Available online at www.sciencedirect.com
SCIENCE DIRECT®
Appl. Comput. Harmon. Anal. 21 (2006) 3
Applied and Computational Harmonic Analysis
www.elsevier.com/locate/acha

Foreword

Special issue on diffusion maps

The goal of our research program on harmonic analysis of data sets in high dimensions, which originated in a "scouting/fishing expedition" that A. Patera and I have been pursuing for DARPA over the last seven years, was to identify new successful mathematical ideas in various scientific and information processing fields. We were looking for tools that would reduce the enormous computational challenges confronting the scientist and engineer in processing and extracting information from massive data sets. This search, conducted through the Defense Science Research Council, involved various workshops on efficient computations and dimensional reduction and various interviews with people from all walks of science: from weather prediction to geology, to data mining, to machine learning, to web searches, and to bio-informatics. It was quite clear that ideas from manifold learning, eigenmaps, and spectral graph theory, as well as ideas involving hierarchical multiscale organizations, are effective approaches.

Our goal was to try to isolate productive common currents to enable cross-pollination between fields. In the context of classical harmonic analysis, these ideas are part of a well-understood theory, relating wavelets to Fourier analysis, in which the "geometry of linear transformations" is directly linked to its spectral properties. In fact, it was a profound insight of A. Zygmund and his students (ca. 1955) that real variable methods, such as multiscale hierarchical organization (for "book keeping" and "folder building"), are directly connected to the properties of Fourier expansions (eigenfunctions) and their localizations. This insight led to the Calderon–Zygmund theory and its generalizations to various abstract settings (see [1,2]).

As a result, we were quite intrigued by the possibility of providing a simple coherent synthetic view to explain the success of the first few eigenfunctions of an operator linking data points, in providing a geometric organization of that data. Working with many collaborators, some of whom contributed to this special issue, we have seen a collection of tools emerge around diffusion (or inference) geometries, which, like differential calculus, can be used to build global inference relations between objects by combining "infinitesimal" (linear) models. It has also become clear that multiscale folder building, or a wavelet paradigm, is a powerful tool for the functional regression and analysis of data.

As this special issue is published, we see applications emerging in sensor fusion, medical diagnostics, machine learning, and other areas. Generally we formalize aspects of the well-known idea that knowledge can be encapsulated by networks of inferences, grouped into folders of related links.

I thank my colleagues Peter Jones and Vladimir Rokhlin for their profound contributions to this effort. While their names are not on the various papers, their ideas are in them. I also acknowledge many fruitful discussions with David Donoho and Yves Meyer concerning everything mentioned here.

We express our gratitude to DARPA and DSRC as well as AFOSR and NSF for their support of this project.

To conclude, we are grateful to Charles Chui for his help and encouragement in publishing this special issue of ACHA.

References

[1] E. Stein, Topics in Harmonic Analysis Related to the Littlewood–Paley Theory, Princeton Univ. Press, Princeton, NJ, 1970.
[2] R. Coifman, G. Weiss, Analyse harmonique noncommutative sur certains espaces homogenes, Springer-Verlag, 1971.

Ronald R. Coifman
Program in Applied Mathematics, Yale University
New Haven, CT, USA

1063-5203/$ – see front matter
doi:10.1016/j.acha.2006.05.011