CRG workshop

Mathematical foundations of scientific machine learning

June 1–3, 2022

Memorial University of Newfoundland, Canada

(Remote participation via Webex possible)

Schedule

Wednesday June 1

Talks are held at **ED–2018A**. Webex webinar: Wednesday link (Password: GTjYFeZf786).

Time	Speaker	Title
Chair: Scott MacLachlan		
9:55–10:00	Alex Bihlo	Welcome
10:00–11:00	Simone Brugiapaglia **(plenary)**	Two case studies in the mathematical foundations of deep learning: rating impossibility and practical existence theorems
11:00–11:15	**Coffee break**	
11:15–12:15	Aaron Berk	Towards generative compressed sensing via random sampling in bounded orthonormal systems
12:15–13:00	**Lunch break**	
Chair: Alison Malcolm		
13:00–14:00	Scott MacLachlan	Optimization and Learning in the Design of Preconditioners
14:00–14:30	Ziad Aldirany	Approximating the operator of the wave equation using deep learning
14:30–15:00	**Coffee break**	
Chair: Yifan Sun		
15:00–16:00	Weiqi Wang	Compressive Fourier collocation methods for high-dimensional diffusion equations with periodic boundary conditions
16:00–17:00	JC Loredo-Osti	Calibrating a stochastic model for an infectious disease outbreak in a small population
17:00–18:00	Nick Dexter (remote)	Efficient sparse recovery and neural network approximation for high-dimensional scientific machine learning

Thursday June 2

Talks are held at **ED–2018A**. Webex webinar: Thursday link (Password: nXacXEbm339).

Time	Speaker	Title
Chair: Hamid Usefi		
10:00–11:00	Yifan Sun **(plenary)**	Continuous time trajectories of optimization methods
11:00–11:15	**Coffee break**	
11:15–11:35	Ince Husain	Understanding machine learning generated molecular representations
11:35–12:00	Sina Mohammad-Taheri	Lasso-inspired variants of weighted orthogonal matching pursuit with applications to sparse high-dimensional approximation
12:00–13:00	**Lunch break**	
Chair: Rudy Brecht		
13:00–14:00	Hamid Usefi	Clustering, multicollinearity, and singular vectors
14:00–14:30	Changxiao Sun	Normalizing flows in seismic full waveform inversion
14:30–15:00	Subedika Debbarma	Using the attention mechanism for understanding frequency contribution in earthquake magnitude estimation
15:00–15:30	**Coffee break**	
Chair: Simone Brugiapaglia		
15:30–16:30	Rudy Brecht	Predicting model uncertainty using deep neural networks
16:30–17:30	Jingjing Zheng	Handling slice permutations and transpose variability in tensor Recovery
17:30–18:00	Peter Lelievre (remote)	Inversion of geophysical data on unstructured meshes using deep learning neural networks

Friday June 3

Talks are held at **ED–2018A**. Webex webinar: Friday link (Password: tPr7R6xb8HN).

Time	Speaker	Title
Chair: Terrence Tricco		
10:00–11:00	Amer El-Samman	Interpretability of graph convolutional neural networks on molecules
11:00–11:15	**Coffee break**	
11:15–12:15	Kyle Nickerson	Deep generative models for creating synthetic transaction sequences and other complex data
12:15–12:35	Ayesha Rathnayake	Robust principal component regression analysis

Abstracts

Simone Brugiapaglia (Concordia University) – plenary

Two case studies in the mathematical foundations of deep learning: rating impossibility and practical existence theorems.

Deep learning is having a profound impact on scientific research. Yet, while deep neural networks continue to be applied to and show impressive performance in a wide variety of fields including scientific computing, their mathematical foundations are far from being well understood. In this talk, we will present recent developments in this direction by discussing two case studies.

First, motivated by applications in cognitive science, we will illustrate some "rating impossibility theorems". These theorems can identify frameworks where neural networks are provably unable to generalize outside the training set in the seemingly simple scenario of learning "identity effects", i.e. classifiying whether two objects are identical or not.

Then, motivated by the construction of surrogate models in scientific computing, we will present so-called "practical existence theorems". These theorems combine approximation-theorical results for deep neural networks with sparse high-dimensional polynomial approximation methods based on compressed sensing. They provide sufficient conditions on the network architecture, the training optimization problem, and the number of samples able to guarantee accurate approximation of (Hilbert-valued) holomorphic functions of many variables.

In both cases, we will illustrate numerical examples that either validate or "challenge" our theorems. We will conclude by discussing current research projects and open questions.

Aaron Berk (McGill University)

Towards generative compressed sensing via random sampling in bounded orthonormal systems

In a work by Bora, Jalal, Price & Dimakis, a mathematical framework was developed for compressed sensing guarantees in the setting where the measurement matrix is Gaussian and the signal structure is the range of an L-Lipschitz function (with applications to generative neural networks). The problem of compressed sensing with generative models has since been extensively analyzed when the measurement matrix and/or network weights follow a (sub-)Gaussian distribution. In this talk, we outline a framework for moving beyond the sub-Gaussian assumption, to measurement matrices that are derived by randomly sampling rows of a unitary matrix corresponding to a bounded orthonormal system. Our framework uses generic chaining ideas clarified by Dirksen (2015) and counting arguments for linear regions of ReLU networks elucidated, for instance, by Naderi & Plan (2021). After providing an overview of our progress so far, we outline possible avenues of continuation and solicit audience discussion.

Scott MacLachlan (Memorial University of Newfoundland)

Optimization and Learning in the Design of Preconditioners

Computer simulation algorithms are a major tool in many areas of science and industry, particularly in areas where the behaviour of fluids or complex materials governs the physical processes of interest. A typical core of these tools is the numerical approximation of the solution to coupled nonlinear systems of partial differential equations, relying on nonlinear and linear solvers, such as Newton's method and preconditioned Krylov iterations. Among the most effective preconditioners for these systems are multigrid and domain decomposition methods, which use multiscale representations of the systems to be solved to achieve linear-scaling complexity for the solution of these linear systems. These preconditioners typically rely on heuristics in their construction, to approximate solutions to underlying combinatorial (and other) optimization problems that specify parameters and other components of the preconditioners, based on the discrete problem to which they are being applied. In this talk, I will discuss the use of advanced optimization and machine learning techniques to approximately solve these optimization problems and the impact these techniques can have on advanced preconditioner design.

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.

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 ℓ^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.

JC Loredo-Osti (Memorial University of Newfoundland)

Calibrating a stochastic model for an infectious disease outbreak in a small population

There are many ways to model the progression of an infectious disease when the process has already started and the population is large. The models go from simple Lotka-Volterra-like models to those highly compartmentalised. These models can be deterministic or stochastic. In this presentation, we present some of the most common stochastic models used to model an infectious disease outbreak and use Newfoundland and Labrador public data to illustrate their performance.

Nick Dexter (Simon Fraser University)

Efficient sparse recovery and neural network approximation for high-dimensional scientific machine learning

Sparse reconstruction techniques from compressed sensing have been successfully applied to many application areas, including signal processing, inverse problems in imaging, and approximation of solutions to parameterized partial differential equations (PDE). Such approaches are capable of exploiting the sparsity of the signal to achieve highly accurate approximations with minimal sample complexity. For problems whose solutions possess a great deal of structure, their recovery properties can be further enhanced through a combination of carefully selected weighting or structured sampling schemes. Recently connections between compressed sensing and deep learning have been explored, and the existence of deep neural network (DNN) architectures which achieve the same sample complexity and accuracy as compressed sensing on function approximation problems have been established. In this work, we further explore these connections and sparse neural network approximation in the context of high-dimensional parameterized PDE problems. We provide a full error analysis for such problems, explicitly accounting for the errors of best approximation (describing DNN expressibility), spatial discretization of the PDE, and the algorithm used in solving the underlying optimization problem. We complement our theoretical contributions with detailed numerical experiments, demonstrating the potential for sparse neural network approximation in scientific machine learning contexts.

Yifan Sun (SUNY Stony Brock) – plenary

Continuous time trajectories of optimization methods

Understanding the behavior of optimization methods is a key component in unraveling the training behavior of machine learning tools, including deep neural networks. In general, convergence proofs are the main standard for giving guarantees on method behavior and robustness; however, these proofs may be very difficult to construct, verify, and use in gaining intuition. Recently, there has been increased interest in using continuous-time analysis, in which optimization methods are viewed as Euler discretizations of continuous flows. The flow itself is then derived and analyzed using standard dynamical systems tools, often allowing for simpler analysis, shorter proofs, and more intuition. Additionally, in cases where the flow rate is faster than the best known method rate, higher order discretization schemes can be used to mitigate discretization error, which can begin to close this gap and produce improved methods, for cheap additional overhead. In this talk, I will walk through the flow analysis in three scenarios: gradient descent on quadratic problems, momentum-based acceleration, and projected gradient descent over general convex problems and sets. While this study is still fledgling, it can hopefully inspire increased interest in using such tools for larger, more complex machine learning scenarios.

Ince Husain (University of New Brunswick)

Understanding machine learning generated molecular representations

With the rise of machine learning (ML) techniques for scientific prediction, computational chemists are interested in understanding how machine learning algorithms represent molecules. Specifically, there is interest in investigating whether ML-generated representations of molecules can be related to those used in computational techniques that are traditional to computational chemistry. In this study, this is examined by comparing machine learning representations of molecules to those used in Density Functional Theory (DFT) calculations. The neural network SchNet generates molecular representations in the form of 'embedding vectors', while DFT represents molecules by their spatial charge densities. These representations were visualized and statistically compared. Results from this study suggest that, although SchNet was trained on DFT datasets, it generates more specific molecular representations than DFT that adhere more strongly to traditional chemical understanding. Ideas for further study to interpret machine learning representations of molecules will be discussed.

Sina Mohammad-Taheri (Concordia University)

Lasso-inspired variants of weighted orthogonal matching pursuit with applications to sparse high-dimensional approximation

Motivated by recent developments in sparse high-dimensional approximation from Monte Carlo sampling, we propose new weighted generalizations of the Orthogonal Matching Pursuit (OMP) algorithm. Greedy algorithms of this type are more computationally efficient than convex optimization-based methods for small values of the target sparsity and offer a promising way to mitigate the curse of dimensionality. In this work, we propose new theoretically-justified greedy selection criteria that are inspired by variants of the LASSO optimization program. A key issue

is the robustness of the optimal choice of the tuning parameter with respect to the measurement noise, which is realized by the square-root LASSO program in the context of convex optimization. We investigate how this property is carried over into the context of LASSO-based OMP methods. Conducting numerical experiments in high-dimensional polynomial approximation, we show the efficacy of the proposed algorithms by studying the recovery error as a function of the algorithm iterations. Moreover, we illustrate settings where the optimal choice of the tuning parameter is more robust against the noise.

Hamid Usefi (Memorial University of Newfoundland)

Clustering, multicollinearity, and singular vectors

In this talk, we shall discuss how to find colinearities between columns of a matrix. Let A be a matrix and denote by $A^\dagger$ its (Moore-Penrose) pseudo-inverse. We show that up to a permutation of columns of A, the projector matrix $P = I - A^\dagger A$ has a block diagonal form. Furthermore, the connected components of the graph associated to P correspond to clusters of columns of A so that columns in a cluster correlate only with the columns in the same cluster. We discuss some applications to solutions of least square problems as well as supervised learning.

Changxiao Sun, Alison Malcolm (Memorial University of Newfoundland) and Rajiv Kumar (Schlumberger)

Normalizing flows in seismic full waveform inversion

The normalizing flows method has gained popularity recently in the machine learning field. Normalizing flows are trained to transfer a simple analytical probability distribution to a more complex distribution through a set of invertible and differential transforms. Compared with Markov Chain Monte Carlo (MCMC) methods, we are able to provide an approximation of the posterior distribution, rather than an ensemble of posterior samples. In this presentation, we show how we use normalizing flows to solve the full waveform inversion (FWI) problem, a high-precision seismic imaging method seismic data. To facilitate the computations, we use a local wave-equation solver that allows us to solve the wave equation in a small subset of our model. This allows us to compute an estimate of the uncertainty in our final images, rather than just a single image.

Subedika Debbarma and Alison Malcolm (Memorial University of Newfoundland)

Using the attention mechanism for understanding frequency contribution in earthquake Magnitude Estimation

The magnitude of an earthquake is defined as the energy it emits during rupture. To measure the magnitude, there are many scales currently in use,the most common being: the Richter scale and the moment magnitude scale. The Richter scale (also known as the local magnitude scale) takes magnitude recordings from seismic stations across the globe. The moment magnitude scale measures the magnitude using the seismic moment or using the source physics of an

earthquake. In this study, we devised a machine learning model to estimate the local magnitude of an earthquake in the frequency domain. The dataset consists of readings from several Wood Anderson instruments (seismometers) stationed in various parts of the world. The time domain data were smoothed using a ten percent Hanning taper and transformed using the Fast Fourier Transform. Our goal is to understand the individual contribution of the frequencies in estimating the magnitude which has not been previously investigated. We replace convolutional neural networks (CNN) used in prior studies with attention based bidirectional recurrent neural networks (RNN) for performing 'informed' dimensionality reduction. Our model estimates are close to those computed using CNNs but give us additional information as to why certain readings were omitted.

Rudy Brecht (University of Bremen)

Predicting model uncertainty using deep neural networks

Ensemble prediction systems are an invaluable tool for weather forecasting. Practically, ensemble predictions are obtained by running several perturbations of the deterministic control forecast. However, ensemble prediction is associated with a high computational cost and often involves statistical post-processing steps to improve its quality. Here we propose to use deep-learning-based algorithms to learn the statistical properties of an ensemble prediction system, the ensemble spread, given only the deterministic control forecast. Thus, once trained, the costly ensemble prediction system will not be needed anymore to obtain future ensemble forecasts, and the statistical properties of the ensemble can be derived from a single deterministic forecast. We adapt the classical `pix2pix` architecture to a three-dimensional model and also experiment with a shared latent space encoder–decoder model, and train them against several years of operational (ensemble) weather forecasts for the 500 hPa geopotential height. The results demonstrate that the trained models indeed allow obtaining a highly accurate ensemble spread from the control forecast only.

Jingjing Zheng (Memorial University of Newfoundland)

Handling slice permutations and transpose variability in tensor recovery

With the growing explosion of high-dimensional tensor data, such as color images and videos, low-rank tensor recovery methods are getting more popular for exploiting intrinsic low-rankness within the data. One interesting question in the tensor recovery is asked: are the ranks of transposed tensors or tensors with slice permutations equivalence? The answer is no! It means the low-rankness prior information within the data cannot be utilized in an effective way when we use tensor recovery to deal with the tensor data under certain transpose operators or slice permutations. These two problems are called Tensor Variability (TV) and Slice Permutations Variability (SPV) of tensor rank. To solve TV, the influence caused by tensor transpose operators for a new tensor rank called Weighted Tensor Average Rank (WTAR) is proposed to learn the low rankness of tensor data from different directions. Besides, SPV of several key tensor recovery problems theoretically and experimentally are discussed. The obtained conclusion shows that there is a huge gap between tensor recovery results by slice permutations. To overcome SPV, a novel tensor recovery algorithm by Minimum Hamiltonian Circle for SPV (TRSPV), is

developed, which exploits low dimensional subspace structures within data tensor more exactly. The experimental results demonstrate the effectiveness of the proposed methods in solving TV and eliminating SPV of tensor recovery.

Peter Lelievre (Mount Allison University)

Inversion of geophysical data on unstructured meshes using deep learning neural networks

We have applied neural networks to the geophysical inverse problem of assessing buried pipeline infrastructure using magnetic survey data. The underground volume of interest was discretized using rectilinear or unstructured tetrahedral meshes, with piecewise magnetization in each mesh cell, which is the typical approach for most geophysical inverse problems. The network was trained with a suite of models with pipelike bodies of different shapes, sizes and magnetizations. The subsequent ML-based inversion was able to recover reasonable representations of the ground truth for synthetic problems, and interpretable results for real data scenarios. In other work, we are developing methods for geophysical inversion that use a fundamentally different discretization and parameterization of the volume of interest: we represent anomalous bodies using wireframe surfaces comprising tessellated triangles, and we define control nodes for changing the geometry of those surfaces. The surfaces represent the interfaces between different material units, for example between different types of rocks, or between soil and pipeline infrastructure. The task of the inversion is to move the surfaces in some estimated initial model to better fit the measured geophysical data. Ongoing work has successfully applied these surface-geometry inversion (SGI) methods to mineral exploration problems using magnetic, gravity and electromagnetic data. Currently, we solve SGI problems using a genetic algorithm, and when computationally feasible we follow this with the Metropolis-Hastings algorithm to estimate uncertainties on the surface geometry. We are now considering the application of ML methods to SGI. The intention of this talk is to not to delve into the ML details of our previous work, but rather to present these applied scenarios and initiate discussion of possibilities for future

Amer El-Samman (University of New Brunswick)

Interpretability of graph convolutional neural networks on molecules

Deep learning is promising an efficient exploration of the enormous chemical space, the endless number of molecular possibilities that one can build from the periodic table of elements. One of the most recent designs, SchNet graph convolutional neural network, uses molecular graphs as input for prediction of molecular properties, such as the total energy. SchNet resolves highly accurate total energies without invoking any physical or chemical principles but simply by looking at the coordinates of each element in a molecule. For this reason, it is unclear if the algorithm is interpretable in terms of fundamental concepts or just a random “black box” fitting tool. In this work, we show that the internal layer activations of SchNet build a highly interpretable and useful representation of chemistry. The representation hinges on the fundamental functional group concept of chemistry. These chemical fragments, a.k.a functional groups, are revealed by the high-dimensional layer activations of the neural network and signify the parts of the molecule that are deemed significant for the algorithm’s decision making. The representation is

highly structured and can be explored with basic linear algebra arithmetic, thus possibly giving an alternative to exploring chemical changes and reactivity using the internal activation vectors of the algorithm. The space also reveals concepts of molecular similarity and may be used to evaluate the notion of "distance" between two molecules. Current work is investigating the structure of this space and how it may be beneficial for understanding chemistry and potential of AI technology.

Kyle Nickerson (Memorial University of Newfoundland)

Deep generative models for creating synthetic transaction sequences and other complex data

Synthetic data are artificially generated data that closely model real-world measurements, and can be a valuable substitute for real data in domains where it is costly to obtain real data, or privacy concerns exist. Synthetic data has traditionally been generated using computational simulations, but deep generative models (DGMs) are increasingly used to generate high-quality synthetic data. While DGMs have produced impressive synthetic samples from many domains — including images, written language, and audio — challenges still exist in generating high-quality synthetic samples in other domains. In this talk, we provide an overview of the capabilities of current DGMs, and introduce Banksformer, a novel DGM that we developed for generating synthetic sequences of banking transactions and other temporal event sequences. We will discuss the unique challenges of modeling this type of data with DGMs, and how we designed Banksformer to tackle these challenges. Additionally, we will discuss challenges in evaluating synthetic data, and present results comparing Banksformer to other state-of-the-art DGMs.

Ayesha Rathnayake and Asokan Variyath (Memorial University of Newfoundland)

Robust Principal Component Regression Analysis

Many real-time applications use large data sets with high dimensions that can be challenging in data analysis. The complexity of the model can be reduced by selecting the most important features that are strongly related to the response variable. Model selection using regression analysis is the most popular technique, but the presence of multicollinearity can adversely influence the selection process. Principal component analysis (PCA) produces independent, uncorrelated features that can be used to perform principal component regression. PCA can also be used as a dimension reduction method for higher-dimensional data while retaining as much variation as possible in the original data set. Outliers in real data can be observed for a variety of reasons. Since both regression analysis and PCA are sensitive to outliers, robust methods can be used to detect the outliers. A large number of simulations conducted to compare different approaches.