Available online at www.sciencedirect.com

SCIENCE DIRECT®

Appl. Comput. Harmon. Anal. 21 (2006) 1–2

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Preface

Special issue on diffusion maps and wavelets

The journal *Applied and Computational Harmonic Analysis* (ACHA) is committed to chronicling important advances in applied harmonic analysis as they occur, and particularly to highlighting research showing the way to broaden new applications and important new theoretical questions. To do so, we occasionally publish special issues compiling some of the central papers driving these new research directions.

This special issue on diffusion maps and wavelets (SIDW) introduces an innovative research direction connecting computational harmonic analysis to a broad range of issues facing those who must exploit high-dimensional data. Such data dominate intellectual activity in fields as diverse as genomics and web searches, and the new tools and viewpoints introduced in this issue are likely to open our field to an extremely broad range of novel applications.

The SIDW has been prepared under the leadership of Editor-in-Chief R.R. Coifman. All papers published in this special issue were rigorously refereed under the same strict journal guidelines and policy that apply to regular issues. In fact, the reviewing process took longer than usual because launching a new research direction places greater demands on readers and referees than does extending an established line of research.

This issue consists of five regular papers and two shorter notes that are published as "Letters to the Editor" with R.R. Coifman as the corresponding editor. This compilation introduces the notion of "diffusion maps" and "geometric harmonics" along with the associated concepts of "diffusion wavelets" and "diffusion wavelet packets," as well as some of the mathematical tools for applications, such as spectral clustering, reaction coordinates, and component analysis.

Motivating this research direction is the idea that high-dimensional data are often viewed by mathematicians as living "near" embedded submanifolds of high-dimensional Euclidean space; however, real data exhibit structures that are significantly more complicated. In general, if there is any geometric structure in high-dimensional data, it can be expected to be highly variable and intermittent in nature, requiring a whole new vocabulary and descriptive toolkit. The work of Peter Jones, followed by that of David and Semmes, pointed the way to tools for finding multiscale structure in high-dimensional data, allowing for data that seem k-dimensional in one region of space and at one scale, while they may seem k'-dimensional at a finer scale ($k' > k$) and k''-dimensional in another part of the data space.

Diffusion maps and diffusion wavelets provide a significant new tool for describing complicated architecture of high-dimensional data. They conveniently reshape many of the ideas of multiscale analysis that have proven so useful in study of one-dimensional signals and two-dimensional images, making natural analogs of these ideas available to describe the complex nature of high-dimensional phenomena.

The first paper, "Diffusion Maps," lays the foundation of this new research direction. The authors, R.R. Coifman and S. Lafon, provide a framework based on certain diffusion process for finding meaningful geometric descriptions of data sets, by showing that eigenfunctions of the Markov matrix corresponding to the kernel of the diffusion process provide coordinates, called diffusion maps, for generating efficient representations of complex geometric structures.

In a continuation paper, titled "Geometric Harmonics: A Novel Tool for Multiscale Out-of-Sample Extension of Empirical Functions" and by the same authors, R.R. Coifman and S. Lafon extend empirical functions on a data set to a larger set by an analog of the Nyström method; the extension process involves construction of functions, called geometric harmonics, that generalize the Slepian prolate spheroidal wave functions of classical one-dimensional signal analysis. The geometric harmonics are optimally concentrated on the extended domain. This approach allows a novel multiscale representation scheme for empirical data coordinates.

1063-5203/$ – see front matter
doi:10.1016/j.acha.2006.05.005

Because diffusion maps give rise to coarse-scale data representations, it is natural and important to introduce and study the corresponding wavelets. In the third paper, "Diffusion Wavelets" by R.R. Coifman and M. Maggioni, diffusion wavelets are studied based on dyadic powers of the diffusion operators, using ideas related to the fast multipole methods as well as to the wavelet analysis of the Calderon–Zygmund and pseudo-differential operators.

The fourth paper, titled "Diffusion Wavelet Packets" by J. Bremer, R.R. Coifman, M. Maggioni, and A. Szlam, extends the Coifman and Maggioni diffusion wavelet paper to construct wavelet packets, particularly on graphs and manifolds, including a discussion of certain "best basis algorithm" and several demonstrative examples.

The fifth paper in this special issue is "Diffusion Maps, Spectral Clustering and the Reaction Coordinates of Dynamic Systems" by B. Nadler, S. Lafon, and R.R. Coifman. In this paper, a viewpoint from stochastic differential equations is adopted, allowing a formal study of diffusion processes defined on high-dimensional measures. Here is a simple example: Suppose a dataset is sampled from a high-dimensional normal distribution. We can view that normal distribution as the underlying measure. So study of eigenfunctions of the continuous-space diffusion is in some sense analogous to studies of the diffusions on point clouds in earlier papers. These eigenfunctions can be computed explicitly using known properties of diffusion operators, and the resulting diffusion coordinates can be interpreted naturally as providing the principal components representation of the normally distributed data. Of course, the paper covers much other territory and can be viewed as offering a unified point of view for the study of problems ranging from "dimension reduction" to "large-scale simulation of complex dynamic systems." Classical topics such as backward Fokker–Planck operator with an underlying potential and eigenanalysis of the Laplace–Beltrami operator of the manifold appear; thus, the new tools of diffusion geometry of data are connected fairly much to classical analysis.

The two letters to the editor are "From Graph to Manifold Laplacian: The Convergence Rate" and "Spectral Independent Component Analysis," both contributed by A. Singer. The first paper is devoted to a discussion of an empirical method that improves the convergence rate in the literature when approximating the Laplace–Beltrami operator. The second paper is concerned with the independent component analysis that is relevant to the topic of diffusion geometry. In particular, it is shown that the methodology for local statistical data analysis automatically reveals the independent components, and that the natural coordinates are precisely the diffusion coordinates.

We thank the authors for submitting their pioneering work in this new research direction for publication in this special issue, and the referees for their conscientious reviews that help maintain the high quality of ACHA. We are truly grateful for the enormous time and effort of these referees in preparing their reviews; as mentioned above, reading through research papers in a new research direction and critiquing them requires tremendous time and efforts. Special thanks are due to R.R. Coifman for suggesting the publication of this important issue.

Charles Chui
David Donoho
Guest Editors
Available online 9 June 2006