Randomized Numerical Linear Algebra
Poster Session Abstracts
February 5, 2026
4:00-5:00pm

The Loss Landscape of Preconditioned Stochastic Gradient Descent
Mitchell Scott, Emory University

Stochastic Gradient Descent (SGD) is the workhorse of machine learning (ML) training. However, SGD is slow, which has lead researchers to develop optimizers-- ADAM, AdaGrad, Sophia, etc-- which can be viewed as preconditioned SGD. These more complicated optimizers coupled with the complexity of ML networks, leads to a complicated loss landscape, and the normal assumptions of numerical optimization cannot be justified. To counteract that, we use standard random matrix theory tools to approximate complicated ML networks with many nodes. Building on results of Pennington, we investigate how preconditioners of various structure affect the loss landscape to better determine the design criteria for these preconditioner. Theoretic results are verified by toy neural networks.

Quantum Algorithm Framework for Driven Dissipative Ground State Preparation
Paul Cazeaux, Virgina Tech

An important subroutine in any quantum computing algorithm is the preparation of the initial useful and relevant quantum state of the machine, preferably efficiently and using a small number of ancilla qubits.
In this talk, we will discuss a novel algorithm for the preparation of the ground state of a given Hamiltonian and its analysis, motivated by the novel approach [1] based on the approximation of Lindblad dynamics often used to model open quantum systems. Under certain assumptions on the eigenvalue distribution of the Hamiltonian and ability to apply certain random unitary operations, explicit mixing time estimates may be obtained showing polynomial scaling in the number of qubits. [1] Z. Ding, C-F. Chen and L. Lin, Single-ancilla ground state preparation via Lindbladians, Phys. Rev. Research 6, 033147, 2024

Sensor Placements in Gaussian Processes Using Column Subset Selection
Jessie Chen, North Carolina State University

Gaussian process regression uses data measured at sensor locations to reconstruct a spatially dependent function with quantified uncertainty. However, if only a limited number of sensors can be deployed, it is important to determine how to optimally place the sensors to minimize a measure of the uncertainty in the reconstruction. We consider the Bayesian D-optimal criterion to determine the optimal sensor locations by choosing k sensors from a candidate set of n sensors. Since this is an NP-hard problem, our approach models sensor placement as a column subset selection problem (CSSP) on the covariance matrix, computed using the kernel function on the candidate sensor points. We propose an algorithm that uses the Golub-Klema-Stewart framework (GKS) to select sensors and provide an analysis of lower bounds on the D-optimality of these sensor placements. To reduce the computational cost in the GKS step, we propose and analyze algorithms for the D-optimal sensor placements using Nystr\"om approximations on the covariance matrix. Moreover, we propose several algorithms that select sensors via Nystr\"om approximation of the covariance matrix, utilizing the randomized Nystr\"om approximation, random pivoted Cholesky and greedy pivoted Cholesky. We demonstrate the performance of our method on two applications: thin liquid film dynamics and sea surface temperature.

Discrepancies are Virtue: Weak-to-Strong Generalization through Lens of Intrinsic Dimension
Yijun Dong, New York University

Weak-to-strong (W2S) generalization is a type of finetuning (FT) where a strong (large) student model is trained on pseudo-labels generated by a weak teacher. Surprisingly, W2S FT often outperforms the weak teacher. We seek to understand this phenomenon through the observation that FT often occurs in intrinsically low-dimensional spaces. Leveraging the low intrinsic dimensionality of FT, we analyze W2S in the ridgeless regression setting from a variance reduction perspective. For a strong student-weak teacher pair with sufficiently expressive low-dimensional feature subspaces $\mathcal{V}_s, \mathcal{V}_w$, we provide an exact characterization of the variance that dominates the generalization error of W2S. This unveils a virtue of discrepancy between the strong and weak models in W2S: the variance of the weak teacher is inherited by the strong student in $\mathcal{V}_s \cap \mathcal{V}_w$, while reduced by a factor of $\dim(\mathcal{V}_s)/N$ in the subspace of discrepancy $\mathcal{V}_w \setminus \mathcal{V}_s$ with N pseudo-labels for W2S. Our analysis further casts light on the sample complexities and the scaling of performance gap recovery in W2S. The analysis is supported by experiments on synthetic regression problems, as well as real vision and NLP tasks.

Superfast and stable divide-and-conquer singular value decomposition for hierarchical rank-structured matrices
Chenyang Cao, Purdue University

This poster gives a superfast divide-and-conquer algorithm for computing the full singular value decomposition (SVD) of hierarchical rank-structured matrices with small off-diagonal ranks. The method achieves nearly linear complexity while delivering all singular values and singular vectors in structured forms. The structured representation of singular vectors enables near-linear operations with reduced storage. In contrast, classical algorithms for the full SVD require cubic computational cost and quadratic storage. The proposed method directly handles nonsymmetric or rectangular matrices by reducing them to a hierarchical block broken arrow form via stable QL factorizations. This form is then repeatedly decomposed through rank-1 SVD updates in the conquering stage. Several stability-preserving mechanisms are incorporated, including deflation, splitting and shifting, and orthogonality-preserving perturbations, to ensure the robustness of this stage. Efficiency is enhanced via fast kernel methods, such as the fast multipole method (FMM). A rigorous numerical error analysis establishes the backward stability of the overall process. Numerical experiments demonstrate significant advantages in both accuracy and efficiency, yielding the first fast and stable SVD for general hierarchical rank-structured matrices.

Randomness as a Resource for Electric Power Networks
Samuel Talkington, Georgia Tech

This poster presents the randomized switching framework for electric power network reconfiguration, a challenging mixed-integer nonlinear program central to grid planning and operations. By optimizing over switching probabilities via conditional gradient methods and applying randomized rounding, we obtain provable approximation guarantees via matrix Chernoff bounds, with practical speedups of 1 to 3 orders of magnitude over commercial solvers. The gradient of our objective is proportional to squared voltage differences across edges, reminiscent of effective resistances and biharmonic distances. This suggests a general leverage score sampling framework for network optimization with hard constraints, e.g., robust connectivity and edge weight constraints. We further discuss ongoing work showing that optimal power flow (OPF) price sensitivities reduce to related Laplacian matrix functionals, enabling applications in energy affordability policymaking. Open problems include: (i) fast solvers for complex-valued and block-structured Laplacian systems arising in power flow physics, (ii) randomized preconditioners for OPF interior point methods, and (iii) constrained spectral sparsification methods.

Structure-Informed Bounds on Matrix Kronecker Rank

Allison Fuller, Arizona State University

In this poster, we present theoretical bounds for the Kronecker rank of block-structured matrices. By utilizing a previously developed method that maps a structured matrix to a tensor and back, the Kronecker rank can be determined as the rank of one unfolding of the mapped tensor. In our analysis, we look at both the original matrix and a second, permuted, form of the original matrix where the outer and inner structures have been swapped. Our bounds are established in terms of matrix subspaces defined from the structures and blocks of both matrices. This is done by relating the dimensions of these subspaces to the rank of the unfolded tensor.

Block-Structured Tensor Train Sketching for High Dimensional Problems

Rodrigo Figueroa, Virginia Tech

Large-scale scientific computing introduces new computational bottlenecks that render many problems infeasible at scale. Randomized numerical linear algebra addresses this challenge through sketching, that is, random dimensional reduction that preserves essential geometric structure, while tensor-network formats such as the tensor-train (TT) decomposition provide compact representations of exponentially large multiway arrays, with impactful applications in quantum physics, parametric PDEs, and modern machine learning. However, existing tensor sketching methods either suffer from exponential dependence on tensor dimension or lack rigorous theoretical guarantees. We propose a novel block-structured sketching framework that avoids exponential scaling while retaining provable embedding properties, admits natural parallelization, and enables efficient algorithms for randomized compression, Krylov eigensolvers, and streaming approximation in tensor formats.

Randomized Algorithms for Compressing and Inverting Rank-Structured Matrices

Anna Yesypenko, The Ohio State University

We present randomized, black-box algorithms for constructing data-sparse representations of rank-structured matrices, with particular emphasis on H^2 matrices. These methods interact with the target matrix exclusively through matrix–vector products, enabling compression and algebraic operations even when explicit matrix entries are unavailable. This makes them especially suitable for tasks such as forming Schur complements and performing matrix–matrix multiplication. The main contribution is Randomized Strong Recursive Skeletonization (RSRS), a novel algorithm that simultaneously compresses and factorizes an H^2 matrix under a strong admissibility criterion within the black-box framework. We demonstrate the effectiveness of RSRS by applying it to the construction of an efficient sparse direct solver for elliptic partial differential equations, achieving substantial reductions in computational cost and memory usage.

The Polar Express: Optimal Matrix Sign Methods and Their Application to the Muon Algorithm

Noah Amsel, NYU

Computing the polar decomposition and the related matrix sign function has been a well-studied problem in numerical analysis for decades. Recently, it has emerged as an important subroutine within the Muon algorithm for training deep neural networks. However, the requirements of this application differ sharply from classical settings: deep learning demands GPU-friendly algorithms that prioritize high throughput over high precision. We introduce Polar Express, a new method for computing the polar decomposition. Like Newton-Schulz and other classical polynomial methods, our approach uses only matrix-matrix multiplications, making it very efficient on GPUs. Inspired by earlier work of Chen & Chow and Nakatsukasa & Freund, Polar Express adapts the update rule at each iteration by solving a minimax optimization problem. We prove that this strategy minimizes error in a worst-case sense, allowing Polar Express to converge as rapidly as possible both in the early iterations and asymptotically. We also address finite-precision issues, making it practical to use in bfloat16. When integrated into the Muon training framework, our method leads to consistent improvements in validation loss when training a GPT-2 model on one billion tokens from the FineWeb dataset, outperforming recent alternatives across a range of learning rates.

Sketchy Multi-Query Finite Element Simulation

Nick Polydorides, University of Edinburgh

We present a low-variance Randomized Numerical Linear Algebra framework for multi-query Finite Element Method (FEM) simulations. By applying Galerkin projection onto a fixed proper orthogonal decomposition (POD) basis and preconditioning the resulting equations, we preserve structural properties while improving the sketching process. Our methodology utilizes Bernoulli leverage-score sampling to sketch reduced coefficient matrices and right-hand sides. To minimize sketching error, we introduce control variates based on a piecewise-constant surrogate and solve a commutation-regularized optimization problem to fuse forward and inverse sketch estimators.

Edge inference from random connectivity interrogations

Omar De La Cruz Cabrera, Kent State University

We seek to infer the presence/absence of edges in a network, that is, 1's or 0's in the adjacency matrix A of a graph, from the communicability matrix C(t) of the network, defined as the matrix exponential of tA, where t>0. The only information available is the result of multiplying C(t) by a set of random vectors (possibly with different values of t). Our approach is based on the use of physics-informed neural networks (PINNs). We present preliminary results.