ARTICLE IN PRESS
YACHA:1180

Appl. Comput. Harmon. Anal. • • • (• • • •) • • •–• • •

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Letter to the Editor

Spectral echolocation via the wave embedding

Alexander Cloninger [a,*], Stefan Steinerberger [b]

[a] *Department of Mathematics, Program in Applied Mathematics, Yale University, New Haven, CT 06510, USA*
[b] *Department of Mathematics, Yale University, 06511 New Haven, CT, USA*

ARTICLE INFO

Article history:
Received 15 July 2016
Received in revised form 10 January 2017
Accepted 14 January 2017
Available online xxxx
Communicated by Charles K. Chui

Keywords:
Spectral embedding
Diffusion maps
Wave equation
Wave kernel
Preprocessing

ABSTRACT

Spectral embedding uses eigenfunctions of the discrete Laplacian on a weighted graph to obtain coordinates for an embedding of an abstract data set into Euclidean space. We propose a new pre-processing step of first using the eigenfunctions to simulate a low-frequency wave moving over the data and using both position as well as change in time of the wave to obtain a refined metric to which classical methods of dimensionality reduction can then applied. This is motivated by the behavior of waves, symmetries of the wave equation and the hunting technique of bats. It is shown to be effective in practice and also works for other partial differential equations – the method yields improved results even for the classical heat equation.

1. Introduction

Spectral embedding methods are based on analyzing Markov chains on a high-dimensional data set $\{x_i\}_{i=1}^n \subset \mathbb{R}^d$. There are a variety of different methods, see e.g. Belkin & Niyogi [1], Coifman & Lafon [2], Coifman & Maggioni [3], Donoho & Grimes [4], Roweis & Saul [6], Tenenbaum, de Silva & Langford [8], and Sahai, Speranzon & Banaszuk [9]. A canonical choice for the weights of the graph is declare that the probability p_{ij} to move from point x_j to x_i to be

$$p_{ij} = \frac{\exp\left(-\frac{1}{\varepsilon}\|x_i - x_j\|^2_{\ell^2(\mathbb{R}^d)}\right)}{\sum_{k=1}^n \exp\left(-\frac{1}{\varepsilon}\|x_k - x_j\|^2_{\ell^2(\mathbb{R}^d)}\right)},$$

where $\varepsilon > 0$ is a parameter that needs to be suitably chosen. This Markov chain can also be interpreted as a weighted graph that arises as the natural discretization of the underlying 'data-manifold'. Seminal results

* Corresponding author.
E-mail addresses: alexander.cloninger@yale.edu (A. Cloninger), stefan.steinerberger@yale.edu (S. Steinerberger).

http://dx.doi.org/10.1016/j.acha.2017.01.002
1063-5203/

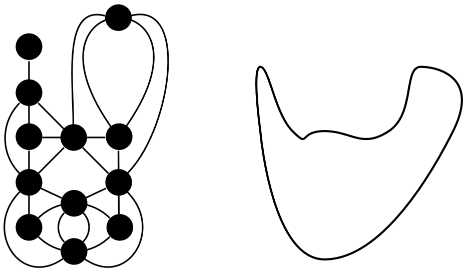

Fig. 1. Graphs that approximate smooth manifolds.

of Jones, Maggioni & Schul [5] justify considering the solutions of

$$-\Delta\phi_n = \lambda_n^2 \phi_n$$

as measuring the intrinsic geometry of the weighted graph. Here we always assume Neumann-boundary conditions whenever such a graph approximates a manifold (see Fig. 1).

The cornerstone of spectral embedding is the realization that the map

$$\begin{aligned} \Phi : \{x_i\}_{i=1}^n &\rightarrow \mathbb{R}^k \\ x &\rightarrow (\phi_1(x), \phi_2(x), \dots, \phi_k(x)), \end{aligned}$$

can be used as an effective way of reducing the dimensionality. One useful explanation that is often given is to observe that the Feynman–Kac formula establishes a link between random walks on the weighted graph and the evolution of the heat equation. We observe that random walks have a tendency to be trapped in clusters and are unlikely to cross over bottlenecks and, simultaneously, that the evolution of the heat equation can be explicitly given as

$$\left[e^{t\Delta} f\right](x) = \sum_{n=1}^{\infty} e^{-\lambda_n^2 t} \left\langle f, \phi_n \right\rangle \phi_n(x).$$

The exponential decay $e^{-\lambda_n^2 t}$ implies that the long-time dynamics is really governed by the low-lying eigenfunctions who then have to be able to somehow reconstruct the random walks' inclination for getting trapped in clusters and should thus be able to reconstruct the cluster. We believe this intuition to be useful and our further exposition will be based on this.

2. The wave equation

2.1. Introduction

Once the eigenfunctions of the Laplacian have been understood, they imply complete control over the Cauchy problem for the wave equation

$$\begin{aligned} (\partial_t^2 - \Delta) u(x,t) &= 0 \\ u(x,0) &= f(x) \\ \partial_t u(x,0) &= g(x) \end{aligned}$$

given by the eigenfunction expansion

Please cite this article in press as: A. Cloninger, S. Steinerberger, Spectral echolocation via the wave embedding, Appl. Comput. Harmon. Anal. (2017), http://dx.doi.org/10.1016/j.acha.2017.01.002