Titles and Abstract

Analytic and Geometric Approaches to Machine Learning

26th-30th July 2021

Monday

Kostas Zygalakis

Title: Connection Between Optimization and Sampling Algorithms

Abstract: Optimization and Sampling problems lie in the heart of Bayesian inverse problems. The ability to solve such inverse problems depends crucially on the efficient calculation of quantities relating to the posterior distribution, giving thus rise to computationally challenging high dimensional optimization and sampling problems. In this talk, we will connect the corresponding optimization and sampling problems to the large time behaviour of solutions to (stochastic) differential equations. Establishing such a connection allows to utilise existing knowledge from the field of numerical analysis of differential equations. In particular, two very important concepts are numerical stability and numerical contractivity. In the case of linear differential equations these two concepts coincide, but with the exception of some very simple Runge-Kutta methods such the Euler method in the non-linear case numerical stability doesn't imply numerical contractivity [1]. However, the recently introduced framework of integral quadratic constraints and Lyapunov functions [2, 3] allows for bridging this gap between linearity and non-linearity in the case of (strongly) convex functions. We will use this framework, to study a large class of strogly convex optimization methods and give an alternative explanation for the good properties of Nesterov method, as well as highlight the reasons behind the failure of the heavy ball method [2]. In addition, using similar ideas [4], we will present a general framework for the non-asymptotic study of the 2-Wasserstein distance between the invariant distribution of an ergodic stochastic differential equation and the distribution of its numerical approximation in the strongly log-concave case. This allows us to study in a unified way a number of different integrators proposed in the literature for the overdamped and underdamped Langevin dynamics. If times allows us we will also talk about the application of these ideas to imaging inverse problems [5, 6].

Future Research: In (strongly log-concave) sampling should you go explicit or implicit?
    I will describe some very recent (and perhaps incomplete) work that aims to study the computational efficiency of explicit methods for sampling problems vs implicit integrators. In particular, the class of strongly convex potentials is ideal as it allows for the explicit computation of the number of function evaluations that one needs to do in order to solve the optimization problem related to performing an implicit numerical step.

References

[1] J. M. Sanz Serna and K. C. Zygalakis, *Contractivity of Runge–Kutta Methods for Convex Gradient Systems*, SIAM Journal on Numerical Analysis 58(4):2079-2092, 2020.

[2] L. Lessard, B. Recht, and A. Packard, *Analysis and design of optimization algorithms via integral quadratic constraints*, SIAM Journal on Optimization, 26(1):57–95, 2016.

[3] M. Fazlyab, A. Ribeiro, M. Morari, and V. M. Preciado, *Analysis of optimization algorithms via integral quadratic constraints: nonstrongly convex problems*, SIAM Journal on Optimization, 28(3):2654–2689, 2018.

[4] J. M. Sanz Serna and K. C. Zygalakis, *Wasserstein distance estimates for the distributions of numerical approximations to ergodic stochastic differential equations*, arXiv:2104.12384, 2021.

[5] Eftekhari, B. Vandereycken, G. Vilmart, and K C. Zygalakis, *Explicit Stabilised Gradient Descent for Faster Strongly Convex Optimisation*, BIT Numerical Mathematics 61:119–139, 2021.

[6] M. Pereyra, L. Vargas Mieles, and K. C. Zygalakis, *Accelerating Proximal Markov Chain Monte Carlo by Using an Explicit Stabilized Method*, SIAM Journal on Imaging Sciences 13(2):905-935, 2020.

Yves van Gennip

Title: Gradient Flows and Thresholding Schemes on Graphs

Abstract: We present unconstrained and constrained gradient flows and thresholding schemes on graphs, which can be used in applications such as image segmentation and data clustering. We study connections between the flows and schemes and investigate discrete-to-continuum limit properties.

Future Research: Besides clustering and the maximum-cut problem, which other combinatorial problems on graphs are amenable to relaxation via differential equations?

Bernhard Schmitzer

Title: Linearization of Balanced and Unbalanced Optimal Transport

Abstract: Optimal transport provides a geometrically intuitive Lagrangian way of comparing distributions by mass rearrangement. The metric can be approximated by representing each sample as deformation of a reference distribution. Formally this corresponds to a local linearization of the underlying Riemannian structure. When combined with subsequent data analysis and machine learning methods this new embedding usually outperforms the standard Eulerian representation. We show how the framework can be extended to the unbalanced Hellinger–Kantorovich distance to improve robustness to noise.

Future Research: For the balanced case the properties of the linearized optimal transport embedding are already understood quite well. In the unbalanaced case we seek to confirm the well-posedness of linearized interpolations and to understand the range in which we can expect the approximation to be accurate. In both cases we look for ways to go beyond a simple one-point-linearization.

Youssef Marzouk

Title: Density Estimation and Conditional Simulation Using Triangular Transport

Abstract: Triangular transformations of measures, such as the Knothe–Rosenblatt rearrangement, underlie many new computational approaches for density estimation and conditional simulation. This talk will discuss several aspects of such constructions.

First we present new approximation results for triangular maps, focusing on analytic densities with bounded support. We show that there exist sparse polynomial approximations and deep ReLU network approximations of such maps that converge exponentially, where error is assessed on the pushforward distributions of such maps in a variety of distances and divergences. We also give an explicit a priori description of the polynomial ansatz space guaranteeing a given error bound.

Second, we discuss the problem of estimating a triangular transformation given a sample from a distribution of interest—and hence, transport-driven density estimation. We present a general functional framework for representing monotone triangular maps between distributions on unbounded domains, and analyze properties of maximum likelihood estimation in this framework. We demonstrate that the associated optimization problem is smooth and, under appropriate conditions, has no spurious local minima. This result provides a foundation for a greedy semi-parametric estimation procedure.

Time permitting, we may also discuss a conditional simulation method that employs a specific composition of maps, derived from the Knothe–Rosenblatt rearrangement, to push forward a joint distribution to any desired conditional. We show that this composed-map approach reduces variability in conditional density estimates and reduces the bias associated with any approximate map representation. For context, and as a pointer to an interesting application domain, we elucidate links between conditional simulation with composed maps and the ensemble Kalman filter used in geophysical data assimilation.

This is joint work with Ricardo Baptista (MIT), Olivier Zahm (INRIA), and Jakob Zech (Heidelberg).

Future Research: Open questions:

1. Removal of dependence, i.e., estimating and minimizing mutual information over a given space of transformations. In other words, how to find a transformation T such that $Z = T(Y, X)$ has minimal dependence with Y? This turns out to be the essential underlying problem of conditional simulation, in my view.

2. Transport-based methods for (approximate) inference are a natural complement to ensemble methods for data assimilation. Yet embedding an approximate inference step into a dynamical system presents many new challenges and considerations. In what sense should inference be made accurate, given the subsequent evolution of the approximate posterior under the dynamics? How do errors in inference affect stability of the associated interacting particle system? How can we use this understanding to design better algorithms?

Tuesday

Coloma Ballester

Title: Generative Methods for Some Inverse Problems in Imaging

Abstract: Generative approaches aim to model the probability distribution of certain data so that we can sample new samples out of the distribution. In this talk, we will discuss them together with its use to tackle several imaging problems. They include a method for out-of-distribution that leverages the learning of the probability distribution of normal data through generative adversarial networks while simultaneously keeping track of the states of the learning to finally estimate an efficient anomaly detector. Also, a method for image colorization will be discussed. It is based on an adversarial strategy that captures geometric, perceptual and semantic information. Finally, the use of generative methods for inpainting will be discussed.

Future Research: An open/interesting problem related to this work is the suitability of generative methods to obtain multiple solutions of problems that do not have a unique solution. Another open problem is how the learning of the underlying distribution of color images is affected by the color space in which those images are given.

Andres Almansa

Title: Bayesian Estimators for Inverse Imaging Problems with Decoupled Learned Priors

Abstract: Decoupled methods derive Minimum Mean Square Error (MMSE) or Maximum A Posteriori (MAP) estimators for inverse problems in imaging by combining an explicit likelihood function with a prior that was previously trained on a large natural image database. In this talk we present recent developments on such decoupled methods for two kinds of learned priors.

Part I: Concentrates on so called Plug & Play (PnP) methods, where the prior is implicitly defined by an image denoising algorithm. In [1] we introduce theory, methods, and a provably convergent algorithm for performing Bayesian inference with PnP priors. We introduce two algorithms: 1) PnP-ULA (Plug & Play Unadjusted Langevin Algorithm) for Monte Carlo sampling and MMSE inference; and 2) PnP-SGD (Plug & Play Stochastic Gradient Descent) for MAP inference. Using recent results on the quantitative convergence of Markov chains, we establish detailed convergence guarantees for these two algorithms under realistic assumptions on the denoising operators used, with special attention to denoisers based on deep neural networks. We also show that these algorithms approximately target a decision-theoretically optimal Bayesian model that is well-posed. The proposed algorithms are demonstrated on several canonical problems such as image deblurring, inpainting, and denoising, where they are used for point estimation as well as for uncertainty visualisation and quantification.

Part II: Concentrates on generative models where the prior in the image domain is expressed as the pushforward measure of a Gaussian distribution in latent space. Whereas previous MAP-based approaches to this problem lead to highly non-convex optimization algorithms, the JPMAP approach proposed in [2] computes the joint (space-latent) MAP that naturally leads to alternate optimization algorithms and to the use of a stochastic encoder to accelerate computations. We show theoretical and experimental evidence that the proposed objective function is quite close to bi-convex. Indeed it satisfies a weak bi-convexity property which is sufficient to guarantee that our optimization scheme converges to a stationary point. Experimental results also show the higher

quality of the solutions obtained by our joint posterior maximization approach with respect to other non-convex MAP approaches which more often get stuck in spurious local optima. Current work also explores how the quasi-bi-convexity property of the joint posterior could be exploited to provide efficient posterior sampling schemes that lead to faster MMSE and uncertainty estimators.

Joint work with: Mario González, Rémi Laumont, Jean Prost, Valentin de Bortoli, Julie Delon, Alain Durmus, Antoine Houdard, Pablo Musé, Nicolas Papadakis, Marcelo Pereyra, Pauline Tan.

Future Research: We are currently exploring theoretical, methodological and applied generalizations of the methods and algorithms described in the abstract.

Resizeable generative priors: PnP priors which are usually implemented as convolutional neural network denoisers that can be trained on image patches of a certain size, and then used on images of any size. Generative priors, in contrast, are usually trained and used on images of a fixed size. This issue significantly limits the applicability of generative priors for general inverse problems in imaging. A few solutions to this issue have been explored: (a) In [3] a generative prior is trained on natural image patches and then the regularization term is applied patch-wise, but this is computationally inefficient and results are not competitive with other PnP approaches. (b) In [4] the architecture of the generative model is fully convolutional which allows the generative model to be resized. The authors show a few relatively convincing applications of this resizability feature but no in-depth theoretical explanation was provided to justify that this operation is well-founded.

Faster posterior sampling schemes: The PnP-ULA scheme only provides stability and convergence guarantees as long as the step-size satisfies certain limits which makes it quite slow. In order to allow for larger step-sizes and faster schemes we can consider semi-implicit or higher-order schemes. These introduce, however, the need for denoisers D such that the operator (D+ a Id) can be efficiently inverted. For generative priors, the pCN scheme [5] can perform posterior sampling based on the generator network only. We are exploring whether this scheme can be accelerated when an encoder network is also available like in the JPMAP approach [2].

Applications: On the applications side, we are exploring the use of PnP posterior sampling techniques for semi-blind restoration in the context of camera shake, as well as for uncertainty

References

[1] R. Laumont, V. de Bortoli, A. Almansa, J. Delon, A. Durmus, and M. Pereyra, *Bayesian imaging using Plug & Play priors: when Langevin meets Tweedie*, arXiv:2103.04715, 2021.

[2] M. González, A. Almansa, and P. Tan, *Solving Inverse Problems by Joint Posterior Maximization with Autoencoding Prior*, arXiv:2103.01648, 2021.

[3] J. Prost, A. Houdard, A. Almansa, and N. Papadakis, *Learning local regularization for variational image restoration*, arXiv:2102.06155, 2021.

[4] D. P. Kingma and P. Dhariwal, *Glow: Generative flow with invertible 1×1 convolutions*, (NeurIPS) Advances in Neural Information Processing Systems pp. 10215–10224, 2018.

[5] M. Holden, M. Pereyra, and K. C. Zygalakis, *Bayesian Imaging With Data-Driven Priors Encoded by Neural Networks: Theory, Methods, and Algorithms*, arXiv:2103.10182, 2021.

Johannes Schmidt-Hieber

Title: Overparametrization and the Bias-Variance Dilemma

Abstract: For several machine learning methods such as neural networks, good generalisation performance has been reported in the overparametrized regime. In view of the classical bias-variance trade-off, this behaviour is highly counterintuitive. The talk summarizes recent theoretical results on overparametrization and the bias-variance trade-off. This is joint work with Alexis Derumigny.

Future Research: In the talk we present universal lower bounds on the bias-variance trade-off. These bounds have also to be obeyed in the overparametrized regime. It is an open problem to derive rate optimal bounds on bias and variance in the overparametrized regime for standard nonparametric and high-dimensional models. Another challenging problem is to extend the results to the trade-off between bias and mean absolute deviation, see the paper arXiv:2006.00278 for more details and a first result in this direction.

Omiros Papaspiliopoulos

Title: Scalable Computation for Bayesian Hierarchical Models

Abstract: I will provide an overview of very recent theory that establishes conditions under which coordinate-wise updating algorithms, in particular Gibbs samplers for posterior sampling, coordinate ascent variational inference for posterior approximation, and backfitting algorithms for MAP estimation, are scalable. Scalability here means that their computational complexity in order to achieve a certain approximation precision scales linearly in the size of the data and the size of the model. The results refer to a very important family of statistical models, known as crossed-effect hierarchical models, which are the canonical predictive/inference models for regressing against categorical predictors. The theory brings together multigrid decompositions, concentration inequalities and random graphs.

Future Research: I will describe some very recent (and perhaps incomplete) work that aims to study the computational efficiency of explicit methods for sampling problems vs implicit integrators. In particular, the class of strongly convex potentials is ideal as it allows for the explicit computation of the number of function evaluations that one needs to do in order to solve the optimization problem related to performing an implicit numerical step.

Wednesday

Mathew Penrose

Title: Optimal Cuts of Random Geometric Graphs

Abstract: Given a "cloud" of n points sampled independently uniformly at random from a Euclidean domain D, one may form a geometric graph by connecting nearby points using a distance parameter $r(n)$. We consider the problem of partitioning the cloud into two pieces to minimise the number of "cut edges" of this graph, subject to a penalty for an unbalanced partition. The optimal score is known as the Cheeger constant of the graph. We discuss convergence of the Cheeger constant (suitably rescaled) for large n with suitably chosen $r(n)$, towards an analogous quantity

defined for the original domain D, and the related problem of optimal bisection into two pieces of equal size.

Future Research: These results are for when the mean degree $nr(n)^d$ grows faster than $logn$; open problems include finding analogous results (i) when the mean degree grows more slowly than this, or (ii) for the k-nearest neighbour graph with $k = k(n)$ tending to infinity, or (iii) extending the results to manifolds.

Partitioning problems are are of interest in themselves when dealing with multidimensional data, and moreover the Cheeger constants provide bounds on Laplacian eigenvalues, both for graphs and in the continuum.

Franca Hoffmann

Title: Discrete-Continuum Interplay: Formulations for Supervised and Semi-Supervised Learning

Abstract: Graph Laplacians encode geometric information contained in data, via the eigenfunctions associated with their small eigenvalues. These spectral properties provide powerful tools in data clustering and data classification. When a large number of data points are available one may consider continuum limits of the graph Laplacian, both to give insight and, potentially, as the basis for numerical methods. We summarize recent insights into the properties of a family of weighted elliptic operators arising in the large data limit of different graph Laplacian normalizations, and the role these operators play, both in the discrete and in the continuum, for clustering and classification algorithms. We show consistency of optimization-based techniques for graph-based semi-supervised learning (SSL), in the limit where the labels have small noise and the underlying unlabelled data is well clustered. This formalism suggests continuous versions of SSL algorithms, making use of these differential operators.

Future Research: How to leverage continuum formulations for algorithm design and implementations? How to evaluate and compare these implementations?

Jeff Calder

Title: PDE Continuum Limits for Prediction with Expert Advice

Abstract: Prediction with expert advice refers to a class of machine learning problems that is concerned with how to optimally combine advice from multiple experts whose prediction qualities may vary greatly. We study a stock prediction problem with history-dependent experts and an adversarial (worst-case) market, posing the problem as a repeated two-player game. The game is a generalization of the Kohn-Serfaty two-player game for curvature motion. We prove that when the game is played for a long time, the discrete value functions converge in the continuum to the solution of a nonlinear parabolic partial differential equation (PDE) and the optimal strategies are characterized by the solution of a Poisson equation over a De Bruijn graph, arising from the history-dependence of the experts. Joint work with Nadejda Drenska (UMN).

Future Research: PDE continuum limits for adversarial multi-armed bandits. Multi-armed bandits is a machine learning problem centered around exploration vs exploitation in an online learning environment, and is used to evaluate strategies for managing large research projects or allocating funds to researchers, among other problems. The user has many possible "arms" to

query (the language relates to pulling an arm of a slot machine), each of which gives the player a randomized reward (following a different distribution for each arm). The user wants to maximize their rewards over the long run, which requires trading off exploration (pulling new arms that may end up giving poor rewards) and exploitation (repeatedly pulling the same arm that is known to give a good reward). The adversarial version of the mulit-armed bandit problem, where the rewards are controlled by an adversary whose goal is to minimize the player's rewards, is very closely related to the prediction from expert advice problem covered in my talk, except that the player has only partial knowledge in the multi-armed bandit setting. An interesting problem would center around determining a PDE continuum limit for the value function of an adversarial multi-armed bandit problem, and using this to construct asymptotically optimal strategies for the player and adversary.

Andrea Bertozzi

Title: TBC

Abstract: TBC

Future Research: TBC

Thursday

Jonas Latz

Title: Stochastic Gradient Descent in Continuous Time: Discrete and Continuous Data

Abstract: Optimisation problems with discrete and continuous data appear in statistical estimation, machine learning, functional data science, robust optimal control, and variational inference. The 'full' target function in such an optimisation problems is given by the integral over a family of parameterised target functions with respect to a discrete or continuous probability measure. Such problems can often be solved by stochastic optimisation methods: performing optimisation steps with respect to the parameterised target function with randomly switched parameter values. In this talk, we discuss a continuous-time variant of the stochastic gradient descent algorithm. This so-called stochastic gradient process couples a gradient flow minimising a parameterised target function and a continuous-time 'index' process which determines the parameter.

   We first briefly introduce the stochastic gradient processes for finite, discrete data which uses pure jump index processes. Then, we move on to continuous data. Here, we allow for very general index processes: reflected diffusions, pure jump processes, as well as other Lévy processes on compact spaces. Thus, we study multiple sampling patterns for the continuous data space. We show that the stochastic gradient process can approximate the gradient flow minimising the full target function at any accuracy. Moreover, we give convexity assumptions under which the stochastic gradient process with constant learning rate is geometrically ergodic. In the same setting, we also obtain ergodicity and convergence to the minimiser of the full target function when the learning rate decreases over time sufficiently slowly. We illustrate the applicability of the stochastic gradient process in a simple polynomial regression problem with noisy functional data, as well as in physics-informed neural networks approximating the solution to certain partial differential equations.

Future Research: We propose a continuous-time optimisation algorithm that needs to be discretised to be used in practice. Hence, we need to find a suitable time-stepping method, for which we have two constraints: (1) a suitable time-stepping method should retain the same/similar ergodic behaviour as the stochastic gradient process; (2) the method should be computationally cheap to be applicable in large scale estimation and learning problems.

Yury Korolev

Title: Approximation Properties of Two-Layer Neural Networks with Values in a Banach Space

Abstract: 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 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.

Future Research:

1. Studying infinitely wide/deep vector-valued networks with a more complex architecture: multilayer networks, ResNets, CNNs. Perhaps also other types of networks such as recurrent neural networks (I know very little about them);

2. Neural networks with a 'regularisation parameter' that could be used to solve inverse problems for different noise levels without retraining (e.g., variable depth, scaling of the bias terms).

Matthias Ehrhardt

Title: Equivariant Neural Networks for Inverse Problems

Abstract: In recent years the use of convolutional layers to encode an inductive bias (translational equivariance) in neural networks has proven to be a very fruitful idea. The successes of this approach have motivated a line of research into incorporating other symmetries into deep learning methods, in the form of group equivariant convolutional neural networks. Much of this work has been focused on roto-translational symmetry of the R^d, but other examples are the scaling symmetry of R^d and rotational symmetry of the sphere. In this work, we demonstrate that group equivariant convolutional operations can naturally be incorporated into learned reconstruction methods for inverse problems that are motivated by the variational regularisation approach. Indeed, if the regularisation functional is invariant under a group symmetry, the corresponding proximal operator will satisfy an equivariance property with respect to the same group symmetry. As a result of this observation, we design learned iterative methods in which the proximal operators are modelled as group equivariant convolutional neural networks. We use roto-translationally equivariant operations in the proposed methodology and apply it to the problems of low-dose computerised tomography reconstruction and subsampled magnetic resonance imaging reconstruction.

The proposed methodology is demonstrated to improve the reconstruction quality of a learned reconstruction method with a little extra computational cost at training time but without any extra cost at test time.

Future Research: In this talk we will make use of the learned proximal gradient method where the "proximal operator" is replaced by a neural network which can be trained either before or after it is being inserted into the algorithm. If these neural networks are indeed proximal operators with the correct scaling, then classical convergence results can be utilized. More interesting in applications is to relax these conditions as much as possible in order to allow for more flexibility and adaptation to data. Some results for averaged operators exist in this direction but these are not yet fully satisfactory. Beyond the convergence of the algorithm it is of interest to mathematically analyze its limit and its regularization properties in an inverse problems context.

Philipp Petersen

Title: Structured Singularities in Deep Learning

Abstract: We will discuss various instances of structured singularities appearing in modern machine learning applications. A structured singularity is a discontinuity or a region where a function is non-smooth, which is itself a (not necessarily smooth) manifold. Such structured singularities appear naturally in classification problems as class boundaries. It can be shown that in this case, the regularity of the singularities is the decisive factor in describing the hardness of the underlying learning problem. In addition, learning classification problems with smooth boundaries is one of the few examples where deep neural network architectures provably outperform classical methods in learning applications. We will describe some partial results quantifying the effect of singularities and discuss the regularity assumption. The second example of structured singularities in learning will be discussed in the framework of inverse problems appearing in imaging. In many inverse problems, based on pseudodifferential operators, one has a precise characterisation of how the forward and inverse operators transform singularities. Due to this, it seems promising to understand how deep neural network architectures that attempt to approximate the inverse operator transform singularities through the layers. Here, however, one runs into the fundamental problem that singularities are a phenomenon of the continuum, whereas neural networks acting on digital images are inherently discrete. We will discuss this question in detail for the inverse tomography problem.

Future Research:

1. How appropriate is the smoothness assumption for classification problems? Is it possible to estimate the regularity of classification problems in real-world scenarios? Is it possible to set up a meaningful numerical example? Can additional implicit assumptions on the decision boundary, such as locally minimal perimeter, be used to obtain reasonable regularity estimates?

2. Assuming a CNN is used to approximate an operator. What is the natural continuum analogue to a CNN when singularity propagation is concerned? What is the correct notion of a singularity in a digital image in a learning-theoretical context? How is this notion of a digital singularity affected by an application of a neural network?

Friday

Christoph Brune

Title: Geometric DL - From Learning ODE Dynamics towards Graph Neural Diffusion

Abstract: In recent years the area of geometric deep learning has attracted significant attention in machine learning theory as well as in data science applications. Two reasons for that development are that graph neural networks can effectively address different non-Euclidean geometric data domains, and that geometric priors can induce appealing analytical properties like symmetry and scale separation for representing physically structured data. Although deep learning has been effectively analyzed and used for learning and discovering dynamics from data, e.g., in the form of neural ODEs, physics-informed NNs or mean field control models, the combination of PDEs with graph neural networks to explain spatio-temporal dynamics has been hardly explored. This talk presents new methods for learning normal forms directly from dynamic data and new mesh graph neural networks useful for physics-informed 4d medical imaging raising an outlook towards graph neural diffusion and beyond.

Future Research: Model-Informed Graph Neural Networks - Are graph neural networks more generalizable for model-informed machine learning problems compared to classical neural networks?

1. Modeling: How to combine graph neural networks properly with PDEs to achieve model-informed GNN?

2. Expressiveness: Are flow-induced function spaces more expressive for GNNs with model-informed constraints than Barron spaces?

3. Generalization: How does the double descent phenomena show up in GNNs when we know more about the dynamics underlying?

Abderrahim Elmoataz

Title: The Nonlocal p-Laplacian Evolution Problem on Sparse Graphs: The Continuum Limit

Abstract: The non-local p-Laplacian evolution equation, governed by a given kernel, has applications invarious areas of science and engineering. In particular, it is modern tools for massive data processing (including signals, images, geometry), and machine learning tasks such as classification. In practice, however, this model is implemented in, time and space, discrete form as a numerical approximation to a continuous problem, where the kernel is replaced by an adjacency matrix of graph. In this work, we propose a far-reaching generalization of the results in [1] to a much more general class of kernels and initial data and the case $p = 1$ which was not handled in [1]. Combining tools from graph theory, convex analysis, non-linear semigroup theory and evolution equations, we give a rigorous interpretation to the continuous limit of the discrete non-local p-Laplacian evolution on sparse graphs. More specifically, we consider a sequence of graphs converging to a so-called limit object known as the L^q-graphon [2, 3]. If the continuous p-Laplacian evolution is properly discretized on this graph sequence, we prove that the solutions of the sequence of discrete problems converge to the solution of the continuous problem governed by L^q-graphon, when the number of graph vertices grows to infinity. Along the way, we provide a consistency/error estimate between the solutions of two evolution problems governed by two kernels and two initial data. For random

graph sequences, using sharp deviation inequalities,we deliver nonasymptotic convergence rates in probability and exhibit the different regimes depending on p and the regularity of the L^q-graphon and the initial condition.

Future Research: By iterating the p-Laplacian operator (for $p = 2$), we get so-called the non-local bilaplacian operator [4], which can be extended to a general operators called the nonlocal p-bilaplacian, for $p \in (1, +\infty)$. It will be very interesting to study the well-posedness of different problems governed by this family of operators e.g. evolution, variational and boundary value problems associated to this operators. It will be also interesting to study their continuum limits on L^1-graphons.

References

[1] Y. HAFIENE, J. FADILI, A. AND ELMOATAZ, *Nonlocal p-Laplacian evolution problems on graphs*, SIAM Journal on Numerical Analysis 56(2):1064-1090, 2018.

[2] C. BORGS, J. T. CHAYES, H. COHN, AND Y. ZHAO, *An L^p theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions*, Transactions of the American Mathematical Society 372(5):3019-3062, 2019.

[3] C. BORGS, J. T. CHAYES, H. COHN, AND Y. ZHAO, *An L^p theory of sparse graph convergence II: LD convergence, quotients and right convergence*, The Annals of Probability 46(1):337-396, 2018.

[4] P. RADU, D. TOUNDYKOV, AND J. TRAGESER, *A nonlocal biharmonic operator and its connection with the classical analogue*, Archive for Rational Mechanics and Analysis 223(2):845-880, 2017.

Daniel Cremers

Title: Self-supervised Learning for 3D Shape Analysis

Abstract: While neural networks have swept the field of computer vision and replaced classical methods in most areas of image analysis and beyond, extending their power to the domain of 3D shape analysis remains an important open challenge. In my presentation, I will focus on the problems of shape matching, correspondence estimation and shape interpolation and develop suitable deep learning approaches to tackle these challenges. In particular, I will focus on the difficult problem of computing correspondence and interpolation for pairs of shapes from different classes – say a human and a horse – where traditional isometry assumptions no longer hold.

Future Research: Open questions:

1. How can we extend the power of deep networks to the field of 3D shape analysis?

2. How can deep learning help to perform camera-based 3D reconstruction of static and dynamic scene?

3. How can deep learning be applied to analyzing 3D shapes – for problems such as estimating correspondence, computing interpolations between shapes and measuring shape similarity?

4. What are the strengths and weaknesses of different representations of shape in the age of deep learning?

Pablo Musé

Title: Non-Uniform Motion Blur Kernel Estimation via Adaptive Decomposition

Abstract: Motion blur estimation remains an important task for scene analysis and image restoration. In recent years, the removal of motion blur in photographs has seen impressive progress in the hands of deep learning-based methods, trained to map directly from blurry to sharp images. Characterization of the motion blur, on the other hand, has received less attention, and progress in model-based methods for deblurring lags behind that of data-driven end-to-end approaches.

In this work we revisit the problem of characterizing dense, non-uniform motion blur in a single image and propose a general non-parametric model for this task. Given a blurry image, a neural network is trained to estimate a set of image-adaptive basis motion kernels as well as the mixing coefficients at the pixel level, producing a per-pixel motion blur field.

When applied to real motion-blurred images, a variational non-uniform blur removal method fed with the estimated blur kernels produces high-quality restored images. Qualitative and quantitative evaluation shows that these results are competitive or superior to results obtained with existing end-to-end deep learning (DL) based methods, thus bridging the gap between model-based and data-driven approaches.

Future Research: We are interested in deriving image restoration methods that combine model-driven and data-driven approaches. In this application, the image formation model is used to learn the local blur kernels from data. Deblurring is performed in a second stage, using a classic non-blind deblurring method. The next step would be to include kernel estimation and deblurring into the same pipeline.