Towards a Theoretical Foundation for Laplacian-Based Manifold Methods

Mikhail Belkin and Partha Niyogi

The University of Chicago, Department of Computer Science
{misha, niyogi}@cs.uchicago.edu

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. We show that under certain conditions the graph Laplacian of a point cloud converges to the Laplace-Beltrami operator on the underlying manifold. Theorem 1 contains the first result showing convergence of a random graph Laplacian to manifold Laplacian in the machine learning context.

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.

In this paper we take the first steps towards a theoretical foundation for manifold-based methods in learning. 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 only applicable, when it can be related to the underlying process generating the data. This is an implicit assumption, which is rarely formalized in the literature.

We will show that under certain conditions the graph Laplacian is directly related to the manifold Laplace-Beltrami operator and converges to it as data goes to infinity.

This paper presents and extends the unpublished results obtained in [1]. A version of Theorem 1 showing empirical convergence of the graph Laplacian to the manifold Laplacian was stated in [19].

1.1 Prior Work

Many manifold and graph-motivated learning methods have been recently proposed, including [22, 27, 3, 12] for visualization and data representation, [30, 29, 9, 23, 4, 2, 26] for partially supervised classification and [25, 28, 24, 18, 14] among others for spectral clustering. A discussion of various spectral methods and their out-of-sample extensions is given in [5].

The problem of estimating geometric and topological invariants from point cloud data has recently attracted some attention. Some of the recent work includes estimating geometric invariants of the manifold, such as homology [31, 20], geodesic distances [6], and comparing point clouds using Gromov-Hausdorff distance [15].

In particular, we note the closely related Ph.D. thesis of Lafon, [16], which generalized the convergence results from [1] to the important case of an arbitrary probability distribution on a manifold. Those results are further generalized and presented with an empirical convergence theorem in the parallel COLT paper [13].

We also note [17], where convergence of a class of graph Laplacians and the associated spectral objects, such as eigenfunctions and eigenvalues, is shown, which in particular, implies consistency of normalized spectral clustering. However connections to geometric objects, such as the Laplace-Beltrami operator, are 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 exact nature of that parallel is usually not made precise.

2 Notation and Preliminaries

Before we can formulate the main result we need to fix some notation. In general, we denote vectors and points on a manifold with bold letters and one-dimensional quantities with ordinary letters. Matrices will be denoted by capital letters, operators on functions by bold capital letters.

A weighted graph $G = (V, E)$ is a set of vertices $v_1, \ldots, v_n \in V$ and weighted edges connecting these vertices represented by an adjacency matrix W. W is a symmetric matrix with nonnegative entries. Recall that the *Laplacian* matrix of a weighted graph G is the matrix $L = D - W$, where D is a diagonal matrix $D(i, i) = \sum_j W(i, j)$.

Given a set of points $\mathcal{S}_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ in $\mathbb{R}^k$, we construct a graph G, whose vertices are data points. We put $W_n^t(i,j) = e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{4t}}$. We will denote the corresponding graph Laplacian by $L_n^t = D_n^t - W_n^t$. Note that we suppress the dependence on $\mathcal{S}_n$ to simplify notation.

We may think of L_n^t as an operator on functions, defined on the graph of data points. If $f : V \to \mathbb{R}$

$$L_n^t f(\mathbf{x}_i) = f(\mathbf{x}_i) \sum_j e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{4t}} - \sum_j f(\mathbf{x}_j) e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{4t}}$$

This operator can be naturally extended to an integral operator (with respect to the empirical measure of the dataset) on functions in $\mathbb{R}^k$:

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

Of course, we have $\mathbf{L}_n^t f(\mathbf{x}_i) = L_n^t f(\mathbf{x}_i)$. We will call $\mathbf{L}_n^t$ the *Laplacian operator associated to the point cloud* $\mathcal{S}_n$.

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[1] k-dimensional differentiable manifold $\mathcal{M}$ isometrically embedded in $\mathbb{R}^N$. We will assume that the data is sampled from a uniform distribution in the induced measure on $\mathcal{M}$.

Given data points $\mathcal{S}_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ in $\mathbb{R}^N$ sampled i.i.d. from this probability distribution we construct the associated Laplacian 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 $\mathbf{p} \in \mathcal{M}$, after appropriate scaling the operator $\mathbf{L}_n^t$ converges to the true Laplace-Beltrami operator on the manifold.

Theorem 1. *Let data points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ 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 there is a constant C, s.t. in probability,*

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

Without going into full details we then outline the proof of the following

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

Theorem 2. *Let data points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be sampled from a uniform distribution on a compact manifold $\mathcal{M} \subset \mathbb{R}^N$. Let $\mathcal{F}$ be the space of functions $f \in C^\infty(\mathcal{M})$, such that Δf is Lipschitz a fixed Lipschitz constant. Then there exists a sequence of real numbers $t_n \to 0$, and a constant C, such that in probability*

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

This stronger uniform result (with, however, a potentially worse rate of convergence) will in our opinion lead to consistency results for various learning algorithms in the future work.

3.1 Laplace Operator and the Heat Equation

We will now recall some results on the heat equation and its connection to the Laplace-Beltrami operator and develop some intuitions about the methods used in the proof.

Now we need to recall some results about the heat equation and heat kernels. Recall that the *Laplace operator* in $\mathbb{R}^k$ is defined as

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

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

$$\frac{\partial}{\partial t} u(\mathbf{x}, t) - \Delta u(\mathbf{x}, t) = 0 \tag{1}$$

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

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

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

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

We summarize this in the following

Theorem 3 (Solution to the heat equation in $\mathbb{R}^k$). *Let $f(\mathbf{x})$ be a sufficiently differentiable bounded function. We then have*

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

$$f(\mathbf{x}) = \lim_{t\to 0} \mathbf{H}^t f(\mathbf{x}) = (4\pi t)^{-\frac{k}{2}} \int_{\mathbb{R}^k} e^{-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{4t}} f(\mathbf{y}) d\mathbf{y} \tag{3}$$

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

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

The heat equation is the key to approximating the Laplace operator. Recalling that a Gaussian integrates to 1, we observe that

$$-\Delta f(\mathbf{x}) = \frac{\partial}{\partial t} \mathbf{H}^t f(\mathbf{x}) \Big|_{t=0} =$$

$$\lim_{t\to 0} \frac{1}{t} \left((4\pi t)^{-\frac{k}{2}} \int_{\mathbb{R}^k} e^{-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{4t}} f(\mathbf{y}) d\mathbf{y} - f(\mathbf{x}) \, (4\pi t)^{-\frac{k}{2}} \int_{\mathbb{R}^k} e^{-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{4t}} d\mathbf{y} \right)$$

This quantity can easily be approximated from a point cloud[2] $\mathbf{x}_1, \ldots, \mathbf{x}_n$ by computing the empirical version of the integrals involved:

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

$$\frac{(4\pi t)^{-\frac{k+2}{2}}}{n} \mathbf{L}_n^t(f)(\mathbf{p})$$

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 $H^t_{\mathcal{M}}(\mathbf{x}, \mathbf{y})$.
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 as certain intrinsic quantities (scalar curvature) make an appearance and ultimately cancel out in the final computation!

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[3] differentiable k-dimensional submanifold

[2] 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 $\mathbf{x}$.

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

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 $\mathbf{x}, \mathbf{y} \in \mathcal{M}$ the *geodesic distance* $\mathrm{dist}_{\mathcal{M}}(\mathbf{x}, \mathbf{y})$ is the length of the shortest curve connecting $\mathbf{x}$ and $\mathbf{y}$. It is clear that $\mathrm{dist}_{\mathcal{M}}(\mathbf{x}, \mathbf{y}) \geq \|\mathbf{x} - \mathbf{y}\|$.

Given a point $\mathbf{p} \in \mathcal{M}$, one can identify the *tangent space* $T_{\mathbf{p}}\mathcal{M}$ with an affine subspace of $\mathbb{R}^N$ passing through $\mathbf{p}$. This space has a natural linear structure with the origin at $\mathbf{p}$. Furthermore it is possible to define the *exponential map* $\exp_{\mathbf{p}} : T_{\mathbf{p}}\mathcal{M} \to \mathcal{M}$. The key property of the exponential map is that it takes lines through origin in $T_{\mathbf{p}}\mathcal{M}$ to geodesics passing through $\mathbf{p}$. The exponential map is a local diffeomorphism and produces a natural system of coordinates for some neighborhood of $\mathbf{p}$. The Hopf-Rinow theorem (see, e.g., [11]) 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 corresponding to the *volume form*, which we will denote as μ. 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 $\mathcal{M}$.

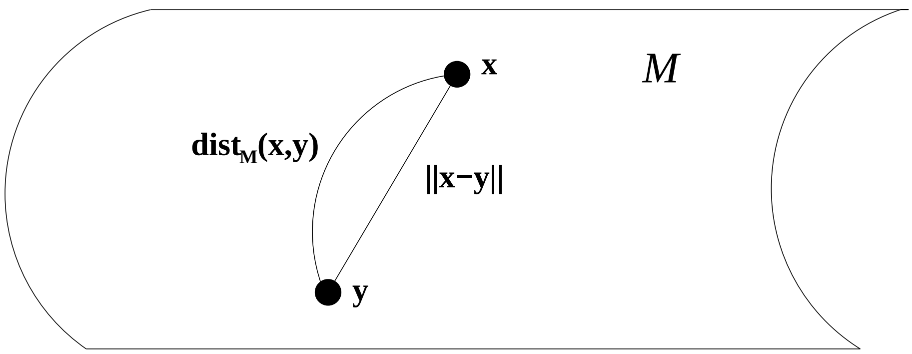

Fig. 1. Geodesic and chordal distance.

Before proceeding with the main proof we state one curious property of geodesics, which will be needed later. It concerns the relationship between $\mathrm{dist}_{\mathcal{M}}(\mathbf{x}, \mathbf{y})$ and $\|\mathbf{x} - \mathbf{y}\|$. The geodesic and chordal distances are shown pictorially in Fig. 1. It is clear that when $\mathbf{x}$ and $\mathbf{y}$ 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. 7) that when the manifold is compact

$$\mathrm{dist}_{\mathcal{M}}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| + O(\|\mathbf{x} - \mathbf{y}\|^3)$$

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:

$$\Delta_{\mathcal{M}}\mathbf{H}^t_{\mathcal{M}}(f) = \frac{\partial}{\partial t}\mathbf{H}^t_{\mathcal{M}}(f) \qquad \text{Heat Equation}$$

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

It can be shown (see, e.g., [21]) that $\mathbf{H}^t_{\mathcal{M}}(f)$ is an integral operator, a convolution with the *heat kernel*. Our proof hinges on the fact that in geodesic coordinates the heat kernel can be approximated by a Gaussian for small values of t and the observations about the geodesics above.

4.2 Main Proof

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

First we note that the quantities

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

and

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

can be empirically estimated from the point cloud.

We will show how the Laplace-Beltrami operator can be estimated using these two empirical quantities. This estimate will provide a connection to $\mathbf{L}^t_n$

The main theorem will be proved in several steps.

Lemma 1. *Given any open set $\mathcal{B}\subset\mathcal{M}$, $\mathbf{p}\in\mathcal{B}$, for any $l\in N$,*

$$\int_{\mathcal{B}\subset\mathcal{M}} e^{-\frac{\|\mathbf{p}-\mathbf{y}\|^2}{4t}} f(\mathbf{y})\; d\mu_{\mathbf{y}} - \int_{\mathcal{M}} e^{-\frac{\|\mathbf{p}-\mathbf{y}\|^2}{4t}} f(\mathbf{y})\; d\mu_{\mathbf{y}} = o(t^l)$$

as $t\to 0$.

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

$$\left|\int_{\mathcal{B}} e^{-\frac{\|\mathbf{p}-\mathbf{y}\|^2}{4t}} f(\mathbf{y})\; d\mu_{\mathbf{y}} - \int_{\mathcal{M}} e^{-\frac{\|\mathbf{p}-\mathbf{y}\|^2}{4t}} f(\mathbf{y})\; d\mu_{\mathbf{y}}\right| \;\le\; M\sup_{\mathbf{x}\in\mathcal{M}}(|f(\mathbf{x})|)e^{-\frac{d^2}{4t}}$$

The first two terms are constant and $e^{-\frac{d^2}{4t}}$ approaches zero faster then any polynomial as t tends to zero.

This Lemma allows us to replace the integral over the manifold by an integral over some small open set around $\mathbf{p}$. We will need it in order to change the coordinates to the standard geodesic coordinate system given by the following equation:

$$\mathbf{y} = \exp_{\mathbf{p}}(\mathbf{x})$$

Given a function $f : \mathcal{M} \to \mathbb{R}$, we rewrite it in geodesic coordinates by putting $\tilde{f}(\mathbf{x}) = f(\exp(\mathbf{x}))$.

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

Lemma 2.

$$\Delta_{\mathcal{M}} f(\mathbf{p}) = \Delta_{\mathbb{R}^k} \tilde{f}(0) \tag{4}$$

Proof. See, e.g., [21], page 90.

This allows one to reduce Laplace-Beltrami operator to a more easily analyzed Laplace operator on $\mathbb{R}^k$.

Since $\exp_{\mathbf{p}} : T\mathcal{M}_{\mathbf{p}} = \mathbb{R}^k \to \mathcal{M}$ is a locally invertible, we can choose an open $\tilde{\mathcal{B}} \subset \mathbb{R}^k$, s.t. $\exp_{\mathbf{p}}$ is a diffeomorphism onto its image $\mathcal{B} \subset \mathcal{M}$.

Lemma 3. *The following change of variable formula holds:*

$$\int_{\mathcal{B}} e^{-\frac{\|\mathbf{p}-\mathbf{y}\|^2}{4t}} f(\mathbf{y})\ d\mu_{\mathbf{y}} = \int_{\tilde{\mathcal{B}}} e^{-\frac{\phi(\mathbf{x})}{4t}} \tilde{f}(\mathbf{x}) \det(d\exp(\mathbf{x}))\ d\mathbf{x} \tag{5}$$

where $\phi(\mathbf{x})$ is a function, such that $\phi(\mathbf{x}) = \|\mathbf{x}^2\| + O(\|\mathbf{x}^4\|)$.

Proof. We obtain the result by applying the usual change of variable formula for manifold integrals and observing the relationship between geodesic and chordal distances from Lemma 7.

Lemma 4. *There exists a constant C, such that*

$$\frac{\partial}{\partial t}\left((4\pi t)^{-\frac{k}{2}} \int_{\mathcal{B}} e^{-\frac{\|\mathbf{p}-\mathbf{y}\|^2}{4t}} f(\mathbf{y})\ d\mu_{\mathbf{y}}\right)\Bigg|_0 = \Delta_{\mathcal{M}} f(\mathbf{p}) + \frac{1}{3} k s(\mathbf{p}) f(\mathbf{p}) + C f(\mathbf{p}) \tag{6}$$

Proof. We first use Eq. 5 from the previous Lemma to rewrite the integral in the geodesic normal coordinates. We then apply Eq. 12 to obtain

$$\frac{\partial}{\partial t}\left((4\pi t)^{-\frac{k}{2}} \int_{\mathcal{B}} e^{-\frac{\|\mathbf{p}-\mathbf{y}\|^2}{4t}} f(\mathbf{y})\ d\mu_{\mathbf{y}}\right)\Bigg|_0 = \Delta_{\mathbb{R}^k}(\tilde{f} \det(d\exp_{\mathbf{p}}))(0) + C\tilde{f}(0) \tag{7}$$

From the asymptotics of the exponential map (Eq. 10), we know that

$$|\Delta_{\mathbb{R}^k} \det(d\exp_{\mathbf{p}}(\mathbf{x}))| = \frac{s(\mathbf{p})}{3} + O(\|\mathbf{x}\|)$$

Using properties of the Laplacian and recalling that $\tilde{f}(0) = f(\mathbf{p})$ yields and that $\det(d\exp_{\mathbf{p}}(\mathbf{x}))|$ has no terms of degree 1 in its Taylor expansion at 0, we have

$$\Delta_{\mathbb{R}^k}(\tilde{f} \det(d\exp_{\mathbf{p}}))(0) = \Delta_{\mathbb{R}^k} \tilde{f}(0) + \frac{1}{3} k s(\mathbf{p}) f(\mathbf{p})$$

Noticing that by Eq. 4 $\Delta_{\mathbb{R}^k} \tilde{f}(0) = \Delta_{\mathcal{M}} f(\mathbf{p})$, we obtain the result.

Thus we get the following

Lemma 5.

$$\lim_{t\to 0}\ (4\pi t)^{-\frac{k+2}{2}}\left(\int_{\mathcal{M}} e^{\frac{\|\mathbf{p}-\mathbf{y}\|^2}{4t}} f(\mathbf{p})\ d\mu_{\mathbf{y}} - \int_{\mathcal{M}} e^{\frac{\|\mathbf{p}-\mathbf{y}\|^2}{4t}} f(\mathbf{y})\ d\mu_{\mathbf{y}}\right) = \Delta_{\mathcal{M}} f(\mathbf{p})$$

Proof. Consider the constant function $g(\mathbf{y}) = f(\mathbf{p})$. By applying the Eq. 6 to this function we obtain

$$\frac{\partial}{\partial t}\left((4\pi t)^{-\frac{k}{2}}\int_{\mathcal{B}} e^{-\frac{\|\mathbf{p}-\mathbf{y}\|^2}{4t}} f(\mathbf{p})\ d\mu_{\mathbf{y}}\right)\Bigg|_0 = \frac{1}{3}ks(\mathbf{p})f(\mathbf{p}) + Cf(\mathbf{p}) \tag{8}$$

To simplify the formulas put $A(t) = (4\pi t)^{-\frac{k+2}{2}}\int_{\mathcal{M}} e^{\frac{\|\mathbf{p}-\mathbf{y}\|^2}{4t}} f(\mathbf{y})\ d\mu_{\mathbf{y}}$. Using the δ-family property of the heat kernel, we see that

$$A(0) = \lim_{t\to 0}\ (4\pi t_n)^{-\frac{k}{2}}\int_{\mathcal{B}} e^{-\frac{\|\mathbf{p}-\mathbf{y}\|^2}{4t}} f(\mathbf{p})\ d\mu_{\mathbf{y}} = f(\mathbf{p})$$

From the definition of the derivative and Eqs. 6,8 we obtain

$$\Delta_{\mathcal{M}} f(\mathbf{p}) = \lim_{t\to 0}\frac{A(t)-A(0)}{t} =$$

$$\lim_{t\to 0}\ (4\pi t_n)^{-\frac{k+2}{2}}\left(\int_{\mathcal{M}} e^{\frac{\|\mathbf{p}-\mathbf{y}\|^2}{4t}} f(\mathbf{p})\ d\mu_{\mathbf{y}} - \int_{\mathcal{M}} e^{\frac{\|\mathbf{p}-\mathbf{y}\|^2}{4t}} f(\mathbf{y})\ d\mu_{\mathbf{y}}\right)$$

Theorem 4. *Let data points $\mathbf{x}_1,\ldots,\mathbf{x}_n$ be sampled in i.i.d. fashion from a uniform distribution on a compact submanifold $\mathcal{M}\subset\mathbb{R}^N$. Fix $\mathbf{p}\in\mathcal{M}$. Let $\mathbf{L}_n^{t_n}$ be the associated operator. Put $t_n = n^{-\frac{1}{k+2+\alpha}}$, where $\alpha > 0, \alpha\in\mathbb{R}$. Then in probability*

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

Proof. Recall that (the extension of) the graph Laplacian $\mathbf{L}_n^t$ applied to f at $\mathbf{p}$ is

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

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

$$\mathbb{E}\,\mathbf{L}_n^t f(\mathbf{p}) = \left(f(\mathbf{p})\int_{\mathcal{M}} e^{-\frac{\|\mathbf{p}-\mathbf{y}\|^2}{4t}}\ d\mathbf{y} - \int_{\mathcal{M}} f(\mathbf{y})e^{-\frac{\|\mathbf{p}-\mathbf{y}\|^2}{4t}}\ d\mathbf{y}\right) \tag{9}$$

By an application of Hoeffding's inequality 6, we have

$$\mathbb{P}\left[(4\pi t)^{-(k+2)/2}|\mathbf{L}_n^t f(\mathbf{p}) - \mathbb{E}\,\mathbf{L}_n^t f(\mathbf{p})| > \epsilon\right] \le e^{-\epsilon^2 n(4\pi t)^{(k+2)}}$$

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 $\epsilon > 0$

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

Noting that by Lemma 5 and Eq. 9

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

we obtain the theorem.

5 Uniform Convergence

For a fixed function f, let

$$A_f(t) = (4\pi t)^{-\frac{k+2}{2}} \left(\int_{\mathcal{M}} e^{-\frac{\|\mathbf{p}-\mathbf{y}\|^2}{4t}} f(\mathbf{p})\ d\mu_{\mathbf{y}} - \int_{\mathcal{M}} e^{-\frac{\|\mathbf{p}-\mathbf{y}\|^2}{4t}} f(\mathbf{y})\ d\mu_{\mathbf{y}}\right)$$

Its empirical version from the point cloud is simply

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

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 1. *Let F be the space of functions $f \in C^\infty(\mathcal{M})$, such that Δf is Lipschitz with Lipschitz constant C. For each fixed t, we have*

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

Proof. Let $F_\gamma \subset F$ be a γ-net in F in the L_∞ topology (guaranteed by the Sobolev embedding theorem) 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

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

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

$$|\hat{A}_f(t) - A_f(t)| \le |\hat{A}_f(t) - \hat{A}_g(t) + \hat{A}_g(t) + A_g(t) - A_g(t) - A_f(t)|$$

$$\leq |\hat{A}_f(t) - \hat{A}_g(t)| + |\hat{A}_g(t) - A_g(t)| + |A_g(t) - A_f(t)|$$

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

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

Therefore

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

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

Now we note from Lemma 5 that for each $f \in F$, we have

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

By an analog of the Arzela-Ascoli Theorem, the uniform convergence over a ball in a suitable Sobolev space over a compact domain can be shown, i.e.,

$$\lim_{t \to 0} \sup_{f \in F}(A_f(t) - \Delta_{\mathcal{M}} f(p)) = 0$$

Therefore, from Proposition 1 and the above fact, 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 5. *Let data points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be sampled from a uniform distribution on a compact manifold $\mathcal{M} \subset \mathbb{R}^N$ and let $\mathcal{F}_C$ be the space of functions $f \in C^\infty(\mathcal{M})$, such that Δf is Lipschitz with Lipschitz constant C. Then there exists a sequence of real numbers t_n, $t_n \to 0$, such that in probability*

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

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

6 Auxiliary and Technical Lemmas

6.1 Exponential Map and Geodesics

Lemma 6. *Asymptotics for the derivative of the* exp.

$$|\Delta_{\mathbb{R}^k} \det(d\exp_{\mathbf{p}}(\mathbf{x}))| = \frac{s(\mathbf{p})}{3} + O(\|\mathbf{x}\|) \tag{10}$$

where $s(\mathbf{p})$ is the scalar curvature of $\mathcal{M}$ at $\mathbf{p}$.

Proof. This fairly standard result of differential geometry follows from properties of Jacobi fields. While the proof goes beyond the scope of this paper, cf. the discussion on page 115 in [11]. The result above follows from Eq. 6 together with some basic linear algebra after writing the curvature tensor in the geodesic normal coordinates.

Lemma 7.

$$\|\exp_{\mathbf{p}}(\mathbf{x})\|^2 = \|\mathbf{x} - \mathbf{p}\|^2 + O(\|\mathbf{x} - \mathbf{p}\|^4)$$

Proof. The geodesic distance from a fixed point $\mathbf{x} \in \mathcal{M}^k$ as a function of $\mathbf{y}$ can be written as

$$\mathrm{dist}_{\mathcal{M}^k}(\mathbf{x}, \mathbf{y}) = \|\mathbf{y} - \mathbf{x}\| + O(\|\mathbf{y} - \mathbf{x}\|^3)$$

where $\|\mathbf{y} - \mathbf{x}\|$ is the ordinary norm in $\mathbb{R}^N$. Thus the geodesic distance can be approximated by Euclidean distance in the ambient space up to terms of order three. We outline the proof. We first prove the statement for the curve length in $\mathbb{R}^2$. Let $f(x)$ be a differentiable function. Without the loss of generality we can assume that $f(0) = 0$, $f'(0) = 0$. Therefore $f(x) = ax^2 + O(x^3)$. Now the length of the curve along the graph of $f(x)$ is given by

$$\mathrm{dist}_{\mathcal{M},0}(t) = \int_0^t \sqrt{1 + (f')^2}\ dx$$

We have $\sqrt{1 + (f')^2} = 1 + 2ax^2 + O(x^3)$. Thus

$$\int_0^t \sqrt{1 + (f')^2}\ dx = t + \frac{2}{3}at^3 + O(t^4)$$

Similarly, we can also see that segment of the line connecting the point t to the origin is equal in length to both the curve length and to t up to some terms of order 3.

In general, we can take a section of the manifold by a 2-dimensional plane through $\mathbf{x}$ and $\mathbf{y}$, such that the plane intersects the manifold at a curve. It is not hard to see, that such a plane always exists.

It is clear that the length of the geodesic is bounded from below by the length of the line segment connecting $\mathbf{x}$ to the $\mathbf{y}$ and from above by the length of the curve formed by intersection of the plane and $\mathcal{M}^k$. By applying the case of $\mathbb{R}^2$, we see that the latter is equal to $\|\mathbf{x} - \mathbf{y}\|$ plus order three terms, which implies the statement.

6.2 Technical Results in $\mathbb{R}^k$

Lemma 8. *Let $\mathcal{B} \in \mathbb{R}^k$ be an open set, such that $\mathbf{x} \in \mathcal{B}$. Then as $t \to 0$*

$$\int_{\mathbb{R}^k - \mathcal{B}} (4\pi t)^{-\frac{k}{2}} e^{-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{4t}}\ d\mathbf{x} = o\left(\frac{1}{t} e^{-\frac{1}{t}}\right)$$

Proof. Without a loss of generality we can assume that $\mathbf{x} = 0$. There exists a cube C_s with side s, such that $0 \in C_s \in \mathcal{B}$. We have $\int\limits_{\mathbb{R}^k - \mathcal{B}} (4\pi t)^{-\frac{k}{2}} e^{-\frac{\|\mathbf{z}\|^2}{4t}} \, d\mathbf{x} <$ $\int\limits_{\mathbb{R}^k - C_s} (4\pi t)^{-\frac{k}{2}} e^{-\frac{\|\mathbf{z}\|^2}{4t}} \, d\mathbf{x}$. Using the standard substitution $\mathbf{z} = \frac{\|\mathbf{z}|}{\sqrt{t}}$, we can rewrite the last integral as

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

The last quantity is the probability that all coordinates of a standard multivariate Gaussian are greater than than $\frac{s}{\sqrt{t}}$ in absolute value and is therefore equal to $2 - 2(1 - \operatorname{Erf}(\frac{s}{\sqrt{t}}))^k < 2k - 2k \operatorname{Erf}(\frac{s}{\sqrt{t}})$. Applying a well-known inequality $1 - \operatorname{Erf}(t) < \frac{1}{t \exp(t^2)}$ yields the statement.

Lemma 9. *Let $\phi : \mathbb{R}^k \to \mathbb{R}^k$ be a differentiable function such that $\phi(\mathbf{x}) = \mathbf{x} + O(\mathbf{x}^3)$, i.e. the Taylor expansion for each coordinate of ϕ does not have any terms of degree 2, $[\phi(\mathbf{x})]_i = x_i + O(\|\mathbf{x}\|^3)$ at the origin. Then for any open set $\mathcal{B}$ containing the origin the following two expressions hold (the first one is true even if ϕ has terms of degree 2.*

$$f(0) = \lim_{t \to 0} (4\pi t)^{-\frac{n}{2}} \int_{\mathcal{B} \subset R^k} e^{-\frac{\phi(\mathbf{y})^2}{4t}} f(\mathbf{y}) \, d\mathbf{y} \tag{11}$$

$$\Delta f(0) = -\frac{\partial}{\partial t} \left((4\pi t)^{-\frac{n}{2}} \int_{\mathcal{B} \subset R^k} e^{-\frac{\phi(\mathbf{y})^2}{4t}} f(\mathbf{y}) \, d\mathbf{y} \right) \Bigg|_0 + C f(0) \tag{12}$$

C here is a constant depending only on ϕ.

Proof. We will concentrate on proving formula (12), formula (11) is a corollary of the computation below. From the previous Lemma, it can be easily seen that the set $\mathcal{B}$ can be replaced by the whole space R^k. For simplicity we will show the formula when $n = 1$. The case of arbitrary n is no different but requires rather cumbersome notation. We can write $f(y) = a_0 + a_1 y + a_2 y^2 + \ldots$ and $\phi(y) = y + b_0 y^3 + \ldots$. Put $y = \sqrt{t} x$. Changing the variable, we get:

$$\frac{1}{\sqrt{t}} \int_{\mathbb{R}} e^{-\frac{\phi(y)^2}{4t}} f(y) dy = \frac{1}{\sqrt{t}} \int_{\mathbb{R}} e^{-\frac{ty^2 + t^2 b_0 y^4 + \ldots}{4t}} f(\sqrt{t} y) \sqrt{t} dy =$$

$$= \int_{\mathbb{R}} e^{-\frac{y^2 + t b_0 y^4 + o(t)}{4}} f(\sqrt{t} y) dy$$

Note that $e^{-\frac{y^2 + t b_0 y^4 + o(t)}{4}} = e^{-\frac{y^2}{4}} e^{-\frac{t b_0 y^4 + o(t)}{4}} = e^{-\frac{y^2}{4}} (1 - t\frac{b_0}{4} y^4 + o(t))$.
Thus the previous integral can be written as

$$\int_{\mathbb{R}} e^{-\frac{x^2}{4}} \left(1 - t\frac{b_0}{4} x^4 + o(t)\right) f(\sqrt{t} x) dx$$

$$= \int_{\mathbb{R}} e^{-\frac{x^2}{4}} \left(1 - t\frac{b_0}{4}x^4 + o(t)\right) \left(a_0 + a_1\sqrt{t}x + a_2 t x^2 + o(t)\right) dx$$

$$= \int_{\mathbb{R}} e^{-\frac{x^2}{4}} \left(a_0 + a_1\sqrt{t}x + t(a_2 x^2 - a_0\frac{b_0}{4}x^4) + o(t)\right) dx$$

Note that the second term $a_1\sqrt{t}x$ is an odd function in x and therefore $\int_{\mathbb{R}} e^{-\frac{x^2}{4}} a_1\sqrt{t}x dx = 0$.

Thus

$$\frac{\partial}{\partial t}\left(\frac{1}{2\sqrt{t\pi}}\int_{\mathbb{R}} e^{-\frac{y^2+y^4\phi(y)}{4t}} f(y)dy\right)\Bigg|_0 = \frac{1}{2\sqrt{\pi}}\int_{\mathbb{R}} e^{-\frac{x^2}{4}} a_2 x^2 dx - \frac{1}{2\sqrt{\pi}}\int_{\mathbb{R}} e^{-\frac{x^2}{4}} a_0\frac{b_0}{4}x^4 dx$$

The first integral in the sum is exactly the Laplacian of f at 0, $\Delta f(0) = \frac{1}{2\sqrt{\pi}}\int_{\mathbb{R}} e^{-\frac{x^2}{4}} a_2 x^2 dx$. The second summand depends only on the value $a_0 = f(0)$ and the function ϕ, which completes the proof.

6.3 Probability

Theorem 6 (Hoeffding). *Let $X_1, \ldots, X_n$ be independent identically distributed random variables, such that $|X_i| \leq K$. Then*

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

Acknowledgements

We thank Matthias Hein for pointing out an error in Claim 4.2.6. in [1] and an earlier version of Lemma 6.

References

1. M. Belkin, *Problems of Learning on Manifolds*, The University of Chicago, Ph.D. Dissertation, 2003.
2. M. Belkin, P. Niyogi, *Using Manifold Structure for Partially Labeled Classification*, NIPS 2002.
3. M. Belkin, P. Niyogi. (2003). *Laplacian Eigenmaps for Dimensionality Reduction and Data Representation*, Neural Computation, Vol. 15, No. 6, 1373-1396.
4. M. Belkin, P. Niyogi, V. Sindhwani, *On Manifold Regularization*, AI Stats 2005.
5. 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*, NIPS 2003.
6. M. Bernstein, V. de Silva, J.C. Langford, J.B. Tenenbaum, *Graph approximations to geodesics on embedded manifolds*, Technical Report, 2000.
7. O. Bousquet, O. Chapelle, M. Hein, *Measure Based Regularization*, NIPS 2003.
8. Y. Bengio, J-F. Paiement, and P. Vincent,*Out-of-Sample Extensions for LLE, Isomap, MDS, Eigenmaps, and Spectral Clustering*, NIPS 2003.

9. Chapelle, O., J. Weston and B. Schoelkopf, *Cluster Kernels for Semi-Supervised Learning*, NIPS 2002.
10. F. R. K. Chung. (1997). *Spectral Graph Theory*. Regional Conference Series in Mathematics, number 92.
11. M. do Carmo, *Riemannian Geometry*, Birkhauser, 1992.
12. D. L. Donoho, C. E. Grimes, *Hessian Eigenmaps: new locally linear embedding techniques for high-dimensional data*, Proceedings of the National Academy of Arts and Sciences vol. 100 pp. 5591-5596.
13. M. Hein, J.-Y. Audibert, U. von Luxburg, *From Graphs to Manifolds – Weak and Strong Pointwise Consistency of Graph Laplacians*, COLT 2005, (to appear).
14. M. Meila, J. Shi, *Learning segmentation by random walks*, NIPS 2000.
15. F. Memoli, G. Sapiro, *Comparing Point Clouds*, IMA Technical Report, 2004.
16. S. Lafon, *Diffusion Maps and Geodesic Harmonics*, Ph. D. Thesis, Yale University, 2004.
17. U. von Luxburg, M. Belkin, O. Bousquet, *Consistency of Spectral Clustering*, Max Planck Institute for Biological Cybernetics Technical Report TR 134, 2004.
18. A. Ng, M. Jordan, Y. Weiss, *On Spectral Clustering: Analysis and an Algorithm*, NIPS 2001.
19. 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.
20. P. Niyogi, S. Smale, S. Weinberger, *Finding the Homology of Submanifolds with High Confidence from Random Samples*, Univ. of Chicago Technical Report TR-2004-08, 2004.
21. S. Rosenberg, *The Laplacian on a Riemannian Manifold*, Cambridge University Press, 1997.
22. Sam T. Roweis, Lawrence K. Saul. (2000). *Nonlinear Dimensionality Reduction by Locally Linear Embedding*, Science, vol 290.
23. A. Smola and R. Kondor, *Kernels and Regularization on Graphs*, COLT/KW 2003.
24. J. Shi, J. Malik, *Normalized Cuts and Image Segmentation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22, 8, 2000.
25. D. Spielman, S. Teng, *Spectral partitioning works: planar graphs and finite element meshes*, FOCS 1996.
26. Martin Szummer, Tommi Jaakkola, *Partially labeled classification with Markov random walks*, NIPS 2001.
27. J.B.Tenenbaum, V. de Silva, J. C. Langford. (2000). *A Global Geometric Framework for Nonlinear Dimensionality Reduction*, Science, Vol 290.
28. R. Kannan, S. Vempala, A. Vetta, *On Clusterings - Good, Bad and Spectral*, Technical Report, Computer Science Department, Yale University, 2000.
29. D. Zhou, O. Bousquet, T.N. Lal, J. Weston and B. Schoelkopf, *Learning with Local and Global Consistency*, NIPS 2003.
30. X. Zhu, J. Lafferty and Z. Ghahramani, *Semi-supervised learning using Gaussian fields and harmonic functions*, ICML 2003.
31. A. Zomorodian, G. Carlsson, *Computing persistent homology*, 20th ACM Symposium on Computational Geometry, 2004.