Efficient Algorithms for Distributional Problems via Sum-of-Squares Proofs

Pravesh Kothari, Institute for Advanced Study, Princeton

Recovering a hidden signal or structure in the presence of random noise is a recurring theme in fundamental problems arising in signal processing, computational complexity, cryptography, machine learning, and statistics. Influential works in the past decade have led to a fine-grained understanding of some fundamental problems of this flavor such as compressive sensing and matrix completion. Such a detailed understanding, however, has not been possible for other fundamental problems such as clustering. This is in part because coming up with and analyzing the guarantees of the "right" convex relaxation usually relies on sophisticated, creative ``leaps" that are problem-specific, such as guessing and analyzing a clever ``dual" solution.

The goal of this talk is to illustrate the ``sum-of-squares method": a recently developed approach that gives a natural and intuitive ``blueprint" for both coming up with and analyzing the ``right" convex relaxation for a number of distributional problems in unsupervised machine learning, computational complexity, cryptography, quantum information and beyond.

Key to this paradigm is a surprising reduction of efficient algorithm design to finding "simple" (in a precise semialgebraic proof system) proofs of ``identifiability" - a proof that uses the given data to _verify_ that a purported solution is indeed the correct one.

In this talk, I'll illustrate the SoS method for parameter estimation by focusing on the example of learning a mixture of spherical Gaussians with information theoretically optimal cluster-separation in quasi-polynomial time. This improves on a classical result of Vempala and Wang (2002). No sub-exponential time algorithm was previously known in this regime.