon^2 n}{2K^2}\right).$$

References

[1] M. Belkin, Problems of learning on manifolds, PhD Dissertation, The University of Chicago, 2003.
[2] M. Belkin, P. Niyogi, Using manifold structure for partially labeled classification, in: NIPS 2002.
[3] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput. 15 (6) (2003) 1373–1396.
[4] M. Belkin, P. Niyogi, Towards a theoretical foundation for Laplacian–Based manifold methods, in: COLT 2005.
[5] M. Belkin, P. Niyogi, V. Sindhwani, On manifold regularization, in: AI Stats 2005.
[6] Y. Bengio, J.-F. Paiement, P. Vincent, O. Delalleau, N. Le Roux, M. Ouimet, Out-of-sample extensions for LLE, isomap, MDS, eigenmaps, and spectral clustering, in: NIPS 2003.
[7] M. Bernstein, V. de Silva, J.C. Langford, J.B. Tenenbaum, Graph approximations to geodesics on embedded manifolds, Technical Report, 2000.
[8] O. Bousquet, O. Chapelle, M. Hein, Measure based regularization, in: NIPS 2003.
[9] O. Chapelle, J. Weston, B. Schoelkopf, Cluster kernels for semi-supervised learning, in: NIPS 2002.
[10] F.R.K. Chung, Spectral Graph Theory, Reg. Conf. Ser. Math., vol. 92, 1997.
[11] R.R. Coifman, S. Lafon, A. Lee, M. Maggioni, B. Nadler, F. Warner, S. Zucker, Geometric diffusions as a tool for harmonic analysis and structure definition of data, Parts I and II, Proc. Natl. Acad. Sci. 102 (May 2005) 7426–7438.
[12] M. do Carmo, Riemannian Geometry, Birkhäuser, 1992.
[13] D.L. Donoho, C.E. Grimes, Hessian eigenmaps: New locally linear embedding techniques for high-dimensional data, Proc. Natl. Acad. Arts Sci. 100 (2003) 5591–5596.
[14] E. Gine, V. Kolchinski, Empirical graph Laplacian approximation of Laplace–Beltrami operators: Large sample results, preprint, 2006.
[15] M. Hein, J.-Y. Audibert, U. von Luxburg, From graphs to manifolds—Weak and strong pointwise consistency of graph Laplacians, in: Proceedings of the 18th Annual Conference on Learning Theory (COLT 2005).
[16] M. Meila, J. Shi, Learning segmentation by random walks, in: NIPS 2000.
[17] F. Memoli, G. Sapiro, Comparing point clouds, IMA Technical Report, 2004.
[18] S. Lafon, Diffusion maps and geodesic harmonics, PhD thesis, Yale University, 2004.
[19] U. von Luxburg, M. Belkin, O. Bousquet, Consistency of spectral clustering, Max Planck Institute for Biological Cybernetics Technical Report TR 134, 2004.
[20] A. Ng, M. Jordan, Y. Weiss, On spectral clustering: Analysis and an algorithm, in: NIPS 2001.
[21] P. Niyogi, Estimating functional maps on Riemannian submanifolds from sampled data, http://www.ipam.ucla.edu/publications/mgaws3/mgaws3_5188.pdf, presented at IPAM Workshop on Multiscale Structures in the Analysis of High-Dimensional Data, 2004.
[22] A. Grigoryan, Heat Kernels on Weighted Manifolds and Applications, Contemp. Math., vol. 398, 2006.
[23] P. Niyogi, S. Smale, S. Weinberger, Finding the homology of submanifolds with high confidence from random samples, Discrete Comput. Geom. 39 (1–3) (2008) 419–441.
[24] S. Rosenberg, The Laplacian on a Riemannian Manifold, Cambridge University Press, 1997.
[25] Sam T. Roweis, Lawrence K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (2000).
[26] A. Smola, R. Kondor, Kernels and regularization on graphs, in: COLT/KW 2003.
[27] J. Shi, J. Malik, Normalized Cuts and Image Segmentation, IEEE Trans. Pattern Anal. Machine Intelligence 22 (2000) 8.
[28] A. Singer, From graph to manifold Laplacian: The convergence rate, Appl. Comput. Harmon. Anal. 21 (1) (2006) 135–144.
[29] D. Spielman, S. Teng, Spectral partitioning works: Planar graphs and finite element meshes, in: FOCS 1996.
[30] Martin Szummer, Tommi Jaakkola, Partially labeled classification with Markov random walks, in: NIPS 2001.
[31] J.B. Tenenbaum, V. de Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (2000).
[32] R. Kannan, S. Vempala, A. Vetta, On clusterings—Good, bad and spectral, Technical Report, Computer Science Department, Yale University, 2000.
[33] I. Kondor, J. Lafferty, Diffusion kernels on graphs and other discrete input spaces, in: Proc. 19th Intl. Conf. on Machine Learning, ICML.
[34] O.G. Smolyanov, H. von Weizsäcker, O. Wittich, Brownian motion on a manifold as limit of stepwise conditioned standard Brownian motions, in: Stochastic Processes, Physics and Geometry: New Interplays. II: A Volume in Honour of S. Albeverio, in: Can. Math. Soc. Conf. Proc., vol. 29, Amer. Math. Soc., 2000, pp. 589–602.

[35] O.G. Smolyanov, H. von Weizsacker, O. Wittich, Chernoff's theorem and discrete time approximations of Brownian motion on manifolds, Potential Anal. 26 (2007) 1–29.
[36] D. Zhou, O. Bousquet, T.N. Lal, J. Weston, B. Schoelkopf, Learning with local and global consistency, in: NIPS 2003.
[37] X. Zhu, J. Lafferty, Z. Ghahramani, Semi-supervised learning using Gaussian fields and harmonic functions, in: ICML 2003.
[38] A. Zomorodian, G. Carlsson, Computing persistent homology, in: 20th ACM Symposium on Computational Geometry, 2004.