Local Geometric Spectral Data Analysis

Gal Mishne, Ph.D.
Gibbs Assistant Professor in Applied Mathematics, Yale

January 7, 2019 • 11 am – 12 pm

M416 (ESAM Conference Room), Tech

Modern technological developments have enabled the acquisition and storage of increasingly large-scale, high-resolution, and high-dimensional data in many fields. Yet in domains such as biomedical data, the complexity of these datasets and the unavailability of ground truth pose significant challenges for data analysis and modeling. In this talk, I present new unsupervised geometric approaches for extracting structure from large-scale high-dimensional data. By looking deep within the spectrum of the graph-Laplacian, we define a new robust measure, the Spectral Embedding Norm, to separate clusters from background, and demonstrate its application to both outlier detection and data visualization. This measure further motivates a new greedy clustering approach based on Local Spectral Viewpoints for identifying high-dimensional overlapping clusters while disregarding noisy clutter. We demonstrate our approach on two-photon calcium imaging data, successfully extracting hundreds of individual cells. Finally, to address the computational complexity of applying spectral approaches to large-scale data, we present a new randomized near-neighbor graph construction. Compared to the traditional k nearest neighbor graph, using our near-neighbor graph for spectral clustering on datasets of a few million points is two orders of magnitude faster, while achieving similar clustering accuracy.

Joint work with Ronald Coifman, Jackie Schiller, Maria Lavzin, Xiuyuan Cheng, George Linderman, Ariel Jaffe, Yuval Kluger and Stefan Steinerberger.