Graph Laplacian: Consistency and Connection with Kernel Learning

Yidan (Eden) Xu, Zeyu (Jerry) Wei
Department of Statistics, University of Washington
yx2516@uw.edu, zwei5@uw.edu

Abstract

It is often of great interest to find meaningful geometric/topological description of data set in virtue of uncovering information otherwise hidden, particularly when a set of data in a high-dimensional Euclidean space is assumed to lie on a lower dimensional manifold. The study of graph Laplacian is therefore of great interest for its connection with graph theory and spectral analysis, which provides both topological and geometric interpretations. In this report of summary, we first focus on the convergence of graph Laplacians, then some interesting applications that incorporate geometric information with the function space are presented.

1 Theory

1.1 Consistency for Graph Constructions

Convergence of the graph Laplacian to the Laplace-Beltrami Operator * (LBO), which analyzes the functions defined on the manifold and hence characterizes the local geometry of the manifold, lies in the heart of topological data analysis. To prove consistency of any graph construction, there is a general framework (Berry and Sauer 2019):

(1) Show that the (discrete) graph Laplacian is a consistent statistical estimator of a (continuous) integral operator (either pointwise or spectrally). Commonly this integral operator is defined as the convolution with some kernel function.

(2) Show that the integral operator in (1) converges to the Laplace-Beltrami operator, which can be achieved by studying the leading order term for the asymptotic expansion of the integral operator or be obtained by studying the difference of the integral operator to some known operators (particularly heat kernel operator) that converge to the LBO.

(3) Show that this estimator has finite variance, which is important in regulating the rate of convergence.

The analysis of the convergence of kernel weighted graph Laplacians to their continuous counterparts was initiated in Belkin and Niyogi 2006, and the theory is latter completed in Belkin and Niyogi 2008, which proved (1) and (2) under uniform sampling on compact manifolds. Coifman and Lafon 2006 latter extended (1) and (2) to non-uniform sampling, but still requires compact manifold. The analysis of variance is completed by later work for the constructions in Belkin and Niyogi 2008 and Coifman and Lafon 2006. The extension of (1) and (2) is also provided by later work Hein 2005. However, analysis of (3) in Berry and Harlim 2014 has shown that pointwise errors can be unbounded on non-compact manifold, and hence additional restrictions need to be imposed to construct desirable graph Laplacians, especially variable bandwidth kernels were required. The construction in Berry

*The Laplace-Beltrami operator (LBO) $\Delta_{\mathcal{M}}$ on a manifold $\mathcal{M}$ is a differential operator $\mathcal{C}^2(\mathcal{M}) \to L^2(\mathcal{M}), f \to -div(\nabla f)$, where ∇f is the gradient and div denotes divergence. $f : \mathcal{M} \to \mathbb{R}$.

Conference on Neural Information Processing Systems (NeurIPS), .

and Sauer 2019 follows the development in theory by adopting a unweighted graph construction with variable bandwidth and hence generalize to non-compact manifolds. Moreover, Berry and Sauer 2019 paper provides the first, and unique, consistency result for unweighted graph construction.

In this report, key constructions and results from Belkin, Niyogi, and Sindhwani 2006 and Coifman and Lafon 2006 are presented as the classical approach to represent manifold with kernel weighted graph construction. Then, the Berry and Sauer 2019 result is discussed with greater detail as the recent development on unweighted graph construction that enjoys the convergence to diffusion operators like the weighted constructions but at the same time has better accessibility for geometric information.

1.2 Convergence of Laplacian Eigenmaps (Belkin, Niyogi, and Sindhwani 2006)

Belkin et al. starts by first defining **Graph Laplacian** L on a graph:
For data points $x_1, \dots, x_n \in \mathcal{M} \subseteq \mathbb{R}^N$, $\mathcal{M}$ is a compact manifold that is infinitely differentiable Riemannian submanifold of $\mathbb{R}^N$ without boundary. Consider the weighted graph with vertex set V and Gaussian weight matrix W with $W_{ij} = \frac{1}{t(4\pi t)^{k/2}} e^{-\frac{||x_j - x_i||^2}{4t}}$ where k is the intrinsic dimension of the manifold $\mathcal{M}$. Let D be the degree matrix. And hence we define the graph Laplacian as :

$$L \equiv D - W$$

The intrinsic dimension may be unknown, but can be estimated through a method proposed in Belkin and Niyogi 2006 using the fact that the Laplacian of the constant function is 0.

The graph Laplacian L can be extended from functions on the graph (V, W) to an empirical operators $\hat{L}_{t,n} : \mathcal{C}(\mathcal{M}) \to \mathcal{C}(\mathcal{M})$ that is maps continuous functions defined on $\mathcal{M}$ to the same function space. This $\hat{L}_{t,n}$ is referred to as the **point cloud Laplacian**
For data points $x_1, \dots, x_n \in \mathcal{M} \subseteq \mathbb{R}^N$:

$$\hat{L}_{t,n}(f)(p) \equiv \frac{1}{t(4\pi t)^{k/2}} \left(\frac{1}{n} \sum_{i=1}^{n} e^{-\frac{||p-x_i||^2}{4t}} \left[f(p) - f(x_i) \right] \right)$$

Let f_v be the restriction of any function $f \in \mathcal{C}(\mathcal{M})$ restricted to V, then $\hat{L}_{t,n} f$ is the same as $L f_v$. Therefore, for small bandwidth t, eigenvalues of L and $\hat{L}_{t,n}$ are the same and the corresponding eigenfunctions (eigenvectors when restricted to V) are naturally related.
So far, the defined L and $\hat{L}_{t,n}$ are all discrete operators. The (continuous) **integral operator**, which serves as a functional approximation to the LBO $\Delta_{\mathcal{M}}$, in step (1) of the proof framework is defined as $L_{t,n} : L^2(\mathcal{M}) \to L^2(\mathcal{M})$

$$L_{t,n}(f)(p) \equiv \frac{1}{t(4\pi t)^{k/2}} \left(\int_{\mathcal{M}} e^{-\frac{||p-y||^2}{4t}} \left[f(p) - f(y) \right] d\mu_y \right)$$

where μ is the uniform measure on $\mathcal{M}$ obtained from the volume form. The uniform distribution is an essential assumption for this result by Belkin and Niyogi 2008 and essentially assumes that there is no information with the density and hence can ignore the coupling of density and geometric structure. In showing that the integral operator $L_{t,n}$ converges to $\Delta_{\mathcal{M}}$, the authors employs the **heat operator** $\mathcal{H}_t : L^2 \to L^2$, which is the convolution with the heat kernel. Previous work has shown $\frac{1-\mathcal{H}}{t}$ and $\Delta_{\mathcal{M}}$ share an eigenbasis, and the proof for $EigL_t \to Eig\Delta_{\mathcal{M}}$ goes by showing the difference $R_t = \frac{1-\mathcal{H}}{t} - L_t$ is relatively bounded.

The main theorem in Belkin, Niyogi, and Sindhwani 2006 paper (Theorem 2.1) can be summarized by a diagram as: $EigL \xrightarrow{extendV \to \mathcal{M}} Eig\hat{L}_{t,n} \xrightarrow{n \to \infty} EigL_t \xrightarrow{t \to 0} Eig\Delta_{\mathcal{M}}$

$$Eig\hat{L}_{t,n} \xrightarrow{n \to \infty \quad t_n \to 0} Eig\Delta_{\mathcal{M}}$$

1.3 Diffusion Maps (Coifman and Lafon 2006)

To relax the uniform distribution assumption as in Belkin and Niyogi 2008, Coifman and Lafon found a way to decouple the interaction between the density/distribution and the geometry of the dataset by constructing a one-parameter family of anisotropic diffusion kernels that capture different features of the data with the parameter varying. Unlike in Belkin and Niyogi 2008, Coifman and Lafon utilized the normalized graph Laplacian, and, more specifically, the graph Laplacian normalization is not applied on a graph with isotropic weights, but rather on a renormalized graph. The construction is as follows:

1. Fix $\alpha \in \mathbb{R}$ and a rotation-invariant (isotropic) kernel $k_\epsilon(x, y) = h(\frac{||x-y||^2}{\epsilon})$
2. Let $q_\epsilon(x) = \int_X k_\epsilon(x, y)q(y)dy$ where $q(y)$ is the density of the points on $\mathcal{M}$. Then form the new kernel as $k_\epsilon^{(\alpha)(x,y)} = \frac{k_\epsilon(x,y)}{q_\epsilon(x)^\alpha q_\epsilon(y)^\alpha}$
3. Appy the weighted graph Laplacian normalization to this kernel by setting $d_\epsilon^{(\alpha)}(x) = \int_X k_\epsilon^{(\alpha)}(x, y)q(y)dy$ and defining the anisotropic transition kernel as $p_{\epsilon,\alpha}(x, y) = \frac{k_\epsilon^{(\alpha)}(x,y)}{d_\epsilon^{(\alpha)}(x)}$

Then, following the framework of proving consistency of graph constructions, the integral operator is defined as$P_{\epsilon,\alpha}f(x) = \int_X p_{\epsilon,\alpha}(x, y)f(y)q(y)dy$, which is later shown to converge spectrally to the symmetric Schrödinger operator, $\Delta\phi - \frac{\Delta(q^{1-\alpha})}{q^{1-\alpha}}\phi,\ \ \phi = fq^{1-\alpha}$.

When $\alpha = 0$, the influence of density is maximal, and the diffusion reduces to computing the normalized graph Laplacian with isotropic (e.g. Gaussian) weights. Particularly, when the density q is uniform on manifold $\mathcal{M}$, the result is consistent with Belkin and Niyogi 2006 that the graph Laplacian approximates the LBO on $\mathcal{M}$. When $\alpha = 1$, the influence of the distribution of the data is eliminated with the renormalized graph Laplacian, i.e. geometry and density are completely decoupled. If the points approximately lie on a submanifold of $\mathbb{R}^n$, then it is an approximation to the LBO, and the heat kernel $e^{-t\Delta}$ can be approximated by $P_{\epsilon,1}^{t/\epsilon}$. With the anisotropic kernel graph construction (renormalized graph), Coifman and Lafon provide a way to fully decouple the distribution and the geometry of points on the data manifold, and therefore the convergence of graph Laplacians can be generalized to cases where uniform distribution doesn't apply. However, the results from Coifman and Lafon 2006 still requires the manifold to be compact, and with the usage of weighted kernels, the geometric features are not easily to be extracted. In finite sample case, the integrals can be approximated by the empirical estimators:
$\bar{q}_\epsilon(x_i) = \sum_{j=1}^m k_\epsilon(x_i, x_j)$ and $\bar{d}_\epsilon^{(\alpha)(x_i)} = \sum_{j=1}^m \frac{k_\epsilon(x,y)}{\bar{q}_\epsilon(x)^\alpha \bar{q}_\epsilon(y)^\alpha}$,
$\bar{p}_{\epsilon,\alpha}(x_i, x_j) = \frac{k_\epsilon(x_i,x_j)}{\bar{d}_\epsilon^{(\alpha)}(x_i)\bar{d}_\epsilon^{(\alpha)}(x_j)}$, and $\bar{P}_{\epsilon,\alpha}f(x_i) = \sum_{j=1}^m \bar{p}_{\epsilon,\alpha}(x_i, x_j)f(x_j)$.

1.4 Recent Development on Manifold Representation (Berry and Sauer 2019)

So far, we've seen in diffusion maps paper by Coifman and Lafon 2006 and eigenmaps paper by Belkin and Niyogi 2008 the approach that use localised kernels to form graphs with *weighted* edges and then use the weighted graphs to produce operators that converge to LBOs with respect to various geometries on the manifold. This approach in general is an outgrowth of Kernel PCA. Different versions of diffusion maps can reconstruct the geometry of the manifold with respect to a desired metric, and hence implicitly include all the topological information about the manifold. However, it is not yet clear how to explicitly establish the the topological or geometric connection between the weighted graph and the underlying manifold.

In contrast to the weighted graph constructions, the unweighted graphs used by persistent homology approach provide better accessibility to geometric features. In persistent homology approach, series of unweighted graphs are generated to reconstruct topology one scale at a time and to compute topological features at different spatial resolutions, where more persistent features are deemed more likely to be true features of the underlying space. The great advantage of an unweighted graph is that a simplicial complex can be built from the graph immediate by the VietoTris-Rips (VR)[†] construction. Despite the explicit connection to geometry, a single unweighted graph cannot be guaranteed to contain topological information at all level, which is doomed by its scale-dependent construction. Consequently, there is not a unified homology in the limit with large data, and consistency results for the persistent homology approach is not possible.

[†]Vietoris-Rips complex is an abstract simplicial complex that can be defined from any metric space M and distance δ by forming a simplex for every finite set of points that has diameter at most δ, hence the connection to unweighted graph, which is constructed with metric, is immediate.

The recent development in Berry and Sauer 2019 proposes a unweighted graph construction that combines the advantages of both the weighted graph approaches and the persistent homology approach. That is, the unnormalized graph Laplacian under the proposed construction has the spectral and pointwise convergence to a LBO on the manifold in large data limit with all the topological information incorporated, and like the unweighted graphs, the manifold's topological information can be extracted.

1.4.1 CkNN Graph Construction

Continuous k-Nearest Neighbors (CkNN) construction connects the points x, y if $d(x,y) < \delta\sqrt{d(x,x_k)d(y,y_k)}$ where d is the given metric on the data points, x_k, y_k are the k-th nearest neighbors of x, y respectively, and δ is a continuous tuning parameter. CkNN can be understood as kNN with varying bandwidth. The introduction of δ allows k to be fixed so that $||x, x_k||$ is an estimator of $q(x)^{-1/m}$ where $q(x)$ is the sampling density and m is the intrinsic dimension of the data. CkNN can be generalized to a broader class of **multi-scale graph constructions**: connect the points if $d(x,y) < \delta\sqrt{\rho(x)\rho(y)}$, where ρ defines the local scaling near the point. It is shown in this paper that the unique multi-scale graph construction that yields a consistent limiting geometry requires $\rho(x) \propto q(x)^{-1/m}$ and hence in this sense CkNN is the unique unweighted graph construction that has consistency.

1.4.2 Explicitly Talking about Manifold Topology from Graph Topology

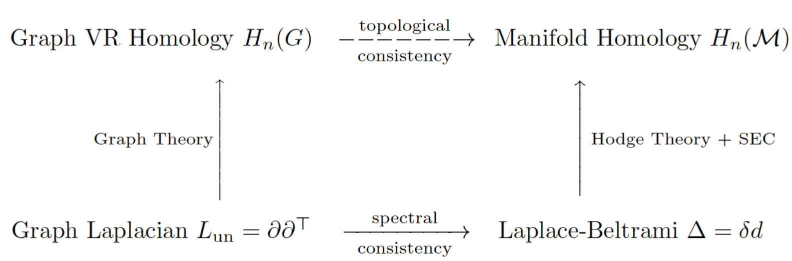

Figure 1: Relation Diagram from (Berry and Sauer 2019).

Berry and Sauer 2019 provides a diagram that explicitly connect graph, graph Laplacian, Laplace-Beltrami operator, and the underlying manifold in Figure 1. So far, there are rich results about the bottom arrow that show consistency for various graph constructions such as in Belkin and Niyogi 2008, Coifman and Lafon 2006, and the CkNN construction in this paper. However, a question arises: Does consistency of the Laplacian operator provide topological consistency in the sense that the homology computed from the VR complex is isomorphic to the homology of the underlying manifold in the large data limit?

The question can be studied first via the connections between the discrete and continuous operators and their corresponding topology (graph and manifold). For the left vertical arrow, the graph determines the entire VR complex, and the graph can be fully reconstructed from the unnormalized graph Laplacian, and so L_{un} completely determines the VR homology of the graph. For the right vertical arrow, it is possible to reconstruct the metric on a Riemannian manifold from the LBO, and the metric completely determines the homology of the manifold.

With the respective connections between operators and topology, the connection is explicitly spectrally for the zero homology, as the zero-homology of a graph corresponds exactly to the zero-eigenspace of the graph Laplacian and analogously the zero-homology of the manifold corresponds to the zero-eigenspace of the Laplace-Beltrami operator. For the isomorphism between higher order homology groups, the authors conjecture the generalization is possible via the more general framework concerning the Laplace-de Rham operators.

With the summary results above, the Berry and Sauer 2019 work makes a great advancement in the theory about manifold representation. The understanding of the data manifold also inspired ways to incorporate geometric information in the setting of statistical learning, and some results along this line are shown in the next section.

2 Applications: Incorporating geometric information into function spaces

As in (Ham et al. 2004) is that, many non-linear dimension reduction algorithms such as Isomap, Laplacian Eigenmap and LLE can be interpreted as kernel PCA, where a local neighbourhood structure on the data is captured by a graph, characterised by the weighted adjacency matrix/ Gram matrix, to globally map the data manifold to a lower dimensional space. The construction of a Gram matrix is equivalent to mapping the data to points $p_i, \ldots, p_n$ in a Hilbert space by a kernel function. Such mapping may be viewed as a warping of input space, defined by the local structure of the data, into a feature space where manifold is flat.

The following series of paper (Belkin, Niyogi, and Sindhwani 2006, Sindhwani, Niyogi, and Belkin 2005, Sindhwani, Chu, and Keerthi 2006) instead follows an approach that is more common in the kernel learning literature, where a kernel is formulated so as to directly construct RKHS warped by the local intrinsic geometry of the data such that the kernel is deformed along a finite subspace spanned by $\{k(x_i, \cdot)\}_{i=1}^{l+u}$[†]. By doing this, it has a natural out-of-sample extension owing to the formulation in the function space, where the graph based algorithms do not exhibit. And the flexible framework of the semi-supervised learning allows for the relaxation to unsupervised, and constraint to fully-supervised setting easily by amending the regularisation problem in (1).

2.1 Manifold Regularisation in RKHS (Belkin, Niyogi, and Sindhwani 2006)

Classical function learning framework (i.e.classification and regression) such as SVM, Ridge regression and splines can be broadly characterised as regularisation problems with different loss functions and complexity constraints, and such constraints can be framed as smoothness conditions in appropriately chosen Reproducing Kernel Hilbert Space (RKHS) with the reproducing kernel $k(\cdot, \cdot)$.

Consider a generative model, it is assumed that there is a probability distribution $P : \mathcal{X} \times \mathbb{R} \to \mathbb{R}$ according to which labelled examples (x, y) are generated from; and unlabelled examples $x \in \mathcal{X}$ are drawn from the marginal distribution $\mathcal{P}_{\mathcal{X}}$ of P. To make the marginal distribution informative regarding the conditional distribution, it is assumed that $\mathcal{P}(y|x)$ varies smoothly along the geodeisic in the intrinsic geometry of $\mathcal{P}_{\mathcal{X}}$. Therefore, the traditional regularisation framework can be extended to (1) by incorporating smoothness penalty of the function in the intrinsic geometry of $\mathcal{P}_{\mathcal{X}}$:

$$f^* = \text{argmin}_{f \in \mathcal{H}_k} \frac{1}{n} \sum_{i=1}^{n} V(x_i, y_i, f) + \gamma_A \|f\|_{\mathcal{H}}^2 + \gamma_I \|f\|_I^2 \tag{1}$$

$\|\cdot\|_{\mathcal{H}}$ is the norm on the ambient space (RKHS) and $\|\cdot\|_I^2$ is the smoothness penalty described above. When the manifold hypothesis holds , the data distribution is supported on some low-dimensional manifold $\mathcal{M}$, the constructed $\|f\|_I^2$ should penalize f along the manifold.

When $\mathcal{M} \subset \mathbb{R}^n$ is a compact, one natural choice for $\|\cdot\|_I^2$ is that

$$\|f\|_I^2 = \int_{x \in \mathcal{M}} \|\nabla_{\mathcal{M}} f\|^2 \mathrm{d}\mathcal{P}_{\mathcal{X}},$$

this provides an estimate that on average how far apart f maps nearby points on the manifold. This choice of $\|f\|_I^2$ also incorporates geometric information through the connection to Laplace-Beltrami Operator (LBO). By Stoke's Theorem the above can be written as

$$\|f\|_I^2 = \int_{x \in \mathcal{M}} \|\nabla_{\mathcal{M}} f\|^2 \ \mathrm{d}\mathcal{P}_{\mathcal{X}} = \int_{x \in \mathcal{M}} f \Delta f \ \mathrm{d}\mathcal{P}_{\mathcal{X}}$$

When restricted to the graph, this can be approximated using data $\{x_i\}_{i=1}^{u+l}$ and graph Laplacian $L = D - W$ with Gaussian weight, and then (1) becomes

$$f^* = \text{argmin}_{f \in \mathcal{H}_k} \frac{1}{n} \sum_{i=1}^{n} V(x_i, y_i, f) + \gamma_A \|f\|_{\mathcal{H}}^2 + \frac{\gamma_I}{(u+l)^2} \underbrace{\sum_{i=1}^{l+u} (f(x_i) - f(x_j))^2 W_{ij}}_{\mathbf{f}^T L \mathbf{f}} \tag{2}$$

[†] l is the number of labelled data and u is the number of unlabelled data, coherent to the notation in the later sections.

Solving the above regularisation problem is equivalent to finding a function $f^* \in \tilde{\mathcal{H}}$, where $\tilde{\mathcal{H}}$ is a new RKHS constructed by deforming the original RKHS to reflect the underlying geometry of the data. This is achieved by reconstructing a Mercer kernel $\tilde{k}(\cdot,\cdot)$ such that it is adpated to the geometry of $\mathcal{P}_{\mathcal{X}}$. The following section illustrates the formulation of such kernel, and therefore by the proof in (Belkin, Niyogi, and Sindhwani 2006), the solution f^* exists and can be written with respect to the data as

$$f^*(x) = \sum_{i=1}^{l} \alpha_i k(x_i, x) + \sum_{j=1}^{l+u} \beta_j(x) k(x_j, x) \tag{3}$$

for some data dependent function $\beta_j(x) \in \mathcal{H}$. Moreover, the Representer Theorem is shown to hold with $\tilde{k}(\cdot,\cdot)$. Therefore the solution can also be found via $f^*(x) = \sum_{i=1}^{l+u} \alpha_i \tilde{k}(x_i, x)$

2.2 Kernel on Data Manifold (Sindhwani, Niyogi, and Belkin 2005)

Given the regularisation problem in (2) the corresponding form of the inner product defined on the new RKHS is formulated as follows

$$\langle f, g \rangle_{\tilde{\mathcal{H}}} = \langle f, g \rangle_{\mathcal{H}} + \langle Sf, Sg \rangle_{\mathcal{V}}$$

where $\mathcal{V} = \mathbb{R}^n$ is chosen along with a positive semi-definite inner product, termed as 'point cloud norm' by the author, such that the induced norm for a evaluation map $S : \mathcal{H} \to \mathbb{R}^n$, $\mathbf{f} = S(f) = (f(x_1), \cdots, f(x_n))$ is defined as:

$$\|Sf\|_{\mathcal{V}^2} = \mathbf{f}^t M \mathbf{f}.$$

where M is a symmetric positive semi-definite matrix. M can be taken as the graph Laplacian $\mathcal{L}$ to be consistent with the formulation in (2); alternatively, for appropriate $\|\cdot\|_I$, $M = \mathcal{L}^p,\ p \in \mathbb{Z}^+$ can also be chosen.

To show that $\tilde{\mathcal{H}}$ is a well-defined RKHS, the authors first obtain that $span\{\tilde{k}(x_i, \cdot)\}_{i=1}^{l+u} = span\{k(x_i, \cdot)\}_{i=1}^{l+u}$. Thus the new kernel function $\tilde{k}$ corresponding to $\tilde{\mathcal{H}}$ can be presented as

$$\tilde{k}(x, \cdot) = k(x, \cdot) + \sum_{j=1}^{l+u} \beta_j(x) k(x_j, \cdot) \tag{4}$$

To find the coefficients $\beta_j(x)$, it solves a system of linear equations generated by evaluating $k(x_i, \cdot)$ at x in the new RKHS, i.e. $k(x_i, x) = \langle k(x_i, \cdot), \tilde{k}(x, \cdot) \rangle_{\tilde{\mathcal{H}}}$ and an explicit form of the new kernel can then be written into a closed form evaluated only at the data,

$$\tilde{k}(x, y) = k(x, y) - \mathbf{k}_x^t (I + MK)^{-1} M \mathbf{k}_y \tag{5}$$

where $\mathbf{k}_x = (k(x_1, x) \ldots k(x_n, x))^t$, $K_{ij} = k(x_i, x_j)$.

Notably, the author pointed out several drawbacks of the such regularisation scheme, where the regularisation parameter γ_A, γ_I for the two penalty terms can be difficult to tune and the 'big-N' problem exists due to the inversion of dense Gram matrix in the inference.

3 Conclusion

In conclusion, we've explored some classical and recent developments on the graph Laplacian. We've seen the graph Laplacians under different graph constructions (with weighted or unweighted edges) converges as operators to the Laplace-Beltrami operator, which encodes the geometric information of the local structure of the underlying manifold. Such geometric information has been widely used for dimension reduction and clustering, and in this report we also present the applications that incorporate the geometric information into function space. Geometry describes the relations of points, lines, surfaces, solids, and higher dimensional analogs, while function is the general notion of association between objects, so it is not surprising the two can be unified in some ways. Essentially, all the works presented in this report concern with constructing functions or functionals that can characterize some features of the data while enjoying nice analytical properties. Such connection is inspiring.

References

Ham, Ji Hun et al. (2004). "A kernel view of the dimensionality reduction of manifolds Recommended Citation". In: *Article Number*, p. 47. DOI: 10.1145/1015330.1015417.

Hein, Matthias (2005). *Geometrical Aspects of Statistical Learning Theory*. Tech. rep.

Sindhwani, Vikas, Partha Niyogi, and Mikhail Belkin (2005). "Beyond the point cloud". In: Association for Computing Machinery (ACM), pp. 824–831. DOI: 10.1145/1102351.1102455.

Belkin, Mikhail and Partha Niyogi (2006). *Laplacian Eigenmaps for Dimensionality Reduction and Data Representation*. Tech. rep.

Belkin, Mikhail, Partha Niyogi, and Vikas Sindhwani (2006). *Manifold Regularization: A Geometric Framework for Learning from Labeled and Unlabeled Examples*. Tech. rep., pp. 2399–2434. URL: http://www.cse.msu.edu/.

Coifman, Ronald R. and Stéphane Lafon (July 2006). "Diffusion maps". In: *Applied and Computational Harmonic Analysis* 21.1, pp. 5–30. ISSN: 10635203. DOI: 10.1016/j.acha.2006.04.006.

Sindhwani, Vikas, Wei Chu, and S Sathiya Keerthi (2006). *Semi-supervised Gaussian Process Classifiers*. Tech. rep.

Belkin, Mikhail and Partha Niyogi (2008). *Convergence of Laplacian Eigenmaps*. Tech. rep.

Berry, Tyrus and John Harlim (2014). *Variable Bandwidth Diffusion Kernels*. Tech. rep.

Berry, Tyrus and Timothy Sauer (2019). "Consistent manifold representation for topological data analysis". In: *Foundations of Data Science* 0.0, pp. 0–0. ISSN: 2639-8001. DOI: 10.3934/fods.2019001. arXiv: 1606.02353.