Neural Networks 24 (2011) 345–359

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

A generalized diffusion frame for parsimonious representation of functions on data defined manifolds

H.N. Mhaskar

Department of Mathematics, California State University, Los Angeles, CA, 90032, USA

ARTICLE INFO

Article history:
Received 27 August 2010
Received in revised form 22 December 2010
Accepted 29 December 2010

Keywords:
Data defined manifolds
Diffusion wavelets
Littlewood–Paley decomposition
Kernel based methods
Local Besov spaces

ABSTRACT

One of the now standard techniques in semi-supervised learning is to think of a high dimensional data as a subset of a low dimensional manifold embedded in a high dimensional ambient space, and to use projections of the data on eigenspaces of a diffusion map. This paper is motivated by a recent work of Coifman and Maggioni on diffusion wavelets to accomplish such projections approximately using iterates of the heat kernel. In greater generality, we consider a quasi-metric measure space $\mathbb{X}$ (in place of the manifold), and a very general operator $\mathcal{T}$ defined on the class of integrable functions on $\mathbb{X}$ (in place of the diffusion map). We develop a representation of functions on $\mathbb{X}$ in terms of linear combinations of iterates of $\mathcal{T}$. Our construction obviates the need to compute the eigenvalues and eigenfunctions of the operator. In addition, the local smoothness of a function f is characterized by the local norm behavior of the terms in our representation of f. This property is similar to that of the classical wavelet representations. Although the operator $\mathcal{T}$ utilizes the values of the target function on the entire space, this ability results in automatic "feature detection", leading to a parsimonious representation of the target function. In the case when $\mathbb{X}$ is a smooth compact manifold (without boundary), our theory allows $\mathcal{T}$ to be any operator that commutes with the heat operator, subject to certain conditions on its eigenvalues. In particular, $\mathcal{T}$ can be chosen to be the heat operator itself, or a Green's operator corresponding to a suitable pseudo-differential operator.

1. Introduction

In the last few years, diffusion/Laplacian eigenmaps have developed into a popular tool for understanding and representing large, high dimensional, unstructured data. Applications of these techniques include document analysis (Coifman & Maggioni, 2006), face recognition (He, Yan, Hu, Niyogi, & Zhang, 2005), semi-supervised learning (Belkin, Matveeva, & Niyogi, 2004; Belkin & Niyogi, 2004), image processing (Donoho & Grimes, 2002), and cataloguing of galaxies (Donoho, Levi, Starck, & Martinez, 2002), to name a few. The special issue of Applied and Computational Harmonic Analysis (Chui & Donoho, 2006) contains several papers that serve as a good introduction to this subject.

At a theoretical level, the idea is to assume that the data lies on a low dimensional compact manifold $\mathbb{X}$ embedded in a high dimensional ambient space. Associated with the manifold is a positive semi-definite differential operator, known as the Laplace–Beltrami operator. This operator has a discrete spectrum, with the sequence of eigenvalues ℓ_k^2 and corresponding eigenfunctions ϕ_k. (For simplicity of exposition, the notation used in the introduction is not necessarily the same as in the rest of the paper.) Projections of the data on the spaces $\Pi_N = \mathrm{span}\{\phi_k : \ell_k \le N\}$ yield important dimensionality reduction advantages, as well as information about the underlying geometric structures present in the data. In Maggioni and Mhaskar (2008), we have coined the term *diffusion polynomial* for a member of $\cup_{N\ge 0} \Pi_N$.

The heat kernel, defined formally by

$$K_t(x, y) = \sum_j \exp(-\ell_j^2 t)\phi_j(x)\overline{\phi_j(y)}, \quad x, y \in \mathbb{X},\ t > 0, \tag{1.1}$$

plays an important role in this theory. The approximation of the heat kernel, respectively the Laplace–Beltrami operator, and the eigenvalues and eigenfunctions from the data itself has been studied by many authors, including Belkin and Niyogi (2008a, 2008b), Lafon (2004) and Singer (2006). On the theoretical side, it is shown in Jones, Maggioni, and Schul (2010) that the heat kernel leads to a local coordinate system on the unknown manifold. In Coifman and Maggioni (2006), the authors develop a multi-resolution analysis based on powers of the operator defined by the heat kernel. Unfortunately, an explicit (theoretical) expression of the heat kernel is usually not known except in the simplest of cases. Therefore, Saito (2008) has proposed the use of other operators which have the same eigenfunctions as the heat kernel, but can be expressed in a closed form. In Saito (2008), the author has illustrated this concept in the context of image processing applications by considering the Green's function for certain boundary value problems in place of the heat kernel.

E-mail address: hmhaska@gmail.com.

0893-6080/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2010.12.007

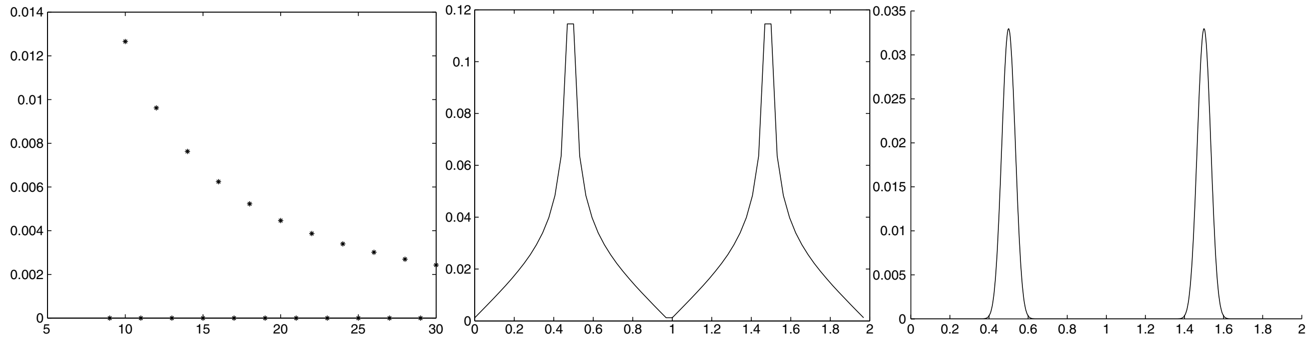

Fig. 1. For the function $f(x) = \sqrt{|\cos x|}$, the leftmost figure shows $|\hat{f}(k)|$, $0 \le k \le 32$, estimated by a 128 point FFT, the middle figure shows the coefficients based on a Haar wavelet based on 128 equidistant samples of f, the rightmost figure shows the graph of $|\sum_{|k|\le 32} h(k/32)\hat{f}(k)e^{ikx} - \sum_{|k|\le 16} h(k/16)\hat{f}(k)e^{ikx}|$, where h is a suitable high-pass filter (cf. Section 3). The x axis in the middle and rightmost figure is $[0, 2]$, representing multiples of π.

The objective of this paper is to consider a very general operator corresponding to a kernel of the form

$$G(x, y) = \sum_j b(\ell_j)\phi_j(x)\overline{\phi_j(y)}, \quad x, y \in \mathbb{X}, \tag{1.2}$$

where b is a function satisfying certain technical conditions to be described later. In particular, our theory includes the following examples: if $b(u) = \exp(-tu^2)$, then the kernel reduces to the heat kernel, and if $b(u) = (u^2 + 1)^{-\beta/2}$ for a sufficiently large $\beta > 0$, then we have the Green's kernel for a pseudo-differential operator. We develop a frame in terms of the powers of the operator corresponding to G, and prove that this frame leads to a sparse representation of functions on the manifold $\mathbb{X}$.

To motivate our ideas, we consider a very simple example from classical Fourier analysis. We start with 128 equidistant samples of the function $f(x) = \sqrt{|\cos x|}$ on $[0, 2\pi]$. The function admits an analytic continuation at every point except at the *singularities* $\pi/2, 3\pi/2$, where the function is not differentiable. The frequency domain description of f in terms of the Fourier coefficients $\hat{f}(k)$, $|k| \le 32$, is shown in the leftmost figure of Fig. 1. It is clear that they do not reveal any features of f, such as the singularities at $\pi/2, 3\pi/2$. In contrast, a simple-minded approximation to the coefficients of f in terms of the Haar wavelet basis is shown in the middle figure of Fig. 1. The discontinuities of f' at $\pi/2, 3\pi/2$ are very clear in terms of the coefficients with a large magnitude. In a reconstruction of f, only these high magnitude coefficients need to be retained, the lower coefficients may be neglected, depending upon the tolerance for error. Thus, a representation of f in terms of the wavelet basis is far more parsimonious than that in terms of the classical Fourier coefficients. A theoretical foundation for this phenomenon is given in terms of a characterization of local function spaces in terms of the wavelet coefficients in a greater generality, cf. (Daubechies, 1992, Theorems 9.2.1, 9.2.2). This sparse representation property is a major reason for the popularity of wavelets and their successors, ridgelets, curvelets, shearlets, etc.

Nevertheless, it is well-known that the sequence of Fourier coefficients contains all the information regarding the target function. In some applications, such as direction finding in phased array antennas (Krim & Viberg, 1996) and solutions of partial differential equations (Tadmor, 1989), it is necessary to find the locations of singularity in a function when the data is in the form of its Fourier coefficients. For this reason, many authors (e.g., Eckhoff, 1995; Gelb & Tadmor, 1999; Mhaskar & Prestin, 2000, 2005b; Tadmor, 1989; Tadmor & Tanner, 2005; Tanner, 2006) have studied methods to extract such information from the Fourier coefficients, as well as obtain more parsimonious representations of a function given the Fourier coefficients rather than wavelet coefficients. The ideas have been extended to contexts more general than the classical trigonometric series; a relatively recent survey is given in Mhaskar and Prestin (2005b). Coming back to our example, the rightmost figure in Fig. 1 shows the frame transform as analyzed in Mhaskar and Prestin (2005a) based on the same Fourier information shown in the leftmost figure. It is clear that our frame transform (coefficients in a discretized form) also detects singularities, and yields a parsimonious representation analogous to wavelets.

The ideas in Mhaskar and Prestin (2005b) were developed further in Maggioni and Mhaskar (2008) to obtain a Littlewood–Paley decomposition of functions on "arbitrary" quasi-metric measure spaces, a Riemannian manifold being a particular case. Let $\mathbb{X}$ be a quasi-metric measure space with the quasi-metric ρ (e.g., the geodesic distance on manifolds) and measure μ (e.g., the volume measure on manifolds). Given a sequence of numbers $\ell_j \uparrow \infty$ and a sequence of continuous functions ϕ_j, orthonormalized with respect to μ, we define the Fourier coefficients of an integrable function f on $\mathbb{X}$ by

$$\hat{f}(j) = \int_{\mathbb{X}} f(y)\overline{\phi_j(y)}d\mu(y), \quad j = 0, 1, \ldots.$$

With a suitably defined compactly supported filter g, we define the frame transform

$$\tau_n(f, x) = \sum_{j=0}^{\infty} g\left(\frac{\ell_j}{2^n}\right)\hat{f}(j)\phi_j(x), \quad n = 0, 1, \ldots, x \in \mathbb{X}. \tag{1.3}$$

We have shown in Maggioni and Mhaskar (2008) that any continuous f on $\mathbb{X}$ can be expressed as a uniformly convergent sum $f = \sum_{n=0}^{\infty} \tau_n(f)$, and the smoothness behavior of f is characterized by the norms of τ_n, analogous to the results in wavelet analysis, except that our characterization is for global function classes.

The objectives of this paper are the following:

- We wish to obtain a complete analogue of the aforementioned results in wavelet analysis for the characterization of *local* function spaces in terms of our frame transform.
- We wish to construct the frame transform using only linear combinations of iterates of the operator corresponding to the kernel (1.2).
  - In order to evaluate the transform as in (1.3) in practice, one needs to compute the eigenfunctions ϕ_j and the eigenvalues ℓ_j. While efficient algorithms exist to accomplish these computations, such a computation for a large number of eigenfunctions remains a bottleneck in the theory. Our construction avoids this bottleneck entirely; we no longer require a computation of either the eigenvalues or eigenfunctions.
- We wish to formulate our assumptions in terms of the heat kernel.