RMT + NLA II, June 16-20, 2025

APPLIED MATHEMATICS
UNIVERSITY of WASHINGTON

Conference Program

Smith Hall Room 205

MC = Minicourse	Monday 6/16	Tuesday 6/17	Wednesday 6/18	Thursday 6/19	Friday 6/20
9:00-9:45 AM	Michael W Mahoney (MC)	Michael W Mahoney (MC)	Govind Menon (MC)	Deanna Needell	Michal Derezinski
9:45-10:30 AM	Michael W Mahoney (MC)	Giorgio Cipolloni (MC)	Ilse Ipsen	Robert Webber	John Peca-Medlin
10:30-11:00 AM	Break (Lewis Hall)	Break (Lewis Hall)	Break (Lewis Hall)	Break (Lewis Hall)	Break (Lewis Hall)
11:00-11:45 AM	Anna Ma	Rishabh Dixit	Ryan Schneider	Zaid Harchaoui	Haixiao Wang
11:45-12:30 PM	Jonathan Novak	Elizaveta Rebrova	Charbel Abi Younes	Anne Greenbaum	
12:30-2:00 PM	Lunch (on your own)	Lunch (on your own)	Lunch (on your own)	Lunch (on your own)	
2:00-2:45 PM	Giorgio Cipolloni (MC)	Govind Menon (MC)	Ken McLaughlin	David Persson	
2:45-3:30 PM	Giorgio Cipolloni (MC)	Govind Menon (MC)	Raphael Meyer	David Renfrew	
3:30-4:00 PM	Break (Lewis Hall)	Break (Lewis Hall)		Break (Lewis Hall)	
4:00-4:45 PM	Poster Session	Jorge Garza Vargas		Yiping Lu	
4:45-5:30 PM		Rajarshi Bhattacharjee		Gunnar Martinsson	

Minicourses

Wigner matrices: A toy model for Quantum Chaos

Giorgio Cipolloni University of Arizona
gcipolloni@arizona.edu

We will begin with discussing Quantum Unique Ergodicity (QUE) and of the Eigenstate Thermalization Hypothesis (ETH), which are fundamental signatures of quantum chaos. However, not much is known about QUE/ETH mathematically for generic quantum systems.

Instead, we study a probabilistic version of these concepts using Wigner matrices, which are more tractable mathematically. The main input to prove QUE/ETH for Wigner matrices are the recently developed multi-resolvent local laws, i.e. compute the deterministic approximation of products of resolvents.

Towards the end of the course, we will discuss some applications of these results/techniques as well as more physically relevant open problems.

Random matrix theory and modern machine learning

Michael W. Mahoney ICSI, LBNL, and Department of Statistics, UC Berkeley
mmahoney@stat.berkeley.edu

Random matrix theory is a large area with a long history, with elegant theory and a wide range of applications. However, the challenges of modern machine learning are forcing us not only to use random matrix theory in new ways, and also to chart out new directions for theory. In this series of presentations, we'll cover several aspects of these developments. This includes challenges in training machine learning models, inference in overparameterized models, diagnostics where heavy-tailed distributions are ubiquitous, and computational-statistical tradeoffs in randomized numerical linear algebra. Addressing these challenges leads to new directions for theory: phenomenology and semi-empirical theory to characterize performance in state-of-the-art neural networks without access to training or testing data; high-dimensional linearizations and deterministic equivalents to go beyond eigenvalue distributions of linear models; very sparse embeddings to perform "algorithmic gaussianization" to speed up core numerical linear algebra problems; new random matrix models that have heavy-tailed spectral structure without having heavy-tailed elements; and using "free compression" ideas in reverse to compute high-quality spectral distributions of so-called impalpable matrices (for which we cannot form or even evaluate with full matrix-vector products).

The geometry of the deep linear network

Govind Menon Brown University
govind_menon@brown.edu

The deep linear network is a matrix model of deep learning. It models the effect of overparameterization for the construction of linear functions. Despite its simplicity, the model has a subtle mathematical structure that yields interesting insights into the training dynamics of deep learning.

We explain a (matrix) geometric perspective for the analysis of the DLN. The heart of the matter is an explicit description of the underlying Riemannian geometry. The use of Riemannian geometry provides unity with the theory of interior point methods for conic programs, and it is helpful to contrast the gradient flows that arise in each setting.

Research Talks

Doubly Noisy Linear Systems and the Kaczmarz Algorithm

Anna Ma University of California, Irvine
anna.ma@uci.edu

Large-scale linear systems, Ax=b, frequently arise in data science and scientific computing at massive scales, thus demanding effective iterative methods to solve them. Often, these systems are noisy due to operational errors or faulty data-collection processes. In the past decade, the randomized Kaczmarz algorithm (RK) was studied extensively as an efficient iterative solver for such systems. However, the convergence study of RK in the noisy regime is limited and considers measurement noise in the right-hand side vector, b. Unfortunately, that is not always the case, and the coefficient matrix A can also be noisy. In this talk, we motivate and discuss doubly noisy linear systems and the performance of the Kaczmarz algorithm applied to such systems. We also present new results on the limit points of Kaczmarz iterates for arbitrary systems. The presented work is joint work with El Houcine Bergou, Soumia Boucherouite, Aritra Dutta, and Xin Li.

Hypergeometric Functions of Huge Matrices

Jonathan Novak UC San Diego
Jinovak@ucsd.edu

Hypergeometric functions of matrix arguments are multivariate generalizations of classical hypergeometric functions which play a key role in random matrix theory and related areas. Approximating hypergeometric functions of huge matrices is one of the most exciting open problems in high-dimensional analysis. In this talk, I will explain how this problem is the discrete counterpart of a much-better understood question, namely that of approximating the partition function of the Hermitian one-matrix model in the high-dimensional limit.

Accelerated gradient methods for nonconvex optimization : asymptotic dynamics and saddle escape

Rishabh Dixit Department of Mathematics, UC San Diego
ridixit@ucsd.edu

This talk focuses on the problem of understanding the behavior of a general class of accelerated gradient methods on smooth nonconvex functions. Motivated by some recent works that have

proposed effective algorithms, based on Polyak's heavy ball method and the Nesterov accelerated gradient method, to achieve convergence to a local minimum of nonconvex functions, we describe a broad class of Nesterov-type accelerated methods and put forth a rigorous study of these methods encompassing the escape from saddle points and convergence to local minima through an asymptotic analysis. In the asymptotic regime, we first answer an open question of whether Nesterov's accelerated gradient method (NAG) with variable momentum parameter avoids strict saddle points almost surely with respect to a random initialization and a random step-size. We then develop two metrics of asymptotic rates of convergence and divergence, and evaluates these two metrics for several popular standard accelerated methods such as the NAG and Nesterov's accelerated gradient with constant momentum (NCM) near strict saddle points. Theoretical results on the asymptotic behavior are demonstrated on the phase retrieval problem.

Randomized sketch-and-project solvers for linear systems with low-rank structure

Liza Rebrova Princeton University
elre@princeton.edu

The sketch-and-project is a unifying framework for many well-known projective iterative methods for solving linear systems, as well as their extensions to nonlinear problems. It generalizes popular algorithms such as randomized Kaczmarz and coordinate descent, and their block variants. In this talk, I present a subspace-constrained version of the sketch-and-project method, where the iterates are restricted to a fixed subspace – for example, the solution space of a selected subsystem. I will show that such subspace constraints can significantly improve expected convergence rates, especially when the linear system has approximately low-rank structure, and describe efficient ways to identify suitable subspaces. On unstructured random data, the constraint acts as a form of dimension reduction that increases the aspect ratio of the system. This effect also makes these methods useful within robust solvers to achieve reliable convergence in challenging settings, when partial information about uncorrupted labels is available.

A new approach to strong convergence: nearly optimal expanders with little randomness

Jorge Garza Vargas Caltech
jgarzav@caltech.edu

In this talk I will discuss the main ideas behind a new method for obtaining sharp norm estimates for random matrix models that, roughly put, possess a lot of symmetry. Among other things, our technique provides a simple way of generating nearly optimal expanders using only a few bits of randomness. This is joint work with Chi-Fang Chen, Joel Tropp, and Ramon van Handel

Improved Spectral Density Estimation via Explicit and Implicit Deflation

Rajarshi Bhattacharjee University of Massachusetts Amherst
rbhattacharj@umass.edu

We study algorithms for approximating the spectral density of a symmetric matrix A that is accessed through matrix-vector products. By combining an existing Chebyshev polynomial moment matching method with a deflation step that approximately projects off the largest magnitude eigendirections of A before estimating the spectral density, we give an $\epsilon\sigma_\ell(A)$ error approximation in the Wasserstein-1 metric using $O(\ell \log n + 1/\epsilon)$ matrix-vector products, where $\sigma_\ell(A)$ is the ℓth largest singular value of A. When A exhibits fast singular value decay, this can be much stronger than prior work, which gives error $\epsilon\sigma_1(A)$ using $O(1/\epsilon)$ matrix-vector products. We also show that this bound is nearly optimal. Additionally, we analyze the Stochastic Lanczos Quadrature (SLQ) method, showing that it achieves nearly the same bound, even though SLQ itself is parameter-free and performs no explicit deflation. This explains the strong practical performance of SLQ, and motivates a simple variant that achieves an even tighter error bound. Our error bound for SLQ leverages an analysis that views it as an implicit polynomial moment matching method, along with recent results on low-rank approximation with single-vector Krylov methods. We use these results to show that SLQ can perform implicit deflation as part of moment matching.

Random Matrix Theory for Stochastic Rounding

Ilse C.F. Ipsen North Carolina State University
ipsen@ncsu.edu

Motivated by the popularity of stochastic rounding in the context of machine learning and the training of large-scale deep neural network models, we consider stochastic nearness rounding of real matrices A with many more rows than columns. We provide novel theoretical evidence, based on random matrix theory, that, with high probability, the smallest singular value of a stochastically rounded matrix is well bounded away from zero - regardless of how close A is to being rank deficient and even if A is rank-deficient. In other words, stochastic rounding implicitly regularizes tall and skinny matrices A so that the rounded version has full column rank. Our proof uses a lower bound, by Dumitriu and Zhu, for the minimum singular value of matrices whose elements are independent zero-mean random numbers that are not necessarily identically distributed. In addition, we exploit the idea that stochastic rounding errors do not concentrate in low-dimensional column spaces.

Randomizing Jacobi's Method

Ryan Schneider University of California Berkeley
ryan.schneider@berkeley.edu

Jacobi's method is the oldest known algorithm for solving the symmetric eigenvalue problem. It is intuitive and simple to implement: Jacobi diagonalizes a symmetric matrix by repeatedly applying orthogonal similarity transformations, each of which zeros out a pair (or block) of off-diagonal entries. Like all solvers for the symmetric eigenvalue problem, it can also be used to compute the SVD of an arbitrary matrix. Nevertheless, Jacobi's method may fail to converge if implemented naively, and (for blocked versions in particular) modifications that guarantee convergence are expensive. We demonstrate that this drawback can be overcome via simple randomization. Moreover, we show that convergent versions of Jacobi's method – both randomized and deterministic – can achieve essentially optimal arithmetic/communication complexity when implemented recursively. This is joint work with James Demmel, Hengrui Luo, and Yifu Wang.

A Lanczos-Based Approach for Spike Detection in Large Sample Covariance Matrices

Charbel Abi Younes University of Washington
cyounes@uw.edu

We introduce a new approach for estimating the number of spikes in a general class of spiked covariance models without directly computing the eigenvalues of the sample covariance matrix. This approach is based on the Lanczos algorithm and the asymptotic properties of the associated Jacobi matrix and its Cholesky factorization. A key aspect of the analysis is interpreting the eigenvector spectral distribution as a perturbation of its asymptotic counterpart. The specific exponential-type asymptotics of the Jacobi matrix enables an efficient approximation of the Stieltjes transform of the asymptotic spectral distribution via a finite continued fraction. As a consequence, we also obtain estimates for the density of the asymptotic distribution and the location of outliers. We provide consistency guarantees for our proposed estimators, proving their convergence in the high-dimensional regime. We demonstrate that, when applied to standard spiked covariance models, our approach outperforms existing methods in computational efficiency and runtime, while still maintaining robustness to exotic population covariances.

Asymptotic analysis of multi-soliton solutions and the formation of soliton gasses

Ken McLaughlin Tulane University
kmclaughlin@tulane.edu

The kinetic theory of solitons and soliton gases originates from the discovery of the complete integrability of the Korteweg–de Vries (KdV) equation in the 1960s, leading to the identification of solitons as fundamental nonlinear phenomena. The concept of a soliton gas was introduced by Zakharov in 1971 and further developed by El in 2003, modeling solitons as interacting particle-like structures. In recent years, rigorous analytical results have been established that provide confirmation of the qualitative theory. In this talk, I will describe some of these advances, including (1) a rigorous derivation of kinetic equations governing soliton gases in KdV-type systems without randomness, as well as (2) the analysis of random collections of solitons, in which both mean behavior and fluctuation results are established. There is a large team of researchers who are collectively investigating these problems, including Manuela Girotti, Aikaterini Gkogkou, Tamara Grava, Robert Jenkins, Guido Mazzuca, Oleksandr Minakov, and Maxim Yattselev. I will hopefully also demonstrate the need for advances in numerical techniques for studying large collections of solitons.

Kronecker Matrix-Vector Complexity is Weird (and exponentially bad)

Raphael Meyer Caltech
ram900@nyu.edu

Matrix-vector complexity has been a fruitful paradigm in the design and analysis of efficient matrix algorithms. For tensor methods, the most natural generalization of the matrix-vector (aka matrix- free aka implicit matrix) model is that of the "Kronecker Matrix-Vector Oracle Model",

and many existing tensor methods belong to this class. In this talk, we will formally introduce and motivate this computational model, study the performance of the Girard-Hutchinson trace estimator in this model, and several information-theoretic lower bounds for the complexity of solving several natural linear algebra problems in this model. Along the way, we will discover several strange observations, including exponential lower bounds, the fact that we should not build random sketches with "small alphabet" distributions, and that sub-gaussianity does not suffice to understand the asymptotic complexity of our algorithms.

Towards Fairness in Machine Learning

Deanna Needell UCLA
deanna@math.ucla.edu

In this talk, we will address areas of recent work centered around the themes of fairness and foundations in machine learning as well as highlight the challenges in this area. We will discuss recent results involving linear algebraic tools for learning, such as methods in non-negative matrix factorization that include tailored approaches for fairness. We will showcase our approach as well as practical applications of those methods. Throughout the talk, we will include example applications from collaborations with community partners, using machine learning to help organizations with fairness and justice goals. This talk includes work joint with Erin George and Lara Kassab.

Adaptive algorithms for data summarization

Robert Webber University of California San Diego, Mathematics
rwebber@ucsd.edu

Unsupervised data analysis techniques such as clustering and nonnegative matrix factorization (NMF) are widely used for data reduction and summarization. Popular algorithms for these tasks like k-means++ (for clustering) and the successive projection algorithm (for NMF) select a small number of prototype elements from a data set that represent or summarize the data. The summarization quality of the selected prototypes is not mathematically optimal, yet the methods are computationally efficient and the suboptimality factor can be bounded in many cases. This work generalizes the prototype-based unsupervised learning approaches. It investigates a broad class of adaptive algorithms that choose prototype elements one at a time based on a randomized or deterministic rule that changes ("adapts") based on the previous selections. Many adaptive algorithms are already in wide use, while other useful algorithms are newly suggested here. The goal of this work is to provide a unified treatment of adaptive algorithms and evaluate their effectiveness both theoretically and empirically.

TBA

Zaid Harchaoui University of Washington
zaid@uw.edu

TBA

When is the Resolvent Like a Rank One Matrix?

Anne Greenbaum University of Washington
greenbau@uw.edu

For a square matrix A, the resolvent of A at a point $z \in \mathbb{C}$ is defined as $(A - zI)^{-1}$. We consider the set of points $z \in \mathbb{C}$ where the relative difference in 2-norm between the resolvent and the nearest rank one matrix is less than a given number $\epsilon \in (0, 1)$. We establish a relationship between this set and the ϵ-pseudospectrum of A, and we derive specific results about this set for Jordan blocks and for a class of large Toeplitz matrices. We also derive disks about the eigenvalues of A that are contained in this set, and this leads to some new results on disks about the eigenvalues that are contained in the ϵ-pseudospectrum of A. In addition, we consider the set of points $z \in \mathbb{C}$ where the absolute value of the inner product of the left and right singular vectors corresponding to the largest singular value of the resolvent is less than ϵ. We demonstrate numerically that this set can be almost as large as the one where the relative difference between the resolvent and the nearest rank one matrix is less than ϵ and we give a partial explanation for this. Some possible applications are discussed.

Singularities in the spectrum of random block matrices

David Renfrew BInghamton University
renfrew@math.binghamton.edu

We consider the density of states of structured Hermitian and non-Hermitian random matrices with a variance profile. As the dimension tends to infinity the associated eigenvalue density can develop a singularity at the origin. The severity of this singularity depends on the relative positions of the zero submatrices. We provide a classification of all possible singularities and determine the exponent in the density blow-up.

Randomized Nyström approximation of non-negative self-adjoint operators

David Persson NYU, Flatiron Institute
dup210@nyu.edu

The randomized singular value decomposition (SVD) has become a popular approach to computing cheap, yet accurate, low-rank approximations to matrices due to its efficiency and strong theoretical guarantees. Recent work by Boullé and Townsend (FoCM, 2023) presents an infinite-dimensional analog of the randomized SVD to approximate Hilbert-Schmidt operators. However, many applications involve computing low-rank approximations to symmetric positive semi-definite matrices. In this setting, it is well-established that the randomized Nyström approximation is usually preferred over the randomized SVD. In this talk, we present an infinite-dimensional extension of the randomized Nyström approximation for computing low-rank approximations to self-adjoint, positive, trace class operators. A significant advantage of the proposed framework is that once a low-rank approximation of the operator is computed, one can use this approximation to compute a low-rank approximation to any discretization of the operator.

What is a Sketch-and-Precondition Derivation for Low-Rank Approximation?

Yiping Lu Northwestern University
yiping.lu@northwestern.edu

Randomized sketching accelerates large-scale numerical linear algebra by reducing computational complexity. While the traditional sketch-and-solve approach reduces the problem size directly through sketching, the sketch-and-precondition method leverages sketching to construct a computational friendly preconditioner. This preconditioner improves the convergence speed of iterative solvers applied to the original problem, maintaining accuracy in the full space. Furthermore, the convergence rate of the solver improves at least linearly with the sketch size. Despite its potential, developing a sketch-and-precondition framework for randomized algorithms in low-rank matrix approximation remains an open challenge. We introduce the Error-Powered Sketched Inverse Iteration (EPSI) Method via run sketched Newton iteration for the Lagrange form as a sketch-and-precondition variant for randomized low-rank approximation. Our method achieves theoretical guarantees, including a convergence rate that improves at least linearly with the sketch size. I'll also demonstrate an online version of the algorithm.

Fast rank adaptive CUR via a recycled small sketch

Per-Gunnar Martinsson UT Austin
pgm@oden.utexas.edu

The talk describes a rank-adaptive technique for computing a low rank approximation to a matrix. It is based on previously described methods for constructing CUR and interpolative decompositions through greedy pivoting on a random sketch of the matrix. In the new approach, a CUR decomposition is built incrementally, with no a priori knowledge of the numerical rank of the matrix required. A small sketch is drawn to find the first set of columns/rows. For subsequent steps, the small sketch is downdated and recycled, leading to significant acceleration. Extensive numerical examples illustrative the highly competitive performance of the method. The talk will also describe how a posteriori error estimation is used to reliably track the approximation error as the algorithm proceeds. (Joint work with Nathaniel Pritchard, Taejun Park, and Yuji Nakatsukasa)

Randomized Algorithms for Linear Systems: Going Beyond Krylov Subspace Methods

Michal Derezinski University of Michigan
michal.derezinski@gmail.com

Solving large systems of linear equations has numerous applications across many areas, from scientific computing to machine learning and beyond. Krylov subspace methods such as conjugate gradient have long been the gold standard in this area, thanks to their ability to exploit clusters and outliers in the spectrum of the input matrix to achieve fast convergence. Yet, recent results have shown that it is possible to improve upon Krylov methods by introducing randomness into the algorithmic procedure via matrix sketching and sub-sampling. In this talk, I will survey the advantages of randomized sketching algorithms in solving large linear systems, highlighting the

random matrix theory challenges that arise along the way. In particular, I will cover our recent work on stochastic iterative solvers based on the sketch-and-project framework (an extension of randomized Kaczmarz), which leverages new spectral guarantees for random orthogonal projection matrices that arise from sketching. This work has led to Kaczmarz++, a practical algorithm that exploits large outlying singular values in the linear system provably better than any Krylov subspace method.

Random permutations from GEPP

John Peca-Medlin UCSD
jpecamedlin@ucsd.edu

Gaussian elimination with partial pivoting (GEPP) remains the most used dense linear solver. For a n x n matrix A, GEPP results in the factorization PA = LU where L and U are lower and upper triangular matrices and P is a permutation matrix. If A is a random matrix, then the associated permutation from the P factor is random. When is this a uniform permutation? How many cycles are in its disjoint cycle decomposition (which equivalently answers how many GEPP pivot movements are needed on A)? What is the length of the longest increasing subsequence of this permutation? We will provide some statistical answers to these questions for select random matrix ensembles and transformations. For particular butterfly permutations, we will present full distributional descriptions for these particular statistics. Moreover, we introduce a random matrix ensemble that induces the Haar measure on a full 2-Sylow subgroup of the symmetric group on a set of size 2.

Critical sparse random rectangular matrices: emergence of spectra outliers

Haixiao Wang University of California, San Diego
h9wang@ucsd.edu

Consider the random bipartite Erdos-Renyi graph G(n, m, p), where each edge with one vertex in $V_1 = [n]$ and the other vertex in $V_2 = [m]$ is connected with probability p with $n \geq m$. For the centered and normalized adjacency matrix H, it is well known that the empirical spectral measure will converge to the Marchenko-Pastur (MP) distribution. However, this does not necessarily imply that the largest (resp. smallest) singular values will converge to the right (resp. left) edge when $p = o(1)$, due to the sparsity assumption. In Dumitriu and Zhu 2024, it was proved that almost surely there are no outliers outside the compact support of the MP law when $np = \omega(\log(n))$. In this paper, we consider the critical sparsity regime with $np = O(\log(n))$, where we denote $p = b\log(n)/\sqrt{mn}$, $\gamma = n/m$ for some positive constants b and γ. For the first time in the literature, we quantitatively characterize the emergence of outlier singular values. When $b > b_\star$, there is no outlier outside the bulk; when $b^\star < b < b_\star$, outlier singular values only appear outside the right edge of the MP law; when $b < b^\star$, outliers appear on both sides. Meanwhile, the locations of those outliers are precisely characterized by some function depending on the largest and smallest degrees of the sampled random graph. The thresholds $b^\star$ and $b_\star$ purely depends on γ. Our results can be extended to sparse random critical matrices with bounded entries.

Posters

Sketch-and-precondition in low rank matrix approximation

Ruihan Xu The university of Chicago
ruihanx@uchicago.edu

Randomized sketching accelerates large-scale numerical linear algebra by reducing computational complexity. While the traditional sketch-and-solve approach reduces the problem size directly through sketching, the sketch-and-precondition method leverages sketching to construct a computational friendly preconditioner. This preconditioner improves the convergence speed of iterative solvers applied to the original problem, maintaining accuracy in the full space. Furthermore, the convergence rate of the solver improves at least linearly with the sketch size. Despite its potential, developing a sketch-and-precondition framework for randomized algorithms in low-rank matrix approximation remains an open challenge. We introduce the Error-Powered Sketched Inverse Iteration (EPSI) Method via run sketched Newton iteration for the Lagrange form as a sketch-and-precondition variant for randomized low-rank approximation. Our method achieves theoretical guarantees, including a convergence rate that improves at least linearly with the sketch size.

Tensor Randomized Kaczmarz Methods for Linear Feasibility Problems

Minxin Zhang University of California, Los Angeles
minxinzhang@g.ucla.edu

The randomized Kaczmarz methods are a popular and effective family of iterative methods for solving large-scale linear systems of equations, which have also been applied to linear feasibility problems. In this work, we propose a new block variant of the randomized Kaczmarz method, B-MRK, for solving linear feasibility problems defined by matrices. We show that B-MRK converges linearly in expectation to the feasible region.Furthermore, we extend the method to solve tensor linear feasibility problems defined under the tensor t-product. A tensor randomized Kaczmarz (TRK) method, TRK-L, is proposed for solving linear feasibility problems that involve mixed equality and inequality constraints. Additionally, we introduce another TRK method, TRK-LB, specifically tailored for cases where the feasible region is defined by linear equality constraints coupled with bound constraints on the variables. We show that both of the TRK methods converge linearly in expectation to the feasible region. Moreover, the effectiveness of our methods is demonstrated through numerical experiments on various Gaussian random data and applications in image deblurring.

A federated Kaczmarz algorithm

Chi-Hao Wu UCLA
chwu93@math.ucla.edu

We propose a federated algorithm for solving large linear systems that is inspired by the classic randomized Kaczmarz algorithm. We provide convergence guarantees of the proposed method, and as a corollary of our analysis, we provide a new proof for the convergence of the classic randomized Kaczmarz method. We also demonstrate experimentally that our approach can be combined with hard thresholding for compressed sensing. This is joint work with Halyun Jeong and Deanna Needell.

Randomized Kaczmarz Methods with Beyond-Krylov Convergence

Jiaming Yang University of Michigan
jiamyang@umich.edu

Randomized Kaczmarz methods form a family of linear system solvers which converge by repeatedly projecting their iterates onto randomly sampled equations. While effective in some contexts, such as highly over-determined least squares, Kaczmarz methods are traditionally deemed secondary to Krylov subspace methods, since this latter family of solvers can exploit outliers in the input's singular value distribution to attain fast convergence on ill-conditioned systems. In this paper, we introduce Kaczmarz++, an accelerated randomized block Kaczmarz algorithm that exploits outlying singular values in the input to attain a fast Krylov-style convergence. Moreover, we show that Kaczmarz++ captures large outlying singular values provably faster than popular Krylov methods, for both over- and under-determined systems. We also develop an optimized variant for positive semidefinite systems, called CD++, demonstrating empirically that it is competitive in arithmetic operations with both CG and GMRES on a collection of benchmark problems. To attain these results, we introduce several novel algorithmic improvements to the Kaczmarz framework, including adaptive momentum acceleration, Tikhonov-regularized projections, and a memoization scheme for reusing information from previously sampled equation blocks.

Some Kernel Matrix Eigenvalue Estimates with Applications

Mikhail Lepilov Emory University
mlepilov@gmail.com

In data science, individual observations are often assumed to come independently from an underlying probability space. Kernel matrices formed from large sets of such observations arise frequently, for example during classification tasks. It is desirable to know the eigenvalue decay properties of these matrices without explicitly forming them, such as when determining if a low-rank approximation is feasible. In this talk, I introduce a new eigenvalue quantile estimation framework for some kernel matrices. This framework gives meaningful bounds for all the eigenvalues of a kernel matrix while avoiding the cost of constructing the full matrix. The kernel matrices under consideration come from a kernel with quick decay away from the diagonal applied to uniformly-distributed sets of points in Euclidean space of any dimension. We prove the efficacy of this framework given certain bounds on the kernel function, and we provide empirical evidence for its accuracy. In the process, we also prove a very general interlacing-type theorem for finite sets of numbers. Additionally, we indicate an application of this framework to the study of the intrinsic dimension of data.

Convergence of AltLS for tensor CP decomposition

Nicholas Hu UCLA
njhu@math.ucla.edu

We prove explicit quantitative local convergence theorems for the CP-AltLS algorithm applied to orthogonally and incoherently decomposable tensors. Specifically, we show that CP-AltLS con-

verges polynomially with order N1 for Nth-order orthogonally decomposable tensors and linearly for incoherently decomposable tensors.

Optimal shrinkage estimation for linear discriminant analysis in ultra-high dimension

Wonjun Seo UC Davis
wseo@ucdavis.edu

Linear discriminant analysis (LDA) faces significant challenges when the number of features (p) exceeds the number of observations (n). While various methods have been proposed to address this issue, most assume n and p are comparable or impose restrictive structural assumptions on the population covariance matrix. In this study, we present a unified framework for LDA based on an optimal shrinkage method designed for ultra-high dimensional data, where p grows polynomially in n. As examples within our framework, we consider two types of shrinkage estimators: a linear shrinker, leading to a regularized LDA, and a nonlinear shrinker under the generalized spiked covariance matrix model. Leveraging recent advances in random matrix theory, we establish theoretical guarantees for our approach by analyzing the asymptotic behavior of outlier eigenvalues and eigenvectors, as well as deriving a quantum unique ergodicity estimate for non-outlier eigenvectors of the spiked sample covariance matrix. These results also reveal a phase transition phenomenon in LDA, allowing us to characterize the conditions under which LDA succeeds or fails based on the magnitude of the mean difference.

Quantile-Based Acceleration of the Random Kaczmarz Method and its Application to the Quantile Random Kaczmarz Method

Emeric Battaglia University of California, Irvine
ebattagl@uci.edu

Inspired by recent work on the quantile-randomized Kaczmarz method for solving a linear system of equations, we propose an acceleration of the randomized Kaczmarz method using quantile information. We show that the method converges faster than the randomized Kaczmarz algorithm when the linear system is consistent. In addition, we demonstrate how this new acceleration may be used in conjunction with the quantile-randomized Kaczmarz algorithm (qRK), without additional computational complexity, to produce both a fast and robust iterative method for solving large, sparsely corrupted linear systems. Additionally, we provide new error horizon bounds for qRK and its acceleration in the setting where the corruption may not be sparse.

Stochastic Iterative Methods for Online Rank Aggregation from Pairwise Comparisons

Alexander Sietsema UCLA
alexsietsema@math.ucla.edu

In this paper, we consider large-scale ranking problems where one is given a set of (possibly non- redundant) pairwise comparisons and the underlying ranking explained by those comparisons is

desired. We show that stochastic gradient descent approaches can be leveraged to offer convergence to a solution that reveals the underlying ranking while requiring low-memory operations. We introduce several variations of this approach that offer a tradeoff in speed and convergence when the pairwise comparisons are noisy (i.e., some comparisons do not respect the underlying ranking). We prove theoretical results for convergence almost surely and study several regimes including those with full observations, partial observations, and noisy observations. Our empirical results give insights into the number of observations required as well as how much noise in those measurements can be tolerated.

Stratified Non-Negative Tensor Factorization

Zerrin Vural UCLA graduate student
zerrin.vural@ucla.edu

Non-negative matrix factorization (NMF) and non-negative tensor factorization (NTF) decompose non- negative high-dimensional data into non-negative low-rank components. NMF and NTF methods are popular for their intrinsic interpretability and effectiveness on large-scale data. Recent work developed Stratified-NMF, which applies NMF to regimes where data may come from different sources (strata) with different underlying distributions, and seeks to recover both strata-dependent information and global topics shared across strata. Applying Stratified-NMF to multi-modal data requires flattening across modes, and therefore loses geometric structure contained implicitly within the tensor. To address this problem, we extend Stratified-NMF to the tensor setting by developing a multiplicative update rule and demonstrating the method on text and image data. We find that Stratified-NTF can identify interpretable topics with lower memory requirements than Stratified-NMF. We also introduce a regularized version of the method and demonstrate its effects on image data.

Optimal shrinkage estimation for linear discriminant analysis in ultra-high dimension

Wonjun Seo UC Davis
wseo@ucdavis.edu

Linear discriminant analysis (LDA) faces significant challenges when the number of features (p) exceeds the number of observations (n). While various methods have been proposed to address this issue, most assume n and p are comparable or impose restrictive structural assumptions on the population covariance matrix. In this study, we present a unified framework for LDA based on an optimal shrinkage method designed for ultra-high dimensional data, where p grows polynomially in n. As examples within our framework, we consider two types of shrinkage estimators: a linear shrinker, leading to a regularized LDA, and a nonlinear shrinker under the generalized spiked covariance matrix model. Leveraging recent advances in random matrix theory, we establish theoretical guarantees for our approach by analyzing the asymptotic behavior of outlier eigenvalues and eigenvectors, as well as deriving a quantum unique ergodicity estimate for non-outlier eigenvectors of the spiked sample covariance matrix. These results also reveal a phase transition phenomenon in LDA, allowing us to characterize the conditions under which LDA succeeds or fails based on the magnitude of the mean difference.

Optimal Oblivious Subspace Embeddings with Near-optimal Sparsity

Shabarish Chenakkod University of Michigan
shabari@umich.edu

A random $m \times n$ matrix S is an oblivious subspace embedding (OSE) with parameters $\epsilon > 0$, $\delta \in (0,1)$ and $d \leq m \leq n$, if for any d-dimensional subspace $W \subseteq \mathbb{R}^n$,

$$\mathbb{P}\big(\forall_{x \in W} \ (1-\epsilon)\|x\| \leq \|Sx\| \leq (1+\epsilon)\|x\| \big) \geq 1 - \delta.$$

OSEs are an important dimension reduction tool in randomized numerical linear algebra and we desire to find a distribution of random matrices S which are sparse, for computational efficiency, and OSEs for the optimal embedding dimension of $m = O(\frac{d}{\epsilon^2})$. I shall discuss recent joint work with Michał Dereziński and Xiaoyu Dong where we show that, given any error bound ϵ, there exists an $m \times n$ random matrix S with optimal embedding dimension $m = O(\frac{d}{\epsilon^2})$ consisting of $s = O(\frac{\log^2(d/\varepsilon\delta)}{\epsilon})$ many non-zero entries in each column which is an OSE with error ϵ. Our result achieves optimal embedding dimension and optimal sparsity of the form $\tilde{O}(1/\epsilon)$ non-zero entries per column as conjectured by Nelson and Nguyen (FOCS 2013). Our proof proceeds by showing the universality of moments of the embedding error of order $\log(d/\varepsilon\delta)$. A specific construction for S enables us to use decoupling to reduce the problem to bounding the moments of a product of independent copies of S, where the corresponding Gaussian counterpart has the correct growth for the embedding error. We show universality via a refinement of the "interpolation using cumulant expansion of the derivative" idea by Brailovskaya-van Handel. This refinement allows us to obtain better bounds for the universality error than a black box application of their spectrum universality result.

On the greatest root statistic for data matrices with Rapidly Increasing Row Dimension

John J. Lopez Santander Tulane University
jlopez12@tulane.edu

The greatest root statistic is defined as the largest eigenvalue of $(A+B)^{-1}B$, where A and B are independent Wishart matrices, $A \sim W_p(I,m)$ and $B \sim W_p(I,n)$. In 2008, Johnstone studied its probability distribution for data matrices where the number of rows m of the underlying data matrix X_A, as well as n in the data matrix X_B, grow proportionally with p. We consider a regime where the number of rows in X_A grows faster, so that A becomes a tall and skinny matrix as the dimension p increases to infinity . More specifically, we study the joint density of eigenvalues of $(A+B)^{-1}B$, which is given by

$$f_p = c_p \prod_{j<k} |\lambda_j - \lambda_k|^2 \prod_{j=1}^{p} (1-\lambda_j)^{m-p} \lambda_j^{n-p},$$

corresponding to the Jacobi ensemble. In our parameter regime, the distribution of the largest eigenvalue can be expressed in terms of Jacobi polynomials with modified Jacobi weights of the form

$$w_m(x) = x^{m\alpha}(1-x)^{m\beta},$$

where α is a positive real number, and the parameter $\beta = \beta(m)$ grows with m as $m \to \infty$, according to $\beta = m^\nu$, with $0 < \nu < 1$. Our approach is based on the Riemann-Hilbert analysis for the associated Jacobi polynomials. This poster is based on joint work with Katerina Gkogkou and Kenneth McLaughlin.