Matematica per l'Intelligenza Artificiale e il Machine Learning
Giovani ricercatori

Workshop Program

Thursday 24th, Morning session

8:30– **Registration**

8:45–9:00 **Opening**

9:00–9:20 **S. Ivan Trapasso**, *Some results on the stability of the scattering networks*

9:20–9:40 **Alessandro Scagliotti**, *Ensemble optimal control: ResNets, diffeomorphisms approximation and Normalizing Flows*

9:40–10:00 **Francesca Bartolucci**, *Understanding neural networks with reproducing kernel Banach spaces*

10:00–10:20 **Andrea Basteri**, *Estimate on Wasserstein distance for Gaussian Neural Networks*

10:20–10:50 **Coffee break**

10:50–11:10 **Andrea Della Vecchia**, *Regularized ERM on random subspaces*

11:10–11:30 **Francesco Iafrate**, *Bridge-type estimation with mixed-rate asymptotics: applications to Stochastic Differential Equations*

11:30–11:50 **Stefano Vigogna**, *Exponential convergence for support vector machines (without hard margin)*

11:50–12:10 **Kasun Fernando**, *A Robust Normalizing Flow using Bernstein-type Polynomials*

12:10–12:30 **Ambrogio Maria Bernardelli**, *A MILP approach to a structured ensemble Binarized Neural Network*

12:30–14:30 **Lunch**

Thursday 24th, Afternoon session

14:30–14:50 **Joshua Wrigley**, *Dynamical objects: Categoric and Topos-theoretic techniques for Deep Neural Networks*

14:50–15:10 **Caterina Graziani**, *Weisfeiler–Lehman goes Dynamic: An Analysis of the Expressive Power of Graph Neural Networks for Attributed and Dynamic Graphs*

15:10–15:30 **Francesco Regazzoni**, *Automatic discovery of low-dimensional dynamics underpinning time-dependent PDEs by means of Neural Networks*

15:30–15:50 **Giulia Bertaglia**, *Asymptotic-Preserving Neural Networks for multiscale hyperbolic models*

15:50–16:10 **Maria Strazzullo**, *Physics-informed Neural Networks for partial differential equations and optimal control in a parametric setting*

16:10–16:40 **Coffee break**

16:40–17:00 **Nicola Rares Franco**, *A theoretical analysis of Deep Learning-based Reduced Order Models*

17:00–17:20 **Stefania Fresca**, *Deep learning-based reduced order models for the real-time solution of parametrized PDEs*

17:20–17:40 **Davide Oberto**, *An invariances-preserving vector basis neural network for the closure of Reynolds-averaged Navier–Stokes equations*

17:40–18:00 **Davide Palitta**, *A simple yet effective tensor-based ODE model for Deep Learning*

18:00–18:20 **Stefano Pagani**, *Physics-based deep learning models in cardiac electrophysiology and cardiovascular flows*

Friday 25th, Morning session

9:00–9:20 **Giacomo Meanti**, *Hyperparameter Tuning for Nystroem Kernel Ridge Regression*

9:20–9:40 **Luca Ratti**, *Learning the optimal regularizer for linear inverse problems*

9:40–10:00 **Konstantin Riedl**, *Consensus-Based Optimization: A Holistic Convergence Analysis and a Machine Learning Application*

10:00–10:20 **Agnese Pacifico**, *Optimal control and Reinforcement Learning - An Algorithm for the LQR Problem with Partially Unknown Dynamics*

10:20–10:50 **Coffee break**

10:50–11:10 **Carlo Marcati**, *Neural network and operator network approximations for elliptic PDEs*

11:10–11:30 **Moreno Pintore**, *Enforcing Dirichlet Boundary Conditions in Physics-Informed Neural Networks and Variational Physics-Informed Neural Networks*

11:30–11:50 **Edoardo Centofanti**, *A DeepOnet Approach to Learning Dynamical Systems in Biomathematics*

11:50–12:10 **Giuseppe Alessio D'Inverno**, *Physics Informed Neural Networks for generation of structured grids on bounded domains*

12:10–12:30 **Caterina Millevoi**, *Neural network application to forward and inverse problems involving differential equations*

12:30–14:30 **Lunch**

Friday 25th, Afternoon session

14:30–14:50 **Gianluca Manzan**, *Efficiency limits of Restricted Boltzmann Machines in teacher-student frameworks*

14:50–15:10 **Filippo Zimmaro**, *Voter model with external biases on a modular network*

15:10–15:30 **Francesco Camilli**, *Inhomogeneity in inference and spin glasses: the deep Boltzmann machine on the Nishimori line*

15:30–15:50 **Francesco Alemanno**, *Supervised Hebbian learning*

15:50–16:20 **Coffee break**

16:20–16:40 **Faraz Ahmad**, *Equivariance, Graphs, and Deep Learning*

16:40–17:00 **Giovanni Bocchi**, *Group Equivariant Non-Expansive Operators: from TDA to Neural Networks*

17:00–17:20 **Maria Sofia Bucarelli**, *A topological description of loss surfaces via Betti Numbers characterization*

17:20–17:40 **Antonio Mastropietro**, *From eXplainable Artificial Intelligence to Game Theory with Coalitional Hodge-Shapley Value*

Saturday 26th, Morning session

9:00–9:20 **Giorgia Franchini**, *A line search based proximal stochastic gradient algorithm with dynamical variance reduction*

9:20–9:40 **Simone Rebegoldi**, *On a stochastic first-order trust-region method for machine learning applications*

9:40–10:00 **Cinzia Bandiziol**, *Adaptive gradient methods in Training Neural Network with Topological Layer*

10:00–10:20 **Giacomo Borghi**, *On stochastic particle systems for complex, global optimization problems*

10:20–10:40 **Eleonora Vercesi**, *Computing Disease-Specific Gene Embeddings via Constrained Optimization*

10:40–11:10 **Coffee break**

11:10–11:30 **Matteo Caldana**, *Accelerating Algebraic Multigrid Methods via Artificial Neural Networks*

11:30–11:50 **Francesco Della Santa**, *Learning Discontinuities via Discontinuous Neural Networks*

11:50–12:10 **Davide Evangelista**, *Stability-Accuracy Trade-off in Neural Networks for Ill-conditioned Inverse Problems*

12:10–12:30 **Danilo Pezzi**, *Unrolling With No Deep: Explainable Bilevel Optimization For The Helsinki Deblur Challenge*

12:30–12:50 **Alberto Fachechi**, *Dreaming Neural Networks - pushing retrieval/learning capabilities to the limit by "sleeping" mechanism*

Abstracts

Faraz Ahmad, *Equivariance, Graphs, and Deep Learning*
University of Bologna

Equivariant operators have proved useful in deep learning. Their success, especially in enabling us to reduce the large dimensionality of data, motivates one to explore the metric and topological properties of the spaces of group equivariant non-expansive operators (GENEOs, for short). One of the most important properties of these spaces is compactness, as it allows one to approximate the space with finitely many operators. We prove that if the data spaces are totally bounded, then under some mild conditions, the data spaces, equivariance groups, and the space of GENEOs under consideration, can all be nicely embedded in compact spaces by means of compatible embeddings, thereby making available a mathematical model for more transparent and interpretable neural networks.

Francesco Alemanno, *Supervised Hebbian learning*
Dip.to di Matematica e Fisica "Ennio de Giorgi", Unisalento, Lecce

In the neural network literature, Hebbian learning traditionally refers to the procedure by which the Hopfield model and its generalisations store the archetypes (i.e. patterns that are experienced only once to form the synaptic matrix). However, the term "Learning" in Machine Learning refers to the machine's ability to extract features from the provided dataset (e.g., consisting of fuzzy examples of these archetypes), in order to create its own representation of the unavailable archetypes. Here, given a sample of examples, we define a supervised learning protocol by which the Hopfield network can infer the archetypes, and identify the correct control parameters (including the size and quality of the dataset) to represent a phase diagram for system performance. We also show that, for structure-free datasets, the Hopfield model equipped with this supervised learning rule is equivalent to a restricted Boltzmann machine and this suggests an optimal and interpretable training routine. Finally, this approach is generalised to structured datasets: we highlight a quasi-ultrametric organisation (reminiscent of replica-symmetry breaking) in the analysed datasets and, as a consequence, introduce an additional hidden broken-replica layer for its (partial) disentanglement, which is shown to improve the MNIST classification from $\sim$75% to $\sim$95%, and offer a new perspective on deep architectures.

Cinzia Bandiziol, *Adaptive gradient methods in Training Neural Network with Topological Layer*
Università di Padova, Dipartimento di Matematica

In the last two decades, Topological Data Analysis (TDA) has emerged as a good choice in analyzing big amount of data. Thanks to its robust basis of algebraic topology and its main tool, the persistent homology, TDA is able to extract intrinsic information about the "shape" of the data and allows to collect them in some graphic viewers, among them the persistent diagram. But dealing with them, one faces inevitably the problem of "vectorization" in order to apply then standard Machine Learning tools. On the other hand, the desire of managing huge amount of data has brought to the relevant growth of Deep Learning, thanks to the complexity of its own Neural Networks. Starting from the widely used Stochastic Gradient Descent, a recent research trend is to develop some variants with the aim to overcome issues like fine-tune process that is often really expensive and slows down the convergence rate of the method itself. As a result, the Adaptive Gradient Methods have been proposed like Adagrad, Adam and RMSProp. Their strength is mainly related to the update of learning rate according to each variable so to encourage larger movements in directions with consistently steep gradients avoiding the fine-tune process. In this talk, we'll describe how to include the vectorization task of persistent diagrams in a specific layer of a Neural Network. Then we'll introduce some recent Adaptive optimizers as Adagradnorm and Padam and will test the architecture with these optimizers

in a classification framework.
Authors: C. Bandiziol, S. Cuomo, S. De Marchi, V.S. Di Cola

References
M. Carriere et al. PersLay: A Neural Network Layer for Persistence Diagrams and New Graph Topological Signatures (2020)
K. Kim et al. PLLay: Efficient Topological Layer based on Persistence Landscapes (2021)
R. Ward et al. AdaGrad Stepsizes: Sharp Convergence Over Nonconvex Landscapes (2019)
J. Chen et al. Closing the Generalization Gap of Adaptive Gradient Methods in Training Deep Neural Networks (2020)

Francesca Bartolucci, *Understanding neural networks with reproducing kernel Banach spaces*
ETH Zürich

Characterizing the function spaces corresponding to neural networks can provide a way to understand their properties. We discuss how the theory of reproducing kernel Banach spaces can be used to tackle this challenge. In particular, we prove a representer theorem for a wide class of reproducing kernel Banach spaces that admit a suitable integral representation and include one hidden layer neural networks of possibly infinite width. Further, we show that, for a suitable class of ReLU activation functions, the norm in the corresponding reproducing kernel Banach space can be characterized in terms of the inverse Radon transform of a bounded real measure, with norm given by the total variation norm of the measure.

Andrea Basteri, *Estimate on Wasserstein distance for Gaussian Neural Networks*
INRIA Paris

Given any fully connected neural network with randomly initialized Gaussian weights, we estimate from the above the Wasserstein distance between the output distribution on the layers of the network and a suitable limit Gaussian process, the Neural Network Gaussian Process. Our inequality is in function of the size of the hidden and output layers, and we recover the well known distributional result when the size of the hidden layers goes to infinity. We support our work with computational experiments that show that our result may not be sharp, and so further studies may be conducted and other directions may be explored.

References
https://arxiv.org/abs/2203.07379v1
https://etd.adm.unipi.it/theses/available/etd-05022022-184611/unrestricted/...
Master_Thesis_Basteri_Andrea_Definitiva.pdf
https://www.cs.toronto.edu/~radford/ftp/pin.pdf
https://arxiv.org/pdf/1711.00165

Ambrogio Maria Bernardelli, *A MILP approach to a structured ensemble Binarized Neural Network*
Università di Pavia

Binarized Neural Networks (BNNs) are receiving increasing attention [1] thanks to their lightweight architecture and ability to run on low-power devices. The current state of the art for training BNNs for classification tasks restricted to few-shot learning is the mixed-integer linear programming (MILP) approach introduced in [2]. In this work, we propose a structured ensemble method based on training one BNN for each possible pair of classes and then using majority voting to predict the final output. The single BNN discriminating between the classes of a given pair is performed by a MILP model that optimizes a bicriteria objective function, which maximizes the margin and minimizes the number of active weights. We validated our model on the MNIST dataset, finding that, with a limited number of training data for class (e.g., 25 images per class),

our structured ensemble BNNs outperform both those trained by stochastic gradient descent [1] and by the MILP-based approach [2].

References

[1] Hubara, I., Courbariaux, M., Soudry, D., El-Yaniv, R. and Bengio, Y., 2016. Binarized neural networks. Advances in Neural Information Processing Systems (NeurIPS), 29.
[2] Toro Icarte, R., Illanes, L., Castro, M.P., Cire, A.A., McIlraith, S.A. and Beck, J.C., 2019, September. Training binarized neural networks using MIP and CP. International Conference on Principles and Practice of Constraint Programming (pp. 401-417). Springer, Cham.

Giulia Bertaglia, *Asymptotic-Preserving Neural Networks for multiscale hyperbolic models*
University of Ferrara

With the rapid advance of Machine Learning techniques and the deep increment of availability of scientific data, data-driven approaches have started to become progressively popular across science, causing a fundamental shift in the scientific method after proving to be powerful tools with a direct impact in many areas of society. Nevertheless, when attempting to study the dynamics of complex multiscale systems, the usage of standard Deep Neural Networks (DNNs) and even standard Physics-Informed Neural Networks (PINNs) may lead to incorrect inferences and predictions, due to the presence of small scales leading to reduced or simplified models in the system that have to be applied consistently during the learning process. In this talk, we will address these issues in light of recent results obtained in the development of Asymptotic-Preserving Neural Networks (APNNs) for multiscale hyperbolic models [1,2]. Several numerical tests show how APNNs provide considerably better results with respect to the different scales of the problem when compared with standard DNNs and PINNs, especially when analyzing scenarios in which only little and scattered information is available.

References

[1] Bertaglia G., Asymptotic-Preserving Neural Networks for hyperbolic systems with diffusive scaling, Preprint, 2022.
[2] Bertaglia G., Lu C., Pareschi L., Zhu X., Asymptotic-Preserving Neural Networks for multiscale hyperbolic models of epidemic spread. Mathematical Models and Methods in Applied Sciences, online ready, 2022.

Giovanni Bocchi, *Group Equivariant Non-Expansive Operators: from TDA to Neural Networks*
Dipartimento di Scienze e Politiche Ambientali, Università di Milano

Equivariance is by now considered a key property when analyzing data that can undergo geometrical transformations. With this belief in mind, Group Equivariant Non-Expansive Operators (GENEOs) are recently emerging as mathematical tools to build new kinds of Neural Networks. To reach this objective, standard neurons can be replaced by parametric families of operators whose equivariance is set with respect to specific groups of transformations, then these new neurons are connected through admissible operations between GENEOs. Moreover, the resulting network can be trained in a gradient descent fashion in order to optimize the values of the parameters. Such models, of which some examples are already available, should be able to reach a high degree of explainability, to incorporate any available prior knowledge and to mitigate the hunger for training examples of classical Neural Networks. Alongside this, the theory of GENEOs is growing with theoretical results which are useful also in the context of applications. A last important fact is that GENEOs are deeply linked with the theory of Topological Data Analysis (TDA) and there are more than a few points of connection: for example GENEOs can help to restrict the intrinsic invariance of TDA, which is insensitive to homeomorphisms, only to proper subgroups of homeomorphisms.

References

[1] Bocchi, G. et al. GENEOnet: A new machine learning paradigm based on Group Equivariant Non-

Expansive Operators. An application to protein pocket detection. Preprint at arXiv.2202.00451 (2022).
[2] Bocchi, G., Botteghi, S., Brasini, M., Frosini, P. and Quercioli, N. On the finite representation of group equivariant operators via permutant measures. Preprint at arXiv:2008.06340 (2022).
[3] Bergomi, M. G., Frosini, P., Giorgi, D. and Quercioli, N. Towards a topological–geometrical theory of group equivariant non-expansive operators for data analysis and machine learning. Nature Machine Intelligence 1, 423–433 (2019).

Giacomo Borghi, *On stochastic particle systems for complex, global optimization problems*
RWTH Aachen University

Learning from data requires to solve a high dimensional, non-convex, non-smooth optimization problem which is typically solved by first-order methods that are likely to find local minima. In this talk, we will consider a class of global, gradient-free methods, the so-called metaheuristic algorithms, where an ensemble of particles explores the search space combining deterministic interactions and a stochastic behaviour. We will review recent results concerning, in particular, the class of consensus-based optimization (CBO) algorithms which can be studied in a mathematically rigorous fashion by means of a mean-field description of the dynamics. Also, by exploiting the interaction between particles and by possibly distributing tasks among them, one can design algorithms to tackle complex optimization problems in an efficient way. We will present how this can be done in the context of constrained optimization, multi-objective optimization and a machine learning application.

References

A. Benfenati, G. Borghi, and L. Pareschi. Binary interaction methods for high dimensional global optimization and machine learning. Applied Mathematics & Optimization, 86(1):9, June 2022.
G. Borghi, M. Herty, and L. Pareschi. Constrained consensus-based optimization. SIAM Journal on Optimization, to appear, 2021.
G. Borghi, M. Herty, and L. Pareschi. An adaptive consensus based method for multi-objective optimization with uniform Pareto front approximation. arXiv pre-print (2022)

Maria Sofia Bucarelli, *A topological description of loss surfaces via Betti Numbers characterization*
Sapienza Università di Roma

In recent years the investigation on searching for an appropriate description, either in analytical or topological fancy ,[Freeman and Bruna 2016], of the surface of the loss function in deep learning models has become a priority, in order to explain why deep neural networks are successfully trained via gradient descent based methods, despite the strong non-convexity of the associated optimization problem. In several works remarkable efforts have been made to find either a description of the spurious minima location for specific architectures (e.g. ReLu [Networks Safran and Shamir 2018]; [Venturi et al. 2018, 2019]) or a characterization of the behaviour of gradient dynamics [Maennel et al. 2018];[Williams et al. 2019]. Our work is an attempt to find a topological characterization of the most used loss functions, optimized on commonly used feedforward neural networks, through a computation of tight bounds on the Betti numbers associated with the loss surface, highlighting how the number of layers can impact on this characterization. This work takes inspiration from Bianchini and Scarselli (2014), in which the same theory is exploited to prove that the geometry of the output set in shallow networks is far less complex than the one in deep networks. This intuition can better explain why the optimization task is carried on on shallow networks more easily than in the deep ones, but on the other hand it shed light over the learning capabilities of deep neural networks.

Matteo Caldana, *Accelerating Algebraic Multigrid Methods via Artificial Neural Networks*
Politecnico di Milano

We present a novel deep learning-based algorithm to accelerate the convergence of Algebraic Multigrid (AMG) methods [Antonietti2021]. AMG methods are a purely matrix-based approach: they construct a sequence of increasingly smaller problems directly from the system matrix based on the value of the strong connection parameter [Stuben2001]. Through the use of artificial neural networks (ANNs), we show that we can tune the strong connection parameter so as to maximize the corresponding convergence factor of the AMG scheme. We propose to apply a procedure analogous to the pooling employed in convolutional neural networks [Krizhevsky2017] to the matrix of the linear system to which the AMG method is applied. To demonstrate the practical capabilities of the proposed algorithm, denoted AMG-ANN, we apply it to the iterative solution of the linear systems of equations stemming from Finite Element discretizations of two-dimensional model problems. Namely, we consider an elliptic equation with a highly heterogeneous diffusion coefficient and a stationary Stokes problem. We present an in-depth analysis of the effects of tuning the strong threshold parameter on the convergence factor of the resulting AMG iterative scheme. Finally, we quantify the improvements with respect to using the value commonly found in the literature.

References

[Antonietti2021] P.F. Antonietti, L. Dede', M. Caldana. Accelerating algebraic multigrid methods via artificial neural networks, 2021. MOX Report 69/2021, arXiv:2111.01629. Submitted.
[Stuben2001] K. Stuben. A review of algebraic multigrid. Numerical Analysis: Historical Developments in the 20th Century, 2001.
[Krizhevsky2017] A. Krizhevsky, I. Sutskever and G. E. Hinton. Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 2012.

Francesco Camilli, *Inhomogeneity in inference and spin glasses: the deep Boltzmann machine on the Nishimori line*
Università di Bologna - École normale supérieure de Paris

Since their birth, spin glasses have been an ever-growing research field, mainly thanks to their universality features that allow to model a remarkable variety of phenomena. One of their applications is finding information theoretical limits for high dimensional inference. In this respect, the contribution will focus on a particular class of inference models, the so-called "spiked models", in which we need to infer a low rank matrix from noisy observations of it. With respect to the previous literature, the model is generalized by introducing an inhomogeneity in the noise. It turns out that its information theoretical limits can be expressed by means of standard thermodynamic quantities, such as the free energy, of a multispecies, or multipartite, spin glass in a particular region of its phase space: the Nishimori line. Moreover, we make a particular choice for the inhomogeneity, so that it breaks a technical convexity requirement on the interactions between particles, or spins, in different species. The spin glass then arising is a Deep Boltzmann Machine (DBM) on the Nishimori line. Being in such region of its phase space, we can prove that the model is replica symmetric, namely it admits a finite dimensional variational principle to express its free energy that we rigorously prove. Finally, we shall see how the geometry of the model, specifically the relative sizes of the species, or layers in the DBM, can influence its phase transition.

References

Alberici, D., Camilli, F., Contucci, P., Mingione, E. The Solution of the Deep Boltzmann Machine on the Nishimori Line. Commun. Math. Phys. 387, 1191–1214 (2021).
Alberici, D., Camilli, F., Contucci, P., Mingione, E. (2022). A Statistical Physics approach to a multi-channel Wigner spiked model. Europhysics Letters, 136(4), 48001.

Edoardo Centofanti, *A DeepOnet Approach to Learning Dynamical Systems in Biomathematics*
Università di Pavia

DeepOnet is a general deep learning framework recently proposed in [1] to learn different continuous nonlinear operators, based on an extension of the Universal Approximation Theorem for Operators. It has been used for learning many different problems in physics, computational chemistry, and biology, such as inferring bubble dynamics [3], learning diffusion-reaction advection problems and stochastic PDEs [1,2]. In this work, we present a strategy to exploit DeepONets to learn ODEs systems governing some ionic models arising in the field of Biomathematics. In particular, some results regarding the FitzHugh- Nagumo ionic model are presented and discussed, as well as some approaches useful to train both the transmembrane potential and the gating and concentration variables related to these models.

References
[1] Lu, Karniadakis et al., Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators, Nature Machine Intelligence, 2021
[2] Lu, Karniadakis et al., A comprehensive and fair comparison of two neural operators (with practical extensions) based on FAIR data, Comput. Methods Appl. Mech. Engrg., 2022
[3] Lin, Karniadakis et al., A seamless multiscale operator neural network for inferring bubble dynamics, J. Fluid Mech., 2021

Giuseppe Alessio D'Inverno, *Physics Informed Neural Networks for generation of structured grids on bounded domains*
Università degli Studi di Siena

The generation of structured grids on bounded domains is a crucial point in the development of numerical models for solving differential problems. In particular, the representation of the given computational domain through a regular parameterization allows us to define a univalent mapping which can be computed as the solution of an elliptic problem, equipped with suitable Dirichlet boundary conditions, [1]. In recent years, Physics Informed Neural Networks (PINNs) have been proved to be a powerful tool to compute the solution of PDEs replacing standard numerical models with deep neural networks : PINNs can be used for predicting the values on simulation grids of different resolutions without the need to be retrained [2]. In this work, we exploit the PINN model in order to solve the PDE associated to the differential problem of the parameterization on both convex and non-convex 2D and 3D domains, for which the describing PDE is known. The final continuous model is then provided by applying a Hermite type quasi-interpolation operator [3] which can guarantee the desired smoothness of the sought parameterization. Finally, some numerical examples are presented.

References
[1] Alfio Quarteroni. Numerical models for differential problems. Springer, 2017.
[2] Stefano Markidis. The Old and the New: Can Physics-Informed Deep-Learning replace traditional linear solvers? Frontiers in big Data, page 92, 2021.
[3] Francesca Mazzia and Alessandra Sestini. The BS class of Hermite spline quasi-interpolants on nonuniform knot distributions. BIT Numerical Mathematics, 49(3):611–628, 2009.

Francesco Della Santa, *Learning Discontinuities via Discontinuous Neural Networks*
Politecnico di Torino - Dipartimento di Scienze Matematiche

In this presentation, we describe a new typology of layers for Neural Networks (NNs) characterized by discontinuities in the function characterizing the layer's action [1]; then, a discontinuous Neural Network is

obtained. To the best of the authors' knowledge, in recent literature do not exist examples of discontinuous NNs, with the few exceptions that study the theoretical consequences of using basic discontinuous activation functions for approximation tasks. Other cases of discontinuous NNs are the first mathematical models of the '50s, inspired by the biological neurons' mechanisms, and nowadays abandoned for practical reasons. We introduce discontinuities in NNs aiming to improve the approximation of discontinuous functions and, at the same time, to detect the discontinuity interfaces. This latter problem is a challenging task, especially for functions with a high-dimensional domain. The information about discontinuities can be quite relevant in several applications (e.g., numerical methods for stochastic collocation in the framework of uncertainty quantification). Actually, methods dedicated to the discontinuity interfaces detection problem exists (e.g., [2]), but they obtain good results only for small dimensions and with particular assumptions on the function. In particular, the new discontinuous layer is characterized by trainable parameters that let the NN "learn" the discontinuities of the approximated function from data. Then, thanks to a sound theoretical analysis of the discontinuous NNs' properties, we define an algorithm that detect the continuity regions and the discontinuity interfaces of the target function, looking at the weights of the trained NN.

References

[1] Della Santa, Pieraccini, "Discontinuous neural networks and discontinuity learning", JCAM, 2023.
[2] Jakeman, Archibald, Xiu, "Characterization of discontinuities in high-dimensional stochastic problems on adaptive sparse grids", JCP, 2011.

Andrea Della Vecchia, *Regularized ERM on random subspaces*
Università degli Studi di Genova

We study a natural extension of classical empirical risk minimization, where the hypothesis space is a random subspace of a given space. In particular, we consider possibly data dependent subspaces spanned by a random subset of the data, recovering as a special case Nyström approaches for kernel methods. Considering random subspaces naturally leads to computational savings, but the question is whether the corresponding learning accuracy is degraded. These statistical-computational tradeoffs have been recently explored for the least squares loss and self-concordant loss functions, such as the logistic loss. Here, we work to extend these results to convex Lipschitz loss functions, that might not be smooth, such as the hinge loss used in support vector machines. This unified analysis requires developing new proofs, that use different technical tools, such as sub-gaussian inputs, to achieve fast rates. Our main results show the existence of different settings, depending on how hard the learning problem is, for which computational efficiency can be improved with no loss in performance.

Davide Evangelista, *Stability-Accuracy Trade-off in Neural Networks for Ill-conditioned Inverse Problems*
University of Bologna

Deep learning algorithms have recently become state-of-art in solving Inverse Problems, overcoming the classical variational methods in terms of both accuracy and efficiency. On the other hand, it is still unclear if neural networks can compete in terms of reliability and a rigorous complete analysis still lacks in the literature. Starting from the brilliant works of N.M.Gottschling, V.Antun (2020) and M.J.Colbrook, V.Antun (2021), we will try to understand the relationship between the accuracy and stability of neural networks for solving ill-conditioned inverse problems, deriving new theoretical results shedding light on the trade-off between accuracy and stability. Following the study of M.Genzel, J.Macdonald (2020), we will find that, under some conditions, neural networks can be more unstable the more they are accurate, and we will propose new regularization techniques with provable increase in stability and minumum accuracy loss. The results will appear soon on a paper in collaboration with J.Nagy, E.Loli Piccolomini and E.Morotti.

Alberto Fachechi, *Dreaming Neural Networks - pushing retrieval/learning capabilities to the limit by "sleeping" mechanism*
Dipartimento di Matematica "Guido Castelnuovo", Sapienza Università di Roma

Hopfield model [1] is the prototype of Associative Neural Networks, in which the information (codified as a set of binary patterns) is distributed in the synaptic couplings according to the Hebb learning rule [2], and corresponds to attractors for the network dynamics. Despite its success, the model is affected by strong limitations, such as the poor critical storage capacity (defined as the ratio between the number of stored patterns and the network size) which turns out to be much smaller than the theoretical upper bound [3]. In this talk, we will show some recent results about an extension of the Hopfield model, called Dreaming Neural Networks [4], exhibiting enhanced capabilities in handling stored information. We will also discuss some relevant results concerning dreaming mechanism in the context of Machine Learning [5].

References
[1] J.J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, Proceedings of the national academy of sciences 79.8 (1982): 2554-2558.
[2] D.O. Hebb, The organization of behavior: A neuropsychological theory, Psychology Press, (1949).
[3] E. Gardner, The space of interactions in neural network models, J. Phys. A 21(1):257, (1988).
[4] A. Fachechi, E. Agliari, A. Barra, Dreaming neural networks: forgetting spurious memories and reinforcing pure ones, Neural Networks 112, pp. 24-40, 2019.
[5] A. Fachechi, A. Barra, E. Agliari, F. Alemanno, IEEE Trans. on Neur. Netw. and Learn. Syst., 2022, doi: 10.1109/TNNLS.2022.3182882.

Kasun Fernando, *A Robust Normalizing Flow using Bernstein-type Polynomials*
Centro De Giorgi, Scuola Normale Superiore di Pisa

Modeling real-world distributions can often be challenging due to sample data subjected to perturbations, e.g., instrumentation errors, or added random noise. Since flow models are typically nonlinear algorithms, they amplify these initial errors, leading to poor generalizations. We propose a framework to construct Normalizing Flows which demonstrate higher robustness against such initial errors using Bernstein-type polynomials. Further, compared to the existing models, our method has other advantages like theoretical upper bounds for the approximation error, better suitability for compactly supported densities, and the ability to use higher degree polynomials (leading to smaller approximation errors) without instability in the training process. We conduct a theoretical analysis and empirically demonstrate the efficacy of the proposed technique using experiments on both real-world and synthetic datasets.

Giorgia Franchini, *A line search based proximal stochastic gradient algorithm with dynamical variance reduction*
UNIMORE

Many optimization problems arising from machine learning applications can be cast as the minimization of the sum of two functions: the first in practice is the empirical risk, and the other one could impose a priori information on the solution. The second term is, in general, convex and could be non-differentiable. In the case of non-differentiability proximal gradient methods are very well suited to face this optimization problems. However, when dealing with large-scale machine learning issues, the computation of the full gradient of the differentiable term can be prohibitively expensive by making these algorithms unsuitable. For this reason, stochastic gradient methods have been extensively studied in the optimization area in the last decades. In this talk I present in both the cases, differentiable [1] and non-differentiable [2], a stochastic gradient algorithm which is based on two main ingredients. We combine a proper technique to dynamically reduce the variance

of the stochastic gradients along the iterative process with a descent condition in expectation for the objective function, aimed to fix the value for the steplength parameter at each iteration. For general objective, proof about a.s. convergence has been shown. The practical implementation of the proposed method needs neither the computation of the exact gradient of the empirical risk during the iterations nor the tuning of an optimal value for the steplength. An extensive numerical experimentation, also with Convolutional Neural Networks, highlights that the proposed approach appears robust with respect to the setting of the hyperparameters and competitive compared to state-of-the-art methods.

References

[1] A line search based proximal stochastic gradient algorithm with dynamical variance reduction F., Porta, Ruggiero, Trombini Submit. J.of Sci.Comp.
[2] Learning rate selection in stochastic gradient methods based on line search strategies F., Porta, Ruggiero, Trombini, Zanni, Submit. Appl. Math. in Sci. and Eng.

Nicola Rares Franco, *A theoretical analysis of Deep Learning-based Reduced Order Models*
MOX, Dipartimento di Matematica, Politecnico di Milano

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 networks 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 new upper-bounds for the approximation error of convolutional neural networks, highlighting how the mathematical properties of the original problem affect the design of DL-ROMs.

Stefania Fresca, *Deep learning-based reduced order models for the real-time solution of parametrized PDEs*
MOX - Dipartimento di Matematica, Politecnico di Milano

The solution of differential problems by means of full order models (FOMs), such as, e.g., the finite element method, entails prohibitive computational costs when it comes to real-time simulations and multi-query routines. The purpose of reduced order modeling is to replace FOMs with reduced order models (ROMs) characterized by much lower complexity but still able to express the physical features of the system under investigation. Conventional ROMs anchored to the assumption of modal linear superimposition, such as proper orthogonal decomposition (POD), may reveal inefficient when dealing with nonlinear time-dependent parametrized PDEs, especially for problems featuring coherent structures propagating over time. To overcome these difficulties, we propose an alternative approach based on deep learning (DL) algorithms, where tools such as convolutional neural networks are used to build an efficient nonlinear surrogate. In the resulting DL-ROM, both the nonlinear trial manifold and the nonlinear reduced dynamics are learned in a non-intrusive way by relying on DL models trained on a set of FOM snapshots, obtained for different parameter values. Accuracy and efficiency of the DL-ROM technique are assessed in several applications, ranging from cardiac electrophysiology to fluid dynamics, showing that new queries to the DL-ROM can be computed in real-time.

References

[1] S. Fresca, A. Manzoni, L. Dede' 2021 A comprehensive deep learning-based approach to reduced order modeling of nonlinear time-dependent parametrized PDEs. Journal of Scientific Computing, 87(2):1- 36.

[2] S. Fresca, A. Manzoni 2022 POD-DL-ROM: enhancing deep learning-based reduced order models for nonlinear parametrized PDEs by proper orthogonal decomposition. Computer Methods in Applied Mechanics and Engineering, 388, 114181.

Caterina Graziani, *Weisfeiler–Lehman goes Dynamic: An Analysis of the Expressive Power of Graph Neural Networks for Attributed and Dynamic Graphs*
Università degli Studi di Siena

Graph Neural Networks (GNNs) are a large class of connectionist models for graph processing. Recent studies on the expressive power of GNNs have focused their attention on two issues. First, it has been proved that GNNs are as powerful as the Weisfeiler-Lehman test (1-WL) in their ability to distinguish graphs. Moreover, it has been shown that the equivalence enforced by 1-WL equals unfolding equivalence. On the other hand, it has been proved that GNNs are universal approximators on graphs modulo the 1-WL/unfolding equivalence. However, those results only apply to Static Undirected Homogeneous Graphs (SUHG) with attributes on nodes. In contrast, real-life application domains often involve much more varied types of graphs, such as, for example, dynamic, heterogeneous, directed graphs, and multigraphs with attributes on nodes and edges. In this paper, we conduct a theoretical analysis of the expressive power of GNNs for two types of graphs that are particularly interesting, namely dynamic graphs and static graphs with attributes also on edges. The former type is widely used in modern applications, and its theoretical analysis requires a new methodology. The latter can act as a standard form for all the graph types since it has been proved that all the graph types can be transformed to SAUHG with attributes on nodes and edges without loss of information. The study considers GNN models for those domains; consequently, appropriate 1-WL tests are proposed. Then, we extend the results on the expressive power of GNNs to the novel graph domains, proving that GNNs have the same capability as the 1-WL test in distinguishing graphs, that the 1-WL equivalence equals unfolding equivalence and that GNNs are universal approximators modulo 1-WL/unfolding equivalence. Moreover, the proof of approximation capability holds for a very general class of graphs and is constructive, allowing to deduce hints on the architecture of GNNs that can achieve the desired accuracy.
Authors: Beddar-Wiesing, D'Inverno, Graziani, Lachi, Moallemy-Oureh, Scarselli, Thomas

Francesco Iafrate, *Bridge-type estimation with mixed-rate asymptotics: applications to Stochastic Differential Equations*
Sapienza Università di Roma

The aim of our work ([1]) is to introduce an adaptive penalized estimator for identifying the true reduced parametric model under the sparsity assumption. In particular, we deal with the framework where the unpenalized estimator of the structural parameters needs simultaneously multiple rates of convergence (i.e., the so-called mixed-rates asymptotic behavior). We introduce a Bridge-type estimator by taking into account penalty functions involving ℓ^q norms ($0 < q \leq 1$). We prove that the proposed regularized estimator satisies the oracle properties. Our approach is useful for the estimation of stochastic diferential equations in the parametric sparse setting. More precisely, under the high-frequency observation scheme, we apply our methodology to an ergodic diffusion process. In real-world applications the estimation task is challenging, especially for $0 < q < 1$ due to the non-convex nature of the problem. We study ([2]) pathwise algorithms that allow to efficiently compute the full solution path for penalized estimation methods as a function of the penalization parameter. Such algorithms rely on proximity operators based non-convex analysis techniques.

References

[1] A. De Gregorio and F. Iafrate. Regularized bridge-type estimation with multiple penalties. Annals of the Institute of Statistical Mathematics, 2020
[2] A. De Gregorio and F. Iafrate. Pathwise optimization for adaptive Bridge-type estimators. Working paper

Gianluca Manzan, *Efficiency limits of Restricted Boltzmann Machines in teacher-student frameworks*
Università di Bologna

Unsupervised Machine learning with Boltzmann machines is the inverse problem of finding a suitable Gibbs measure to approximate an unknown probability distribution from a training set consisting of a large amount of samples. The minimum size of the training set necessary for a good estimation depends on both the properties of the data and the machine. We investigate this problem in a controlled environment where a Teacher Restricted Boltzmann machine (T-RBM) is used to generate the dataset and another student machine (S-RBM) is trained with it. We consider different classes of units priors and weight regularizers and we analyze both the informed and mismatched scenarios depending on the amount of information the student receives about the teacher model. We describe the problem in terms of phase transitions in the posterior distribution, interpreted as a statistical mechanics system.

References
Phase diagram of restricted Boltzmann machines and generalized Hopfield networks with arbitrary priors. Barra, A., Genovese, G., Sollich, P., & Tantari, D. (2018). Physical Review E, 97(2), 022310.
Phase transitions in restricted Boltzmann machines with generic prios. Barra, A., Genovese, G., Sollich, P., & Tantari, D. (2017). Physical Review E, 96(4), 042156.
Restricted Boltzmann machine: Recent advances and mean-field theory. Decelle, A., & Furtlehner, C. (2021). Chinese Physics B, 30(4), 040202.
On the equivalence of Hopfield networks and Boltzmann Machines. Barra, A., Bernacchia, A., Santucci, E., & Contucci, P. (2012). Neural Networks, 34, 1-9.
Statistical Mechanics of Neural Networks-Huang, H. (2022). Statistical Mechanics of Neural Networks.

Carlo Marcati, *Neural network and operator network approximations for elliptic PDEs*
Università di Pavia

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 (deep) ReLU NN approximation theory. I will first show how NNs accurately approximate functions with isolated singularities, for example the solutions to elliptic problems in polygons and polyhedra, or the eigenfunctions of problems with singular potentials that arise in quantum chemistry [1]. I will then introduce operator networks, which approximate the solution operator of 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) [2]. When the coefficients are smooth, the size of the networks can then be bounded with respect to the H^1 norm of the error, uniformly over the parameter set. The proofs of our approximation rates combine elliptic regularity, classical and recent results in numerical analysis, and tools from NN approximation theory.

References
[1] Carlo Marcati, Joost A. A. Opschoor, Philipp C. Petersen, and Christoph Schwab. Exponential ReLU neural network approximation rates for point and edge singularities. Found. Comp. Math. (2022)
[2] Carlo Marcati and Christoph Schwab. Exponential Convergence of Deep Operator Networks for Elliptic Partial Differential Equations. Submitted, arXiv:2112.08125 (2021)

Antonio Mastropietro, *From eXplainable Artificial Intelligence to Game Theory with Coalitional Hodge-Shapley Value*
EURECOM

In cooperative game theory, a set of players or decision makers should negotiate to decide how to allocate the worth gained by the coalition composed by all the players. A value is a solution concept that suggests the negotiation outcome among players. Among the many alternatives, the Shapley value solution concept is very popular: it has the property of being a fair allocation, where the fairness is described by a set of desirable properties, or axioms. The axioms characterize the Shapley value in the sense that it is the unique value satisfying those properties; in addition, axioms allow deriving a simple explicit combinatorial formula to compute the Shapley value. In our approach, coalitions are the main subjects of cooperation, instead of single players. Inspired by the Shapley value, the goal is to derive a fair allocation to coalitions. The methodology uses the Hodge decomposition of the simplicial complex associated to the poset of the subsets of the set of players ordered by inclusion, like the one proposed by [1]. We will motivate this investigation in terms of eXplainable Artificial Intelligence (XAI), and link our work with the SHAP algorithm [2].

References
[1] Stern, A. and Tettenhorst, A., 2019. Hodge decomposition and the Shapley value of a cooperative game. Games and Economic Behavior, 113, pp.186-198.
[2] Lundberg, S.M. and Lee, S.I., 2017. A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.

Giacomo Meanti, *Hyperparameter Tuning for Nystroem Kernel Ridge Regression*
Università di Genova

Kernel methods are among the most commonly used approaches to nonparametric learning. Although the number of hyperparameters is often kept small, the job of tuning them is left to the user to perform. Furthermore, a higher number of hyperparameters could improve the accuracy of learned models (for example using different kernel lengthscales for each dimension). We review and contrast a number of hyperparameter tuning strategies, and proposed a complexity regularization criterion based on a data dependent penalty, and discuss its efficient optimization.

Caterina Millevoi, *Neural network application to forward and inverse problems involving differential equations*
Università degli Studi di Padova

Neural networks, which have achieved success in several applications, are beginning to play a significant role in advancing scientific research in fields that are traditionally dominated by discretization methods for the numerical solution of PDEs [1]. Such models are particularly promising in processes that are not completely understood, or where it is computationally infeasible to run simulations at desired resolutions, or (noisy) data can be recorded in addition to boundary conditions. For complex scientific applications, a possible strategy is to combine data-driven machine learning models with traditional ones. Therefore, methods such as Physics-Informed Neural Networks (PINNs) have been proposed [2], which in addition to data consider information from the governing equations by including the residual as a constraint in the training. We present different applications of PINNs for the solution of both forward and inverse problems involving PDEs. First, a study of the hyperparameter influence in PINN setup in coupled hydro-poromechanics has been performed to identify the most appropriate architecture, and a “sensor condition” strategy is proposed to solve forward problems with the aid of data located inside the domain to simulate the availability of sensor recordings [3]. The goal is to validate the approach, thus laying the foundation of the method in real-world

complex hydro-poromechanical applications. Second, a PINN model has been analyzed on a synthetic SIR compartmental model of the COVID-19 pandemic. Its ability to approximate Susceptible, Infectious and Recovered classes of population, while simultaneously solving an ill-posed inverse problem by estimating the reproduction number based on infectious data only, has been explored. Establishing the technique in epidemiology is the aim, towards the application to more sophisticated models able to provide fast and reliable answers for helping decision-making processes at the rise of pandemic.

References

[1] Willard, J., Jia, X., Xu, S., Steinbach, M., and Kumar, V. Integrating physics-based modeling with machine learning: A survey. arXiv preprint arXiv:2003.04919, 1(1), 1-34 (2020).
[2] Raissi, M., Perdikaris, P., and Karniadakis, G. E., Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 378:686–707, 2019.
[3] Millevoi, C., Spiezia, N., and Ferronato, M. On Physics-Informed Neural Networks architecture for coupled hydro-poromechanics problems. Available at SSRN: https://ssrn.com/abstract=4074416

Davide Oberto, *An invariances-preserving vector basis neural network for the closure of Reynolds-averaged Navier–Stokes equations*
Dipartimento di Scienze Matematiche, Politecnico di Torino

Most of the relevant flows in the engineering world are turbulent and so performing reliable simulations is very challenging due to the multiscale nature of turbulence. Usually, Navier-Stokes (NS) equations are not directly tackled because this would require very fine computational domains in both space and time. An approach to overcome this issue is to deal with the so-called Reynolds-averaged Navier-Stokes (RANS) equations. The RANS equations are formally equal to the NS equations with an additional term associated to the divergence of the Reynolds stress tensor that accounts the effects of turbulence on the averaged velocity and pressure fields. Classical approaches to approximate the Reynolds stress tensor are to solve additional Partial Differential Equations associated to turbulent kinetic energy and to its dissipation rate. Recently, supervised Machine Learning techniques trained with high fidelity simulations have been used to this scope. The aim of this work is to define a neural network that predicts the divergence of the Reynolds stress tensor and, at the same time, preserves physical properties of the system such as Galilean and frame-reference invariances. To do that, the divergence of the Reynolds stress tensor is expanded as sum of vector quantities and the proposed Vector Basis Neural Network (VBNN) predicts the coefficients of such an expansion. Dealing with the divergence of the Reynolds stress tensor instead of the tensor itself makes the training easier and more accurate because this quantity is less affected by statistical errors in high fidelity simulations. Numerical investigations with classic turbulent flows in simple geometries are performed to show the advantages of this approach.

References

S. Berrone and D. Oberto, *An invariances-preserving vector basis neural network for the closure of Reynolds-averaged Navier–Stokes equations by the divergence of the Reynolds stress tensor*, Physics of Fluids 34, 095136 (2022)

Agnese Pacifico, *Optimal control and Reinforcement Learning - An Algorithm for the LQR Problem with Partially Unknown Dynamics*
Sapienza Università di Roma

Reinforcement Learning is a Machine Learning paradigm. Many works have highlighted that it has strong connections with Optimal Control, since the problem that Reinforcement Learning seeks to solve can be seen as an Optimal Control problem where the dynamics is an unknown function. In [4] a framework is proposed to study control problems in uncertain environments. Uncertainty in the dynamics is modelled as a probability

distribution over functions. Recent works ([2],[3]) have proved some convergence results for problems defined in this framework. In [1] the particular case of an LQR optimal control problem with partially unknown dynamics is considered. A model-based online algorithm is proposed to obtain an approximation of the dynamics and of the control at the same time during a single simulation. The set of possible dynamics is a set of matrices in this linear case, and the uncertain knowledge is represented as a probability distribution on it. At each step, the control is chosen according to the present knowledge of the dynamics, solving a Riccati equation. Then the probability distribution is updated according to the prior knowledge and the observed behaviour of the system by means of Bayesian formulas. A new description of the dynamics is obtained and then used in the following step.

References

[1] A. Pacifico, A. Pesare, M. Falcone (2022). A New Algorithm for the LQR Problem with Partially Unknown Dynamics. Lecture Notes in Computer Science, vol 13127, Springer.
[2] A. Pesare, M. Palladino, M. Falcone (2021). Convergence results for an averaged LQR problem with applications to reinforcement learning. Mathematics of Control, Signals and Systems 33, 379–411.
[3] A. Pesare, M. Palladino, and M. Falcone (2022). Convergence of the value function in optimal control problems with unknown dynamics, in 2021 European Control Conference (ECC), IEEE.
[4] R. Murray, M. Palladino (2018). A model for system uncertainty in reinforcement learning. Systems & Control Letters, 122, 24.

Stefano Pagani, *Physics-based deep learning models in cardiac electrophysiology and cardiovascular flows*
MOX - Dipartimento di Matematica, Politecnico di Milano

Precision medicine aims to develop new personalized approaches to diagnosis and treatment based on patients' individual characteristics. In this context, models based on learning techniques are particularly suitable as they allow efficient feedback following the patient's evolution. However, these models' accuracy and ability to generalize hinge on the availability of large volumes of high-resolution data for their training, which are not always available. In this talk, we present some hybrid learning approaches that exploit physical knowledge of the phenomenon, encoded through differential models, to counterbalance the presence of limited or noisy data. In particular, we show how to combine multi-fidelity and dimensionality reduction techniques with neural networks to improve their performance both in terms of efficiency and accuracy. Differential models can be integrated directly into the deep learning framework by adding the residual of the partial or ordinary differential equations to the loss function (physics-informed neural network) or by incorporating projection-based tensorial solvers into the architecture (physics-aware neural network). Our numerical results show how these physics-based learning approaches improve efficiency and accuracy in solving inverse problems related to cardiac electrophysiology, even with incomplete or noisy data. Finally, we show how to construct surrogate models based on neural networks that encode geometric variability across patients, allowing reconstruction of distributed quantities such as the velocity and pressure fields in coronary bifurcations. This work is in collaboration with F. Regazzoni (Politecnico di Milano), R. Tenderini (EPFL), S. Deparis (EPFL), and A. Quarteroni (Politecnico di Milano-EPFL).

Davide Palitta, *A simple yet effective tensor-based ODE model for Deep Learning*
Alma Mater Studiorum, Università di Bologna

In the past few years, applied mathematicians started looking at the forward propagation step of deep learning techniques in terms of discretization methods, e.g., forward Euler, applied to an unkown, underlying differential operator. Each layer of the network is seen as a time step of the discretization method. This point of view paved the way to the so-called neural ordinary differential equations (ODE). In the latter framework, the deep learning process is modeled by an ODE: Inputs are translated into initial values whereas outputs are viewed as the ODE solution evaluated at the final time step. Information propagates along the ODE flow in place of the net so that the extremely problem-dependent design of the latter is no longer needed. The

training phase is now employed to learn the parameters definining the neural ODE. In this talk we present a novel tensor-based neural ODE, namely an ODE defined by tensors, to model a deep learning process. Preliminary results on classification problems show the potential of such new tool. This is a joint work with Jemima Tabeart (Oxford) and Nick Vannieuwenhoven (Leuven).

Danilo Pezzi, *Unrolling With No Deep: Explainable Bilevel Optimization For The Helsinki Deblur Challenge*
Università di Parma

In this talk a bilevel optimization scheme for solving a general image deblurring problem is presented. With the use of machine learning tools, a variational approach with automatically learned parameters is able to achieve a high quality reconstructed image, while maintaining the theoretical background and interpretability of variational models. The features of the bilevel scheme are tailored for the Helsinki Deblur Challenge 2021, which tasked the participants with restoring out-of-focus text images. In the lower level problem, a fixed number of FISTA iterations is applied to an edge preserving energy functional. On the other hand, the parameters of the model are learned either through a similarity index or a support vector machine strategy.

Moreno Pintore, *Enforcing Dirichlet Boundary Conditions in Physics-Informed Neural Networks and Variational Physics-Informed Neural Networks*
Politecnico di Torino

Physics Informed Neural Networks (PINNs) are neural networks designed to solve partial differential equations (PDEs). Thanks to the nonlinear and flexible structure of the neural network, PINNs can be used to solve nonlinear, high-dimensional, parametric and inverse problems. Moreover, several architectures and loss functions have been devised to improve the accuracy and the efficiency of the original model. In this talk we consider two types of PINNs: the original model proposed in [1] and the interpolated Variational Physics Informed Neural Network (VPINN) presented in [2]. The latter is a PINN trained using the variational formulation of the PDE and stabilized by means of an additional interpolation operator. In particular, we discuss and compare four approaches to enforce Dirichlet Boundary conditions on PINNs and VPINNs. Such conditions are usually imposed using additional penalization terms in the loss function. Instead, we propose two alternative ways to exactly enforce them by changing the neural network output and a modified loss function to variationally impose them. The models are tested on several problems to analyze the effect of the considered approaches. We observe that the most efficient way to obtain an accurate solution is to modify the neural network output using an approximate distance function.

References
[1] M. Raissi, P. Perdikaris, G.E. Karniadakis Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics 378 (2019) 686-707.
[2] S. Berrone, C. Canuto, M.Pintore. Variational physics informed neural networks: the role of quadratures and test functions. Journal of Scientific Computing 92 (2022) 1-27.

Luca Ratti, *Learning the optimal regularizer for linear inverse problems*
Università degli Studi di Genova

The task of an inverse problem is to determine an unknown function from measurements obtained through a forward operator, possibly corrupted by noise. Such problems are usually unstable: small perturbations in the observed measurements may cause large deviations in the reconstructed solutions. Variational regularization provides a stable approximation of the inverse map by means of continuous operators. From a statistical learning perspective, one might take advantage of partial knowledge of the joint distribution of unknowns and

measurements (e.g. by means of a training set) to design data-driven regularization operators. In this talk, I will consider a linear inverse problem defined on infinite-dimensional Hilbert spaces, and try to learn the optimal operator in the family of generalized Tikhonov regularizers. After setting a statistical framework for the proposed learning problem, allowing to consider a sufficiently large class of noise models, I will characterize the optimal regularizer, showing that it is completely independent of the forward operator. Then, I will propose a supervised-learning strategy and an unsupervised-learning one to learn such a regularizer from a finite training set: in both cases, I will prove theoretical bounds on the excess risk. The results are also validated by means of a set of numerical examples, both related to a denoising problem and to an ill-posed problem. Finally, I will discuss the extension of this framework from the Tikhonov case to different families of regularization functional, with a particular interest for sparsity-promoting regularization.
This is based on joint work with G. S. Alberti, E. De Vito, M. Santacesaria (University of Genoa), and M. Lassas (University of Helsinki)

Simone Rebegoldi, *On a stochastic first-order trust-region method for machine learning applications*
Università degli studi di Firenze

Several problems in machine learning and data fitting are equivalent to minimizing the sum of a very large number of continuously differentiable functions. In recent years, the stochastic gradient method and their more contemporary variants (SVRG,SAGA,ADAM and others) have become the state-of-the-art methods for solving such problems, thanks to their simplicity and low cost per-iteration. They do not call for function evaluations, however they require the tuning of the learning rate and other hyper-parameters such as the mini-batch size. More recently, a series of works have considered the adoption of stochastic linesearches or trust-region strategies to adaptively choose the parameters and avoid tuning efforts. Although these adaptive stochastic methods are promising, their application is still limited in practice, as they require that both function and gradient estimates are sufficiently accurate with sufficiently high probability, leading to an increasing computational cost as the iterations proceed. In this talk, we propose a stochastic first-order trust-region method for solving finite-sum minimization problems. The proposed method approximates the function and the gradient by sampling, choosing adaptively the sample sizes by means of a deterministic rule inspired by the inexact restoration method. The trust-region step is computed by minimizing a random linear model, and it is accepted or rejected according to a sufficient decrease condition on a suitable merit function. We provide the complexity analysis in expectation of our proposed method, by imposing some probabilistic accuracy requirements on random function and gradients that are less stringent than the corresponding ones in the literature. We validate the proposed method on nonconvex classification and regression problems, showing that it performs well in terms of cost and accuracy and allows to reduce the burdensome tuning of the hyper-parameters involved.

Francesco Regazzoni, *Automatic discovery of low-dimensional dynamics underpinning time-dependent PDEs by means of Neural Networks*
MOX - Dipartimento di Matematica, Politecnico di Milano

In this talk, we present a novel Machine Learning technique able to learn differential equations that surrogate the solution of time-dependent problems. Our method exploits a finite number of latent variables, which provide a compact representation of the state of the system. We propose an algorithm, based on Artificial Neural Networks (ANNs) that learns in a simultaneous way the low dimensional encoding and the reduced dynamics. Remarkably, our method allows building, in a fully non-intrusive manner, surrogate models for evolutionary equations, accounting for the dependence on parameters and time-dependent inputs. Numerical tests for both parabolic and hyperbolic PDEs show that very accurate numerical results are obtained, even for a reduced size of the ANNs, in a computationally efficient manner. Finally, we show applications to global sensitivity analysis and Bayesian parameter estimation of multiphysics models, contexts in which our technique becomes an enabling technology for use cases whose computational cost would otherwise be prohibitive.

References
[1] F. Regazzoni, L. Dede', A. Quarteroni "Machine learning for fast and reliable solution of time-dependent differential equations", Journal of Computational Physics (2019) 397, 108852.
[2] F. Regazzoni, D. Chapelle, P. Moireau "Combining data assimilation and machine learning to build data-driven models for unknown long-time dynamics—Applications in cardiovascular modeling", International Journal for Numerical Methods in Biomedical Engineering (2021) 37(7), e3471.
[3] F. Regazzoni, L. Dede', A. Quarteroni "Machine learning of active force generation models for the efficient multiscale simulation of the cardiac function", Computer Methods in Applied Mechanics and Engineering (2020) 370, 113268.
[4] F. Regazzoni, M. Salvador, L. Dede', A. Quarteroni "A Machine Learning method for real-time numerical simulations of cardiac electromechanics", Computer Methods in Applied Mechanics and Engineering (2022) 393, 114825.

Konstantin Riedl, *Consensus-Based Optimization: A Holistic Convergence Analysis and a Machine Learning Application*
Technical University of Munich, School of Computation, Information and Technology

Consensus-based optimization (CBO) is a multi-particle derivative-free optimization method capable of globally minimizing high-dimensional nonconvex and nonsmooth functions, a task which is ubiquitous in machine learning. In this talk we shed a light on the internal mechanisms of CBO, which are responsible for its success. Based on an experimentally supported intuition that, in the mean-field limit (as the number of particles goes to infinity), CBO always performs a gradient descent of the squared Euclidean distance to the global minimizer, we develop a novel technique for proving global convergence in mean-field law for a rich class of objective functions. In particular, we prove that CBO performs a convexification of a very large class of optimization problems as the number of optimizing particles tends to infinity. From this result it becomes apparent that the hardness of any global optimization problem is necessarily encoded in the mean-field approximation, or, more precisely, in the way how the empirical measure of the finite particle dynamics is used to approximate the mean-field limit. In consideration of the central significance of such approximation with regards to the overall computational complexity of the implemented numerical scheme, we establish a novel probabilistic quantitative result about the convergence of the interacting particle system towards the corresponding mean-field dynamics. By combining both results we provide the first holistic convergence proof of CBO methods on the plane. To demonstrate its practicability we provide experiments for well-understood machine learning problems.

References
[1] M. Fornasier, T. Klock, and K. Riedl. Consensus-Based Optimization Methods Converge Globally. arXiv:2103.15130 (2022).
[2] M. Fornasier, T. Klock, and K. Riedl. Convergence of Anisotropic Consensus-Based Optimization in Mean-Field Law. International Conference on the Applications of Evolutionary Computation (Part of EvoStar). Springer, Cham, 2022.

Alessandro Scagliotti, *Ensemble optimal control: ResNets, diffeomorphisms approximation and Normalizing Flows*
Technical University of Munich. Munich Center for Machine Learning

In the last years it was observed that Residual Neural Networks (ResNets) can be interpreted as discretizations of control systems, bridging ResNets (and, more generally, Deep Learning) with Control Theory. In the first part of this seminar we formulate the task of a data-driven reconstruction of a diffeomorphism as an ensemble optimal control problem. In the second part we adapt this machinery to address the problem of Normalizing Flows: after observing some samplings of an unknown probability measure, we want to (approximately) construct a transport map that brings a "simple" distribution (e.g., a Gaussian) onto the unknown target

distribution. In both problems we use tools from Γ-convergence to study the limiting case when the size of the data-set tends to infinity.

References
[1] A. Scagliotti. Deep Learning approximation of diffeomorphisms via linear-control systems. Math. Control Rel. Fields (2022).
[2] S. Farinelli, A. Scagliotti. An optimal control approach to Normalizing Flows. In preparation.

Maria Strazzullo, *Physics-informed Neural Networks for partial differential equations and optimal control in a parametric setting*
DISMA - Politecnico di Torino

This talk focuses on physics informed supervised learning approach to partial differential equations in a parametric setting. We use the physics informed paradigm to provide different solutions to these problems in a small amount of time since standard simulations can be unbearable in a real-time and many-query context. The physics information is exploited not only in the loss function but also as an augmented input (extra feature) and as a guiding tool to build an effective architecture for the network (physics informed architecture). The combination of these aspects will lead to reliable prediction with faster training. We propose the application of this problem also in an optimal control framework.

S. Ivan Trapasso, *Some results on the stability of the scattering networks*
Politecnico di Torino

Within the mathematical analysis of deep convolutional neural networks, the wavelet scattering transform introduced by S. Mallat [CPAM, 2012] is a unique example of how the ideas of multiscale analysis can be combined with a cascade of modulus nonlinearities to build a nonexpansive, translation invariant signal representation with provable geometric stability properties, namely Lipschitz continuity to the action of small C^2 diffeomorphisms – a remarkable result for both theoretical and practical purposes, inherently depending on the choice of the filters and their arrangement into a hierarchical architecture. In the recent notes [arXiv:2104.11977, arXiv:2205.11142], we further investigated some aspects concerning the intimate relationship between the scattering structure and the regularity of the deformation. In particular, considering the Hölder regularity scale C^α, $\alpha > 0$, as a reference, we were able to precisely identify the stability threshold, proving that stability is still achievable for deformations of class C^α, $\alpha > 1$, whereas instability phenomena can occur at lower regularity levels modelled by C^α, $0 \leq \alpha < 1$. While the behaviour at the threshold given by Lipschitz (or even C^1) regularity remains beyond reach, we are able to prove a stability bound in that case, up to ε losses. We also study the stability problem for irregular deformation fields in $L^\infty(\mathbb{R}^d;\mathbb{R}^d)$, and we prove that for signals in multiresolution approximation spaces U_s at scale s, whenever the network is Lipschitz continuous (regardless of its architecture), stability in L^2 holds in the regime $\|\tau\|_{L^\infty}/s \ll 1$, essentially as a consequence of the uncertainty principle. On the other hand, if $\|\tau\|_{L^\infty}/s \gg 1$ instability can occur even for well-structured DCNNs such as wavelet scattering networks, and we provide a sharp upper bound for the asymptotic growth rate.

Eleonora Vercesi, *Computing Disease-Specific Gene Embeddings via Constrained Optimization*
Università degli Studi di Pavia

This paper presents an optimization-based approach to computing disease-specific gene embeddings. In particular, we focused on Acute Myeloid Leukemia (AML), a blood cancer presenting several subgroups related to the severity of the disease, characterized by major differences at the genetic level. Instead of using Deep Learning techniques, we formulate a non-convex continuous optimization problem whose solution

embeds genes in a Euclidean space of dimension 14. This problem is solved using the Minimum Distortion Embedding solver [1] that implements a projected-gradient quasi-Newton algorithm. This framework allows us to express the relation between genes in different ways, and in particular, we exploit both the information contained in gene expression profiles and gene ontologies. We investigate the quality of the final embedding, which depends on the input data and the choice of the penalty functions. To validate our gene embeddings, we perform standard sanity checks, and we use two independent datasets (i.e., CancerSEA [2] and KEGG [1]) containing functional correlations among genes. Preliminary computational results show that our approach computes a gene embedding that correctly encodes functional similarities between pairs of genes for AML and is competitive with the existing gene2vec [4] embedding. This is a joint work with S. Gualandi (Dept. of Mathematics, UNIPV), and G. Guidetti (Dept. of Biology & Biochemistry, UNIPV).

References

[1] AGRAWAL, Akshay, et al. Minimum-distortion embedding. Foundations and Trends® in Machine Learning, 2021, 14.3: 211-378.
[2] YUAN, Huating, et al. CancerSEA: a cancer single-cell state atlas. Nucleic acids research, 2019, 47.D1: D900-D908.
[3] KANEHISA, Minoru; GOTO, Susumu. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic acids research, 2000, 28.1: 27-30.
[4] ZOU, Quan, et al. Gene2vec: gene subsequence embedding for prediction of mammalian N6-methyladenosine sites from mRNA. Rna, 2019, 25.2: 205-218.

Stefano Vigogna, *Exponential convergence for support vector machines (without hard margin)*
Università degli Studi di Roma Tor Vergata

Optimizing the misclassification risk is in general NP-hard. Tractable solvers can be obtained by considering a surrogate regression problem. While convergence to the regression function is typically sublinear, the corresponding classification error can decay much faster. Fast and super fast rates (up to exponential) have been established for general smooth losses on problems where a hard margin is present between the classes. This leaves out models based on non-smooth losses such as support vector machines, and problems where there is no hard margin, begging several questions. Are such models incapable of fast convergence? Are they therefore structurally inferior? Is the hard margin condition really necessary to obtain exponential convergence? Developing a new strategy, we provide an answer to these questions. In particular, we show not only that support vector machines can indeed converge exponentially fast, but also that they can do so even without hard margin.

Joshua Wrigley, *Dynamical objects: Categoric and Topos-theoretic techniques for Deep Neural Networks*
Università degli Studi dell'Insubria

Recent work by Belfiore and Bennequin in [1] has inspired speculation into the further applications of categorical methods, especially topos-theoretic methods, in the field of machine learning. This presentation is intended as a mere taste of some of the potential possibilities of the approach highlighted in [1]. We will review the construction of the dynamical objects of a deep neural network (DNN) as presheaves over the layers of the DNN viewed as a category. Such a presheaf consists of assigning a set to each layer and, importantly, a transition map for each connection between the layers. In this way, we are able to define a presheaf of possible activations, possible weights, and possible states of the DNN. Pursuing a categorical formulation of a DNN allows the efficient handling of this data, as evidenced when we observe that backpropagation can be understood as a series of natural transformations of the presheaf of possible weights. We will also observe that extra structure can easily be defined on the dynamical objects by requiring that the presheaves factor through certain subcategories, e.g. smooth manifolds. Finally, we present an application of this formulation to the problem of limiting the search space of possible weights for a DNN. As seen in the practical experiments presented in [2], the logical of the behaviour of the DNN, according to some propositions one can make about

the system, play an important role. We will observe that the search space of the possible weights can be restricted to only those satisfying some chosen logical formulae via the ubiquitous categorical construction of a pullback.

References

[1] Jean-Claude Belfiore, Daniel Bennequin, Topos and Stacks of Deep Neural Networks, arXiv:2106.14587 (2021).
[2] Jean-Claude Belfiore, Daniel Bennequin, Xavier Giraud, Logical Information Cells I, arXiv:2108.04751 (2021).

Filippo Zimmaro, *Voter model with external biases on a modular network*
University of Bologna/ University of Pisa

To maximize the time spent on the platform, AI-based social network algorithms often hide dissonant opinions and conversely promote the interactions with agents close to the agent's current, or usual, belief. This reinforces a mechanism of homophily that motivates us to study, by techniques of statistical physics and dynamical systems, an extension of the voter model on a modular graph with community-based external biases. The model on the fully connected network was introduced as a generalization of the Abrams-Strogatz model for language death by Masuda and Redner (called Partisanship Voter Model), and was recently adapted to the case of one biased population (Czaplicka et al., 2022). We focus in particular on the transition between consensus and polarized states, as a function of the composition of the system, of the connectivities of the block model and of the strengths of the external biases, which reflect the historical prejudices of the communities. The results suggest that polarization may appear also for extremely low values of the biases, if the populations are sufficiently close-minded. The study is an upcoming work.