The Alan Turing Institute

Mathematics of data
Structured representations for sensing, approximation and learning
29 – 31 May 2019

Abstracts

Approximation with deep networks
Remi Gribonval (Inria, France)
10:15 – 11:00, 29 May

We study the expressivity of deep neural networks. Measuring a network's complexity by its number of connections or by its number of neurons, we consider the class of functions for which the error of best approximation with networks of a given complexity decays at a certain rate when increasing the complexity budget. Using results from classical approximation theory, we show that this class can be endowed with a (quasi)-norm that makes it a linear function space, called approximation space. We establish that allowing the networks to have certain types of "skip connections" does not change the resulting approximation spaces. We also discuss the role of the network's nonlinearity (also known as activation function) on the resulting spaces, as well as the role of depth. For the popular ReLU nonlinearity and its powers, we relate the newly constructed spaces to classical Besov spaces. The established embeddings highlight that some functions of very low Besov smoothness can nevertheless be well approximated by neural networks, if these networks are sufficiently deep.

Joint work with Gitta Kutyniok (TU Berlin), Morten Nielsen (Aalborg University) and Felix Voigtlaender (KU Eichstätt).

Nonlinear approximation by deep ReLU networks
Ron DeVore (Texas A&M, USA)
11:00 – 11:45, 29 May

Despite its demonstrable success in learning environments, even it proponents agree that there is not an adequate theoretical foundation for understanding the advantages of deep neural networks. In the case of ReLU activation, these networks generate piecewise linear functions which are used to approximate a target function. If we consider networks with fixed width W but varying depth L, the output of such networks depend on roughly $n:=W^2L$ parameters. Thus, the output of these networks is a nonlinear manifold depending on n parameters. This talk will discuss what we know about the approximation properties of these deep ReLU networks and in particular in what sense are they more powerful then more traditional methods of manifold approximation depending on n parameters. In particular, we are interested in what sense they are more powerful than their shallow network cousins. The talk will concentrate on the approximation of univariate functions where there is the best chance for definitive results. This is a joint collaboration with Ingrid Daubechies, Simon Foucart, Boris Hanin, and Guergana Petrova.

Two decentralized learning problems: Sketching and policy evaluation
Justin Romberg (Georgia Institute of Technology, USA)
12:15 – 13:00, 29 May

We study how two fundamental problems in approximation and learning can be solved in a distributed setting. In the first, we consider the problem of sketching a low rank matrix when the columns are distributed across multiple computational nodes. We show that if each subset of columns is sketched independently, then the matrix can be recovered at a central location by solving low-rank recovery problem. We show that a novel convex relaxation of this problem results in optimal sample complexity bounds. This relaxation is tailored to the form of the matrix compression scheme being considered and also provides some general guidelines for relaxing the problem for other compression schemes.

In the second problem, we consider the policy evaluation problem in multi-agent reinforcement learning. While distributed reinforcement learning algorithms have been presented in the literature, almost nothing is known about their convergence rate. We provide such a rate for the well-known time differencing ("TD0") algorithm in the distributed setting.

Clustering and classification - From the core to the edge
Thomas Strohmer (University of California, Davis, USA)
14:30 – 15:15, 29 May

Organizing data into meaningful groups is one of the most fundamental tasks in data analysis and machine learning. The first part of my talk is devoted to the theoretical foundations of spectral clustering and graph cuts. Spectral clustering has arguably become the most popular clustering technique when the structure of the individual clusters is non-convex and/or highly anisotropic. Yet, despite its tremendous popularity, a rigorous and meaningful theoretic justification is still elusive. I will discuss a convex relaxation approach, which gives rise to a rigorous theoretical analysis of spectral clustering. We do this by deriving deterministic bounds of finding optimal graph cuts via a natural and intuitive spectral proximity condition. Moreover, our framework provides theoretical guarantees for community detection.
The second part of my talk could be entitled "What Happens on the Edge, Stays on the Edge". Here, we are motivated by the imperative to improve privacy in data analysis, and by the practical need to run AI on edge devices. These devices, which are often very limited in terms of memory, computing power, and battery life, are expected to form a significant component of the future Internet-of-Things. Inspired by ideas from compressive sensing and using mathematical concepts of structured subspace approximation, we propose a hybrid hardware-software framework that can substantially reduce the computational complexity of image classification.

The mother of all representer theorems for inverse problems and machine learning
Michael Unser (EPFL, Switzerland)
15:15 – 16:00, 29 May

Regularization addresses the ill-posedness of the training problem in machine learning or the reconstruction of a signal from a limited number of measurements. The standard strategy consists in augmenting the original cost functional by an energy that penalizes solutions with undesirable behavior. The effect of regularization is very well understood when the penalty involves a Hilbertian norm. Another popular configuration is the use of an l1-norm (or some

variant thereof) that favors sparse solutions. In this presentation, we present a general representer theorem that characterizes the solutions of a remarkably broad class of optimization problems. We then use our theorem to retrieve a number of known results in the literature such as the celebrated representer theorem of machine leaning for RKHS, Tikhonov regularization, representer theorems for sparsity promoting functionals, the recovery of spikes, as well as a few new ones.

From shallow to deep learning for inverse imaging problems: some recent approaches
Carola-Bibiane Schönlieb (University of Cambridge, UK)
16:30 – 17:15, 29 May

In this talk we discuss the idea of data-driven regularisers for inverse imaging problems. We are in particular interested in the combination of model-based and purely data-driven image processing approaches. In this context we will make a journey from “shallow” learning for computing optimal parameters for variational regularisation models by bilevel optimization to the investigation of different approaches that use deep neural networks for solving inverse imaging problems. Alongside all approaches that are being discussed, their numerical solution and available solution guarantees will be stated.
This talk is based on S. Arridge, P. Maass, O. □Oktem, C.-B. Sch□onlieb, Solving inverse problems using data-driven models, to appear in Acta Numerica, 178 p, 2019.

SketchySVD
Joel Tropp (California Institute of Technology, USA)
10:15 – 11:00, 30 May

This talk asserts that randomized linear algebra is a natural tool for on-the-fly compression of data matrices that arise from large-scale scientific simulations and data collection. The technical contribution consists in a new algorithm for constructing an accurate low-rank approximation of a huge matrix from streaming data. Among other applications, we show how the SVD of a large-scale sea surface temperature dataset exposes features of the global climate.

Optimal transport for machine learning
Gabriel Peyre (Ecole Normale Superieure, France)
11:00 – 11:45, 30 May

Optimal transport (OT) has become a fundamental mathematical tool at the interface between calculus of variations, partial differential equations and probability. It took however much more time for this notion to become mainstream in numerical applications. This situation is in large part due to the high computational cost of the underlying optimization problems. There is a recent wave of activity on the use of OT-related methods in fields as diverse as image processing, computer vision, computer graphics, statistical inference, machine learning. In this talk, I will review an emerging class of numerical approaches for the approximate resolution of OT-based optimization problems. This offers a new perspective for the application of OT in high dimension, to solve supervised (learning with transportation loss function) and unsupervised (generative network training) machine learning problems. More information and references can be found on the website of our book "Computational Optimal Transport" https://optimaltransport.github.io/

On the (unreasonable) effectiveness of compressive imaging
Ben Adcock (Simon Fraser University of Canada)
12:15 – 13:00, 30 May

Compressive imaging is the process of reconstructing images from highly incomplete data using techniques such as sparse regularization, or most recently, neural networks. Since the advent of compressed sensing in 2005 it has proved an extremely fruitful area of innovation, with many different modalities seeing substantial gains in acquisition time, cost or image fidelity. This talk seeks to explain why this has been the case. In particular, we address the following paradox: why is the (seemingly optimal) strategy of Gaussian random sensing a relatively poor approach in compressive imaging, when structured sensing matrices (e.g. Fourier or Hadamard) provide far higher-quality reconstructions in practice? Using the language of approximation theory, we present a new result showing that Fourier sampling provides a near-optimal sensing strategy for recovering piecewise regular functions, provably outperforming random Gaussian sampling. The key component of this a novel theory of compressed sensing based on local structure. Using this theory, we also introduce a suite of new sampling schemes for 2D and 3D compressive imaging. We illustrate their effectiveness over more standard approaches, including learned sampling strategies. Time permitting, we will discuss how, despite their derivation via compressed sensing, these sampling schemes are in some senses ubiquitous. In particular, they can be used as part of a neural network architecture which is provably stable and accurate.

Deep dictionary learning approaches for image super-resolution
Pier Luigi Dragotti (Imperial College, UK)
14:30 – 15:15, 30 May

Single-image super-resolution refers to the problem of obtaining a high-resolution (HR) version of a single low-resolution (LR) image. This problem is highly ill-posed since it is possible to find many high-resolution images that can lead to the same low-resolution one.

Current strategies to solve the single-image super-resolution problem are learning-based and the model that maps the LR image to the HR image is learned from external image datasets.

Originally, learning-based approaches were built around the idea that both the LR and HR images admit a sparse representation in proper dictionaries and that the sparsity patterns of the two representations can be shared when the design of the two dictionaries is properly coupled. More recently, deep neural network (DNN) architectures have led to state-of-the-art results.

Inspired by the recent success of deep neural networks and the recent effort to develop multi-layer sparse models, we propose an approach based on deep dictionary learning. The proposed architecture contains several layers of analysis dictionaries to extract high-level features and one synthesis dictionary which is designed to optimize the reconstruction task. Each analysis dictionary contains two sub-dictionaries: an information preserving analysis dictionary (IPAD) and a clustering analysis dictionary (CAD). The IPAD with its corresponding thresholds passes the key information from the previous layer, while the CAD with its properly designed thresholds provides a sparse representation of input data that facilitates discrimination of key features.

We then look at the multi-modal case and use the dictionary learning framework as a tool to model dependency across modality, to dictate the architecture of a deep neural network and

to initialize the parameters of the network. Numerical results show that this approach leads to state-of-the-art results.

Mad Max: Affine spline insights into deep learning
Richard Baraniuk (Rice University USA)
15:15 – 16:00, 30 May

We build a rigorous bridge between deep networks (DNs) and approximation theory via spline functions and operators. Our key result is that a large class of DNs can be written as a composition of max-affine spline operators (MASOs), which provide a powerful portal through which to view and analyze their inner workings. For instance, conditioned on the input signal, the output of a MASO DN can be written as a simple affine transformation of the input. This implies that a DN constructs a set of signal-dependent, class-specific templates against which the signal is compared via a simple inner product; we explore the links to the classical theory of optimal classification via matched filters and the effects of data memorization. Going further, we propose a simple penalty term that can be added to the cost function of any DN learning algorithm to force the templates to be orthogonal with each other; this leads to significantly improved classification performance and reduced overfitting with no change to the DN architecture. The spline partition of the input signal space that is implicitly induced by a MASO directly links DNs to the theory of vector quantization (VQ) and K-means clustering, which opens up new geometric avenue to study how DNs organize signals in a hierarchical fashion. To validate the utility of the VQ interpretation, we develop and validate a new distance metric for signals and images that quantifies the difference between their VQ encodings.

Modelling networks and network populations via graph distances
Sofia Olhede (EPFL)
10:15 – 11:00, 31 May

Networks have become a key data analysis tool. They are a simple method of characterising dependence between nodes or actors. Understanding the difference between two networks is also challenging unless they share nodes and are of the same size. We shall discuss how we may compare networks and also consider the regime where more than one network is observed.

We shall also discuss how to parametrize a distribution on labelled graphs in terms of a Frechét mean graph (which depends on a user-specified choice of metric or graph distance) and a parameter that controls the concentration of this distribution about its mean. Entropy is the natural parameter for such control, varying from a point mass concentrated on the Frechét mean itself to a uniform distribution over all graphs on a given vertex set.

Networks present many new statistical challenges. We shall discuss how to resolve these challenges respecting the non-Euclidean nature of network observations.

Talk title TBC
Michael Bronstein (Imperial College, UK)
11:00 – 11:45, 31 May

Building and validating causal inference models - theory and algorithms
Mihaela van der Schaar (University of Cambridge, UK)
12:15 – 13:00, 31 May