Sparse Recovery, Learning, and Neural Networks
Récupération clairsemée, apprentissage et réseaux neuronaux
(Org: **Ben Adcock** (Simon Fraser University), **Simone Brugiapaglia** (Simon Fraser University), **Yaniv Plan** (University of British Columbia) and/et **Ozgur Yilmaz** (University of British Columbia))

CHARLES DELAHUNT, University of Washington, Seattle
Biological mechanisms for learning: experiments and mathematical questions

Biological neural systems excel at fast and adaptable learning. Our experiments with computational models of the moth olfactory network suggest this learning depends on a triad of sparsity, Hebbian plasticity, and neuromodulatory signals. We also find that a bug brain can outperform machine learning methods in certain rapid-learning regimes. So given sufficient mathematical characterization, these biological mechanisms could yield high-value tools applicable to machine learning and especially neural nets. We'll describe some experimental results, and point to some current gaps in the mathematical analysis of these mechanisms.

NICK DEXTER, Simon Fraser University
Energy norm regularized sparse simultaneous reconstruction of solutions to parameterized PDEs

We present and analyze a novel sparse polynomial approximation method for the solution of PDEs with stochastic and parametric inputs. Our approach treats the parameterized problem as a problem of sparse signal reconstruction through a reformulation of the standard basis pursuit denoising problem. To achieve global reconstruction of solutions to the parameterized elliptic PDE, we employ a novel mixed ℓ_1 and energy norm-based method of regularization. Combined with the standard measurement scheme developed for compressed sensing-based sparse polynomial recovery, this approach allows for simultaneous global approximations of the solution over both physical and parametric domains. In addition, we are able to show that, with minimal sample complexity, error estimates comparable to the best s-term approximation are achievable, while requiring only *a priori* bounds on polynomial truncation error in energy norms. While the current work focuses on sparse approximation of solutions to parameterized elliptic PDE models, we note that the developments herein may be readily applied to any parameterized systems satisfying a sparsity assumption. Finally, we perform extensive numerical experiments on several high-dimensional parameterized elliptic PDE models to demonstrate the superior recovery properties of the proposed approach.

SINAN GUNTURK, Courant Institute, NYU
Bounded total variation of Kaczmarz trajectories for arbitrary control sequences

The Kaczmarz method is a popular iterative algorithm for solving linear systems of equations where the next iterate is determined by projecting the current iterate onto the solution hyperplane associated with any one of the individual equations. The sequence of the chosen hyperplanes is called the control sequence. For consistent systems, this iteration is known to always converge, and in particular, to a point in the intersection of the hyperplanes that are visited infinitely often. Linear convergence rates can be achieved with control sequences that are sufficiently regular, but for arbitrary control sequences the convergence can also be arbitrarily slow. In this talk (in joint work with Thao Nguyen), we show that regardless of the control sequence, the convergence takes place in a stronger sense in that the trajectories always have bounded variation (this is sometimes called absolute convergence), and their total variation can be bounded uniformly over all control sequences. Our result also holds for a more general version of the Kaczmarz method that incorporates relaxation.

XIAOWEI LI, University of British Columbia
Bernstein's Inequality and Sharp Bounds for Concentration of Euclidean Norm

We present a new Bernstein's inequality for sum of mean-zero independent sub-exponential random variables with absolutely bounded first absolute moment. We use this to prove a tight concentration bound for the Euclidean norm of sub-gaussian

random vectors. Then we apply the result to sub-gaussian random matrices on geometric sets, where the bounded first absolute moment condition comes naturally from the isotropic condition of random matrices. As an application, we discuss the implications for dimensionality reduction and Johnson-Lindenstrauss transforms. Lastly, we will talk about the possibility of extending this new Bernstein's inequality to second order chaos (Hanson-Wright inequality).

CHRIS LIAW, University of British Columbia
Nearly-Tight Sample Complexity Bounds for Learning Mixtures of Gaussians

Estimating distributions from observed data is a fundamental task in statistics that has been studied for over a century. We consider such a problem where the distribution is a mixture of k Gaussians in $\mathbb{R}^d$. The objective is density estimation: given i.i.d. samples from the (unknown) distribution, produce a distribution whose total variation distance from the unknown distribution is at most ϵ.

We prove that $\tilde{\Theta}(kd^2/\epsilon^2)$ samples are necessary and sufficient for this task, suppressing logarithmic factors. This improves both the known upper bound and lower bound for this problem.

SHUYANG LING, New York University
Certifying Global Optimality of Graph Cuts via Semidefinite Relaxation: A Performance Guarantee for Spectral Clustering

Spectral clustering has become one of the most widely used clustering techniques when the structure of the individual clusters is non-convex or highly anisotropic. Yet, despite its immense popularity, there exists fairly little theory about performance guarantees for spectral clustering. This issue is partly due to the fact that spectral clustering typically involves two steps which complicated its theoretical analysis: first, the eigenvectors of the associated graph Laplacian are used to embed the dataset, and second, the k-means clustering algorithm is applied to the embedded dataset to get the labels. This paper is devoted to the theoretical foundations of spectral clustering and graph cuts. We consider a convex relaxation of graph cuts, namely ratio cuts and normalized cuts, that makes the usual two-step approach of spectral clustering obsolete and at the same time gives rise to a rigorous theoretical analysis of graph cuts and spectral clustering. We derive deterministic bounds for successful spectral clustering via a spectral proximity condition that naturally depends on the algebraic connectivity of each cluster and the inter-cluster connectivity. Moreover, we demonstrate by means of some popular examples that our bounds can achieve near-optimality. Our findings are also fundamental to the theoretical understanding of kernel k-means.

DUSTIN G. MIXON, The Ohio State University
SqueezeFit: Label-aware dimensionality reduction by semidefinite programming

Given labeled points in a high-dimensional vector space, we seek a low-dimensional subspace such that that projecting onto this subspace maintains some prescribed distance between points of differing labels. Intended applications include compressive classification. This talk will introduce a semidefinite relaxation of this problem, along with various performance guarantees. (Joint work with Culver McWhirter (OSU) and Soledad Villar (NYU).)

ERIC PRICE, UT Austin
Compressed Sensing and Generative Models

The goal of compressed sensing is make use of image structure to estimate an image from a small number of linear measurements. The structure is typically represented by sparsity in a well-chosen basis. We show how to achieve guarantees similar to standard compressed sensing but without employing sparsity at all − instead, we suppose that vectors lie near the range of a generative model $G : \mathbb{R}^k \to \mathbb{R}^n$.

This talk will discuss two aspects of this problem. First, given a Lipschitz generative model that represents the data, we show that relatively few Gaussian linear measurements allow for robust estimation. Second, we consider how to learn such a generative model (as a generative adversarial network) from incomplete and inaccurate measurements.

This talk is based on joint work with Ashish Bora, Alex Dimakis, and Ajil Jalal.

YANIV ROMANO, Stanford University
Classification Stability for Sparse-Modeled Signals

Despite their impressive performance, deep convolutional neural networks (CNNs) have been shown to be sensitive to small adversarial perturbations. These nuisances, which one can barely notice, are powerful enough to fool sophisticated and well performing classifiers, leading to ridiculous misclassification results. We study the stability of state-of-the-art classification machines to adversarial perturbations by assuming that the signals belong to the (possibly multi-layer) sparse representation model. We start with convolutional sparsity and then proceed to its multi-layered version, which is tightly connected to CNNs. Our claims can be translated to a practical regularization term that provides a new interpretation to the robustness of Parseval Networks. Also, the proposed theory justifies the increased stability of the recently emerging layered basis pursuit architectures, when compared to the classic forward-pass.

MAHDI SOLTANOLKOTABI, University of Southern California
From Shallow to Deep: Rigorous Guarantees for Training Neural Networks

Neural network architectures (a.k.a. deep learning) have recently emerged as powerful tools for automatic knowledge extraction from data, leading to major breakthroughs in a multitude of applications. Despite their wide empirical use the mathematical success of these architectures remains a mystery. A major challenge is that training neural networks correspond to extremely high-dimensional and nonconvex optimization problems and it is not clear how to provably solve them to global optimality. While training neural networks is known to be intractable in general, simple local search heuristics are often surprisingly effective at finding global/high quality optima on real or randomly generated data. In this talk I will discuss some results explaining the success of these heuristics. First, I will discuss results characterizing the training landscape of single hidden layer networks demonstrating that when the number of hidden units are sufficiently large then the optimization landscape has favorable properties that guarantees global convergence of (stochastic) gradient descent to a model with zero training error. Second, I introduce a de-biased variant of gradient descent called Centered Gradient Descent (CGD). I will show that unlike gradient descent, CGD enjoys fast convergence guarantees for arbitrarily deep convolutional neural networks with large stride lengths. The second part of the talk is joint work with Samet Oymak.

HAU TIENG-WU, Duke University
Recover fetal electrocardiogram morphology via optimal shrinkage

I will discuss recent progress in dealing with the single-channel blind source separation by combining time-frequency analysis, random matrix theory, and diffusion geometry and its application to recover fetal electrocardiogram (ECG) morphology from the single-channel trans-abdominal ECG signal. The main challenge for the long term fetal ECG analysis is the limited physiological channels with complicated statistical features, like time varying amplitude, frequency and non-sinusoidal pattern, and the signal quality is often impaired by non-stationary noise including uterine contraction and motion artifacts. If time permits, I will report a recently finished clinical trial outcome that detects maternal stress based on the fetal ECG analysis.

GIANG TRAN, University of Waterloo
Sparse Recovery and Outliers Detection for Dependent Data

Learning non-linear systems from noisy, limited, and/or dependent data is an important task across various scientific fields including statistics, engineering, computer science, mathematics, and many more. In this work, we study the problem of learning nonlinear functions from sparse corrupted and dependent data. The learning problem is recast as a sparse linear regression problem where we incorporate both the unknown coefficients and the corruptions in a basis pursuit framework. The main contribution of our paper is to provide a reconstruction guarantee for the associated ℓ_1-optimization problem where the sampling matrix is formed from dependent data. Specifically, we prove that the sampling matrix satisfies the null space property and the stable null space property, provided that the data is compact and satisfies a suitable concentration inequality. We show that our recovery results are applicable to various types of dependent data such as exponentially strongly α-mixing data, geometrically $\mathcal{C}$-mixing data, and uniformly ergodic Markov chain.

ROMAN VERSHYNIN, University of California, Irvine
Neural capacity

How many Boolean functions can a given neural network compute? We find an formula that approximately determines this number for any fully-connected, feedforward network with any number of layers, virtually any sizes of layers, and with the threshold activation function. This capacity formula can be used to identify networks that achieve maximal capacity under various natural constraints. At the heart of our analysis is a fundamental question: how many different threshold functions are defined on a given subset S in $\mathbb{R}^n$?

RONGRONG WANG, Michigan State University
Sigma Delta quantization on wide-band signals

Abstract: Sigma Delta quantization has been known to work efficiently on low-frequency signals or on Gaussian samples of sparse signals, in the sense that the rate-distortion decreases quickly as the number of samples increases. Extending this result to the more practical spectrally sparse or wavelet domain sparse signals faces two issues. 1. the high-frequency components in the measurements cannot be well preserved by Sigma Delta quantization. 2. very little is known about the singular-vectors of the high order finite difference matrices which play a crucial role in the Sigma Delta analysis. In this talk, we will address these problems and subsequently prove the reconstruction guarantee under partial Fourier measurements and Haar basis. (This is joint work with Mark Iwen, Rayan Saab, and Wei-husan Yu).