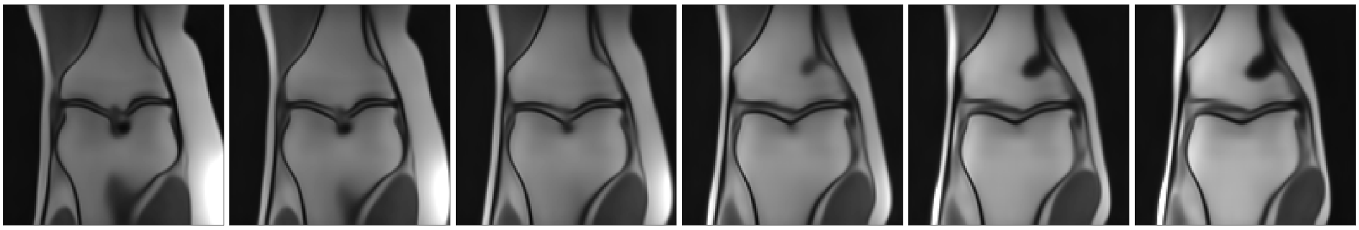
Mathematics for Deep Learning Opening Workshop



Program & abstracts

Contents

Abstracts of talks: **Thursday 13:30 - 15:00** **3**
- Andrew Fitzgibbon . . . 3
- Kweku Abraham . . . 4
- Riccardo Barbano . . . 4
- Teo Deveney . . . 5

Abstracts of talks: **Thursday 15:30 - 17:10** **6**
- Gitta Kutyniok . . . 6
- Subhadip Mukherjee . . . 6
- Kwinten van Weverberg . . . 7
- Lisa Kreusser . . . 7

Abstracts of talks: **Friday 9:30 - 10:30** **8**
- Jong Chul Ye . . . 8
- Keith Briggs . . . 8

Abstracts of talks: **Friday 11:00 - 12:40** **9**
- Weinan E . . . 9
- Yury Korolev . . . 9
- Margaret Duff . . . 10
- Sandeep Kaushik, Carolin Pirkl and Timo Schirmer . . . 10

Abstracts of talks: Thursday 13:30 - 15:00

Andrew Fitzgibbon

Graphcore

Provably Correct, Asymptotically Efficient, Higher-Order Reverse-Mode Automatic Differentiation

Automatic differentiation is a crucial part of modern machine learning, and has proven its worth in numerous implementations and papers. However, to date we know of no account that has all these properties:

- Asymptotically efficient. It runs in linear time, with only a constant-factor slow-down relative to the original (or "primal") program.
- Expressive – accepts a higher-order language with first-class anonymous functions; with sum types as well as products and arrays; and with functions and data types.
- Compositional. If a program consists of many function definitions, we want to differentiate them separately, and compose the results.
- Unified. Computes reverse and forward derivatives in the same formulation.
- Provably correct.

To attain these goals, we show how, starting with a simple but inefficient forward-mode formulation, we can exhibit a series of simple steps that take us to a weird-but-very-efficient system to produce reverse derivatives.

Joint work with: Faustyna Krawiec, Simon Peyton Jones, Neel Krishnaswami, Tom Ellis, and Richard Eisenberg.

Kweku Abraham

University of Cambridge

Deep Gaussian processes for PDE inverse problems

Inspired by successes of deep neural networks, researchers in Bayesian statistics have started considering ways of introducing depth to priors, for example by composing Gaussian processes. Recent works have shown the "deep GPs" thus created can perform well in direct problems. Given an expanding body of work demonstrating the effectiveness of Bayesian methods in PDE inverse problems, it is interesting to assess in such settings whether deep GPs offer something more than single GP priors. I'll outline some progress towards results in this direction, based on joint work in progress with Neil Deo.

Riccardo Barbano

UCL

Scalable Uncertainty Quantification in Unsupervised Deep Learning Methods for CT Reconstruction

Existing deep-learning based tomographic image reconstruction methods do not provide accurate estimates of reconstruction uncertainty, hindering their real-world deployment. To address this limitation, we construct a Bayesian prior for tomographic reconstruction, which combines the classical total variation (TV) regulariser with the modern deep image prior (DIP). Specifically, we use a change of variables to connect our prior beliefs on the image TV semi-norm with the hyper-parameters of the DIP network. For the inference, we develop an approach based on the linearised Laplace method, which is scalable to high-dimensional settings. The resulting framework provides pixelwise uncertainty estimates and a marginal likelihood objective for hyperparameter optimisation. We demonstrate the method on synthetic and real-measured high-resolution CT data, and show that it provides superior calibration of uncertainty estimates relative to previous probabilistic formulations of the DIP.

Teo Deveney

University of Bath

Adaptive deep learning PDE surrogates to accelerate Bayesian inference with guaranteed accuracy

Neural networks have proven remarkably capable in the approximation of solutions to PDEs, with several deep learning based methods demonstrating impressive results even in very high dimensions. Parallel to this, attention has also been paid to the approximation of solutions to parametric PDEs, where the neural network defines a mapping from the inputs of a PDE and its parameters, to (an approximation of) the solution corresponding to those inputs and parameters. Such a mapping can be used as a cheap surrogate in MCMC for Bayesian inverse problems over high-dimensional parameter spaces, however the accuracy of the approximated posterior is not typically easy to determine using the information provided during training. In this talk we outline an adaptive hybrid method for Bayesian inference. This method simultaneously approximates the posterior distribution and uses this as a weighting over which to train the forward solution. We show that this adaptive training scheme results in more accurate approximations to the forward and inverse problem when used as a surrogate for MCMC. We then use delayed acceptance MCMC to combine this surrogate with a finite difference (FD) solver and sample from the posterior. This is achieved such that the surrogate is applied to efficiently generate decorrelated MCMC proposals, while FD is used as a form of validation for promising proposals that ensures the accuracy of the approximated posterior is guaranteed at the level of the FD solver. We highlight that this hybrid approach provides significant computational savings when compared to methods based solely on traditional solvers whilst guaranteeing the same level of accuracy.

Abstracts of talks: Thursday 15:30 - 17:10

Gitta Kutyniok

LMU

The World of Graph Neural Networks: From the Mystery of Generalization to Foundational Limitations

The tremendous importance of graph structured data due to recommender systems or social networks led to the introduction of graph neural networks (GNNs). After a general introduction to GNNs, we will discuss results about their amazing generalization capabilities. We will study the more specialized question to which extent GNNs are able to generalize to graphs, which describe a similar phenomenon as present in the training data set, as well as the fully general problem. We will present results for both message passing and spectral GNNs. We will finish with a word of caution when training GNNs on classical digital hardware, and present fundamental limitations.

Subhadip Mukherjee

University of Cambridge

Adversarial regularization for imaging inverse problems

Inverse problems are encountered in applications where an unknown image needs to be recovered from its noisy and indirect observation. Variational regularization with handcrafted regularizers has by far been the most widely adopted approach for imaging inverse problems. While such regularizers are well-understood mathematically, they are generic and cannot adapt well to a specific application. The success of deep learning has motivated the quest for data-adaptive regularizers with tunable parameters that can be learned from available training images. Data-driven regularizers have been shown to outperform their analytical variants in different applications. This talk will focus on the recent advances in data-driven variational regularization, both from algorithmic and theoretical perspectives. In particular, the adversarial regularization framework will be introduced and the corresponding learning protocols and network constructions will be discussed. The talk will highlight the trade-off between mathematical guarantees and empirical performance using the examples of convex adversarial regularizers and a recent optimal transport-based approach that combines adversarial regularizers with end-to-end algorithm unrolling.

Kwinten van Weverberg

Met Office

Machine Learning Applications in the Met Office Clouds and Radiation Team

In this talk, we will go through a number of recent machine learning applications that are being explored in the Atmospheric Processes and Parametrizations Cloud and Radiation Team at the Met Office. Any process that is not resolved by a weather prediction or climate model's grid (typically in the order of kilometres to hundreds of kilometres), but has an impact on the resolved state of the atmosphere, needs to be accounted for by a so-called parametrization. This is for instance the case for the formation of clouds and precipitation, or the interaction between clouds and radiation. Most parametrizations build on a combination of theory and empirical relations between the small, unresolved and the large-scale, resolved state. For many processes that require parametrizations in models, theory is not very well defined. Other parametrizations have a good foundation in theory, but are very complex and expensive to incorporate in operational simulations. Hence, parametrizations are good candidates for machine learning applications to either rationalise computation time, or fill in gaps in the knowlegde of uncertain physical processes. We will explore a number of parametrizations or relations with parametrizations that can be replaced by machine learning algorithms, trained either on long model integrations, or on high-resolution observational data.

Lisa Kreusser

University of Bath

Generalised Eikonal equations on graphs with applications to semi-supervised learning

Many computational methods for semi-supervised and unsupervised classification are based on variational models and PDEs. Since shortest path graph distances are widely used in data science and machine learning, it is natural to introduce the concept of information propagation to data classification and semi-supervised learning. The success of eikonal equations in the continuum setting motivates the development of similar tools on graphs. We propose and unify classes of different models for information propagation over graphs, and prove equivalences between them. Motivated by the connection between first arrival time model and the eikonal equation in the continuum setting, we derive mean field limits for graphs based on uniform grids in Euclidean space under grid refinement. For a specific parameter setting, we demonstrate that the solution on the grid approximates the Euclidean distance. Finally, we illustrate the use of front propagation on graphs to semi-supervised learning.

Abstracts of talks: Friday 9:30 - 10:30

Jong Chul Ye

KAIST

Accelerating Conditional Diffusion Models for Inverse Problems through Stochastic Contraction

Diffusion models have recently attained significant interest within the community owing to their strong performance as generative models. Furthermore, its application to inverse problems has demonstrated state-of-the-art performance. Unfortunately, diffusion models have a critical downside - they are inherently slow to sample from, needing a few thousand steps of iteration to generate images from pure Gaussian noise. In this work, we show that starting from Gaussian noise is unnecessary. Instead, starting from a single forward diffusion with better initialization significantly reduces the number of sampling steps in reverse conditional diffusion. This phenomenon is formally explained by the contraction theory of the stochastic difference equations like our conditional diffusion strategy - the alternating applications of reverse diffusion followed by a non-expansive data consistency step. The new sampling strategy, dubbed Come-Closer-Diffuse-Faster (CCDF), also reveals new insight into how the existing feed-forward neural network approaches for inverse problems can be synergistically combined with the diffusion models. Experimental results with super-resolution, image inpainting, and compressed sensing MRI demonstrate that our method can achieve state-of-the-art reconstruction performance at significantly reduced sampling steps.

Keith Briggs

BT

Applying AI to the management of 5G radio systems

In the AIMM project (a multi-partner collaboration under the CELTIC-Next scheme, initiated and partly led by BT), we are researching the integration of AI techniques into the management and control of 5G cellular networks. In this talk I will describe some of the AIMM activities, with a focus on simulation techniques for evaluating competing algorithms.
https://aimm.celticnext.eu/

Abstracts of talks: Friday 11:00 - 12:40

Weinan E

Peking University and Princeton University

The Mathematical Theory of Neural Network-based Machine Learning

The task of supervised learning is to approximate a function using a given set of data. In low dimensions, its mathematical theory has been established in classical numerical analysis and approximation theory in which the function spaces of interest (the Sobolev or Besov spaces), the order of the error and the convergence rate of the gradient-based algorithms are all well-understood. Direct extension of such a theory to high dimensions leads to estimates that suffer from the curse of dimensionality as well as degeneracy in the over-parametrized regime. In this talk, we attempt to put forward a unified mathematical framework for analyzing neural network-based machine learning in high dimension (and the over-parametrized regime). We illustrate this framework using kernel methods, shallow network models and deep network models. For each of these methods, we identify the right function spaces (for which the optimal complexity estimates and direct and inverse approximation theorems hold), prove optimal generalization error estimates and study the behavior of gradient decent dynamics.

Yury Korolev

University of Cambridge

Approximation properties of two-layer neural networks with values in a Banach space

Approximation properties of infinitely wide neural networks have been studied by several authors in the last few years. New function spaces have been introduced that consist of functions that can be efficiently (i.e., with dimension-independent rates) approximated by neural networks of finite width. Typically, these functions are supposed to act between Euclidean spaces, typically with a high-dimensional input space and a lower-dimensional output space. As neural networks gain popularity in inherently infinite-dimensional settings such as inverse problems and imaging, it becomes necessary to analyse the properties of neural networks as nonlinear operators acting between infinite-dimensional spaces. In this talk, I will present dimension-independent Monte-Carlo rates for neural networks acting between Banach spaces with a partial order (vector lattices), where the ReLU no linearity will be interpreted as the lattice operation of taking the positive part.

Margaret Duff

University of Bath

Regularising Inverse Problems with Generative Machine Learning Models

Deep neural network approaches to inverse imaging problems have produced impressive results in the last few years. Here we consider the use of generative models in a variational regularisation approach to inverse problems. Generative models learn, from observations, approximations to high-dimensional data distributions. The considered regularisers penalise images that are far from the range of a generative model that has learned to produce images similar to a training dataset. We name this family generative regularisers. In contrast to other data-driven approaches, generative regularisers do not require paired training data and are learned independently of the forward model. This makes the method very flexible in real-world scenarios where noise levels and forward model parameters may change. The success of generative regularisers depends on the quality of the generative model and so we propose a set of desired criteria to assess models and highlight avenues for future research.

Sandeep Kaushik, Carolin Pirkl and Timo Schirmer

GE Healthcare

DL - challenges and opportunities for medical MR imaging

Deep learning has offered a variety of new challenges and opportunities in medical imaging. While the quality of medical images has been subject to continuous incremental improvements over time, the introduction of DL has accelerated this process by orders of magnitude. The speed and depth of change require major adjustments of the path from idea to clinical practice. Examples of DL in MR Imaging: As MRI is an inherently slow technique, one major goal of current research efforts is to keep the MRI scan times tolerable short. One approach to this is compressive sensing by exploiting sparsity e.g. in the spatial image or temporal domain to reduce the number of required sampling points. In those a priori ill-posed reconstruction tasks, MR physics-informed DL has been shown to be a promising alternative to conventional model-based approaches, enabling higher acceleration rates, faster reconstruction times and/or higher image qualities. Besides finding a tolerable compromise in-between invested time and quality of data, extracting the relevant data from the images is another area of development. As an example, DL has shown the potential to convert MR images into CT-like images, which would allow for workflow improvements, cost savings and reduction of ionizing radiation.