Probabilistic techniques and algorithms

Workshop Program as of April 3, 2012

Friday, April 6, 2012

8.50 - 9.00 Opening
9.00 - 9.25 Adam Klivans
   Learning Convex Sets
   Algorithms for learning convex sets are some of the most important tools in machine learning. We will survey some recent developments giving the first *provably* efficient algorithms for learning convex sets with respect to various probability distributions (e.g., Gaussian or Log-Concave).

9.30 - 9.55 Mark Iwen
   A Symbol-based Bar Code Decoding Algorithm
   We investigate the problem of decoding bar codes from a signal measured with a hand-held laser-based scanner. We develop a sparse representation of the UPC bar code and relate it to the measured signal. A greedy symbol-based algorithm for decoding a bar code is proposed. We show that the algorithm is robust to noise, and to unknown parameters in the scanning device. Numerical examples further illustrate the behavior of the method.
   Joint work with Rachel Ward and Fadil Santosa

10.00 - 10.25 Lorenzo Rosasco
   Unsupervised learning and regularization

11.00 - 11.25 Pablo Parillo
   tba

11.30 - 11.55 Michael Mahoney
   Approximate computation and implicit regularization in large-scale data analysis
   Traditional methods of worst-case analysis often fail to provide even qualitative guidance as to what algorithms are appropriate in many large-scale data analysis applications. This fact is usually blamed on the exclusive focus on running time for arbitrary input. A deeper reason is that most problems arising in modern data applications are intractable and since the relaxations and embeddings underlying most worst-case approximation algorithms are not "close enough" to the data to provide such qualitative guidance. Recent empirical work on very large social and information networks, however, has clearly demonstrated that by exploiting the geometric properties associated with these relaxations and embeddings, one can extract insight and perform inference from large, sparse, and noisy graphs in a scalable and robust way. Performing such inference is typically the domain of a statistical technique known as regularization, which usually involves modifying the original objective with a geometric constraint and then exactly optimizing the modified objective. Thus, these results demonstrate a manner in which approximate computation can implicitly lead to statistical regularization, i.e., in which the solutions from worst-case approximation algorithms are not only faster but also "better" for downstream applications than the solutions to their intractable counterparts. This will be illustrated in the context of using spectral versus flow-based approximation algorithms for the graph partitioning problem; and it will be explained in terms of how the geometric embeddings underlying spectral methods versus flow methods provide a form of capacity control to "smooth" or "regularize" the input graph in very different ways. The implications of these results for worst-case approximation algorithms more generally will be discussed.
12.00 - 12.25 Piotr Indyk  
*Faster Algorithms for Sparse Fourier Transform*  
The Fast Fourier Transform (FFT) is one of the most fundamental numerical algorithms. It computes the Discrete Fourier Transform (DFT) of an n-dimensional signal in $O(n \log n)$ time. The algorithm plays a key role in many areas. In many applications (e.g., audio, image or video compression), most of the Fourier coefficients of a signal are "small" or equal to zero, i.e., the output of the transform is (approximately) sparse. In this case, there are algorithms that enable computing the non-zero coefficients faster than the FFT. However, in practice, the exponents in the runtime of these algorithms and their complex structure have limited their applicability to only very sparse signals. In this talk, I will describe a new set of algorithms for sparse Fourier Transform. Their key feature is simplicity, which leads to efficient running time with low overhead, both in theory and in practice. In particular, one of the algorithms achieves a runtime of $O(k \log n)$, where $k$ is the number of non-zero Fourier coefficients of the signal. This improves over the runtime of the FFT for any $k = o(n)$.
Joint work with Haitham Hassanieh, Dina Katabi and Eric Price.

Lunch  
2.30 - 2.55 Joel Tropp  
*How to Find a Needle in a Haystack*  
Imagine that you are given 20,010 points unlabeled points in a 10,000-dimensional Euclidean space. Ten points fall on a line through the origin, while the remaining 20,000 points are distractors that have no linear structure. This talk explains how to find the direction of the "needle" exactly, in spite of the fact it is buried in a "haystack." The same method can also find sheets of paper, Rubik’s cubes, and other low-dimensional objects.

3.00 - 3.25 Rayan Saab  
*Sigma-Delta quantization of Gaussian frame expansions: root-exponential accuracy*  
It has been recently shown that by quantizing the $N$-dimensional Harmonic frame coefficients of signals in $\mathbb{R}^d$ using Sigma-Delta quantization schemes, it is possible to achieve root-exponential accuracy in the oversampling rate $\lambda := N/d$ (even when one bit per measurement is used). Similarly, root-exponential accuracy is attainable when using the so-called Sobolev self-dual frames. In both of these cases, the structure of the frame plays a crucial role in establishing the result. We show a somewhat surprising result whereby random Gaussian frames achieve similar, albeit slightly weaker guarantees (with high probability).

3.30 - 3.55 Raj Rao Nadakuditi  
*Some surprises in the estimation of low rank, matrix-valued random variables*  
Principal component analysis or equivalently the truncated singular value decomposition is widely used for the estimation of low-rank signal matrices buried in noise. The Eckart-Young-Mirsky (EYM) theorem seemingly blesses this choice. We revisit low-rank signal matrix estimation in noise using the lens of random matrix theory and uncover two surprises: 1) the EYM solution is provably sub-optimal and 2) there are many (practically motivated) settings where middle components might be more informative than principal components.

Coffee Break
4.30 - 4.55 Ben Recht

System Identification with Atomic Norm Regularization

Building accurate models of dynamical systems from noisy observation of their input-output behavior is of fundamental importance to data analysis in fields as diverse as process engineering, aerospace design, and systems biology. Yet after fifty years of fruitful research in system identification, there is still no agreed upon method that balances statistical consistency, algorithmic simplicity, and dynamical accuracy. This talk attempts to strike such a balance by uniting tools from operator theory, systems theory, and high-dimensional statistics.

I will propose a new algorithm for identifying linear dynamical systems by regularizing prediction error with an atomic norm induced by the set of single pole filters. I will show that the associated atomic norm is equivalent to the nuclear norm of the Hankel operator of the dynamical system and can be approximated efficiently on large-scale problems. For several different measurement models, I will provide both statistical guarantees and numerical experiments demonstrating the efficacy of this new method.

Joint work with Badri Narayan Bhaskar, Gongguo Tang, and Parikshit Shah.

5.00 - 5.25 Maryam Fazel

The Convex Geometry of Linear Inverse Problems

Deducing the state or structure of a system from partial, noisy measurements is a fundamental task throughout the sciences and engineering. The resulting inverse problems are often ill-posed because there are fewer measurements available than the ambient dimension of the model to be estimated. In practice, however, many interesting signals or models contain few degrees of freedom relative to their ambient dimension: a small number of genes may constitute the signature of a disease, very few parameters may specify the correlation structure of a time series, or a sparse collection of geometric constraints may determine a molecular configuration. Discovering, leveraging, or recognizing such low-dimensional structure plays an important role in making inverse problems well-posed. Examples of structured models include previously studied cases such as sparse signals and low-rank matrices, as well as others such as low-rank tensors, binary vectors, orthogonal matrices, and matrices formed as the sum of a few permutations. Inspired by the success of the L1-norm and nuclear-norm heuristics, we propose a general convex relaxation framework to recover such simple structured models from partial information. We provide sharp estimates of the number of generic measurements required for exact and robust recovery in a variety of settings. These estimates are based on computing certain Gaussian statistics related to the underlying model geometry. Thus our work extends the catalog of objects and structures (beyond sparse vectors and low-rank matrices) that can be recovered from limited information via tractable convex programming.

Joint work with Benjamin Recht, Pablo Parrilo, and Alan Willsky.

Workshop dinner
Saturday, April 7, 2012

9.00 - 9.25 Yaniv Plan

*Structured signal recovery from single-bit measurements*

1-bit compressed sensing was introduced by Boufounos and Baraniuk to model extreme quantization in compressed sensing; in this model each measurement consists of a single bit. Signal recovery from 1-bit measurements poses an elegant mathematical problem, and it has practical applications in analog-to-digital conversion and bit-constrained compressed sensing amid high noise. We review recent results which show that a simple convex program may be used to estimate the signal. In particular, an s-sparse signal in $\mathbb{R}^n$ may be accurately recovered from $O(s \log(n/s))$ single-bit measurements. This remains true when nearly half of the measurements are randomly flipped. We describe the connection to related statistical models including sparse logistic regression. Other signal structures aside from sparsity can be also easily incorporated.

9.30 - 9.55 Simon Foucart

*Stability and robustness of $\ell_1$-minimizations with Weibull matrices and redundant dictionaries*

We investigate the recovery of almost $s$-sparse vectors $x \in \mathbb{C}^N$ from undersampled and inaccurate data $y = Ax + e \in \mathbb{C}^m$ by means of minimizing $\|z\|_1$ subject to the equality constraints $Az = y$. If $m \gtrsim s \ln(N/s)$ and if Gaussian random matrices $A \in \mathbb{R}^{m \times N}$ are used, this equality-constrained $\ell_1$-minimization is known to be stable with respect to sparsity defects and robust with respect to measurement errors. If $m \gtrsim s \ln(N/s)$ and if Weibull random matrices are used, we prove here that the equality-constrained $\ell_1$-minimization remains stable and robust. The arguments are based on two key ingredients, namely the robust null space property and the quotient property. The robust null space property relies on a variant of the classical restricted isometry property where the inner norm is replaced by the $\ell_1$-norm and the outer norm is replaced by a norm comparable to the $\ell_2$-norm. For the $\ell_1$-minimization subject to inequality constraints, this yields stability and robustness results that are also valid when considering sparsity relative to a redundant dictionary. As for the quotient property, it relies on lower estimates for the tail probability of sums of independent Weibull random variables.

10.00 - 10.25 Thomas Strohmer

*Phase Retrieval, Random Matrices, and Convex Optimization*

Phase retrieval is the problem of reconstructing a function, such as a signal or image, from intensity measurements, typically from the modulus of the diffracted wave. Phase retrieval problems - which arise in numerous areas including X-ray crystallography, astronomy, diffraction imaging, and quantum physics, are notoriously difficult to solve numerically. I will introduce a novel framework for phase retrieval, which comprises tools from optimization, random matrix theory, and compressive sensing. I will prove that for certain types of random measurements a signal or image can be recovered exactly with high probability by solving a convenient semidefinite program - a trace-norm minimization problem, without any assumption about the signal whatsoever and under a mild condition on the number of measurements. Our method is also provably stable vis-a-vis noise. If time permits, I will then discuss how this approach carries over to the classical phase retrieval setting using multiple structured illuminations.

*Coffee Break*
11.00 - 1.25 Nir Ailon

Efficient Adaptive Querying Strategies for Clustering and Ordering Problems

I will discuss two learning problems for which iid sampling techniques, as well as known adaptive sampling techniques developed for general "active learning" problems are grossly suboptimal. The two problems are k-Correlation Clustering, and Minimum Feedback Arc-Set in Tournaments. In both problems the instance space is the set of pairs of elements over a ground set V of n elements. In the first, the objective is to learn a partition of V into k disjoint sets, in other words, to learn a graph containing k disjoint cliques covering V. In the second, the objective is to learn an ordering of V, in other words, a transitive tournament graph. Our setting is agnostic in the sense that the observations are noisy, and we compete with the best fit. Our sampling strategy is iterative, producing an improved solution at each iteration. Each "current" solution defines a distribution, from which we sample a batch of (noisy) edges. The underlying empirical process allows uniform estimation of the difference between the cost of any solution and the current one to within an error of epsilon times the distance between the solutions, for any epsilon. Minimizing this empirical process defines the next solution, and allows a fast learning rate.

I will describe this strategy and show some connections with the theory of matrix completion.

Joint work with Esther Ezra and Ron Begleiter.

11.30 - 11.55 Luis Rademacher

Recent developments in column subset selection and volume sampling

I will discuss recent algorithmic developments for the classical problem of approximating a given matrix by a low-rank matrix. This is motivated by certain applications that want the approximating matrix to have rows living in the span of only a few columns of the original matrix, which adds a combinatorial twist to the problem. The novel algorithms are based on sampling columns randomly (but non-uniformly) and random projection, from which a low rank approximation can be computed.

12.00 - 12.25 Nikhil Srivastava

Lunch

2.30 - 2.55 Mauro Maggioni

Multiscale geometric methods for noisy point clouds in high dimensions

We discuss techniques for analyzing at different scales the geometry of intrinsically low-dimensional point clouds perturbed by high-dimensional noise. We first show how such techniques may be used to estimate the intrinsic dimension, approximate tangent planes, and certain stable notions of curvatures of data sets. We then introduce a novel geometric multiscale transform, based on what we call geometric multi-resolution analysis, that leads to efficient approximation schemes for point clouds, as well as new dictionary learning methods and density estimation for data sets. Applications to dictionaries for images, anomaly detection hyperspectral images, and compressive sampling will be discussed.

3.00 - 3.25 Laurent Demanet

Lunch
3.30 - 3.55 Deanna Needell

Robust image recovery via total variation minimization

Discrete images, composed of patches of slowly-varying pixel values, have sparse or compressible wavelet representations which allow the techniques from compressed sensing such as L1-minimization to be utilized. In addition, such images also have sparse or compressible discrete derivatives which motivate the use of total variation minimization for image reconstruction. Although image compression is a primary motivation for compressed sensing, stability results for total-variation minimization do not follow directly from the standard theory. In this talk, we present numerical studies showing the benefits of total variation approaches and provable near-optimal reconstruction guarantees for total-variation minimization using properties of the bivariate Haar transform.

This is joint work with Rachel Ward.

Coffee Break

4.30 - 4.55 Sujay Sanghavi

Learning Stochastic Dynamical Systems with Latent Variables

We consider the problem of learning, from samples, the dependency structure of a system of linear stochastic differential equations, when some of the variables are latent. In particular, we observe the time evolution of some variables, and never observe other variables; from this, we would like to find the dependency structure between the observed variables – separating out the spurious interactions caused by the (marginalizing out of the) latent variables’ time series.

We develop a new method, based on convex optimization, to do so in the case when the number of latent variables is smaller than the number of observed ones. For the case when the dependency structure between the observed variables is sparse, we theoretically establish a high-dimensional scaling result for structure recovery. We verify our theoretical result with both synthetic and real data (from the stock market).

5.00 - 5.25 Massimo Fornasier

Learning functions of few arbitrary linear parameters in high dimensions

In this talk we summarize the results of our recent work [1, 2]. Let us assume that \( f \) is a continuous function defined on a convex body in \( \mathbb{R}^d \), of the form \( f(x) = g(Ax) \), where \( A \) is a \( k \times d \) matrix and \( g \) is a function of \( k \) variables for \( k \ll d \). Using only a limited number of point evaluations \( f(x_i) \), we would like to construct a uniform approximation of \( f \). Under certain smoothness and variation assumptions on the function \( g \), and an arbitrary choice of the matrix \( A \), we present a randomized algorithm, where the sampling points \( \{x_i\} \) are drawn at random and which recovers a uniform approximation of \( f \) with high probability. Under the arbitrariness of \( A \), the choice of the sampling points will be according to suitable random distributions and our results hold with overwhelming probability. Our approach uses tools taken from the compressed sensing framework, recent Chernoff bounds for sums of positive-semidefinite matrices, and classical stability bounds for invariant subspaces of singular value decompositions. This is a joint work with Karin Schnass and Jan Vybiral.


5.30 - 5.40 Closing