HAUSDORFF CENTER FOR MATHEMATICS

UNIVERSITÄT BONN

Hausdorff Workshop

"HCM Workshop: Synergies between Data Sciences and PDE Analysis"

June 13 - 17, 2022

organized by
Franca Hoffmann, Leon Bungert

Abstracts for Keynote Talks

Gitta Kutyniok (LMU Munich)

The Impact of Artificial Intelligence on Parametric Partial Differential Equations: From Successes to Limitations

Abstract: High-dimensional parametric partial differential equations (PDEs) appear in various contexts including control and optimization problems, inverse problems, risk assessment, and uncertainty quantification. Recently, numerical experiments demonstrated the remarkable efficiency of using deep neural networks to solve parametric problems. In this talk, after an introduction into deep learning, we will provide a theoretical justification for this class of approaches in term of approximation-theoretical results. Moreover, we will present a comprehensive numerical study of the effect of such results for neural networks on practical learning problems. We will finish with a word of caution when training neural networks for solving PDEs on classical digital hardware, and present fundamental limitations.

Michael Unser (EPFL)

Neural networks and minimum-norm ridge splines

Abstract: A powerful framework for supervised learning is the minimization of a cost that consists of a data fidelity term plus a regularization functional. In this talk, I investigate a Radon-domain regularization functional that depends on a generic operator L. The proposed formulation yields a solution that takes the form of a two layer neural network with an activation function that is determined by the regularization operator. In particular, one retrieves the popular ReLU networks by taking L to be the Laplacian. The proposed setting offers guarantees of universal approximation for a broad family of regularization operators or, equivalently, for a wide variety of shallow neural networks including cases (such as ReLU) where the activation function is increasing polynomially. It also explains the favorable role of bias and skip connections in neural architectures.

Eldad Haber (University of British Columbia)

Graphs, differential equations and neural networks

Abstract: Graph neural networks are increasingly becoming the go-to approach in various fields such as computer vision, computational biology and chemistry, where data are naturally explained by graphs. However, unlike traditional convolutional neural networks, deep graph networks do not necessarily yield better performance than shallow graph networks. This behavior usually stems from the over-smoothing phenomenon. In this work, we propose a family of architectures to control this behavior by design. Our networks are motivated by numerical methods for solving Partial Differential Equations (PDEs) on manifolds, and as such, their behavior can be explained by similar analysis. Moreover, as we demonstrate using an extensive set of experiments, our PDE-motivated networks can generalize and be effective for various types of problems from different fields. Our architectures obtain better or on par with the current state-of-the-art results for problems that are typically approached using different architectures.

Abstracts for Invited Talks

Lisa Maria Kreusser (University of Bath)

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

Abstract: 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.

Tim Roith (University of Erlangen)

Uniform Convergence Rates for Lipschitz Learning Down to Graph Connectivity

Abstract: Discrete to continuum convergence results for graph-based learning have seen an increased interest in the last years. In particular, the connections between discrete machine learning and continuum partial differential equations or variational problems, lead to new insights and better algorithms. This talk considers Lipschitz learning—which is the limit of p-Laplacian learning for p to infinity—and introduces new proof strategies for the discrete to continuum limit. Our framework provides a convergence result in a sparse graph regime and additionally yields convergence rates. Employing a homogenized non-local operator with a much larger bandwidth allows us to extend uniform convergence rates to any graph length scale strictly above graph connectivity. We will sketch the ideas of the proof and indicate how the approach may be used in other problems, like spectral convergence of the graph Laplacian. Joint work with Leon Bungert and Jeff Calder.

Dejan Slepčev (CMU Pittsburgh)

Quantization of Measures via Interacting Particle-based Optimization

Abstract: Approximating a probability distribution can be cast as an optimization problem where the objective functional measures the dissimilarity to the target. This optimization can be addressed by approximating Wasserstein and related gradient flows. In practice, these are simulated by interacting particle systems, whose stationary states define an empirical measure approximating the target distribution. This approach has been popularized recently to design sampling algorithms, e.g. Stein Variational Gradient Descent, or by minimizing the Maximum Mean or Kernel Stein Discrepancy. However, little is known about quantization properties of these approaches, i.e. how well is the target approximated by a finite number particles. We investigate this question theoretically and numerically. In particular, we prove general upper bounds on the quantization error of MMD and KSD with various kernels. Furthermore we introduce a Normalized Stein Variational Gradient Descent and argue in favor of adaptive kernels. The talk is based on joint work with Anna Korba and Lantian Xu.

Bamdad Hosseini (University of Washington)

Solving and Learning Nonlinear PDEs with Gaussian Processes

Abstract: In this talk I present a simple, rigorous, and interpretable framework for solution of nonlinear PDEs based on the framework of Gaussian Processes. The proposed approach provides a natural generalization of kernel methods to nonlinear PDEs; has guaranteed convergence; and inherits the state-of-the-art computational complexity of linear solvers for dense kernel matrices. I will outline our approach by focusing on an example nonlinear elliptic PDE followed by further numerical examples. I will also briefly comment on extending our approach to solving inverse problems.

Stephan Wojtowytsch (Texas A&M University)

Mean field analysis of shallow neural networks

Abstract: In the 'mean field regime', gradient flows of shallow neural networks can be understood as Wasserstein gradient flows of the distribution of neurons of the neural network, considered as interchangeable particles. This connection provides tools from PDE theory for the training of neural networks and allows us to prove conditional convergence results and analyze important aspects of initialization.

Anna Korba (ENSAE ParisTech)

Variational Inference of overparameterized Bayesian Neural Networks: a theoretical and empirical study

Abstract: We study the Variational Inference (VI) used for training Bayesian Neural Networks (BNN) in the overparameterized regime, i.e., when the number of neurons tends to infinity. More specifically, we consider overparameterized two-layer BNN and point out a critical issue in the mean-field VI training. This problem arises from the decomposition of the lower bound on the evidence (ELBO) into two terms: one corresponding to the likelihood function of the model and the second to the Kullback-Leibler (KL) divergence between the prior distribution and the variational posterior. In particular, we show both theoretically and empirically that there is a trade-off between these two

terms in the overparameterized regime only when the KL is appropriately re-scaled with respect to the ratio between the the number of observations and neurons. We also illustrate our theoretical results with numerical experiments that highlight the critical choice of this ratio.

Philipp Petersen (University of Vienna)

Optimal representation and learning of classifier functions

Abstract: Deep learning has established itself as, by far, the most successful machine learning approach in sufficiently complex tasks. Nowadays, it is used in a wide range of highly complex applications such as natural language processing or even scientific applications. Its first major breakthrough, however, was achieved by shattering the state-of-the-art in image classification. We revisit the problem of classification by deep neural networks and attempt to find an answer to why deep networks are remarkably effective in this regime. We will interpret the learning of classifiers as finding piecewise constant functions from labelled samples. We then precisely link the hardness of the learning problem to the complexity of the regions. Concretely, we will establish fundamental lower bounds on the learnability of certain regions. Finally, we will show that in many cases, these optimal bounds can be achieved by deep-neural-network-based learning. In quite realistic settings, we will observe that deep neural networks can learn high-dimensional classifiers without a strong dependence of the learning rates on the dimension.

Nicolás García Trillos (University of Wisconsin Madison)

The multimarginal optimal transport formulation of adversarial multiclass classification

Abstract: Adversarial training is a framework widely used by machine learning practitioners to enforce robustness of learning models. Despite the development of several computational strategies for adversarial training and some theoretical development in the broader distributionally robust optimization literature, there are still several theoretical questions about adversarial training that remain relatively unexplored. In this talk, I will discuss an equivalence between adversarial training in the context of non-parametric multiclass classification problems and multimarginal optimal transport problems. This is another analytical interpretation of adversarial training that expands recently studied connections to perimeter minimization problems. One of the implications of the connection discussed during the talk is computational: to solve a certain adversarial problem, we may as well solve a multimarginal optimal transport problem. We will discuss many of the nuances of this interpretation and of its computational consequences. This is joint work with my student Jakwang Kim (UW) and my colleague Matt Jacobs (Purdue).

Clarice Poon (University of Bath)

Smooth over-parametrized solvers for non-smooth structured optimisation.

Abstract: Non-smooth optimization is a core ingredient of many imaging or machine learning pipelines. Non-smoothness encodes structural constraints on the solutions, such as sparsity, group sparsity, low-rank and sharp edges. It is also the basis for the definition of robust loss functions such as the square-root lasso. Standard approaches to deal with non-smoothness leverage either proximal splitting or coordinate descent. The effectiveness of their usage typically depend on proper parameter tuning, preconditioning or some sort of support pruning. In this work, we advocate and study a different route. By over-parameterization and marginalising on certain variables (Variable Projection), we show how many popular non-smooth structured problems can be written as smooth optimization

problems. The result is that one can then take advantage of quasi-Newton solvers such as L-BFGS and this, in practice, can lead to substantial performance gains. Another interesting aspect of our proposed solver is its efficiency when handling imaging problems that arise from fine discretizations (unlike proximal methods such as ISTA whose convergence is known to have exponential dependency on dimension). On a theoretical level, one can connect gradient descent on our over-parameterized formulation with mirror descent with a varying Hessian metric. This observation can then be used to derive dimension free convergence bounds and explains the efficiency of our method in the fine-grids regime.

Matthew Thorpe (University of Manchester)

Understanding Bias in Graph-Based Semi-Supervised Learning at Low Label Rates: From Laplace to Poisson Learning

Abstract: Laplacian learning is a semi-supervised method that finds missing labels via minimising a Dirichlet energy. It is well known that Laplacian learning is (asymptotically) ill-posed at low labelling rates. In this talk I will identify the bias of Laplace learning and show how this can be corrected leading to significant improvement in performance. The correction in the bias leads one to a Poisson equation.

Yury Korolev (University of Cambridge)

Laplace learning in Hilbert spaces

Abstract: Graph-based approaches, such as Laplace learning, have been successfully used in semi-supervised learning for finite dimensional data. In this talk we formulate the Laplace learning problem in a functional setting, that is where the feature vectors live in an infinite-dimensional space. We show that the method is consistent in the large data limit (as the number of feature vectors goes to infinity) under appropriate conditions on the graph connectivity.

Josiah Park (Texas A&M University)

Approximation and Optimization via Neural Networks

Abstract: There is a large amount of interest surrounding the 'unreasonable effectiveness' of neural networks in deep learning and other applications. A combination of major hardware advances and increasing sophistication of techniques for learning in recent years have brought impressive advances in performance for tasks in speech processing, translation, image recognition, among other domains. Practitioners have worked out a rule of thumb for how to use networks in the sciences, but several fundamental questions remain about the effectiveness of these models. Using tools from approximation theory, we can examine the success of neural networks by understanding functions which may be approximated well by neural networks. We describe some classes of functions which may be approximated with accuracy exponentially decaying in the number of parameters used, including 'refinable functions', which play a role in the construction of wavelet systems. We also formulate some alternative proofs for approximation theorems concerning networks, and finally draw parallels between optimization procedures used in learning and others used in various energy/packing problems. This talk represents joint work with Ingrid Daubechies, Ronald DeVore, Nadav Dym, Shira Faigenbaum-Golovin, Shahar Z. Kovalsky, Kung-Ching Lin, Guergana Petrova, Barak Sober, Carlos Trevino, Stephan Wojtowytsch, and Ming Zhong.

Yunan Yang (ETH Zurich)

Benefits of Weighted Training in Machine Learning and PDE-based Inverse Problems

Abstract: Many models in machine learning and PDE-based inverse problems exhibit intrinsic spectral properties, which have been used to explain the generalization capacity and the ill-posedness of such problems. In this talk, we discuss weighted training for computational learning and inversion with noisy data. The highlight of the proposed framework is that we allow weighting in both the parameter space and the data space. The weighting scheme encodes both a priori knowledge of the object to be learned and a strategy to weight the contribution of training data in the loss function. We demonstrate that appropriate weighting from prior knowledge can improve the generalization capability of the learned model in both machine learning and PDE-based inverse problems.

Carlo Marcati (University of Pavia)

Operator network approximations for elliptic PDEs

Abstract: The application of neural networks (NNs) to the numerical solution of PDEs has seen growing popularity in the last five years: NNs have been used as an ansatz space for the solutions, with different training approaches (PINNs, deep Ritz methods, etc.); they have also been used to infer discretization parameters and strategies. In this talk, I will focus on the convergence of operator networks that approximate the solution operator of linear elliptic PDEs. I will, in particular, consider operator networks that, given a fixed right-hand side, map sets of diffusion-reaction coefficients into the space of solutions (coefficient-to-solution map). When the coefficients are smooth and with periodic boundary conditions, the size of the networks can be bounded with respect to the H^1 norm of the error, uniformly over the parameter set. Specifically, the number of non zero weights grows at most poly-logarithmically with respect to the error. The proofs of our approximation rates combine elliptic regularity, classical and recent results in numerical analysis, and tools from NN approximation theory. Using the same techniques, we extend the analysis to linear elasticity and parametric problems.

Tamara Großmann (University of Cambridge)

Unsupervised Learned Total Variation Flow

Abstract: The total variation (TV) flow generates a scale-space representation of an image based on the TV functional. This gradient flow observes desirable features for images such as sharp edges and enables spectral, scale, and texture analysis. The standard numerical approach for TV flow requires solving multiple non-smooth optimisation problems. Even with state-of-the-art convex optimisation techniques, this is often prohibitively expensive and strongly motivates the use of alternative, faster approaches. Inspired by and extending the framework of physics-informed neural networks (PINNs), we propose the TVflowNET, a neural network approach to compute the solution of the TV flow given an initial image and a time instance. We significantly speed up the computation time by more than one order of magnitude and show that the TVflowNET approximates the TV flow solution with high fidelity.

Nikolas Nüsken (King's College London)

Stein variational gradient descent: Gradient Flows, Optimal Transport and Large Deviations

Abstract: Sampling or approximating high-dimensional probability distributions is a key challenge in computational statistics and machine learning. This talk will present connections to gradient flow PDEs, optimal transport and interacting particle systems, focusing on the recently introduced Stein variational gradient descent methodology and some variations. The construction induces a novel geometrical structure on the set of probability distributions related to a positive definite kernel function. We discuss the corresponding geodesic equations, infinitesimal optimal transport maps, as well as large deviation functionals. This is joint work with A. Duncan (Imperial College London), L. Szpruch (University of Edinburgh) and M. Renger (Weierstrass Institute Berlin).

Arnulf Jentzen (University of Münster)

Overcoming the curse of dimensionality: from nonlinear Monte Carlo to deep learning

Abstract: Partial differential equations (PDEs) are among the most universal tools used in modelling problems in nature and man-made complex systems. For example, stochastic PDEs are a fundamental ingredient in models for nonlinear filtering problems in chemical engineering and weather forecasting, deterministic Schroedinger PDEs describe the wave function in a quantum physical system, deterministic Hamiltonian-Jacobi-Bellman PDEs are employed in operations research to describe optimal control problems where companys aim to minimise their costs, and deterministic Black-Scholes-type PDEs are highly employed in portfolio optimization models as well as in state-of-the-art pricing and hedging models for financial derivatives. The PDEs appearing in such models are often high-dimensional as the number of dimensions, roughly speaking, corresponds to the number of all involved interacting substances, particles, resources, agents, or assets in the model. For instance, in the case of the above mentioned financial engineering models the dimensionality of the PDE often corresponds to the number of financial assets in the involved hedging portfolio. Such PDEs can typically not be solved explicitly and it is one of the most challenging tasks in applied mathematics to develop approximation algorithms which are able to approximatively compute solutions of high-dimensional PDEs. Nearly all approximation algorithms for PDEs in the literature suffer from the so-called "curse of dimensionality" in the sense that the number of required computational operations of the approximation algorithm to achieve a given approximation accuracy grows exponentially in the dimension of the considered PDE. With such algorithms it is impossible to approximatively compute solutions of high-dimensional PDEs even when the fastest currently available computers are used. In the case of linear parabolic PDEs and approximations at a fixed space-time point, the curse of dimensionality can be overcome by means of Monte Carlo approximation algorithms and the Feynman-Kac formula. In the first part of this talk we prove that suitable deep artificial neural network (ANN) approximations do indeed overcome the curse of dimensionality in the case of a general class of semilinear parabolic PDEs and we thereby show, for the first time, that a general semilinear parabolic PDE can be solved approximatively without the curse of dimensionality. In the second part of this talk we present some recent results regarding the mathematical analysis of the training of ANNs by means of gradient descent optimization methods.

Matthias J. Ehrhardt (University of Bath)

Bilevel Learning for Inverse Problems

Abstract: Variational regularization techniques are dominant in the field of inverse problems. A drawback of these techniques is that they are dependent on a number of parameters which have to be

set by the user. This issue can be approached by machine learning where we estimate these parameters from data. This is known as "Bilevel Learning" and has been successfully applied to many tasks, some as small-dimensional as learning a regularization parameter, others as high-dimensional as learning a sampling pattern in MRI. While mathematically appealing this strategy leads to a nested optimization problem which is computationally difficult to handle. In this talk we discuss several applications of bilevel learning for imaging as well as new computational approaches. There are quite a few open problems in this relatively recent field of study, some of which I will highlight along the way.

Urbain Vaes (Inria Paris)

Consensus-based sampling

Abstract: In the first part of this talk, I will present background material on Bayesian inverse problems, the associated challenges at the numerical level, and gradient-free sampling and optimization approaches for solving them. In the second part, I will present recent work [1] on a novel gradient-free sampling method that is well suited for Bayesian inverse problems. The method is inspired by consensus-based methods in optimization and based on a stochastic interacting particle system. We demonstrate its potential in regimes where the target distribution is unimodal and close to Gaussian. More precisely, we prove that consensus-based sampling enables to recover a Laplace approximation of the measure in certain parametric regimes and provide numerical evidence that this Laplace approximation attracts a large set of initial conditions in a number of examples.
[1] Carrillo, J.A., Hoffmann, F., Stuart, A.M. and Vaes, U., 2022. Consensus-based sampling. Studies in Applied Mathematics, 148(3), pp.1069-1140.

Abstracts for Contributed Talks

Imad El Bouchairi (University of Caen Normandy)

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 in various 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 non-asymptotic convergence rates in probability and exhibit the different regimes depending on p and the regularity of the L^q-graphon and the initial condition. Joint work with: Jalal Fadili and Abderrahim Elmoataz.
[1] Hafiene, Y., Fadili, J. and Elmoataz, A., 2018. Nonlocal p-Laplacian evolution problems on graphs. SIAM Journal on Numerical Analysis, 56(2), pp.1064-1090.
[2] Borgs, C., Chayes, J., Cohn, H. and Zhao, Y., 2019. 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), pp.3019-3062.
[3] Borgs, C., Chayes, J.T., Cohn, H. and Zhao, Y., 2018. An L^p theory of sparse graph convergence II: LD convergence, quotients and right convergence. The Annals of Probability, 46(1), pp.337-396.

Jeremy Budd (University of Bonn)

Image Processing with "PDEs on Graphs" and their Continuum Limits

Abstract: Over the last decade, an emerging technique for solving image processing tasks has been that of "PDEs on Graphs", stemming from work such as (Bertozzi, Flenner, 2012). This talk will discuss some applications of this technique. First, I will discuss how these techniques can be applied to image segmentation, the task of identifying the key features of an image. Next, I will discuss how these techniques can be extended to (joint) reconstruction-segmentation, whereby the image to be segmented must also be reconstructed from incomplete/noisy measurements. Finally, there has been significant recent work examining the continuum limits of these graph PDEs, and I will discuss how these continuum limits might be useful in applications.

Yahya Saleh (University of Hamburg)

Deep spectral methods for solving variational problems arising from differential equations

Abstract: Neural networks (NN) are being increasingly investigated as approximators to solutions of (partial) differential equations in high dimensions. Numerical simulations show enhanced approximation capabilities of these methods and few numerical analysis results show that such capabilities can be dimension-independent. Here, we investigate approximate solutions to variational problems in a NN-based spectral Galerkin framework. In particular, we seek approximate solutions in the linear span of augmented basis sets. These sets are constructed using deep normalizing flows where the base distributions are standard basis sets. We report convergence guarantees and simulations of this approach for approximating Schwartz functions as well as for approximating eigenfunctions to Schrödinger operators. The proposed method shows more stability during training than typical NN-based nonlinear methods. It has a better scaling to higher dimensions and enables more precise approximations.

Aizhan Issagali (FU Berlin)

On tensor-based training of neural networks

Abstract: Neural networks have been successfully used in many applications. However, there are also cases of failure, for instance, in image classification tasks, when a hardly perceptible perturbation of an image leads to an incorrect classification. In this work by resorting to the continuous model of a shallow neural network we might shed some light on this issue of the success and failure by observing the ill-posedness of the underlying continuous problem, i.e. Fredholm integral equation of the first kind. More specifically, we use the regularised moment collocation method and the generalisation of tensor trains comparable to the so-called functional tensor trains to find an approximation of the density of the parameter distribution. This allows us to reconstruct a neural network with the help of sampling/quadrature, and results in the novel training algorithm.

Julian Suk (University of Twente)

Geometric deep learning of time-dependent hemodynamics on surface meshes with boundary conditions

Abstract: In biomedical engineering, computational fluid dynamics (CFD) is the gold standard for solving the Navier-Stokes equations on realistic blood vessel shapes. To this end, the domain is discretised in finite elements yielding tetrahedral volume and triangluar surface meshes. We observe that every surface mesh induces a graph by discarding the information which vertices belong to which triangle. As a consequence, we can use graph neural networks (GNN) to perform operations on signals defined on these shapes. The graph structure enables us to use SO(3)-equivariant neural networks, a recent advance in the field of geometric deep learning. We show that this framework is powerful enough to predict vector-valued hemodynamic quantities that vary over time and depend on a given flow boundary condition. Furthermore, we can make a connection between the discretisation of the Navier-Stokes equations and the graph which our network uses to process information.

Lihan Wang (CMU Pittsburgh)

Accelerate Sampling Using Birth-Death Dynamics

Abstract: In this talk, I will discuss the birth-death dynamics for sampling multimodal probability distributions, which is the spherical Fisher-Rao gradient flow of relative entropy. The advantage of the birth-death dynamics is that, unlike any local dynamics such as the overdamped Langevin dynamics, it allows global movement of mass directly from one mode to another, without the difficulty of going through low probability regions. We prove that the birth death dynamics converges to the unique invariant measure with a uniform rate under some mild conditions, showing its potential of overcoming metastability. To further illustrate the strengths and weaknesses of the birth-death dynamics, we will also discuss other possible extensions and variants of the dynamics, including implementations using kernel density estimators, as well as different choices of energy functional. Joint work with Yulong Lu (UMass Amherst) and Dejan Slepcev (CMU).

Jakwang Kim (University of Wisconsin Madison)

On generalized barycenter problems and their connections to adversarial learning

Abstract: In a recent paper we study the connection between multimarginal optimal transport and adversarial learning, and their connections to a geometric problem that we refer to as generalized barycenter problem. Our motivation to study these connections is to understand the geometry of adversarial learning in multiclass classification. The study of this geometric problem, i.e. the generalized Barycenter problem, not only gives an understanding of adversarial learning, but also provides new directions for research of interest on their own right. In this talk, I'll discuss this family of problems and provide very concrete, but still enlightening, examples illustrating some of our findings. This is a joint work with Nicolás García Trillos and Matt Jacobs.

Andrea Manzoni (Polytechnic University of Milan)

POD-DL-ROM: a comprehensive deep learning-based approach to reduced order modeling of nonlinear time-dependent parametrized PDEs

Abstract: Conventional reduced order modeling techniques such as the reduced basis method (relying, e.g., on proper orthogonal decomposition (POD)) may incur in severe limitations when dealing with nonlinear time-dependent parametrized PDEs, as these are strongly anchored to the assumption of modal linear superimposition they are based on. These might be related to (i) the need to deal with projections onto high dimensional linear approximating trial manifolds, (ii) expensive hyper-reduction strategies, or (iii) the intrinsic difficulty to handle physical complexity with a linear superimposition of modes. To overcome these limitations, we propose a new nonlinear approach to set reduced order models (ROMs) by exploiting deep learning (DL) algorithms. In the resulting nonlinear ROM, which we refer to as DL-ROM, both the nonlinear trial manifold (corresponding to the set of basis functions in a linear ROM) and the nonlinear reduced dynamics (corresponding to the projection stage in a linear ROM) are learned in a non-intru ive way, by relying on deep (e.g., feedforward, convolutional, autoencoder) neural networks; these latter are trained on a set of full order model (FOM) solutions obtained for different parameter values. Although extremely efficient at testing time, when evaluating the PDE solution for any new testing-parameter instance, DL-ROMs require an expensive training stage, because of the extremely large number of network parameters to be estimated. We show that a substantial speed up in the training stage of DL-ROMs can be achieved by (i) performing a prior dimensionality reduction through POD, and (ii) relying on a multi-fidelity pretraining stage, where different physical models can be efficiently combined. The resulting POD-DL-ROM strategy is finally

tested on several different cases (such as, e.g., nonlinear diffusion-reaction, nonlinear elastodynamics, unsteady Navier-Stokes equations and fluid-structure interaction problems) to show the generality of this approach and its remarkable computational savings. In pa ticular, numerical results indicate that POD-DL-ROMs whose dimension is equal to the intrinsic dimensionality of the solutions' manifold are able to efficiently approximate the solution of parametrized systems, especially in those cases for which a huge number of POD modes would have been necessary to achieve the same degree of accuracy. Joint work with Stefania Fresca, Politecnico di Milano

---

Nicola Rares Franco (Polytechnic University of Milan)

A theoretical analysis of Deep Learning-based Reduced Order Models

Abstract: In the context of parameter dependent PDEs, Deep Learning-based Reduced Order Models (DL-ROMs) are a class of model surrogates that can efficiently approximate the parameter-to-solution map. Differently from classical state-of-the-art approaches, such as the Reduced Basis method, DL-ROMs provide an appealing alternative for tackling problems with a slow decay in the Kolmogorov n-width. In fact, by carrying out a nonlinear dimensionality reduction of the solution manifold through the use of deep autoencoders, DL-ROMs are able to overcome the classical difficulties of linear methods. Here, we focus on the theoretical properties of DL-ROMs, assessing both their approximation power and their complexity. First, starting from a nonlinear version of the Kolmogorov n-width, we introduce the concept of a minimal latent dimension, which we later link with the more familiar notion of latent dimension for autoencoders. Then, we proceed to bound the DL-ROM approximation error in terms of the complexity of the neural netwo s involved. Finally, we argue the importance of using specific architectures, such as convolutional layers, in the design process of deep autoencoders. To this end, we present some preliminary error bounds that show how the mathematical properties of the original problem affect the design of DL-ROMs.

Hendrik Kleikamp (University of Münster)

A certified and adaptive RB-ML-ROM surrogate approach for parametrized PDEs

Abstract: In this talk, we present a new surrogate modeling approach to approximate input-output maps occurring in the context of parametrized PDEs. The algorithm is based on a full order model (FOM), reduced order model (ROM) and machine learning (ML) model chain. By applying a posteriori error estimates to the ROM and ML model, we obtain certified results. The model is therefore able to fulfill fixed or adaptively chosen error tolerances for every requested parameter. The reduced order and machine learning models are adapted during the algorithm based on the a posteriori error estimates. In particular, the algorithm provides ML-based results with guaranteed quality statements and error control. Numerical experiments showcase the efficiency of our approach in different scenarios, for instance a parameter optimization problem and uncertainty quantification. In our examples, the reduced basis (RB) method is used as ROM. Further, we employ kernel methods and deep neural networks to generate the ML models, but a wide range of ML algorithms is applicable within the modeling chain.
