ICERM Topical Workshop

Modern Applied Computational Analysis

Poster Session

Tuesday, June 27, 2023

Near-Optimality Guarantees for Approximating Rational Matrix Functions by the Lanczos Method

Noah Amsel, New York University

We study the Lanczos method for approximating the action of a symmetric matrix function f(A) on a vector b (Lanczos-FA). Lanczos-FA iteratively constructs its approximation from the Krylov subspace K_k = span{b, Ab, ..., A^{k-1}b} and can be implemented efficiently using only matrix-vector products with A, making it particularly popular for high dimensional sparse or structured problems in computational science.

For the function classes exp(tA)b, A^{-1}b, and matrix polynomials of reasonable degree, previous work has shown that the error of the Lanczos-FA approximation is within a small multiplicative factor of the error of the best approximation from the Krylov subspace.

In practice, Lanczos-FA consistently achieves similarly high accuracy on a wide variety of other functions too, even outperforming alternative methods for which stronger theoretical guarantees are known.

We aim to narrow this understanding gap by considering rational functions with no poles in the interval containing A's eigenvalues. We prove that for these functions, the error of infinite-precision Lanczos-FA is within a poly(kappa) factor of the optimal Krylov approximation, where kappa is the condition number of A. While we believe the dependence on kappa is loose, our bound qualitatively matches the convergence behavior of the algorithm better than previous analyses.

Our result also provides insight on functions that are well approximated by rational functions, such as A^{±1/2}. Finally, the problem of tightening the dependence of our bound on kappa gives rise to several interesting questions in approximation theory for functions on a real interval.

NerualFIM for learning Fisher Information Metrics from point cloud data

Dami Fasina, Yale University

Although data diffusion embeddings are ubiquitous in unsupervised learning and have proven to be a viable technique for uncovering the underlying intrinsic geometry of data, diffusion embeddings are inherently limited due to their discrete nature. To this end, we propose neural FIM, a method for computing the Fisher information metric (FIM) from point cloud data - allowing for a continuous manifold model for the data. Neural FIM creates an extensible metric space from discrete point cloud data such that information from the metric can inform us of manifold characteristics such as volume and geodesics.
We demonstrate Neural FIM's utility in selecting parameters for the PHATE visualization method as well as its ability to obtain information pertaining to local volume illuminating branching points and cluster centers embeddings of a toy dataset and two single-cell datasets of IPSC reprogramming and PBMCs (immune cells).

Data-efficient matrix recovery and PDE learning

Diana Halikias, Cornell University Department of Mathematics

Can one recover the entries of a matrix from only matrix-vector products? If so, how many are needed? I will present my research relating to this problem, including randomized algorithms to recover various structured matrices, as well as theoretical results which bound the query complexity of these structured families. Moreover, a continuous generalization of query complexity describes how many pairs of forcing terms and solutions are needed to uniquely identify a Green's function corresponding to the solution operator of an elliptic PDE. I will present a recent main result, which is a theoretical guarantee on the number of input-output pairs required in elliptic PDE learning problems with high probability. The proof of this result is constructive, and relies on a randomized algorithm which leverages insights from numerical linear algebra and PDE theory. Finally, I will discuss future research directions investigating when querying the adjoint of an operator is essential to attain optimal query complexity.

Measuring Electromagnetic Green's Functions via Field Correlations

Jonas Katona, Applied Mathematics Program, Yale University

Relations between the field correlations generated by noise and Green's functions are central to correlation-based imaging techniques. Such relations are well-known for applications in geophysics, acoustics, and scalar wave optics, but have not yet been generalized to electromagnetic waves, i.e., solutions to Maxwell's equations. In this work, we derive formulae for deriving the real and imaginary parts of the electromagnetic Green's tensor from the electric fields generated by suitable random surface and volume noises for the following boundary conditions: 1) boundary conditions for a perfect electric conductor (where the tangential component of the electric field is zero), 2) boundary conditions for a perfect magnetic conductor (where the tangential component of the magnetic field is zero), and 3) Sommerfeld radiation conditions on a sufficiently large ball. We illustrate the validity of our generalizations via numerical experiments.

2D Analytic Signals on Bounded Domains

Brian Knight, UC Davis

By viewing a 1D signal as the boundary value of a harmonic function in the unit disc in $\mathbb{C}$, one can obtain a multiscale analytic signal representation by supplementing its conjugate counterpart and viewing the function on the disc at fixed radii. This can be done in the upper-half plane as well, and the two approaches are related via a conformal transformation. For nD signals there are several common approaches to generalizing this, namely, solving the Riemann-Hilbert problem in several complex variables, or forming the so-called monogenic signal on upper-half space via Clifford analysis. These are no longer related via any nice transformation. The former is naturally formed on a polydisc in $\mathbb{C}^n$, and is easily seen as an extension of the unit disc case in $\mathbb{C}$, whereas the latter is formed on the upper-half space $\mathbb{R}^{n}_{+}$ and is an extension of the upper-half plane case, though Felsberg et. all introduced a method to study the monogenic scale space on a bounded domain. We compare and contrast these methods with concern to their application in image processing tasks and discuss future directions of development. This is joint work with Dr. Naoki Saito.

Hyperbolic Diffusion Embedding and Distance for Hierarchical Representation Learning
Ya-Wei Eileen LIN, Technion - Israel Institute of Technology

Finding meaningful representations and distances of hierarchical data is important in many fields. This paper presents a new method for hierarchical data embedding and distance. Our method relies on combining diffusion geometry, a central approach to manifold learning, and hyperbolic geometry.
Specifically, using diffusion geometry, we build multi-scale densities on the data, aimed to reveal their hierarchical structure, and then embed them into a product of hyperbolic spaces. We show theoretically that our embedding and distance recover the underlying hierarchical structure. In addition, we demonstrate the efficacy of the proposed method and its advantages compared to existing methods on graph embedding benchmarks and hierarchical datasets.

This is a joint work with Ronald R. Coifman, Gal Mishne, and Ronen Talmon

Sample Complexity for Scientific Machine Learning
Yiping Lu, Stanford->Courant->Northwestern

Massive data collection and computational capabilities have enabled data-driven scientific discoveries and control of engineering systems. However, there are still several questions that should be answered to understand the fundamental limits of just how much can be discovered with data and what is the value of additional information. For example, 1) How can we learn a physics law or economic principle purely from data? 2) How hard is this task, both computationally and statistically? 3) What's the impact on hardness when we add further information (e.g., adding data, model information)? I'll answer these three questions in this talk in two learning tasks. A key insight in both two cases is that using direct plug-in estimators can result in statistically suboptimal inference.

The first learning task I'll discuss is linear operator learning/functional data analysis, which has wide applications in causal inference, time series modeling, and conditional probability learning. We build the first min-max lower bound for this problem. The min-max rate has a particular structure where the more challenging parts of the input and output spaces determine the hardness of learning a linear operator. Our analysis also shows that an intuitive discretization of the infinite-dimensional operator could lead to a sub-optimal statistical learning rate. Then, I'll discuss how, by suitably trading-off bias and variance, we can construct an estimator with an optimal learning rate for learning a linear operator between infinite dimension spaces. We also illustrate how this theory can inspire a multilevel machine-learning algorithm of potential practical use.

For the second learning task, we focus on variational formulations for differential equation models. We discuss a prototypical Poisson equation. We provide a minimax lower bound for this problem. Based on the lower bounds, we discover that the variance in the direct plug-in estimator makes sample complexity suboptimal. We also consider the optimization dynamic for different variational forms. Finally, based on our theory, we explain an implicit acceleration of using a Sobolev norm as the objective function for training.

If time permitted, I'll also briefly talk about the statistical limit of debiasing machine learning algorithm for scientific computing and how it's relates to rare event.

Wavelet Galerkin Method for an Electromagnetic Scattering Problem
Michelle Michelle, Purdue University

The Helmholtz equation is challenging to solve numerically due to the pollution effect, which often results in a huge ill-conditioned linear system. We present a high order wavelet Galerkin method to numerically solve an electromagnetic scattering from a large cavity problem modeled by the 2D Helmholtz equation. The high approximation order and the sparse stable linear system offered by wavelets are useful in dealing with the pollution effect. By using the direct approach presented in our past work, we present various optimized spline biorthogonal wavelets on a bounded interval. We provide a self-contained proof to show that the tensor product of such wavelets forms a 2D Riesz wavelet in the appropriate Sobolev space. Compared to the coefficient matrix of a standard Galerkin method, when an iterative scheme is applied to the coefficient matrix of our wavelet Galerkin method, much fewer iterations are needed for the relative residuals to be within a tolerance level. Furthermore, for a fixed wavenumber, the number of required iterations is practically independent of the size of the wavelet coefficient matrix. In contrast, when an iterative scheme is applied to the coefficient matrix of a standard Galerkin method, the number of required iterations doubles as the mesh size for each axis is halved. The implementation can also be done conveniently thanks to the simple structure, the refinability property, and the analytic expression of our wavelet bases.

Central-Upwind Schemes for Weakly Compressible Two-layer Shallow-Water Flows
Sarswati Shah, National Autonomous University of Mexico

We formulate a weakly compressible two-layer shallow water flows in channels with arbitrary cross sections. The standard approach for those flows results in a conditionally hyperbolic balance law with non-conservative products while the current model is unconditionally hyperbolic. A detailed description of the properties of the model is provided, including entropy inequalities and entropy stability. Furthermore, a high-resolution, non-oscillatory semi-discrete central-upwind scheme is presented. The scheme extends existing central-upwind semi-discrete numerical methods for hyperbolic balance laws. Along with the description of the scheme, we present several numerical experiments that demonstrate the robustness of the numerical algorithm

Extracting the geometry of low-dimensional secondary features of data Bogdan Toader, Yale University

Data often displays multiple sources of variability and there is a rich literature focusing on extracting useful low-dimensional features from such datasets. For example, multiple views of the data can be used to extract common features, while each view displays secondary low-dimensional features that remain unexplored using standard methods. In this work, we propose an autoencoder that extracts the geometry of such private features in an unsupervised manner. We show a number of simple examples that capture the main properties of this method, as well as its generalization power.

The Law of Parsimony in Gradient Descent for Learning Deep Linear Networks. Peng Wang, The University of Michigan

Over the past few years, an extensively studied phenomenon in training deep networks is the implicit bias of gradient descent towards parsimonious solutions. In this work, we investigate this phenomenon by narrowing our focus to deep linear networks. Through our analysis, we reveal a surprising ``law of parsimony'' in the learning dynamics when the data possesses low-dimensional structures. Specifically, we show that the evolution of gradient descent starting from orthogonal initialization only affects a minimal portion of singular vector spaces across all weight matrices. In other words, the learning process happens only within a small invariant subspace of each weight matrix, despite the fact that all weight parameters are updated throughout training. This simplicity allows us to better understand deep representation learning by elucidating the linear progressive separation and concentration of representations from shallow to deep layers.

Optimization on Manifolds via Graph Gaussian Processes

Ruiyi Yang, Princeton University

Optimization problems on smooth manifolds are ubiquitous in science and engineering. Oftentimes the manifolds are not known analytically and only available as an unstructured point cloud, so that gradient-based methods are not directly applicable. In this poster, we shall discuss a Bayesian optimization approach, which exploits a Gaussian process over the point cloud and an acquisition function to sequentially search for the global optimizer. Regret bounds are established and several numerical examples demonstrate the effectiveness of our method.

Learning Collective Behaviors from Observation

Ming Zhong, Illinois Institute of Technology

We present a family of learning methods to study collective behaviors such as clustering, flocking, swarming, synchronization. Our learning methods can derive physically meaningful dynamical systems from observation data. Furthermore, our methods are efficient (able to handle high-dimensional data in linear time and learn even the feature variables), effective (able to learn many different kinds of dynamical systems), and convergent. We present extensive numerical experiments to support our theoretical claims as well as a real data study into how AI can discover laws of planetary motion from NASA JPL's Horizon data of our solar system.