High-Order Regularization on Graphs

Dengyong Zhou DENGYONG.ZHOU@MICROSOFT.COM
Microsoft Research, One Microsoft Way, Redmond, WA 98052 USA

Christopher J.C. Burges CBURGES@MICROSOFT.COM
Microsoft Research, One Microsoft Way, Redmond, WA 98052 USA

Keywords: discrete Laplace-de Rham operator, discrete Laplace-Beltrami operator, discrete regularization

Abstract

The Laplace-Beltrami operator for graphs has been been widely used in many machine learning issues, such as spectral clustering and transductive inference. Functions on the nodes of a graph with vanishing Laplacian are called harmonic functions. In differential geometry, the Laplace-de Rham operator generalizes the Laplace-Beltrami operator. It is a differential operator on the exterior algebra of a differentiable manifold, and it is equivalent to the Laplace-Beltrami operator when acting on a scalar function. In this paper, we develop a discrete analogue of the Laplace-de Rham operator, which naturally generalizes the discrete Laplace-Beltrami operator. The discrete Laplace-de Rham operator can then be used to define harmonic functions on arbitrary paths in a graph, in particular, functions on edges. Consequently, we build discrete regularization using the discrete Laplace-de Rham operator, and validate it on real-world web categorization tasks.

1. Introduction

In many machine learning issues, we are only concerned with a finite set of objects rather than a continuous space, such as clustering and transductive inference. Without any assumption on the relationships among the given objects, we can cluster or classify them in an arbitrary way. Typically, it is assumed that there is a graph defined on the object set (Figure 1). The graph can be undirected or directed. For instance, in image segmentation, an image can be regarded as an undirected graph, in which each vertex represents a pixel, and each edge represents the similarity between two pixels (Shi & Malik, 2000); and, in web categorization, a set of web pages can be regarded as a directed graph (Zhou et al., 2005), in which each each vertex represents a web page, and each edge represents a hyperlink between two web pages.

There have been many graph based machine learning approaches, in which graph Laplacians play important roles, although the definitions of graph Laplacians vary subtly across literature. In 1970s, Fielder began to investigate algebraic connectivity of graphs via the second smallest eigenvector of a discrete analogue of the Laplacian (Fielder, 1973), which is now widely called the unnormalized graph Laplacian. Fielder's work provided the theoretical justification for the use of graph Laplacians in partitioning, in particular, the ratio cut (Hagen & Kahng, 1992), which divides the vertices of a graph into two subsets such that the number of vertices in each subset is as equal as possible, and the number of the edges which are cut is as small as possible. If we ask the volume of each subset instead of the number of vertices to be as equal as possible, then we obtain the normalized cut (Shi & Malik, 2000), and the normalized graph Laplacian takes the role of its unnormalized counterpart in partitioning. The normalized cut has been extended to directed graphs (Zhou et al., 2005) and hyergraphs (Zhou et al., 2007). For classification issues, graph Laplacians have been used in kernel design (Chapelle et al., 2003; Smola & Kondor, 2003; Ando & Zhang, 2007) and in transductive inference (Belkin & Niyogi, 2004; Joachims, 2003; Zhu et al., 2003; Zhou et al., 2004). The convergence properties of graph Laplacians are addressed in (Hein et al., 2005) and the references therein.

Graph Laplacians so far are restricted to a discrete

Appearing in the 6^{th} *International Workshop on Mining and Learning with Graphs*, Helsinki, Finland, 2008.

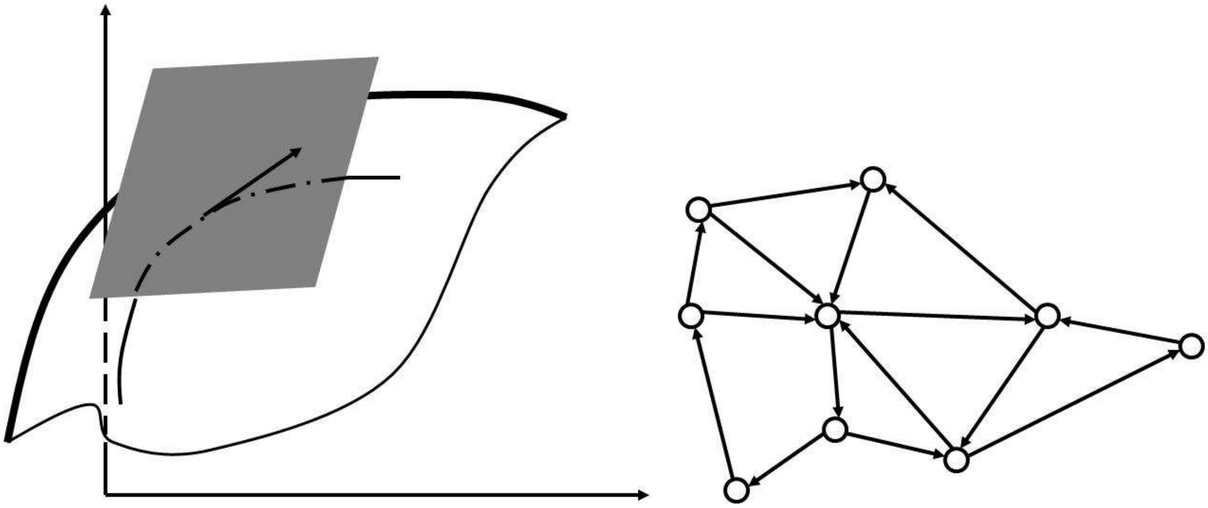


Figure 1. Manifold and graph. A graph can be considered as a discrete approximation to a manifold; on the other hand, a manifold can be considered as a continuous approximation to a graph. In this sense, it is hard to say which one is more fundamental.

analogue of the Laplace-Beltrami operator on Riemannian manifolds however. In differential geometry, the Laplace-de Rham operator generalizes the Laplace-Beltrami operator. It is a differential operator on the exterior algebra of a differentiable manifold, and it is equivalent to the Laplace-Beltrami operator when acting on a scalar function. Hence, we consider developing a discrete analogue of the Laplace-de Rham operator such that it can naturally generalize the discrete Laplace-Beltrami operator. Then the discrete Laplace-de Rham operator can be used to define harmonic functions on arbitrary paths in a graph, in particular, functions on edges.

The present work is a further development of the discrete analysis over graphs in (Zhou et al., 2005), where the discrete analogues of differential operators including the divergence, the Laplacian and the p-Laplacian have been constructed in a coordinate-free fashion. However, the discrete analogue of the gradient in (Zhou et al., 2005) is coordinate-dependent. This observation motivates us to construct an intrinsic definition of the discrete gradient instead, and a discrete analogue of the Laplace-de Rham operator is consequently obtained. The improvement over the work in (Zhou et al., 2005) is based on noncommutative geometry, discrete field theories and quantum mechanics (Connes, 1994; Dimakis & Müller-Hoissen, 1992; Noyes, 1996). It is worth mentioning that there has been much effort devoted to developing discrete exterior calculus over triangle meshes or simplicial complexes (Forman, 1999; Mercat, 2001; Leok, 2004) instead of over graphs, and it has been applied to computer graphics (Fisher et al., 2007) and computational mechanics (Leok, 2004), where the manifold of interest is discretized into simplicial complexes. For developing discrete differential operators over graphs, however, we have to develop new discretizing techniques.

The paper is organized as follows. In Section 2, we present some basic notions in differential geometry such that the reader who is not familiar with differential geometry can find the continuous roots of the definitions of the discrete differential operators in the later sections. In Section 3, we introduce discrete analogues of the differential, the codifferential, and the Laplace-de Rham operator in a coordinate-free fashion almost by simply copying the definitions of their continuous counterparts. In Section 4, the above discrete differential operators are computed with respect to a specific chosen basis. In other words, they were expressed in a coordinate-dependent fashion. In Section 5, we propose an optimization framework for regularizing the functions on arbitrary paths of a graph. This method is validated on a task of web spam detection in Section 6. A web site is judged to be spam or not via checking the quality of its links. Finally, we conclude the paper in Section 7.

2. Laplace-de Rham Operator on Manifolds

We review some related notions in differential geometry. For a comprehensive introduction to differential geometry, we refer the reader to (Jost, 2002).

Let $\mathcal{M}$ be a n-dimensional smooth manifold (Figure 1). For any $p \in \mathcal{M}$, two curves $\gamma_1 :]-\epsilon_1, \epsilon_1[\to \mathcal{M}$ and $\gamma_2 :]-\epsilon_2, \epsilon_2[\to \mathcal{M}$ through p (i.e. $\gamma_1(0) = \gamma_2(0) = p$) are equivalent iff there is some chart (U, φ) at p so that $(\varphi \circ \gamma_1)'(0) = (\varphi \circ \gamma_2)'(0)$. A tangent vector at p is any equivalent class of smooth curves though p modulo the equivalence. The set of all tangent vectors at p is denoted by $T_p\mathcal{M}$. It is obvious that $T_p\mathcal{M}$ is a vector space. The dual of $T_p\mathcal{M}$ is denoted by $T_p^*\mathcal{M}$. It is called the cotangent space, and its elements are called cotangent vectors.

Given a vector space V, the exterior or Grassmann algebra of V over a field K is an associative algebra which contains V as a subspace. Its multiplication is known as the wedge product denoted by $\wedge$. The wedge product is associative and bilinear, and it satisfies that $v \wedge v = 0$ for all $v \in V$. The property implies that $u \wedge v = -v \wedge u$ for all $u, v \in V$. The r-th exterior power of V is denoted by $\wedge^r V$. In particular, $\wedge^0 V = \mathbb{R}$ and $\wedge^1 V = V$. Let $\{e_i\}$ be a basis of V. Then

$$\{e_{i_1} \wedge \cdots \wedge e_{i_r}, 1 \leq i_1 < \cdots < i_r \leq n\}$$

forms a basis for $\wedge^r V$. Given an element $\omega \in \wedge^r V$, its degree is denoted by $\deg \omega = r$. The direct sum

$$\wedge(V) = \bigoplus_{r \geq 0} \wedge^r V$$

forms a graded associative algebra, which is closed with respect to the wedge product. This algebra is called the exterior algebra of V.

The disjoint union of the tangent spaces $T_p\mathcal{M}$ is called the tangent bundle denoted by $T\mathcal{M}$. A vector field is a section of $T\mathcal{M}$. The dual bundle of the tangent bundle is called the cotangent bundle denoted by $T^*\mathcal{M}$. A 1-form is a section of $T^*\mathcal{M}$, while a differential r-form is a section of $\wedge^r T^*\mathcal{M}$. A 0-form means a smooth function on $\mathcal{M}$. Let $\mathcal{A}$ denote the algebra of smooth functions on $\mathcal{M}$, and let $\Omega^r(\mathcal{A})$ denote the $\mathcal{A}$-bimodule of differential r-forms. Let

$$\Omega(\mathcal{A}) = \bigoplus_{r \geq 0} \Omega^r(\mathcal{A})$$

denote the vector space of all differential forms, where $\Omega^0(\mathcal{A}) = \mathcal{A}$. The exterior derivative is a differential operator $d : \Omega^r(\mathcal{A}) \to \Omega^{r+1}(\mathcal{A})$, that can be defined as the unique linear mapping satisfying the Leibniz rule

$$d(\omega \wedge \eta) = (d\omega) \wedge \eta + (-1)^{\deg \omega}(\omega \wedge d\eta),$$

and

$$d(d\omega) = 0,$$

for $\omega,\ \eta \in \Omega(\mathcal{A})$, and $(df)(\xi) = \xi f$, where ξ is a vector field, and $f \in \mathcal{A}$.

Suppose V to be an oriented inner product space, and $\{e_i\}$ an oriented orthonormal basis. For $0 \leq k \leq n$, the Hodge star operator on V is a linear operator on $\wedge(V)$ with the property

$$*(e_1 \wedge_2 \wedge \cdots \wedge e_k) = e_{k+1} \wedge_{k+2} \wedge \cdots \wedge e_n.$$

This operator induces an inner product on $\wedge^r V$. Given $\omega, \eta \in \wedge^r V$, one has

$$\omega \wedge *\eta = (\omega, \eta)\sigma,$$

where σ is the normalized volume form. One can repeat the construction above for each tangent space of an oriented Riemannian manifold such that, given $\zeta, \eta \in \Omega^r(\mathcal{A})$,

$$(\omega, \eta) = \int_{\mathcal{M}} \omega \wedge *\eta.$$

The codifferential $\delta : \Omega^r(\mathcal{A}) \to \Omega^{r-1}(\mathcal{A})$ is the adjoint of the exterior derivative, that is,

$$(\omega, d\eta) = (\delta\omega, \eta),$$

which is actually the generalized Stokes' theorem. The Laplace-de Rham operator $\Delta : \Omega^r(\mathcal{A}) \to \Omega^r(\mathcal{A})$ is given by

$$\Delta = \delta d + d\delta.$$

It lies at the heart of Hodge theory. The elements in the space

$$\mathcal{H}_\Delta^r(\mathcal{A}) = \{\omega \in \Omega^r(\mathcal{A}) | \Delta\omega = 0\}$$

are called harmonic forms. In particular, the elements in the space $\mathcal{H}_\Delta^0(\mathcal{A})$ are called harmonic functions. The dimension of the space $\mathcal{H}_\Delta^r(\mathcal{A})$ is called the r-th Betti number.

3. Laplace-de Rham Operator on Graphs: Coordinate-Free

Let $\mathcal{M}$ be a finite set. An associative algebra $\mathcal{A}$ can be formed over all real-valued functions on $\mathcal{M}$ via introducing a bilinear multiplication

$$(fg)(i) = f(i)g(i)$$

for all $i \in \mathcal{M}$. It is obvious that the constant function $\mathbb{I}(i) = 1$ for all $i \in \mathcal{M}$ is an identity element in $\mathcal{A}$.

We can extend $\mathcal{A}$ to a differential graded algebra

$$\Omega(\mathcal{A}) = \bigoplus_{r \geq 0} \Omega^r(\mathcal{A}),$$

where $\Omega^0(\mathcal{A}) = \mathcal{A}$, and $\Omega^r(\mathcal{A})$ consists of $\mathcal{A}$-bimodules. Given an element $\omega \in \Omega^r(\mathcal{A})$, its degree is denoted by

$\deg \omega = r$. The differential graded algebra is equipped with a linear map $d : \Omega^r(\mathcal{A}) \to \Omega^{r+1}(\mathcal{A})$ which satisfies the Leibniz rule

$$d(\omega\eta) = (d\omega)\eta + (-1)^{\deg \omega}(\omega d\eta), \tag{1}$$

and

$$d(d\omega) = 0, \tag{2}$$

for $\omega,\ \eta \in \Omega(\mathcal{A})$. By following the terminologies in differential geometry, the map d is called the exterior derivative or differential, and the elements in $\Omega^r(\mathcal{A})$ are called r-forms.

Assume an inner product defined on $\Omega(\mathcal{A})$. The codifferential $\delta : \Omega^r(\mathcal{A}) \to \Omega^{r-1}(\mathcal{A})$ is the adjoint of the exterior derivative with respect to the inner product, that is

$$(\omega, d\eta) = (\delta\omega, \eta), \tag{3}$$

where $\omega \in \Omega^r(\mathcal{A}),\ \eta \in \Omega^{r-1}(\mathcal{A})$. Moreover, we define $\delta f = 0$ for any $f \in \mathcal{A}$. It is obvious that $\delta(\delta\omega) = 0$ for any $\omega \in \Omega(\mathcal{A})$.

As in the continuous case, now we can construct a discrete analogue of Laplace-de Rham operator $\Delta : \Omega^r(\mathcal{A}) \to \Omega^r(\mathcal{A})$ as

$$\Delta = d\delta + \delta d. \tag{4}$$

It can be verified that

$$(\Delta\omega, \eta) = (\omega, \Delta\eta)$$

and

$$(\Delta\omega, \omega) \geq 0.$$

The spaces of harmonic forms are defined by

$$\mathcal{H}^r_\Delta(\mathcal{A}) = \{\omega \in \Omega^r(\mathcal{A}) | \Delta\omega = 0\}.$$

In particular, the elements in the space $\mathcal{H}^0_\Delta(\mathcal{A})$ are called harmonic functions. The r-th Betti number is then given by

$$b_r = \dim \mathcal{H}^r_\Delta(\mathcal{A}).$$

4. Laplace-de Rham Operator on Graphs: Coordinate-Dependent

Using the Kronecker delta, we define a set of functions $e_i \in \mathcal{A}$ as $e_i(j) = \delta_{ij}$ for any $i, j \in \mathcal{M}$. Obviously, the function set $\{e_i\}$ forms a basis of $\mathcal{A}$. Then each $f \in \mathcal{A}$ can be expressed as $f = \sum_i f(i)e_i$. In particular, $\mathbb{I} = \sum_i e_i$.

4.1. Discrete Differential

Let $e_{ij} = e_i de_j,\ i \neq j$. It can be shown that $\{e_{ij}\}_{i \neq j}$ is a basis of $\Omega^1(\mathcal{A})$. That means each $\omega \in \Omega^1(\mathcal{A})$ can be written as

$$\omega = \sum_{i \neq j} \omega_{ij} e_{ij}.$$

Note that e_{ii} has not been defined. We may set $e_{ii} = 0$. Then

$$\omega = \sum_{i,j} \omega_{ij} e_{ij}.$$

Through a straightforward computation based on Equations (1) and (2), we can obtain

$$df = \sum_{i,j} [f(j) - f(i)] e_{ij} \tag{5}$$

Since e_{ij} is a basis of $\Omega^1(\mathcal{A})$, a function $f \in \Omega^0(\mathcal{A})$ is constant iff $df = 0$. Generally, define

$$e_{i_1,\ldots,i_r} = e_{i_1 i_2} e_{i_2 i_3} \cdots e_{i_{r-1} i_r}.$$

Then $\{e_{i_1 \ldots i_r}\}$ forms a basis of $\Omega^{r-1}(\mathcal{A})$. Hence, any $\omega \in \Omega^{r-1}(\mathcal{A})$ can be written as

$$\omega = \sum_{i_1, \cdots, i_r} \omega_{i_1 \cdots i_r} e_{i_1 \ldots i_r}.$$

Similarly, we have

$$d\omega = \sum_{i_1, \ldots, i_{r+1}} e_{i_1 \ldots i_{r+1}} \sum_{k=1}^{r+1} (-1)^{k+1} \omega_{i_1 \cdots \hat{i}_k \cdots i_{r+1}}. \tag{6}$$

4.2. Discrete Codifferential

Let us associate the finite set $\mathcal{M}$ with an ergodic Markov chain. Denote by p_{ij} for all $i, j \in \mathcal{M}$ the transition probabilities in which $p_{ii} = 0$, and denote by π_i for all $i \in \mathcal{M}$ the stationary probabilities. An inner product on $\Omega^0(\mathcal{A})$ can be defined by

$$(e_i, e_j) = 0, i \neq j;\ \text{and}\ (e_i, e_i) = \pi_i.$$

Let $c_{ij} = \pi_i p_{ij}$. An inner product on $\Omega^1(\mathcal{A})$ can be defined by

$$(e_{ij}, e_{kl}) = 0, i \neq j,\ k \neq l;\ \text{and}\ (e_{ij}, e_{ij}) = c_{ij}.$$

For any $\omega \in \Omega^1(\mathcal{A})$, with respect to the above inner product, the codifferential δ defined in Equation (3) is computed as

$$\delta\omega = \sum_i e_i \sum_j \frac{c_{ji}\omega_{ji} - c_{ij}\omega_{ij}}{\pi_i}. \tag{7}$$

Generally, let

$$c_{i_1 \cdots i_r} = \pi_{i_1} p_{i_1 i_2} \cdots p_{i_{r-1} i_r}$$

for all $r > 1$. An inner product on $\Omega^{r-1}(\mathcal{A})$ is defined by

$$(e_{i_1\cdots i_r}, e_{j_1\cdots j_r}) = \delta_{i_1 j_1}\cdots\delta_{i_r j_r} c_{i_1\cdots i_r}.$$

Then, for any $\omega \in \Omega^r(\mathcal{A})$, we have

$$\delta\omega = \sum_{i_1,\cdots,i_r} e_{i_1\cdots i_r} \sum_j \sum_{k=1}^{r} \frac{1}{c_{i_1\cdots i_r}} \cdot(-1)^{k+1} c_{i_1\cdots i_{k-1} j i_k \cdots i_r} \omega_{i_1\cdots i_{k-1} j i_k \cdots i_r} \quad (8)$$

4.3. Discrete Laplace-de Rham Operator

By substituting Equations (5) and (7) into Equation (4), the discrete Laplace-de Rham operator acts on $f \in \Omega^0(\mathcal{A})$ as

$$\Delta f = \sum_i \left[2f(i) - \sum_j \frac{c_{ji} + c_{ij}}{\pi_i} f(j)\right] e_i, \quad (9)$$

and it acts on $\omega \in \Omega^1(\mathcal{A})$ as

$$\begin{aligned}\Delta\omega = & \sum_{i,j} 2\omega_{ij} e_{ij} \\ & + \sum_{i,j} \frac{1}{c_{ij}} \sum_k c_{ijk}(\omega_{jk} - \omega_{ik}) e_{ij} \\ & - \sum_{i,j} \frac{1}{c_{ij}} \sum_k c_{ikj}(\omega_{kj} - \omega_{ij} + \omega_{ik}) e_{ij} \\ & + \sum_{i,j} \frac{1}{c_{ij}} \sum_k c_{kij}(\omega_{ki} - \omega_{kj}) e_{ij} \\ & + \sum_{i,j} \left(\sum_k \frac{c_{kj}\omega_{kj} - c_{jk}\omega_{jk}}{\pi_j}\right) e_{ij} \\ & - \sum_{i,j} \left(\sum_k \frac{c_{ki}\omega_{ki} - c_{ik}\omega_{ik}}{\pi_i}\right) e_{ij}. \quad (10)\end{aligned}$$

For the sake of simplicity, we omit the discussion of the coordinate-dependent representation of the discrete Laplace-de Rham operator acting on $\Omega^r(\mathcal{A})$ with $r > 1$.

Given a finite set $\mathcal{M}$ equipped with a function $w : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ which is nonnegative and symmetric, a Markov chain over this finite set can be naturally defined by

$$p_{ij} = \frac{w(i,j)}{d_i},$$

where

$$d_i = \sum_j w(i,j),$$

and it has a closed-form stationary distribution

$$\pi_i = \frac{d_i}{\sum_j d_j}.$$

Equation (9) with respect to this particular Markov chain will be

$$\Delta f = 2\sum_i \left[f(i) - \sum_j \frac{w(i,j)}{d_i} f(j)\right] e_i.$$

The matrix form of the above expression has been widely used as the definition of Laplacian for undirected graphs (Chung et al., 2000) up to factor 2. The strong pointwise consistency of this undirected graph Laplacian to the weighted Laplace-Beltrami operator (Chung et al., 2000) is established in (Hein et al., 2005).

5. High-Order Regularization

We use the discrete analogue of the Laplace-de Rham operator to regularize the functions on paths of a graph or forms. For the sake of simplicity, the discussion is restricted to functions on edges or 1-forms.

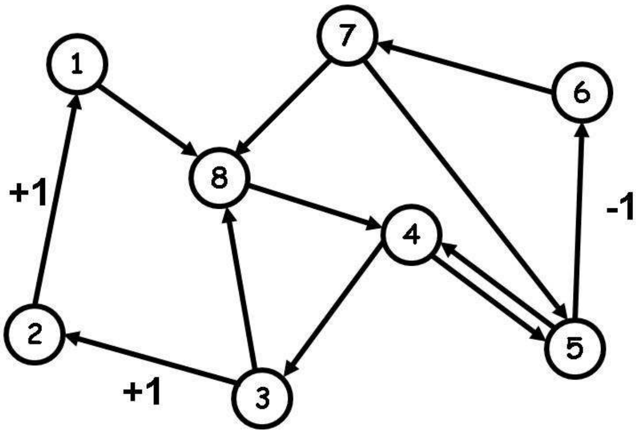
Given a directed graph $\mathcal{G} = (\mathcal{M}, \mathcal{E})$ with vertex set $\mathcal{M}$ and edge set $\mathcal{E}$, the edges in $\mathcal{E}$ belong to two different classes denote by a discrete set $\mathcal{Y} = \{-1, 1\}$. For instance, in a social network, the relationships between two individuals can be roughly classified as `trust` and `not trust`. We assume that the labels of the edges in a subset of $\mathcal{E}$ have been given. The task is to predict the labels of the remaining unclassified edges (Figure 2). Let ω be a 1-form in $\Omega^1(\mathcal{A})$, which is used to classify the edges. Let η be another 1-form, which encodes the label information that is provided in advance. We suppose there is an ergodic Markov chain associated with this directed graph. The Markov chain is used to form the inner product in $\Omega(\mathcal{A})$. The unclassified edges can then be classified via solving the optimization problem

$$\underset{\omega\in\Omega^1(\mathcal{A})}{\operatorname{argmin}} \left\{(\omega, \Delta\omega) + C\|\omega - \eta\|^2\right\}, \quad (11)$$

where $C > 0$ is the regularization parameter. The first term in the objective function requires the solution to be as as harmonic as possible while the second term requires the solution to be as close to the original form as possible. A trade-off between these two terms is made via the regularization parameter.

Let us take the standard basis $\{e_i\}$. From Equation (10), we have

$$\begin{aligned}\Delta e_{ij} = & \left(2 + \sum_k \frac{c_{ikj}}{c_{ij}}\right) e_{ij} \\ & - \sum_k \left(\frac{c_{ikj} + c_{ijk}}{c_{ik}} - \frac{c_{ij}}{\pi_i}\right) e_{ik} \\ & - \sum_k \left(\frac{c_{ikj} + c_{kij}}{c_{kj}} - \frac{c_{ij}}{\pi_j}\right) e_{kj}\end{aligned}$$



0	0	0	0	0	0	0	1
1	0	0	0	0	0	0	0
0	1	0	0	0	0	0	1
0	0	1	0	1	0	0	0
0	0	0	0	1	1	0	0
0	0	0	0	0	0	1	0
0	0	0	0	1	0	0	1
0	0	0	1	0	0	0	0

Figure 2. Classification for the edges of a directed graph. Left panel: a directed graph with some edges are labeled; right panel: the adjacency matrix of the directed graph. This kind of classification issue can be solved through the discrete Laplace-de Rham operator based high-order regularization.

This can be used to represent the discrete Laplace-de Rham operator Δ as a $|\mathcal{E}| \times |\mathcal{E}|$ matrix with respect to the standard basis. Obviously, this matrix is not symmetric.

6. Experiments

We address the spam detection issue using the high-order regularization on graph. In most web search engines, the more hyperlinks that point to a web page, the more important the web page. Thus, web spammers will try to create a large number of links to their web pages by creating lots of their own pages and web sites and linking them all together to fool the web search engines in which the hyperlink structure is considered for ranking web pages. Our experiments are based on the spam data set `webspam-uk2006-1.2` (Castillo et al., 2006). The hosts in this data set have been manually labeled as `normal`, `borderline`, `spam`, and `cannot judge`.

A directed graph over hosts is constructed as follows. Each host can be regarded a collection of web pages. Given two hosts, if there exists at least one hyperlink from the web pages on one host to the web pages on the other host, then we say that there is a directed link from one to the other. We take a strongly connected component of the host graph for a case study. The subgraph contains 556 hosts and $9,573$ links, and 183 hosts are spam.

We compare two approaches respectively based on the zero-order and the first-order discrete Laplace-de Rham operators. As discussed in Section 4.3, the zero-order discrete Laplace-de Rham operator is just the discrete Laplace-Beltrami operator, and it has shown good performance on web categorization (Zhou et al., 2005). For applying the discrete higher-order Laplace-Beltrami operator based regularization to the same task, however, the situation becomes somewhat tricky. In the spam data set, what we know is the labels of hosts rather than the labels of links. So we have to consider assigning labels to hyperlinks with respect the labels of hosts. In our experiments, if a link points to a spam host, then this link will be regarded as spam; otherwise, it will be regarded as normal. Consequently, we obtain the labels for a subset of links according to the labels of a subset of hosts which are provided as training examples. The labels of the subset of links are then used in the high-order regularization, and the solution is a 1-form defined on the whole set of links. The value of the 1-form on each link can be considered as the measure of its quality. For classifying a host to be spam or normal, we simply sum those values on both its inlinks and outlinks, and the sum is correspondingly considered as the measure of the quality of the host. In other words, if the inlinks and the outlinks of a host are of low quality, it is likely that the host is spam.

Spam detection is an unbalanced classification issue. In the chosen subgraph, 32.91% hosts are labeled as `spam`. Hence, we consider measuring algorithmic performances via precision and recall, rather than classification accuracy. Precision is the ratio of the number of retrieved and relevant documents to the number of documents retrieved, and recall is the proportion of the number of relevant documents that are retrieved to the total number of the relevant documents available. In addition, classifying a normal host into spam is much worse than classifying a spam host into nor-

Table 1. Precisions for the two classification approaches respectively based on the zero-order and the first-order discrete Laplace-de Rham operators. Recall is fixed at 50%. The zero-order discrete Laplace-de Rham operator is just the discrete Laplace-Beltrami operator. The numbers in the first line show the proportion of labeled instances. Each precision result is averaged over 20 trials.

Labeled instances (%)	10	15	20	25
Laplace-de Rham ($r = 0$)	86.69 ± 2.35	94.66 ± 0.95	98.89 ± 0.47	99.99 ± 0.00
Laplace-de Rham ($r = 1$)	71.09 ± 2.84	73.73 ± 1.47	81.41 ± 1.43	82.82 ± 1.25

mal. That means precision is more crucial than recall. Consequently, comparing precision with low recall is more significant than comparing precision with high recall.

The experimental results are summarized in Table 1. The regularization approach based on the zero-order Laplace-de Rham operator performs better than the regularization approach based on the first-order Laplace-de Rham operator in this classification task. We think the reason is that the the first-order approach has to classify the vertices in an indirect way. In other tasks like classifying the pairwise relationships among the individuals of a social networks, the first-order approach can be expected to perform better than the zero-order approach. Unfortunately, so far, we have not seen such a benchmark publicly available.

7. Conclusion and Discussion

We proposed a discrete analogue of the Laplace-de Rham operator. The discrete analogue acts on the functions on the paths of any length in a directed graph. When acting on the functions on vertices, it naturally reduces to the usual graph Laplacian, a discrete analogue of the Laplace-Beltrami operator. The discrete Laplace-de Rham operator was used to develop high-order regularization on graphs. The basic methodology of the present work is to consider the algebra of the functions on the vertices of a graph. We do not attempt directly manipulating the combinatorial structure of a graph. This kind of methodology is actually pretty common is modern mathematics. It turns out that the geometrical structure on a space is always completely expressible at the language of the associated (commutative) algebra of the appropriate complex-valued functions on the space.

There are a number of interesting further directions suggested by this work. We will restrict ourselves to three such directions. First, we want to develop more discrete differential operators, and then use them to construct a general regularization theory for graphs. Although the graph Laplacian has been successively applied to many machine learning issues, so far we have no seen any reason of that the graph Laplacian is the only choice in forming a graph regularization. Second, it is worth exploring the convergence properties of the discrete Laplace-de Rham operator when the elements of a finite set are sampled from a manifold. This study will be greatly helpful in identifying which definition is more reasonable in the sense of convergence once we have other choices of defining a discrete analogue. Perhaps we would be able to obtain a unified approach of showing the convergence of all of those discrete analogues such that a bridge from discrete to continuous mathematics would be built. Third, we are interested in applying this new graph-based method to a variety of real-world problems where the data are generally represented as graphs.

References

Ando, A., & Zhang, T. (2007). Learning on graph with Laplacian regularization. *Advances in Neural Information Processing Systems 19*. MIT Press, Cambridge, MA.

Belkin, M., & Niyogi, P. (2004). Semi-supervised learning on manifolds. *Machine Learning*, *56*, 209–239.

Castillo, C., Donato, D., Becchetti, L., Boldi, P., Santini, M., & Vigna, S. (2006). A reference collection for web spam. *SIGIR Forum*, *40*.

Chapelle, O., Weston, J., & Schölkopf, B. (2003). Cluster kernels for semi-supervised learning. *Advances in Neural Information Processing Systems 15*. MIT Press, Cambridge, MA.

Chung, F., Grigoryan, A., & Yau, S.-T. (2000). Higher eigenvalues and isoperimetric inequalities on Riemannian manifolds and graphs. *Communications on Analysis and Geometry*, *8*, 969–1026.

Connes, A. (1994). *Noncommutative geometry*. New York: Academic Press.

Dimakis, A., & Müller-Hoissen, F. (1992). Quantum mechanics on a lattice and q-deformations. *Physics Letters B*, *295*, 242–248.

Fielder, M. (1973). Algebraic connectivity of graphs. *Czech. Math. J.*, *23*, 298–305.

Fisher, M., Schröder, P., Desbrun, M., & Hoppe, H. (2007). Design of tangent vector fields. *Proc. 34th International ACM SIGGRAPH Conference on Computer Graphics and Interactive Techniques.*

Forman, R. (1999). Combinatorial differential topology and geometry. *New Perspectives in Algebraic Combinatorics*, *38*, 177–206.

Hagen, L., & Kahng, A. (1992). New spectral methods for ratio cut partitioning and clustering. *IEEE. Trans. on Computed Aided Desgin*, *11*, 1074–1085.

Hein, M., Audibert, J., & von Luxburg, U. (2005). From graphs to manifolds - weak and strong pointwise consistency of graph Laplacians. *Proc. 18th Annual Conference on Learning Theory* (pp. 470–485).

Joachims, T. (2003). Transductive learning via spectral graph partitioning. *Proc. 20th International Conference on Machine Learning.*

Jost, J. (2002). *Riemannian geometry and geometric analysis.* Berlin-Heidelberg: Springer-Verlag. Third edition.

Leok, M. (2004). *Foundations of computational geometric mechanics.* Doctoral dissertation, California Institute of Technology.

Mercat, C. (2001). Discrete Riemann surfaces and the Ising model. *Communications in Mathematical Physics*, *218*, 177–216.

Noyes, H. (1996). *Some thoughts on discrete physics and the reconstruction of quantum mechanics* (Technical Report SLAC-PUB-7145). Stanford Linear Accelerator Center.

Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*, 888–905.

Smola, A., & Kondor, R. (2003). Kernels and regularization on graphs. *Proc. 16th Annual Conference on Learning Theory.*

Zhou, D., Bousquet, O., Lal, T., Weston, J., & Schölkopf, B. (2004). Learning with local and global consistency. *Advances in Neural Information Processing Systems 16.* MIT Press, Cambridge, MA.

Zhou, D., Huang, J., & Schölkopf, B. (2005). Learning from labeled and unlabeled data on a directed graph. *Proc. 22th International Conference on Machine Learning.*

Zhou, D., Huang, J., & Schölkopf, B. (2007). Learning with hypergraphs: Clustering, classification, and embedding. *Advances in Neural Information Processing Systems.* MIT Press, Cambridge, MA.

Zhu, X., Ghahramani, Z., & Lafferty, J. (2003). Semi-supervised learning using Gaussian fields and harmonic functions. *Proc. 20th International Conference on Machine Learning.*