

---

**Scientific Machine Learning**  
**Apprentissage automatique scientifique**

(Org: **Ben Adcock** (Simon Fraser), **Alex Bihlo** (Memorial University), **Simone Brugiapaglia** (Concordia) and/et **Hamid Usefi** (Memorial University))

---

---

**ZIAD ALDIRANY**, Polytechnique Montreal

*Approximating the Operator of the Wave Equation using Deep Learning*

The solution of the wave equation is required in a wide variety of fields, such as seismology, electromagnetism, and acoustics. In the last few years, a number of deep learning methods have been developed for the solution of PDE-based problems, with the objective of producing techniques that are more flexible and faster than the traditional FEM, FD, FV approaches. Deep operator networks (DeepONet) attempt to solve PDEs by learning the inverse of the differential operator for a wide class of initial data, rather than learn a single solution. However, this approach is especially expensive for problems containing high frequencies, such as those with the linear wave equation.

For the approximation of the homogeneous wave equation, we present a neural network architecture that is based on the integral representation formula of the wave equation. This architecture yields a faster learning and a better generalization error when compared to the classical DeepONet architecture. Moreover, with the proposed architecture, a trained network can be retrained for solutions with higher frequencies which results in an efficient learning strategy for high frequency functions. Numerical results in 1D and 2D will be presented to analyze frequency dependent convergence of the proposed approach.

---

**JAVAD RAHIMPOUR ANARAKI**, University of Toronto

*Personalized Classifier Selection in EEG-based Brain-Computer Interfaces*

Cerebral palsy, affecting nearly 1 in 500 children globally, is a dominant cause of severe movement disabilities in children. Electroencephalogram (EEG) signals have successfully been utilized to provide alternative communication pathways through brain-computer interface (BCI) systems to help those children communicate their needs. The crucial component of a BCI is a classifier, which works in real-time to translate EEG signals into meaningful words or wheelchair commands, and its accuracy and speed are critical to the utility of BCI devices. However, there is significant inter-subject variability in the data; moreover, this variability affects classification accuracy and the choice of the best classifier for different individuals over time. This calls for a personalized medicine approach, with classifier selection automatically tailored to individuals and their current needs.

---

**AARON BERK**, McGill University

*On Lipschitzness of the solution mapping for LASSO programs*

Compressed sensing theory explains why LASSO programs recover structured high-dimensional signals with minimax order-optimal error. Yet, the optimal choice of the program's tuning parameter is often unknown in practice. It has not been fully clarified how variation of the governing parameter impacts recovery error in compressed sensing, which is otherwise provably stable and robust. We present a novel upper bound on the Lipschitz constant for the solution mapping of the unconstrained LASSO with respect to its tuning parameter, using tools from variational analysis. We show how this bound behaves in the setting of subgaussian measurement matrices with gaussian noise and contrast it against recent asymptotic results for parameter sensitivity in constrained LASSO and basis pursuit. In particular, we demonstrate that informed selection of a LASSO program can avoid sensitivity issues.

---

**RUDY BRECHT**, University of Bremen

*Deep learning for ensemble forecasting*

For today's weather prediction it is necessary to generate several perturbed numerical simulations for an ensemble prediction. However, running a sufficient number of such simulations to produce an ensemble is computationally costly. Moreover, statistical

post-processing is needed to further improve the ensemble quality. Thus, here we propose to learn the statistical properties of the ensemble using deep learning and demonstrate first results.

---

**ASHKAN EBADI**, National Research Council Canada  
*Few-Shot Detection of COVID-19 Infection from Medical Images*

Since the beginning of 2020, the COVID-19 pandemic has had an enormous impact on global healthcare systems, and there has not been a region or domain that has not felt its impact in one way or another. The gold standard of COVID-19 screening is the reverse transcription-polymerase chain reaction (RT-PCR) test. With RT-PCR being laborious and time-consuming, much work has gone into exploring other possible screening tools to observe abnormalities in medical images using deep neural network architectures. But, such deep neural network-based solutions require a large amount of labelled data for training. In this talk, I'll first briefly introduce the few-shot learning approach in which models are built such that they can adapt to novel tasks based on small numbers of training examples. Next, we will see its application in a real-life example where we used few-shot learning strategies to build a model sensitive to COVID-19 positive cases, using a very limited set of annotated data. The model can generalize from a few examples by employing a unique structure to rank similarities between inputs without necessitating extensive retraining.

---

**MANUELA GIROTTI**, Saint Mary's University, Mila Institute  
*Convergence Analysis and Implicit Regularization of Feedback Alignment for Deep Linear Networks*

We consider the Feedback Alignment algorithm, an bio-plausible alternative to backpropagation for training neural networks, and we analyze (1) convergence rates for deep linear networks and (2) incremental learning phenomena for shallow linear networks. Interestingly, depending on the initialization, the principal components of the model may be learned first (implicit regularization) or after (implicit anti-regularization) the negligible ones, thus affecting the effectiveness of the learning process.

---

**CRAIG GROSS**, Michigan State University  
*Sparsifying high-dimensional, multiscale Fourier spectral methods*

Fast Fourier transforms (FFT) have made Fourier spectral methods extremely popular for solving partial differential equations (PDE). However, when very fine frequency scales of PDE data and solutions need to be resolved, the superlinear dependence on bandwidth in FFTs' computational complexity makes traditional spectral methods infeasible. The exponential dependence on the spatial dimension of the problem only exacerbates this computational intractability.

Sparse Fourier transforms (SFT) on the other hand have enjoyed great success at computing univariate functions' most significant frequency data while running with computational complexity sublinear in the bandwidth. This talk will discuss the extension of SFTs to high-dimensions, where the emphasis on sparsity both bypasses the curse of dimensionality and superlinear dependence on wide frequency bands. These techniques then allow for the sparsification of a traditional spectral method. We present an adaptive algorithm for quickly solving extremely high-dimensional, multiscale diffusion equations. The algorithm is furnished with error guarantees on the solution in terms of the Fourier-compressibility of the PDE data and the ellipticity of the problem.

---

**ARMIN HATEFI**, Memorial University of Newfoundland  
*Unsupervised Shrinkage Estimation Methods for Mixture of Regression Models*

In many applications (e.g., medical studies), the population of interest (e.g., disease status) comprises heterogeneous subpopulations. The mixture of probabilistic regression models is one of the most common model-based techniques to incorporate the information of covariates into learning of the population heterogeneity. Despite the flexibility, the model leads to misleading and unreliable estimates in the presence of a high multicollinearity problem. We developed two shrinkage methods through an unsupervised learning approach to estimate the model parameters even in the presence of multicollinearity issues. The performance of the developed methods is evaluated via classification and stochastic versions of EM algorithms. The numerical

studies show the proposed methods outperform their maximum likelihood counterparts. Finally, the developed methods are applied to analyze the bone mineral data of women aged 50 and older.

---

**GEOFFREY MCGREGOR**, University of Northern British Columbia  
*Conservative Hamiltonian Monte Carlo*

Markov Chain Monte Carlo (MCMC) methods enable meaningful extraction of statistics from complex distributions, frequently appearing in applications such as parameter estimation, Bayesian statistics, statistical mechanics and machine learning. The main goal of MCMC is to generate a sequence of samples which converges to a specified stationary distribution. However, as the dimensionality of the target distribution increases, the convergence rate of typical MCMC sequences toward the stationary distribution slows down dramatically. This has led to recent developments in computational techniques, such as Hamiltonian Monte Carlo (HMC), to improve on the performance in convergence and acceptance rate of proposed samples by solving specific Hamiltonian systems using symplectic methods. Nonetheless, maintaining high acceptance rates using HMC in high dimensions remains a significant challenge. In this talk, we introduce the Conservative Hamiltonian Monte Carlo (CHMC) method, which alternatively utilizes an energy-preserving numerical method, known as the Discrete Multiplier Method. We show that CHMC converges to the correct stationary distribution under appropriate conditions, and provide numerical examples showcasing improvements in acceptance rates and in scaling for high dimensional problems.

This is joint work with Andy Wan from the University of Northern British Columbia.

---

**PARDIS SEMNANI**, University of British Columbia  
*Log-concave Density Estimation in Undirected Graphical Models*

We study the problem of maximum likelihood estimation of densities that have a log-concave factorization according to a given undirected graph  $G$ . We show that the maximum likelihood estimate (MLE) exists and is unique with probability 1 as long as the number of samples is larger than the smallest size of a maximal clique in a chordal cover of the graph  $G$ . Furthermore, we show that the MLE is the product of the exponentials of several tent functions, one for each maximal clique of the graph. While the set of log-concave densities in a graphical model is infinite-dimensional, our results imply that the MLE can be found by solving a finite-dimensional convex optimization problem. Finally, we discuss the conditions under which a log-concave function, which factorizes according to the maximal cliques of  $G$ , can be factorized in the same manner with log-concave clique potentials.

This talk is based on a joint work with Kaie Kubjas, Olga Kuznetsova, Elina Robeva, and Luca Sodomaco.

---

**YIFAN SUN**, Stony Brook University  
*Using flow analysis to accelerate the Frank-Wolfe method*

The Frank-Wolfe (FW) method is popular in sparse constrained optimization, due to its fast per-iteration complexity. However, the tradeoff is that its worst case global convergence is comparatively slow; without line search or specialized steps, the vanilla method converges at a rate of  $O(1/k)$ , even if the objective is strongly convex. However, we show that when the method is viewed as an Euler discretization of an underlying flow, the flow rate itself may be arbitrarily fast, reaching  $O(1/k^c)$  for any  $c > 0$ . Therefore, the slowdown of the FW method can be attributed directly to discretization error, which we show can be mitigated using two strategies: multistep methods, and weighted averaging. In the latter approach we prove an overall global convergence rate of  $O(1/k^p)$ , where  $0 \leq p \leq 1$ , which accelerates empirically to  $O(1/k^q)$  once the sparse manifold has been identified, for  $q \geq 3/2$ . In practice we also observe that the method achieves this accelerated rate from a very early stage, suggesting a promising mode of acceleration for this family of methods.

---

**GIANG TRAN**, University of Waterloo  
*SRMD: Sparse Random Mode Decomposition*

Signal decomposition and multiscale signal analysis provide many useful tools for time-frequency analysis. We proposed a random feature method for analyzing time-series data by constructing a sparse approximation to the spectrogram. The

randomization is both in the time window locations and the frequency sampling, which lowers the overall sampling and computational cost. The sparsification of the spectrogram leads to a sharp separation between time-frequency clusters which makes it easier to identify intrinsic modes, and thus leads to a new data-driven mode decomposition. The applications include signal representation, outlier removal, and mode decomposition. On the benchmark tests, we show that our approach outperforms other state-of-the-art decomposition methods.

---

**TERRENCE TRICCO**, Memorial University

*Synthetic generation of multi-modal discrete time series using transformers*

Synthetic data is artificially generated data that contains the same properties as the real data it mimics. There are many benefits to synthetic data. For instance, it allows for the generation of significant quantities of data, more than may be possible through real data collection. Similarly, rare classes or labels can be generated which may be useful for data augmentation or transfer learning. Synthetic data may also be valuable where there exists privacy or security concerns. In this work, we have used a transformer-based model to generate multi-modal, discrete time series data with application to personal financial data. Our data contains multiple classes of events in a highly irregular temporal sequence, where each event may operate on its own timescale with simultaneous dependence upon other events. We have built our architecture and encoded our features specifically to handle multiple patterns found within date and time features. Our transformer-based results have been compared with results from the generative adversarial network (GAN) model, DoppleGANger.

---

**WEIQI WANG**, Concordia University

*Compressive Fourier collocation methods for high-dimensional diffusion equations with periodic boundary conditions*

High-dimensional Partial Differential Equations (PDEs) are a popular mathematical modelling tool, with applications ranging from finance to computational chemistry. However, standard numerical techniques for solving these PDEs are typically affected by the curse of dimensionality. In this work, we tackle this challenge while focusing on stationary diffusion equations defined over a high-dimensional domain with periodic boundary conditions. Inspired by recent progress in sparse function approximation in high dimensions, we propose a new method called compressive Fourier collocation. Combining ideas from compressive sensing and spectral collocation, our method replaces the use of structured collocation grids with Monte Carlo sampling and employs sparse recovery techniques, such as orthogonal matching pursuit and  $\ell^1$  minimization, to approximate the Fourier coefficients of the PDE solution. We conduct a rigorous theoretical analysis showing that the approximation error of the proposed method is comparable with the best  $s$ -term approximation (with respect to the Fourier basis) to the solution. Using the recently introduced framework of random sampling in bounded Riesz systems, our analysis shows that the compressive Fourier collocation method mitigates the curse of dimensionality with respect to the number of collocation points under sufficient conditions on the regularity of the diffusion coefficient. We also present numerical experiments that illustrate the accuracy and stability of the method for the approximation of sparse and compressible solutions.