



THE WEIZMANN INSTITUTE OF SCIENCE
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

Machine Learning and Statistics Seminar

Room 1 ,Ziskind Building
on Wednesday, Nov 29, 2017
at 11:15

Roy Lederman
Princeton University

Inverse Problems and Unsupervised Learning with applications to Cryo-Electron
Microscopy (cryo-EM)

Abstract:

Cryo-EM is an imaging technology that is revolutionizing structural biology; the Nobel Prize in Chemistry 2017 was recently awarded to Jacques Dubochet, Joachim Frank and Richard Henderson "for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution".

Cryo-electron microscopes produce a large number of very noisy two-dimensional projection images of individual frozen molecules. Unlike related methods, such as computed tomography (CT), the viewing direction of each image is unknown. The unknown directions, together with extreme levels of noise and additional technical factors, make the determination of the structure of molecules challenging.

While other methods for structure determination, such as x-ray crystallography and nuclear magnetic resonance (NMR), measure ensembles of molecules together, cryo-EM produces measurements of individual molecules. Therefore, cryo-EM could potentially be used to study mixtures of different conformations of molecules. Indeed, current algorithms have been very successful at analyzing homogeneous samples, and can recover some distinct conformations mixed in solutions, but, the determination of multiple conformations, and in particular, continuums of similar conformations (continuous heterogeneity), remains one of the open problems in cryo-EM.

I will discuss a one-dimensional discrete model problem, Heterogeneous Multireference Alignment, which captures many of the group properties and other mathematical properties of the cryo-EM problem. I will then discuss different components which we are introducing in order to address the problem of continuous heterogeneity in cryo-EM: 1. "hyper-molecules", the first mathematical formulation of truly continuously heterogeneous molecules, 2. The optimal representation of objects that are highly concentrated in both the spatial domain and the frequency domain using high-dimensional prolate spheroidal functions, and 3. Bayesian algorithms for inverse problems with an unsupervised-learning component for recovering such hyper-molecules in cryo-EM.