

Letter to the Editor

From graph to manifold Laplacian: The convergence rate

A. Singer

Department of Mathematics, Yale University, 10 Hillhouse Ave., P.O. Box 208283, New Haven, CT 06520-8283, USA

Available online 26 May 2006

Communicated by Ronald R. Coifman on 12 March 2006

Abstract

The convergence of the discrete graph Laplacian to the continuous manifold Laplacian in the limit of sample size $N \rightarrow \infty$ while the kernel bandwidth $\varepsilon \rightarrow 0$, is the justification for the success of Laplacian based algorithms in machine learning, such as dimensionality reduction, semi-supervised learning and spectral clustering. In this paper we improve the convergence rate of the variance term recently obtained by Hein et al. [From graphs to manifolds—Weak and strong pointwise consistency of graph Laplacians, in: P. Auer, R. Meir (Eds.), Proc. 18th Conf. Learning Theory (COLT), Lecture Notes Comput. Sci., vol. 3559, Springer-Verlag, Berlin, 2005, pp. 470–485], improve the bias term error, and find an optimal criteria to determine the parameter ε given N . © 2006 Elsevier Inc. All rights reserved.

1. Introduction

Graph Laplacians are widely used in machine learning for dimensionality reduction, semi-supervised learning and spectral clustering ([3–9] and references therein). In these setups one is usually given a set of N data points $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N \in \mathcal{M}$, where $\mathcal{M} \subset \mathbb{R}^m$ is a Riemannian manifold with $\dim \mathcal{M} = d < m$. The points are given as vectors in the ambient space \mathbb{R}^m and the task is to find the unknown underlying manifold \mathcal{M} , its geometry and its low-dimensional representation.

The starting point of spectral methods is to extract an $N \times N$ weight matrix W from a suitable semi-positive kernel k as follows

$$W_{ij} = k(\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\varepsilon), \quad (1.1)$$

where $\|\cdot\|$ is the Euclidean distance in the ambient space \mathbb{R}^m and $\varepsilon > 0$ is the bandwidth of the kernel. A popular choice of kernel is the exponential kernel $k(x) = e^{-x}$, though other choices are also possible.

The weight matrix W is then normalized to be row stochastic, by dividing it by a diagonal matrix D whose elements are the row sums of W

$$D_{ii} = \sum_{j=1}^N W_{ij}, \quad (1.2)$$

and the (negative defined) graph Laplacian L is given by

E-mail address: amit.singer@yale.edu.

$$L = D^{-1}W - I, \tag{1.3}$$

where I is an $N \times N$ identity matrix.

In the case where the data points $\{\mathbf{x}_i\}_{i=1}^N$ are independently uniformly distributed over the manifold \mathcal{M} the graph Laplacian converges to the continuous Laplace–Beltrami operator Δ_M of the manifold. This statement has two manifestations. First, if $f : \mathcal{M} \rightarrow \mathbb{R}$ is a smooth function, then the following result concerning the bias term has been established by various authors [1,5,6] (among others)

$$\frac{1}{\varepsilon} \lim_{N \rightarrow \infty} \sum_{j=1}^N L_{ij} f(\mathbf{x}_j) = \frac{1}{2} \Delta_M f(\mathbf{x}_i) + O(\varepsilon^{1/2}). \tag{1.4}$$

Second, Hein et al. [1] established a uniform estimate of the error in both N and ε (the variance term), that decreases as $1/\sqrt{N}$ as $N \rightarrow \infty$, but increases as $\frac{1}{\varepsilon^{1+d/4}}$ as $\varepsilon \rightarrow 0$,¹

$$\frac{1}{\varepsilon} \sum_{j=1}^N L_{ij} f(\mathbf{x}_j) = \frac{1}{2} \Delta_M f(\mathbf{x}_i) + O\left(\frac{1}{N^{1/2} \varepsilon^{1+d/4}}, \varepsilon^{1/2}\right). \tag{1.5}$$

In practice, one cannot choose ε to be too small, even though smaller ε decreases the bias error (1.4), because $\frac{1}{N^{1/2} \varepsilon^{1+d/4}}$ diverges with ε . It was reasoned [1] that the convergence rate $\frac{1}{N^{1/2} \varepsilon^{1+d/4}}$ is unlikely to be improved, since the rates for estimating second derivatives in nonparametric regression are the same.

In this paper we show that the variance error is $O\left(\frac{1}{N^{1/2} \varepsilon^{1/2+d/4}}\right)$, which improves the convergence rate (1.5) by an asymptotic factor $\sqrt{\varepsilon}$. This improvement stems from the observation that the noise terms of Wf and Df are highly correlated, with a correlation coefficient of $1 - O(\varepsilon)$. At points where $\nabla_M f = 0$ the convergence rate is even faster. Moreover, we refine the bias error to be $O(\varepsilon)$ instead of $O(\varepsilon^{1/2})$ in (1.4), by a symmetry argument. Finally, balancing the two error terms leads to a natural optimal choice of ε ,

$$\varepsilon = \frac{C(\mathcal{M})}{N^{1/(3+d/2)}}, \tag{1.6}$$

that gives a minimal error between the continuous Laplace–Beltrami and its discrete graph Laplacian approximation. The constant $C(\mathcal{M})$ is a function of the manifold, that depends on its geometry, for example, its dimension, curvature and volume, but is independent of the number of sample points N . The variance error depends on the intrinsic dimension of the manifold and its volume, that can be estimated [11] with $O(1/N)$ accuracy, rather than $O(1/\sqrt{N})$. However, the bias error depends on the local curvature of the manifold which is unknown in advance. Therefore, when dealing with real life applications one still needs to use some trial and error experimentation in order to find the optimal ε for a given N , because $C(\mathcal{M})$ is unknown. When a different N is introduced, the parameter ε could be chosen according to (1.6).

Our results are summarized in the following:

$$\frac{1}{\varepsilon} \sum_{j=1}^N L_{ij} f(\mathbf{x}_j) = \frac{1}{2} \Delta_M f(\mathbf{x}_i) + O\left(\frac{1}{N^{1/2} \varepsilon^{1/2+d/4}}, \varepsilon\right). \tag{1.7}$$

We remark that if the points are not uniformly distributed, then Lafon et al. [6,7] showed that the limiting operator contains an additional drift term, and suggested a different kernel normalization that separates the manifold geometry from the distribution of points on it, and recovers the Laplace–Beltrami operator. Therefore, our analysis assumes that the initial data set is already uniformly distributed. Finally, we illustrate the improved convergence rates by computer simulations of spherical harmonics.

¹ The notation $O(\cdot, \cdot)$ means that there exist positive constants C_1, C_2 (independent of N and ε) such that $|O(\frac{1}{N^{1/2} \varepsilon^{1+d/4}}, \varepsilon^{1/2})| \leq C_1 \frac{1}{N^{1/2} \varepsilon^{1+d/4}} + C_2 \varepsilon^{1/2}$ for large N and small ε .

2. The bias term

Let $\mathbf{x}_1, \dots, \mathbf{x}_N$ be N independent uniformly distributed points over a d -dimensional compact Riemannian manifold $\mathcal{M} \subset \mathbb{R}^m$ of volume $\text{vol}(\mathcal{M})$. Suppose $f: \mathcal{M} \rightarrow \mathbb{R}$ is a smooth function. In this section we investigate limiting properties of the graph Laplacian (1.1)–(1.3) at a given fixed point $\mathbf{x} = \mathbf{x}_i$

$$(Lf)(\mathbf{x}) = \sum_{j=1}^N L_{ij} f(\mathbf{x}_j) = \frac{\sum_{j=1}^N \exp\{-\|\mathbf{x} - \mathbf{x}_j\|^2/2\varepsilon\} f(\mathbf{x}_j)}{\sum_{j=1}^N \exp\{-\|\mathbf{x} - \mathbf{x}_j\|^2/2\varepsilon\}} - f(\mathbf{x}). \quad (2.1)$$

We rewrite the graph Laplacian (2.1) as

$$(Lf)(\mathbf{x}) = \frac{\sum_{j=1}^N F(\mathbf{x}_j)}{\sum_{j=1}^N G(\mathbf{x}_j)} - f(\mathbf{x}), \quad (2.2)$$

where

$$F(\mathbf{x}_j) = \exp\left\{-\frac{\|\mathbf{x} - \mathbf{x}_j\|^2}{2\varepsilon}\right\} f(\mathbf{x}_j), \quad (2.3)$$

$$G(\mathbf{x}_j) = \exp\left\{-\frac{\|\mathbf{x} - \mathbf{x}_j\|^2}{2\varepsilon}\right\}. \quad (2.4)$$

We shall see that excluding the diagonal terms $j = i$ from the sums in (2.2) results in an $O\left(\frac{1}{N\varepsilon^{d/2}}\right)$ error, which is even smaller than the variance error squared (see Eq. (1.7)), therefore it is negligible

$$(Lf)(\mathbf{x}) = \frac{\sum_{j \neq i}^N F(\mathbf{x}_j)}{\sum_{j \neq i}^N G(\mathbf{x}_j)} - f(\mathbf{x}) \left(1 + O\left(\frac{1}{N\varepsilon^{d/2}}\right)\right). \quad (2.5)$$

The points \mathbf{x}_j are independent identically distributed (i.i.d.), therefore, $F(\mathbf{x}_j)$ ($j \neq i$) are also i.i.d., and by the law of large numbers one should expect

$$\frac{\sum_{j \neq i}^N F(\mathbf{x}_j)}{\sum_{j \neq i}^N G(\mathbf{x}_j)} \approx \frac{\mathbb{E}F}{\mathbb{E}G}, \quad (2.6)$$

where

$$\mathbb{E}F = \frac{1}{\text{vol}(\mathcal{M})} \int_{\mathcal{M}} \exp\left\{-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\varepsilon}\right\} f(\mathbf{y}) d\mathbf{y}, \quad (2.7)$$

$$\mathbb{E}G = \frac{1}{\text{vol}(\mathcal{M})} \int_{\mathcal{M}} \exp\left\{-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\varepsilon}\right\} d\mathbf{y}, \quad (2.8)$$

are the expected value of F and G , respectively. Our goal is to make Eq. (2.6) precise by obtaining the error estimate as a function of N and ε .

The expansion of the integral (2.7) in a Taylor series in $\sqrt{\varepsilon}$ is given by [2,6,10]

$$\frac{1}{(2\pi\varepsilon)^{d/2}} \int_{\mathcal{M}} \exp\left\{-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\varepsilon}\right\} f(\mathbf{y}) d\mathbf{y} = f(\mathbf{x}) + \frac{\varepsilon}{2} [E(\mathbf{x})f(\mathbf{x}) + \Delta_M f(\mathbf{x})] + O(\varepsilon^{3/2}), \quad (2.9)$$

where $E(\mathbf{x})$ is a scalar function of the curvature of \mathcal{M} at \mathbf{x} . Hence,

$$\frac{\mathbb{E}F}{\mathbb{E}G} - f(\mathbf{x}) = \frac{\varepsilon}{2} \Delta_M f(\mathbf{x}) + O(\varepsilon^{3/2}). \quad (2.10)$$

In other words, the finite sums in (2.5) are Monte Carlo estimators of the kernel integrals in (2.7) and (2.8), whose deterministic expansion in ε gives the Laplace–Beltrami operator as the leading order term. Moreover, Eqs. (2.7) and (2.9) mean that the error in excluding the diagonal terms $i = j$ is indeed $O\left(\frac{1}{N\varepsilon^{d/2}}\right)$.

Our first observation is that for smooth manifolds and smooth functions, Taylor expansions of the form (2.9) contain only integer powers of ε , whereas fractional powers, such as, $\varepsilon^{3/2}$ must vanish. The kernel integral (2.9) is evaluated by changing the integration variables to the local coordinates, and the fractional powers of ε that appear in the asymptotic expansion are due to odd monomials. However, the integrals of odd monomials vanish, because the kernel is symmetric. We conclude that a finer error estimate of (2.9) is

$$\frac{1}{(2\pi\varepsilon)^{d/2}} \int_{\mathcal{M}} \exp\left\{-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\varepsilon}\right\} f(\mathbf{y}) d\mathbf{y} = f(\mathbf{x}) + \frac{\varepsilon}{2}[E(\mathbf{x})f(\mathbf{x}) + \Delta_M f(\mathbf{x})] + O(\varepsilon^2), \tag{2.11}$$

where the $O(\varepsilon^2)$ depends on the curvature of the manifold, but the $O(\varepsilon^{3/2})$ vanishes as long as the manifold is smooth.

3. The variance error

The error term in Monte Carlo integration stems from the variance (noise) term which we evaluate below. Interestingly, the noise terms in the integrals of F and G are highly correlated, so we obtain a faster convergence rate.

In what follows we show that the estimator $\frac{\sum_{j \neq i}^N F(\mathbf{x}_j)}{\sum_{j \neq i}^N G(\mathbf{x}_j)} \rightarrow \frac{\mathbb{E}F}{\mathbb{E}G}$ in probability faster than what could be expected [1], because F and G are highly correlated random variables. To this end, we use the Chernoff inequality to establish an upper bound for the probability $p(N, \alpha)$ of having an α -error

$$p(N, \alpha) \equiv \Pr\left\{\frac{\sum_{j \neq i}^N F(\mathbf{x}_j)}{\sum_{j \neq i}^N G(\mathbf{x}_j)} - \frac{\mathbb{E}F}{\mathbb{E}G} > \alpha\right\}, \tag{3.1}$$

while an upper bound for $\Pr\left\{\frac{\sum_{j \neq i}^N F(\mathbf{x}_j)}{\sum_{j \neq i}^N G(\mathbf{x}_j)} - \frac{\mathbb{E}F}{\mathbb{E}G} < -\alpha\right\}$ can be obtained similarly.

The random variables $G(\mathbf{x}_j)$ are positive, thus

$$p(N, \alpha) = \Pr\left\{\sum_{j \neq i}^N [(\mathbb{E}G)F(\mathbf{x}_j) - (\mathbb{E}F + \alpha\mathbb{E}G)G(\mathbf{x}_j)] > 0\right\}, \tag{3.2}$$

which we rewrite as

$$p(N, \alpha) = \Pr\left\{\sum_{j \neq i}^N Y_j > (N - 1)\alpha(\mathbb{E}G)^2\right\}, \tag{3.3}$$

where

$$Y_j = (\mathbb{E}G)F(\mathbf{x}_j) - (\mathbb{E}F)G(\mathbf{x}_j) + \alpha(\mathbb{E}G)(\mathbb{E}G - G(\mathbf{x}_j)) \tag{3.4}$$

are zero mean ($\mathbb{E}Y_j = 0$) i.i.d. random variables. To ease the notation we replace $F(\mathbf{x}_j)$ and $G(\mathbf{x}_j)$ by F and G , respectively, because these are identical random variables.

The variance of Y_j depends on the second moments $\mathbb{E}(F^2)$, $\mathbb{E}(G^2)$ and $\mathbb{E}(FG)$

$$\begin{aligned} \mathbb{E}Y_j^2 &= (\mathbb{E}G)^2\mathbb{E}(F^2) - 2(\mathbb{E}G)(\mathbb{E}F)\mathbb{E}(FG) + (\mathbb{E}F)^2\mathbb{E}(G^2) \\ &\quad + 2\alpha(\mathbb{E}G)[(\mathbb{E}F)\mathbb{E}(G^2) - (\mathbb{E}G)\mathbb{E}(FG)] + \alpha^2(\mathbb{E}G)^2[\mathbb{E}(G^2) - (\mathbb{E}G)^2], \end{aligned} \tag{3.5}$$

which are calculated by Eq. (2.11)

$$\mathbb{E}F = \frac{(2\pi\varepsilon)^{d/2}}{\text{vol}(\mathcal{M})} \left\{ f(\mathbf{x}) + \frac{\varepsilon}{2}[E(\mathbf{x})f(\mathbf{x}) + \Delta_M f(\mathbf{x})] + O(\varepsilon^2) \right\}, \tag{3.6}$$

$$\mathbb{E}G = \frac{(2\pi\varepsilon)^{d/2}}{\text{vol}(\mathcal{M})} \left\{ 1 + \frac{\varepsilon}{2}E(\mathbf{x}) + O(\varepsilon^2) \right\}, \tag{3.7}$$

$$\mathbb{E}(F^2) = \frac{(\pi\varepsilon)^{d/2}}{\text{vol}(\mathcal{M})} \left\{ f^2(\mathbf{x}) + \frac{\varepsilon}{4}[E(\mathbf{x})f^2(\mathbf{x}) + \Delta_M f^2(\mathbf{x})] + O(\varepsilon^2) \right\}, \tag{3.8}$$

$$\mathbb{E}(G^2) = \frac{(\pi\varepsilon)^{d/2}}{\text{vol}(\mathcal{M})} \left\{ 1 + \frac{\varepsilon}{4} E(\mathbf{x}) + O(\varepsilon^2) \right\}, \quad (3.9)$$

$$\mathbb{E}(FG) = \frac{(\pi\varepsilon)^{d/2}}{\text{vol}(\mathcal{M})} \left\{ f(\mathbf{x}) + \frac{\varepsilon}{4} [E(\mathbf{x})f(\mathbf{x}) + \Delta_M f(\mathbf{x})] + O(\varepsilon^2) \right\}. \quad (3.10)$$

Therefore, the variance of Y_j is

$$\begin{aligned} \mathbb{E}Y_j^2 &= \frac{2^d(\pi\varepsilon)^{3d/2}}{\text{vol}(\mathcal{M})^3} \frac{\varepsilon}{4} [\Delta_M f^2(\mathbf{x}) - 2f(\mathbf{x})\Delta_M f(\mathbf{x}) + O(\varepsilon)] \\ &\quad + 2\alpha \frac{2^d(\pi\varepsilon)^{3d/2}}{\text{vol}(\mathcal{M})^3} \frac{\varepsilon}{4} [\Delta_M f(\mathbf{x}) + O(\varepsilon)] + \alpha^2 \frac{2^d(\pi\varepsilon)^{3d/2}}{\text{vol}(\mathcal{M})^3} [1 + O(\varepsilon, \varepsilon^{d/2})]. \end{aligned} \quad (3.11)$$

Clearly, we put our focus in the regime where $\alpha \ll \varepsilon$, because we are estimating an $O(\varepsilon)$ quantity, namely $\frac{\varepsilon}{2}\Delta_M f(\mathbf{x})$, so there is no point of having an error α which is larger than the estimate itself. Therefore, the α and α^2 terms of the variance (3.11) are negligible, and we obtain

$$\mathbb{E}Y_j^2 = \frac{2^d(\pi\varepsilon)^{3d/2}}{\text{vol}(\mathcal{M})^3} \frac{\varepsilon}{4} [\Delta_M f^2(\mathbf{x}) - 2f(\mathbf{x})\Delta_M f(\mathbf{x}) + O(\varepsilon)]. \quad (3.12)$$

Furthermore,

$$\Delta_M f^2 - 2f\Delta_M f = 2\|\nabla_M f\|^2,$$

hence

$$\mathbb{E}Y_j^2 = \frac{2^d(\pi\varepsilon)^{3d/2}}{\text{vol}(\mathcal{M})^3} \frac{\varepsilon}{2} [\|\nabla_M f\|^2 + O(\varepsilon)]. \quad (3.13)$$

Note that Y_j are bounded random variables, because F and G are bounded (2.3)–(2.4), and the definition of Y_j (3.4). Therefore, the inequality of Chernoff gives

$$p(N, \alpha) \leq 2 \exp \left\{ -\frac{(N-1)\alpha^2 2^d (\pi\varepsilon)^{d/2} \text{vol}(\mathcal{M})}{2\varepsilon [\|\nabla_M f\|^2 + O(\varepsilon)]} \right\}. \quad (3.14)$$

The practical meaning of the inequality (3.14) is that the noise error in the estimation of $\frac{\varepsilon}{2}\Delta_M f(\mathbf{x})$ is of the order of

$$\alpha \approx \frac{\sqrt{2\varepsilon} \sqrt{\|\nabla_M f\|^2 + O(\varepsilon)}}{\sqrt{N} 2^{d/2} (\pi\varepsilon)^{d/4} \sqrt{\text{vol}(\mathcal{M})}}. \quad (3.15)$$

To extract the Laplacian itself we further divide by ε (see Eq. (1.4)), therefore, the noise error is

$$\frac{\alpha}{\varepsilon} \approx \frac{\sqrt{\|\nabla_M f\|^2 + O(\varepsilon)}}{\sqrt{N} 2^{d/2-1} \pi^{d/4} \varepsilon^{1/2+d/4} \sqrt{\text{vol}(\mathcal{M})}} = O\left(\frac{1}{N^{1/2} \varepsilon^{1/2+d/4}}\right). \quad (3.16)$$

This improves the convergence rate of [1] by the asymptotic factor $\varepsilon^{1/2}$. Moreover, if

$$\nabla_M f(\mathbf{x}) = 0, \quad (3.17)$$

Eq. (3.15) implies that the convergence rate is even faster by another asymptotic factor of $\varepsilon^{1/2}$, and is given by $O\left(\frac{1}{N^{1/2} \varepsilon^{d/4}}\right)$.

4. Example: spherical harmonics

In this section we illustrate the convergence rate (1.7) for the case of the unit sphere $\mathcal{M} = S^2 \subset \mathbb{R}^3$, a two-dimensional compact manifold ($d = 2$). The eigenfunctions of the Laplacian on the unit sphere are known as the spherical harmonics. The first nontrivial spherical harmonic function is $f(x, y, z) = z$ which satisfies $\Delta_M f = -2f$. In particular, at the north pole $\Delta_M f|_{(0,0,1)} = -2$. Furthermore, the tangent plane at the north pole is the xy -plane, so $\nabla_M f|_{(0,0,1)} = 0$. The gradient condition (3.17) implies that the convergence rate should be $\varepsilon^{-1/2}$, and indeed the computer simulation yielded $\varepsilon^{-0.52}$ (see Figs. 1 and 2). Next, we examined the function $g = z + x$ at the north pole. Since x is also a spherical harmonic, and $x = 0$ at the north pole, it follows that $\Delta_M g|_{(0,0,1)} = -2$ as well. However, $\nabla_M g \neq 0$, because g changes at the x direction. Therefore, by Eq. (1.7) we expect the convergence rate to be ε^{-1} , and indeed a linear fit gives a slope -0.998 (see Fig. 3).

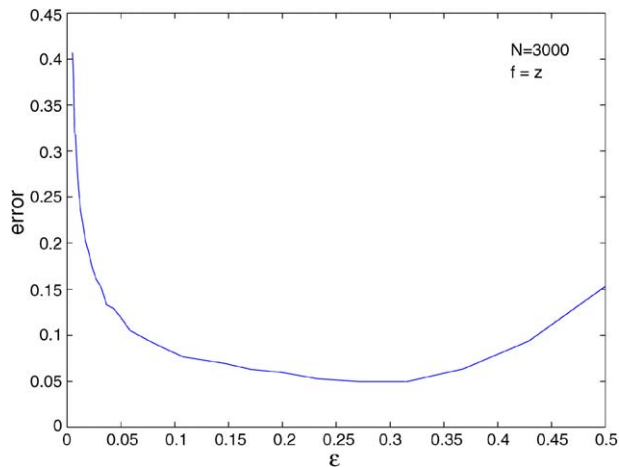


Fig. 1. The error between the discrete and continuous Laplace–Beltrami of the unit sphere S^2 of the spherical harmonic $f = z$ at the north pole as a function of ε . $N = 3000$ points were uniformly distributed over the sphere, and the error is averaged over $m = 1000$ independent trials (for each ε) as $\text{error} = \sqrt{\frac{1}{m} \sum_{i=1}^m \left(\frac{2}{\varepsilon} Lf - (-2)\right)^2}$. For $\varepsilon \approx 0.3$ a minimal error is obtained, where the sum of the bias and the variance errors is minimal.

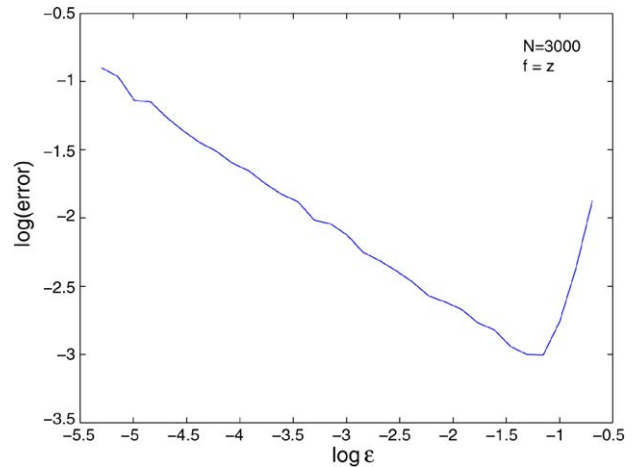


Fig. 2. A logarithmic plot of Fig. 1. A linear fit of the small ε region, where the variance error is dominant, gives a slope -0.52 in agreement with the gradient condition (3.17) that predicts $\varepsilon^{-1/2}$ asymptotic.

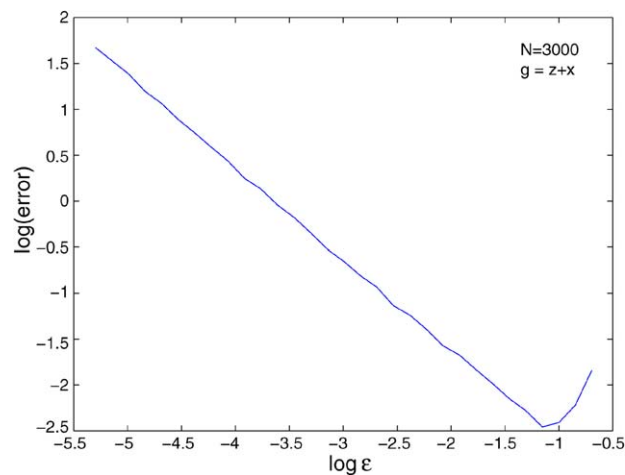


Fig. 3. A logarithmic plot of the graph Laplacian error against ε , for $g = z + x$ at the north pole. A linear fit in the variance error dominant regime gives a slope -0.998 in accordance with the expected ε^{-1} (Eq. (1.7)).

References

- [1] M. Hein, J. Audibert, U. von Luxburg, From graphs to manifolds—Weak and strong pointwise consistency of graph Laplacians, in: P. Auer, R. Meir (Eds.), Proc. 18th Conf. Learning Theory (COLT), Lecture Notes Comput. Sci., vol. 3559, Springer-Verlag, Berlin, 2005, pp. 470–485.
- [2] M. Belkin, Problems of learning on manifolds, Ph.D. dissertation, University of Chicago, 2003.
- [3] M. Belkin, P. Niyogi, Laplacian eigenmaps and spectral techniques for embedding and clustering, in: T.G. Dietterich, S. Becker, Z. Ghahramani (Eds.), Adv. Neural Inform. Process. Syst., vol. 14, MIT Press, Cambridge, MA, 2002.
- [4] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput. 15 (6) (2003) 1373–1396.
- [5] M. Belkin, P. Niyogi, Towards a theoretical foundation for Laplacian-based manifold methods, in: P. Auer, R. Meir (Eds.), Proc. 18th Conf. Learning Theory (COLT), Lecture Notes Comput. Sci., vol. 3559, Springer-Verlag, Berlin, 2005, pp. 486–500.
- [6] S. Lafon, Diffusion maps and geometric harmonics, Ph.D. dissertation, Yale University, 2004.
- [7] B. Nadler, S. Lafon, R.R. Coifman, I.G. Kevrekidis, Diffusion maps, spectral clustering and eigenfunctions of Fokker–Planck operators, in: Y. Weiss, B. Schölkopf, J. Platt (Eds.), Adv. Neural Inform. Process. Syst. (NIPS), vol. 18, MIT Press, Cambridge, MA, 2006.

- [8] R.R. Coifman, S. Lafon, A.B. Lee, M. Maggioni, B. Nadler, F. Warner, S.W. Zucker, Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps, *Proc. Natl. Acad. Sci.* 102 (21) (2005) 7426–7431.
- [9] R.R. Coifman, S. Lafon, A.B. Lee, M. Maggioni, B. Nadler, F. Warner, S.W. Zucker, Geometric diffusions as a tool for harmonic analysis and structure definition of data: Multiscale methods, *Proc. Natl. Acad. Sci.* 102 (21) (2005) 7432–7437.
- [10] O.G. Smolyanov, H.V. Weizsäcker, O. Wittich, Brownian motion on a manifold as limit of stepwise conditioned standard Brownian motions, in: F. Gesztesy (Ed.), *Stochastic Processes, Physics and Geometry, New Interplays, II*, CMS Conf. Proc., vol. 29, Amer. Math. Soc., Providence, RI, 2000.
- [11] M. Hein, Y. Audibert, Intrinsic dimensionality estimation of submanifolds in R^d , in: L. De Raedt, S. Wrobel (Eds.), *Proc. 22nd Int. Conf. Machine Learning*, ACM, 2005, pp. 289–296.