Book of Abstracts

3rd Workshop of UMI Group

Mathematics for Artificial Intelligence and Machine Learning

Math4AIML
Bari, Italy
29-31 January 2025

Department of Mathematics – University of Bari "Aldo Moro"
https://umi-math4aiml2025.uniba.it/
midasgroup.math@gmail.com

Contents

About MATH4AIML-2025

The MiδAs research group of the Department of Mathematics of the University of Bari Aldo Moro, in collaboration with the Italian Mathematical Union (UMI) research group on "Mathematics for Artificial Intelligence and Machine Learning", organized at the Department of Mathematics, the third edition of the Workshop *Mathematics for Artificial Intelligence and Machine Learning – MATH4AIML2025*.

The main aim of the Mathematics for Artificial Intelligence and Machine Learning meetings is to provide a platform for early-career researchers working on topics within the broad research interests to present their work and network with peers and future collaborators.

The first two editions of the workshop, held respectively at the Polytechnic University of Turin (in November 2022) and at the Bocconi University of Milan (in January 2024), were attended by a large audience of researchers, also from outside the UMI group, and proved to be a platform for sharing the work of many young researchers on emerging issues and mathematical aspects of Artificial Intelligence, Machine Learning, and Optimization.

We have the honor of hosting the third edition of this workshop at UniBA, in the year of its first centenary, as a corollary of scientific and cultural events organized in 2024. The three-day event will be attended by more than 160 people, including young PhD students, researchers, and senior members of the scientific community, and will feature three plenary lectures, eight keynote speakers, two parallel sessions of contributed presentations and posters by young PhD students and researchers.

In addition, this third edition of MATH4AIML will see the organization of a round table bringing together the academic, corporate, and research worlds to discuss and explore the interactions between MATH, ML and AI in addressing challenges and innovative applications for industry and science.

Finally, there will also be opportunities to exchange ideas and opinions: all participants are invited to take advantage of the social opportunities offered by the coffee breaks, lunch breaks, and poster sessions.

We would like to thank the supporters of this edition, whose help was essential for the organization of the workshop. In particular, we thank the industrial partners Pirelli S.p.A and Planetek Italia S.r.l, for supporting this workshop, University of Bari Aldo Moro, ERC Seeds Uniba project "Biomes Data Integration with Low-Rank Models" (CUP H93C23000720001) and Piano Nazionale di Ripresa e Resilienza (PNRR), Missione 4 "Istruzione e Ricerca"-Componente C2 Investimento 1.1, "Fondo per il Programma Nazionale di Ricerca e Progetti di Rilevante Interesse Nazionale", Progetto PRIN-2022 PNRR, P2022BLN38, Computational approaches for the integration of multi-omics data. CUP: H53D23008870001.

The organizing committee

Nicoletta Del Buono
Flavia Esposito

Supported by

Plenary Speakers

From single omics dataset to multi-omics and multi-datasets integration through a statistical learning perspective and beyond

Claudia Angelini
Istituto per le Applicazioni del Calcolo "M. Picone", Napoli, Italy

claudia.angelini@cnr.it

The widespread availability of high-throughput instruments for collecting omics data has opened new avenues in personalized medicines, disease etiology understanding, and bio-marker discoveries. However, analyzing this data presents several challenges, including high dimensionality, distribution heterogeneity, and elevated noise levels. Various statistical and machine-learning methods have been proposed to address these issues in contexts such as classification, clustering, survival analysis, and network inference. Recently, data collection efforts have evolved from focusing on a single omics dataset (e.g., gene expression) to gathering multiple datasets from different individuals on specific omics or datasets encompassing multiple omics from the same individuals (e.g., gene expression, methylation, and gene structural variants). The availability of such a large amount of data, including single-cell resolution data, can enhance the accuracy of predictions when combined with appropriate computational approaches, to cite only a few examples.

This work first provides an overview of our recent methods for analyzing single omics, such as gene expression, within the context of survival analysis [1, 2]. Subsequently, we discuss how such statistical methods can be generalized to accommodate scenarios with multiple datasets or multiple omics. Therefore, we will present our recent extension of the cooperative learning approach [3] to survival analysis and our latest methods for network inference, such as [4, 5]. Finally, we provide insights into how artificial intelligence methods can further move steps ahead in extracting valuable knowledge and improving performance.

Acknowledgments This work is part of an extended collaboration with several colleagues and is partially supported by the PRIN 2022 PNRR P2022BLN38 project, “Computational approaches for the integration of multi-omics data” funded by European Union - Next Generation EU, CUP B53D23027810001.

References

[1] C. Angelini, D. De Canditiis, I. De Feis, A. Iuliano *A Network-Constrain Weibull AFT Model for Biomarkers Discovery*, Biometrical Journal 2024, 66 (7), e202300272

[2] A. Iuliano, A. Occhipinti, C. Angelini, I. De Feis, P. Liò *COSMONET: An R Package for Survival Analysis Using Screening-Network Methods* , Mathematics 2021, 9, 3262

[3] D.Y. Ding, S. Li, B. Narasimhan, R. Tibshirani *Cooperative learning for multiview analysis*, Proceedings of the National Academy of Sciences 119 (38), e2202113119

[4] C. Angelini, D. De Canditiis, A. Plaksienko *Jewel 2.0: An Improved Joint Estimation Method for Multiple Gaussian Graphical Models* , Mathematics 2022, 10, 21, 3983

[5] V. Policastro, M. Magnani, C. Angelini, A. Carissimo *INet for network integration*, Computational Statistics, 2024, 1-23

Short Bio Dr. Claudia Angelini graduated in Mathematics in 1994 at the University of Naples "Federico II"; where she also obtained her Ph.D. in Applied Mathematics and Computer Science in 2002. Since 2001, she has worked as a permanent Researcher at the Institute for Applied Calculus (IAC-CNR). She became a Senior researcher in 2019, and since January 2020, she has held the position of Director of Research. Moreover, since July 2024, she has been acting as head of the Naples branch of the Institute for Applied Calculus. Her main research activity is devoted to developing new statistical and machine learning methods to analyze complex

data, focusing on the analysis and integration of omics data. She has been the scientific coordinator of the IAC-CNR research unit in several scientific and industrial projects at national and international levels. She has co-authored more than 100 full articles in ISI peer-reviewed international journals and numerous other international publications in conference proceedings and book chapters. Over the years, she has supervised the research activities of several Ph.D. students, Master students, and research fellows. She also gave courses and seminars in Statistics and Computational Biology at several universities for Master's and Ph.D. students and was a member of the evaluation committee for several projects, including European projects.

Current and Future Trends in Recommender Systems

Tommaso Di Noia
Politecnico di Bari, Italy.

tommaso.dinoia@poliba.it

Recommender systems have become an integral component of modern digital ecosystems, shaping user experiences across various domains such as e-commerce, social media, and streaming services. This talk will explore the current landscape of recommender systems and address emerging trends and future directions in the field. We will give an overview of current trends in recommender systems research and discuss potential evolutions of the recommendation problem.

Short Bio Tommaso Di Noia is a Professor of Computer Science at Politecnico di Bari (Italy). His research activities, mainly focused on Artificial Intelligence and Data Management, were initially devoted to theoretical and practical issues in knowledge representation and automated reasoning. In these fields, he proposed innovative solutions to knowledge-aware resource retrieval and matching by exploiting non-monotonic automated reasoning techniques. Then, he moved to study how to apply knowledge representation techniques and tools both to automated negotiations among rational agents with preferences and to mobile and ubiquitous computing scenarios and protocols. Following these ideas, he started to study applications of knowledge graphs and Linked Open Data datasets to user modeling and recommender systems. He has recently been publishing many works covering theoretical, algorithmic, and experimental aspects on the subject of recommender systems. During the last years, he has also focused on security and privacy issues related to recommender systems with a specific emphasis on adversarial and federated machine learning. Tommaso Di Noia has published many papers in international journals, conferences, and book chapters related to his research interests. Some of them have been awarded the Best Paper Award in different conferences. He is a recipient of IBM Ph.D. Fellowship in 2015 and HP Labs Innovation Research Program Award in 2011 and 2012.

Optimization, the philosophical background of artificial intelligence

Yurii Nesterov
UCLouvain, Belgium.

yurii.nesterov@uclouvain.be

We discuss new challenges in the modern Science, created by Artificial Intelligence (AI). Indeed, AI requires a system of new sciences, mainly based on computational models. Its development has already started by the progress in Computational Mathematics. In this new reality, Optimization plays an important role, helping the other fields with finding tractable models and efficient methods, and significantly increasing their predictive power. We support our conclusions by several examples of efficient optimization schemes related to human activity.

Short Bio Yuri Nesterov is a renowned mathematician and one of the leading experts in optimization theory. He is a professor at the Université catholique de Louvain (UCLouvain) in Belgium, where he has made groundbreaking contributions to the field of convex optimization, particularly in the development of fast gradient methods. Prof. Nesterov is best known for introducing Nesterov's Accelerated Gradient (NAG) method, a cornerstone of modern optimization algorithms widely used in machine learning and artificial intelligence. His work spans convex and non-convex optimization, large-scale optimization, and polynomial optimization, with profound impacts on both theoretical and applied aspects of the field. He has authored several influential books, including Introductory Lectures on Convex Optimization: A Basic Course, and has received numerous prestigious awards, such as the John von Neumann Theory Prize, for his significant contributions to optimization and mathematical sciences.

Keynotes Speakers

MGProx: A nonsmooth multigrid proximal gradient method with adaptive restriction for strongly convex optimization

Andersen Ang

School of Electronics and Computer Science University of Southampton, UK

andersen.ang@soton.ac.uk

We study the combination of proximal gradient descent with multigrid for solving a class of possibly nonsmooth strongly convex optimization problems. We propose a multigrid proximal gradient method called MGProx, which accelerates the proximal gradient method by multigrid, based on utilizing hierarchical information of the optimization problem. MGProx applies a newly introduced adaptive restriction operator to simplify the Minkowski sum of subdifferentials of the nondifferentiable objective function across different levels. We provide a theoretical characterization of MGProx. First we show that variables at all levels exhibit a fixed-point property at convergence. Next, we show that the coarse correction is a descent direction for the fine variable in the general nonsmooth case. Lastly, under some mild assumptions we provide the convergence rate for the algorithm, such as the classical sub-linear rate and also the linear rate. By treating the multigrid proximal gradient iteration as a black-box, we also proposed a fast MGProx with Nesterov's acceleration, together with the classical rate. In the numerical experiments, we show that MGProx has a significantly faster convergence speed than proximal gradient descent and proximal gradient descent with Nesterov's acceleration on nonsmooth convex optimization problems such as the Elastic Obstacle Problem, which the restriction operator is well known.

Graph and Hypergraph Learning via Complex- and Quaternion-Valued Spectral Convolutional Operators

Stefano Coniglio

Department of Economics, University of Bergamo stefano.coniglio@unibg.it

In many learning problems, graphs and hypergraphs are powerful abstractions that can used to model various types of interactions among the elements of a given dataset. Over the past years, these structures have been attracting a growing interest in the deep-learning literature thanks to many successful applications in several fields, including key ones in chemistry and biology. Hypergraphs, in particular, are crucial for their capability of representing real-world phenomena involving polyadic (many-to-many) relations between the elements, generalizing the simpler diadic (pairwise) relationships that are classically captured by a graph. While the possibility of capturing asymmetric relationships (either diadic or polyadic) within a dataset is crucial in many applications, (hyper)edge directions are often ignored in many state-of-the-art works that rely on a convolutional operator of spectral type, i.e., one grounded in graph-signal theory.

In this presentation, we survey recent results in directed (hyper)graph learning based on the construction of complex- or quaternion-valued graph Laplacian matrices which are suitably designed to capture the (hyper)edge directions while being amenable for the construction of spectral convolutional operator. In particular, we present the Sign-Magnetic Laplacian and SigMaNet, a generalized Graph Convolutional Network (GCN) capable of handling both undirected and directed graphs with weights not restricted in sign nor magnitude; a quaternion-valued extension of the Sign-Magnetic Laplacian which is suitable for graphs involving digons (antiparallel edges) of asymmetric weights and its associated GCN QuaterGCN; the Generalized Directed Laplacian and GeDi-HNN, a Hypergraph Neural Network (HNN) suitable for hypergraph-learning tasks involving hyperedge directions; and the Directed Line Graph Laplacian and its associated HNN DLGNet, which are designed to tackle chemical-reaction classification problems by a suitably-designed transformation of the input directed hypergraph to a directed line graph with complex-valued edge weights.

References

[1] Stefano Fiorini, Stefano Coniglio, Michele Ciavotta, Enza Messina *SigMaNet: One Laplacian to Rule Them All*, Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37, no. 11, pp. 7568-7576, 2023.

[2] Stefano Fiorini, Stefano Coniglio, Michele Ciavotta, Enza Messina *Graph Learning in 4D: A Quaternion-Valued Laplacian to Enhance Spectral GCNs*, Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, no. 11, pp. 12006-12015, 2024.

[3] Stefano Fiorini, Stefano Coniglio, Michele Ciavotta, Alessio Del Bue *Let There be Direction in Hypergraph Neural Networks*, Transactions on Machine Learning Research, 2024.

[4] Stefano Fiorini, Giulia M. Bovolenta, Stefano Coniglio, Michele Ciavotta, Pietro Morerio, Michele Parrinello, Alessio Del Bue *DLGNet: Hyperedge Classification through Directed Line Graphs for Chemical Reactions*, arXiv preprint arXiv:2410.06969, October 2024.

Trained quantum neural networks are Gaussian processes

Giacomo De Palma

University of Bologna, Department of Mathematics, Piazza di Porta San Donato 5, 40126 Bologna BO, Italy giacomo.depalma@unibo.it

Filippo Girardi

Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa PI, Italy filippo.girardi@sns.it

Quantum neural networks represent the quantum analog of deep neural networks, leveraging the unique properties of quantum mechanics to potentially enhance machine-learning algorithms. Despite their promise, quantum neural networks currently lack a solid mathematical foundation. This work aims to establish such foundation.

We investigate quantum neural networks for supervised learning, constructed with parametric one-qubit gates and fixed two-qubit gates, where the output function is the expectation value of the sum of single-qubit observables across all qubits.

First, we demonstrate that the probability distribution of the function generated by untrained quantum neural networks with randomly initialized parameters converges in distribution to a Gaussian process in the limit of infinite width, provided that each measured qubit is correlated with only a few other qubits.

Then, we analytically characterize the gradient-descent training dynamics of the network in the limit of infinite width. We prove that the loss function decays exponentially in the training time, and therefore that the trained network can perfectly fit the training set. Moreover, we prove that during the whole training, the probability distribution of the generated function still converges in distribution to a Gaussian process. The proof of such a result relies on proving that training occurs in the lazy regime, i.e., that the maximum variation of each parameter vanishes in the limit of infinite width.

Finally, we address the statistical noise in measurements at the output of the network, proving that number of measurements growing polynomially with the number of qubits is sufficient to ensure the convergence to a Gaussian process, and therefore that the network can be trained in polynomial time.

References

[1] Filippo Girardi, Giacomo De Palma *Trained quantum neural networks are Gaussian processes*, arXiv:2402.08726

Latent Dynamics Models

Stefania Fresca

Address stefania.fresca@polimi.it

Nicola Farenga, Simone Brivio, Andrea Manzoni

Address {nicola.farenga, simone.brivio, andrea1.manzoni}@polimi.it

Solving differential problems using full order models (FOMs), such as the finite element method, usually results in prohibitive computational costs, particularly in real-time simulations and multi-query routines. Reduced order modeling aims 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. Within this context, deep learning-based reduced order models (DL-ROMs) have emerged as a novel and comprehensive approach, offering efficient and accurate surrogates for solving parametrized time-dependent nonlinear PDEs. By leveraging the mathematical properties of the system, the accuracy and generalization capabilities of DL-based ROMs can be further enhanced.

In this respect, latent dynamics models (LDMs) represent a novel mathematical framework in which the latent state is constrained to evolve according to an (unknown) ODE. A time-continuous setting is employed to derive error and stability estimates for the LDM approximation of the FOM solution. The impact of using an explicit Runge-Kutta scheme in a time-discrete setting is then analyzed, resulting in the ΔLDM formulation. Additionally, the learnable setting, ΔLDM_θ, is explored, where deep neural networks approximate the discrete LDM components, ensuring a bounded approximation error with respect to the high-fidelity solution. Moreover, the framework demonstrates the capability to achieve a time-continuous approximation of the FOM solution in a multi-query context, thus being able to compute the LDM approximation at any given time instance while retaining a prescribed level of accuracy.

References

[1] N. Farenga, S. Fresca, S. Brivio, A. Manzoni *On latent dynamics learning in nonlinear reduced order modeling*, arXiv preprint arXiv:2408.15183, 2024.

[2] S. Fresca, A. Manzoni *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, 2022.

Formal Analysis of Data-Aware Processes via Symbolic AI

Alessandro Gianola

INESC-ID/Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

alessandro.gianola@tecnico.ulisboa.pt

Contemporary organizations are complex organisms involving multiple actors that use multiple resources to perform activities, interact with data objects and take decisions based on this interaction. This inherent complexity highlights the growing need for advanced modeling and analysis of business processes using modern and efficient techniques from various areas of computer science, and for the automatic regulation of their internal work by exploiting Artificial Intelligence (AI) methods. Business process management (BPM [5]) has emerged as a well-established research field and industry-oriented discipline at the crossroads of operations management, computer science, data science, and software and systems engineering. Its primary goal is to support managers, analysts, ICT professionals, and domain experts in designing, deploying, enacting, and continuously improving processes to meet organizational objectives. Addressing the intricate nature of modern business processes requires safe and trustworthy systems that stakeholders and practitioners can depend on, posing significant challenges for both modeling and analysis.

The complexity intensifies when business processes are analyzed not only through their control flow but also by examining their interaction with data [4]. The data dimension can take various forms, such as case variables that encapsulate data objects or more intricate persistent storage systems like relational databases. Recently, considerable research across different fields has focused on integrating data and processes [3] to gain a deeper understanding of their dynamic interaction. This integration necessitates exploring how data influences process behavior and, conversely, how the control flow of the process affects the data it accesses and modifies. We call such complex systems *data-aware processes*.

Overall, the integration of BPM and AI is transforming the development of intelligent and reliable information systems, in particular when these systems integrate processes and data. On the one hand, BPM raises novel and challenging questions about processes and the event data they generate during execution. On the other hand, AI provides a set of robust techniques that require continuous adaptation and refinement to address these questions successfully. In this talk, I will explore how the integration of BPM and AI paves the way for innovative systems capable of managing organizational complexities while meeting operational goals. A particular emphasis will be placed on advanced techniques for the analysis of data-aware processes [6], considering tasks such as formal verification [2] and conformance checking [1]. Specifically, I will argue how symbolic AI and formal methods provide rigorous foundational approaches and powerful tools to precisely specify and analyze generic relational dynamic systems for capturing data-aware processes, ensuring both reliability and robustness.

References

[1] W. M. P. van der Aalst. *Process Mining - Data Science in Action, Second Edition*, Springer, 2016

[2] C. Baier, J.-P. Katoen. *Principle of Model Checking*, MIT Press, 2008

[3] D. Calvanese, G. De Giacomo, M. Montali. *Foundations of data-aware process analysis: a database theory perspective*, In Proceedings of PODS, 2013

[4] M. Dumas. *On the convergence of data and process engineering*, In Proceedings of ADBIS 2011, volume 6909 of LNCS. Springer, 2011

[5] M. Dumas, M. La Rosa, J. Mendling, H. A. Reijers. *Fundamentals of Business Process Management, Second Edition*, Springer, 2018

[6] A. Gianola. *Verification of Data-Aware Processes via Satisfiability Modulo Theories*, Lecture Notes in Business Information Processing 470, Springer, 2023

Stochastic (but *structured*) zeroth order optimization

Cesare Molinari

Università di Genova, MaLGa Center; cecio.molinari@gmail.com

M. Rando, C. Traoré, D. Kozak, L. Rosasco, S. Villa

Finite-difference methods are a class of algorithms designed to solve black-box optimization problems by approximating a gradient of the target function on a set of directions. In black-box optimization, the non-smooth setting is particularly relevant since, in practice, differentiability and smoothness assumptions cannot be verified. To cope with non-smoothness, several authors use a smooth approximation of the target function and show that finite difference methods approximate its gradient. Recently, it has been proved that imposing a structure in the directions allows improving performance. However, only the smooth setting was considered. To close this gap, we introduce and analyze O-ZD, the first structured finite-difference algorithm for non-smooth black-box optimization. Our method exploits a smooth approximation of the target function and we prove that it approximates its gradient on a subset of random *orthogonal* directions. We analyze the convergence of O-ZD under different assumptions. For non-smooth convex functions, we obtain the optimal complexity. In the non-smooth non-convex setting, we characterize the number of iterations needed to bound the expected norm of the smoothed gradient. For smooth functions, our analysis recovers existing results for structured zeroth-order methods for the convex case and extends them to the non-convex setting. We conclude with numerical simulations where the assumptions are satisfied, observing that our algorithm has very good practical performance.

References

[1] D. Kozak, C. M., S. Villa, L. Rosasco, L. Tenorio: *Zeroth-order optimization with orthogonal random directions*, Mathematical Programming, 199 (1-2), 1179-1219 (2023)

[2] M. Rando, C. M., L. Rosasco, S. Villa: *An Optimal Structured Zeroth-order Algorithm for Non-smooth Optimization*, Advances in Neural Information Processing Systems 36 (2023)

[3] M. Rando, C. M., S. Villa, L. Rosasco: *Stochastic Zeroth order Descent with Structured Directions* Computational Optimization and Applications, 1-37 (2024)

[4] M. Rando, C. Traoré, C. M., L. Rosasco, S. Villa: *A Structured Proximal Stochastic Variance Reduced Zeroth-order Algorithm*, in preparation

Bures-Wasserstein gradient-based learning of covariance operators in Gaussian processes

Katerina Papagiannouli

University of Pisa & Max Planck Institute Mathematics in the Sciences;

aikaterini.papagiannouli@unipi.it

We study gradient-based learning of covariance operators in Gaussian processes, emphasizing low-rank approximations. Based on the Bures-Wasserstein gradient flow framework, we propose methods to learn covariance eigenvalues and eigenfunctions feature-by-feature. Utilizing the dynamics of the Kullback-Leibler (KL) divergence within the framework of Gaussian distributions equipped with Bures-Wasserstein geometry, we show convergence guarantees for eigenvalue learning in various settings, including neural network architectures. Our approach extends to neural network-based parametrizations, enabling scalable and efficient learning for complex data distributions.

References

[1] K. P., P. Bréchet, j. An, G. Montufar *Critical Points and Convergence Analysis of Generative Deep Linear Networks Trained with Bures-Wasserstein Loss* ICML (2023)

[2] K. P, P. Brechét, A. Agazzi: *Learning covariance operators feature by feature Gradient-based low rank approximation of Gaussian processes*, in preparation

Whiteness-based learning of parameters in inverse imaging problems

Monica Pragliola

Department of Mathematics and Applications, University of Naples Federico II

monica.pragliola@unina.it

Variational methods for ill-posed imaging inverse problems aim to minimize a functional which is sum of a fidelity term and of a regularization term, the two terms being balanced by the so-called regularization parameter. It is well-established that flexible models are characterized by highly-parametrized regularizers, and it is thus crucial to design robust methods for the selection of the possibly high number of parameters arising in the models of interest. In this talk, we take a journey through the different instances of the Residual Whiteness Principle, an unsupervised approach that has been originally introduced for the estimation of the single regularization parameter in variational models [3]. In its seminal version, the RWP is applied to white-noise corrupted data and it amounts to maximize the whiteness of the residual image, i.e. to minimize the autocorrelation of its entries. We will discuss how the RWP can be extended so as to be applied to non-white yet whitenable noise statistics, such as, e.g. Poisson noise and mixed Poisson-Gaussian noise [2, 1]. Moreover, we will show how the bilevel optimization task expressing the RWP can be tackled so as to reduce the computational costs and to make it possible to employ the whiteness-based unsupervised principle for the estimation of a general large number of unknown parameters [4, 1].

This talk summarizes the results achieved with several co-authors: Francesca Bevilacqua, Alessandro Lanza, Fiorella Sgallari, Luca Calatroni, Marco Donatelli, Carlo Santambrogio.

References

[1] Bevilacqua F., Lanza A., Pragliola M., Sgallari F. *A general framework for whiteness-based parameters selection in variational models* , Computational Optimization and Applications (2024)

[2] Bevilacqua F., Lanza A., Pragliola M., Sgallari F. *Whiteness-based parameter selection for Poisson data in variational image processing* , Applied Mathematical Modelling, 117 (2023)

[3] Lanza A., Pragliola M., Sgallari F. *Residual whiteness principle for parameter-free image restoration*, Electronic Transactions on Numerical Analysis, 53 (2020)

[4] Santambrogio C., Pragliola M., Lanza A., Donatelli M., Calatroni L. *Whiteness-based bilevel learning of regularization parameters in imaging*, European Signal Processing Conference (2024)

Contributed Talks

Boolean SK model

Linda Albanese

University of Salento, linda.albanese@unisalento.it

Andrea Alessandrelli

University of Pisa, andrea.alessandrelli@studenti.unipi.it

In recent years, the rapid development of Artificial Intelligence (AI) solutions has profoundly influenced contemporary scientific research. Its impact is reshaping the scope of applied disciplines [1, 2] while simultaneously inspiring theoretical interest in automated systems across fields such as neuroscience, statistics, complex systems physics, engineering, and information theory.

The statistical mechanics of spin glasses has traditionally served as a paradigm for modelling and interpreting diverse phenomena, spanning from quantitative biology to computer science. Despite the substantial body of research in this field, there remains a notable gap concerning the substitution of Ising spins with Boolean spins; given the role of Boolean variables as binary units in Machine Learning, addressing this gap is now essential.

In this presentation, we will discuss an approach to filling this lacuna for the mean-field model with Boolean variables and disordered couplings governed by a Gaussian distribution. Given the similarities with the Sherrington-Kirkpatrick (SK) model [3, 4] – a foundational framework for mean-field spin glasses – this model is naturally referred to as the *Boolean SK model*. Due to time constraints, our focus will be on the application of Guerra's interpolation method [5] to derive the thermodynamic expression of the quenched statistical pressure under both the Replica Symmetric and first-step Replica Symmetry Breaking assumptions.

However, despite the structural similarities, the Boolean SK model exhibits distinct characteristics compared to the original SK model. Specifically, due to the breaking of spin-flip symmetry, it exhibits an inherent magnetisation, and the overlap (an analogue for the SK model) lacks the conventional phase transition. Instead, the system transitions continuously from a random state to a disordered phase. All theoretical results are substantiated by numerical analyses.

This work may serve as a foundation for a series of studies aimed at understanding other network models where Ising spins are replaced by Boolean spins.

This research is inspired by joint work with Andrea Alessandrelli (University of Pisa) [6].

References

[1] J. Leskovec, A. Rajaraman, J. D. Ullman *Mining of massive datasets* , Cambridge University Press, 2014.

[2] K. K. Jain *Personalized medicine* , Current Opinion in Molecular Therapeutics, 4(6):548-558, 2002.

[3] Sherrington, D., Kirkpatrick, S. *Solvable model of a spin-glass* , Physical review letters 35.26 (1975): 1792.

[4] Mézard, M., Parisi, G., Virasoro, M. A. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications* , Vol. 9. World Scientific Publishing Company, 1987.

[5] Guerra, Francesco *Broken replica symmetry bounds in the mean field spin glass model* , Communications in mathematical physics 233 (2003): 1-12

[6] Albanese, L., Alessandrelli, A. *Boolean mean field spin glass model: rigorous results* , arXiv preprint arXiv:2409.08693 (2024).

Networks of neural networks: disentanglement of overlapping inputs

Andrea Alessandrelli

Università di Pisa andrea.alessandrelli@phd.unipi.it

Elena Agliari

Sapienza Università di Roma elena.agliari@uniroma1.it

Adriano Barra

Sapienza Università di Roma adriano.barra@uniroma1.it

Martino S. Centonze

Università di Bologna martino.centonze@unibo.it

Federico Ricci-Tersenghi

Sapienza Università di Roma federico.ricci@uniroma1.it

This work investigates the intersection of Artificial Intelligence and Statistical Mechanics, focusing on the hetero-associative extension of the classic Hopfield network [1]. Indeed, we present an extended version of the Bidirectional Associative Memory (BAM) [3] that can concurrently process three or more patterns [2].

Our analysis shows that an ensemble of BAM models exhibits emergent capabilities absent in a single network. Specifically, we design a layered associative Hebbian network that not only performs standard pattern recognition but also achieves pattern disentanglement. For instance, when we present a composite input – such as a musical chord – the network can extract the individual elements constituting it, i.e. the distinct notes. In our investigation, we restrict to notes represented as Rademacher vectors and chords constructed as their mixtures, analogous to the spurious states in a Hopfield model. Through a statistical-mechanical analysis (both analytical and computational), we derive the conditions on the model parameters that enable successful pattern disentanglement.

Leveraging statistical mechanics, interpolation techniques, and phase diagrams, we characterize critical computational features and optimize network configurations. Numerical experiments on hierarchical synthetic datasets confirm the model's capability for input disentanglement, with theoretical predictions aligning closely with the empirical results. This statistical-mechanical framework not only enables optimized network parameterization but also provides a pathway for *a priori* optimization of deep learning architectures, aligning network structure with the intrinsic organization of the data under analysis.

References

[1] J. J. Hopfield, *Neural networks and physical systems with emergent collective computational abilities*, Proceedings of the National Academy of Sciences of the United States of America, 79:2554–2558 (1982).

[2] E. Agliari, A. Alessandrelli, A. Barra, M.S. Centonze, F. Ricci-Tersenghi, *Generalized hetero-associative neural networks*, arXiv preprint arXiv:2409.08151 (2024).

[3] B. Kosko, *Bidirectional associative memories*, IEEE Transactions on Systems, man, and Cybernetics,18(1):49–60 (1988).

Exploring Deep Learning in Seismology for Early Warning systems

Antonioreneè Barletta

Università degli studi di Napoli Federico II antoniorenee.barletta@gmail.com

S. Cuomo, G. Milano

Università degli studi di Napoli Federico II salvatore.cuomo@unina.it,

Spici srl milano@spici.eu

One of the major challenges in seismology is the development of fast, precise, and robust solutions for early warning (EW) systems. EW involves methodologies for detecting and rapidly analyzing an earthquake's initial, non-damaging primary (P) waves. These approaches aim to estimate critical parameters such as the earthquake's epicenter, magnitude, and potential impact, allowing alerts to be issued before the arrival of the slower, destructive secondary (S) waves. In this domain, seismic data are collected from sensors (typically seismographs) and recorded as time series. These data capture essential characteristics of seismic waves, including their amplitude, frequency, and timing, providing crucial information for accurate analysis and interpretation. The research literature demonstrates the effectiveness of various machine learning approaches for EW applications, for example, models such as random forests, gradient boosting algorithms, and Support Vector Machines (SVMs) have been widely explored due to their robustness and reliability. More recently, the emergence of deep learning, driven by advancements in high-performance hardware like GPUs and TPUs, has revolutionized this research field. Techniques involving Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTMs), and Convolutional Neural Networks (CNNs) have shown excellent performance in seismology-related tasks. In this study, we explore the application of a Temporal Convolutional Network (TCN) for analyzing earthquake seismograms in the context of an EW system. Our investigation focuses on leveraging the unique capabilities of TCNs to enhance the speed and accuracy of seismic data analysis, oriented to the development of applications and services designed for EW scenarios.

References

[1] C. Satriano, Y. Wu, A. Zollo, H. Kanamori *Earthquake early warning: Concepts, methods and physical grounds*, Soil Dynamics and Earthquake Engineering 31 (2), 106-118.

[2] W. Zhu, G. C. Beroza *PhaseNet: a deep-neural-network-based seismic arrival-time picking method*, Geophysical Journal International, Volume 216, Issue 1, January 2019, Pages 261–273.

[3] R. Rea, S. Colombelli, L. Elia, A. Zollo *Retrospective performance analysis of a ground shaking early warning system for the 2023 Turkey–Syria earthquake*, Communications Earth & Environment - Nature, 5 (1), 332.

[4] X. Liu, T, Ren, H. Chen, G. M. Dimirovski, F. Meng, P. Wang *Earthquake magnitude estimation using a two-step convolutional neural network*, Journal of Seismology - Springer, 2024

[5] F. Piccialli, S. Cuomo, F. Giampaolo, E. Prezioso *Prediction method and related system*, US Patent App. 17/815,737.

Simulations of Water Distribution Systems via Radial Basis Function Neural Networks

Vittorio Bauduin

University of Campania L. Vanvitelli, email: vittorio.bauduin@unicampania.it

Salvatore Cuomo

University of Naples Federico II, email: salvatore.cuomo@unina.it

Water Distribution Networks (WDNs) are critical infrastructures consisting of thousands of interconnected elements, characterized by a meshed and irregular topology shaped by urban layouts. Their structural robustness is inherently linked to the operational efficiency of both the networks themselves and associated infrastructures. Analyzing such complex systems necessitates a balanced methodology that integrates holistic and reductionist perspectives, supported by tools to examine network topology, behavior, and dynamic evolution using advanced analytical frameworks. This study adopts a predictive approach by integrating Radial Basis Function Neural Networks (RBF-NNs) with real-time sensor data to enhance understanding and management of WDNs. Numerous methods for approximating multivariate functions have their own benefits but struggle with high dimensionality, common in water distribution networks (WDN) applications. This "curse of dimensionality" limits traditional methods' effectiveness due to computational challenges. As a result, more specialized approaches like reduced-rank techniques, sparse grid approximations, or neural networks are preferred for their computational efficiency and accuracy in high-dimensional contexts. RBF-NNs emerge as a robust solution to address these challenges, as capturing the local-to-global temporal inference inherent to the dynamic behavior of these networks. Their efficacy has been validated through numerical testing in this study, with the primary objective of developing an interpolative model capable of accurately predicting pressures and flows across the entire WDN using available sensor data. The proposed model demonstrated excellent performance metrics, achieving a Loss Function value and Mean Squared Error (MSE) of approximately 10^{-12}, alongside a Mean Absolute Error (MAE) of around 10^{-7}. These results underline the high accuracy and reliability of the predictive approach, validating its potential for effective WDN analysis and management.

The work was carried out in collaboration with Prof. Giovanni Francesco Santonastaso from the Department of Engineering at the University of Campania "L. Vanvitelli", who served as the tutor for the research project PRIN 2022 "SMART RENEW" - Rehabilitation of Water Distribution Networks through a Data-Driven Approach *(CUP: B53D23006080006).*

References

[1] A. Di Nardo, C. Giudicianni, R. Greco, M. Herrera, G. F. Santonastaso, *Applications of Graph Spectral Techniques to Water Distribution Network Management*, Water, vol. 10, no. 1, p. 45, Jan. 2018, doi: 10.3390/w10010045.

[2] M. D. Buhmann, *Radial basis functions*, Acta Numerica, vol. 9, pp. 1–38, Jan. 2000, doi: 10.1017/s0962492900000015.

[3] V. Bauduin, S. Cuomo, V. Schiano Di Cola, *Constraint Satisfaction approach for Neuron Configurations in Neural Networks*,(submitted for pubblication)

An efficient matheuristic for nurse rostering problems

Cristian Belfiore

de-Health Lab - Laboratory of Decision Engineering for Health Care Services Dep. of Mechanical, Energy and Management Engineering, University of Calabria, Ponte Pietro Bucci Cubo 41C, 87036 Arcavacata di Rende (CS), Italy cristian.belfiore@unical.it

Rosita Guido, Domenico Conforti

de-Health Lab - Laboratory of Decision Engineering for Health Care Services Dep. of Mechanical, Energy and Management Engineering, University of Calabria, Ponte Pietro Bucci Cubo 41C, 87036 Arcavacata di Rende (CS), Italy rosita.guido@unical.it domenico.conforti@unical.it

Matheuristics represent a novel approach to problem-solving, combining elements of both exact and heuristic methodologies. FiNeMath is a matheuristic that employs a combination of a Large Neighbourhood Search approach and a Destroy-and-Reconstruct strategy. In this study, we adapt FiNeMath to address the well-known nurse rostering problem. The goal of this problem is to assign a set of nurses to shifts within a predefined period. Nurses are classified according to one or more skills, are engaged under a contract that governs their work, and may express preferences regarding days-off, shifts-off, or shifts-on. For each shift in the planning period, a preferred level of staffing is defined. This value represents the number of nurses considered optimal by the hospital. The assignments must comply with a set of constraints derived from the hospital's internal rules and legal requirements, as well as the preferences expressed by the nurses. The objective is to obtain shifts coverage while ensuring staff satisfaction and workload balance.
In our approach, we start by constructing an initial feasible solution to the problem by using a Fix-and-Optimize strategy. Then, we attempt to iteratively improve this solution by partially destroying it and then reconstruct it through a solver. The ruined portion is identified by a destruction operator that varies iteration to iteration. The experimental campaign conducted on benchmark instances available in the literature demonstrated promising outcomes for our solution approach.

References

[1] Cristian Belfiore. *An effective matheuristic approach to solve Nurse Rostering Problem*, 7AYW-8AYW· Operations Research Beyond Frontier – Proceedings of the 7th and the 8th AIROYoung Workshops. AIRO Springer Series. (2024) [in press]

[2] Guido, R., Groccia, M. C., Conforti, D. *An efficient matheuristic for offline patient-to-bed assignment problems* , European Journal of Operational Research, 268, 2, 486-503, (2018).

[3] Curtois, T., Qu, R. *Computational results on new staff scheduling benchmark instances* , Technical Report, ASAP Research Group, School of Computer Science, University of Nottingham, NG8 1BB, Nottingham, UK. (2014).

Semi-Supervised Learning for Time Series Clustering Using Copulas

Alessia Benevento

Dipartimento di Matematica e Fisica "Ennio De Giorgi", Università del Salento, Lecce, Italy

alessia.benevento@unisalento.it

Fabrizio Durante

Dipartimento di Matematica e Fisica "Ennio De Giorgi", Università del Salento, Lecce, Italy

fabrizio.durante@unisalento.it

Roberta Pappadà

Dipartimento di Scienze Economiche, Aziendali, Matematiche e Statistiche "B. de Finetti", Università degli Studi di Trieste, Trieste, Italy rpappada@units.it

Time-series data containing one or multiple variables that vary with time is extensively recorded and analyzed in various fields, such as science, engineering, medicine, economics, and finance. Clustering is a powerful data mining technique for classifying these temporal data into related groups in the absence of sufficient prior knowledge of the groups. Clustering methods for time series are typically performed in unsupervised learning settings, where the aim is to uncover hidden structures in the data. However, if the data comes with additional background information, such as pairwise positive/negative relationships with associated degrees among the time series, this can impose constraints on the clustering process. In such cases, the approach is more accurately described as semi-supervised learning. The first goal of this presentation is to review certain aspects of dissimilarity-based clustering methods that have been introduced within a copula framework.

Additionally, in many applications, the identification of clusters among time series is complicated by the presence of spatial constraints and the need to capture complex dependence structures, including tail dependencies. This talk presents a novel semi-supervised learning framework for clustering time series based on copula models, inspired by the methodologies introduced in [1]. We leverage copula-based measures to model temporal dependence structures and tail behaviors. The semi-supervised approach lead to the clustering of the time-series while taking into account spatial proximities. We demonstrate the method's efficacy through simulated and real-world datasets, highlighting its applicability in fields such as environmental monitoring.

References

[1] Benevento, A., Durante, F., and Pappadà, R. *Tail-dependence clustering of time series with spatial constraints* , Environmental and Ecological Statistics (2024): 1-17

Graph distinction through GENEOs and Permutants

Giovanni Bocchi

Address giovanni.bocchi1@unimi.it

Massimo Ferri, Patrizio Frosini

Address massimo.ferri@unibo.it, patrizio.frosini@unipi.it

The theory of Group Equivariant Non-Expansive Operators (GENEOs) was initially developed in Topological Data Analysis for the geometric approximation of data observers, including their invariances and symmetries. In this work we depart from that line of research and explore the use of GENEOs for distinguishing graphs up to isomorphisms. In doing so, we aim to test the capabilities and flexibility of the operators obtained exploiting Generalized Permutants specifically designed to search for interesting subgraph structures in the graphs to be tested. Our experiments show that the isomorphism test we obtained using a minimal number of GENEOs learned from data offers the best compromise between efficiency and computational costs when tested on the comparison r-regular graphs. In addition, the actions on data of the learned operators are easily interpretable. This helps to support the idea that GENEOs could be a general-purpose approach to discriminative problems in Machine Learning when some structural information about data and observers is explicitly given.

References

[1] Giovanni Bocchi, Patrizio Frosini, Alessandra Micheletti, et al. *A geometric XAI approach to protein pocket detection*, xAI-2024 Late-breaking Work, Demos and Doctoral Consortium Joint Proceedings - The 2nd World Conference on eXplainable Artificial Intelligence. CEUR https://ceur-ws.org/Vol-3793/ (2024).

[2] Giovanni Bocchi, Stefano Botteghi, Martina Brasini, et al *On the finite representation of linear group equivariant operators via permutant measures*, Annals of Mathematics and Artificial Intelligence 91.4 (2023), pp. 465-487. ISSN: 1012-2443. DOI: 10.1007/s10472-022-09830-1.

[3] Faraz Ahmad, Massimo Ferri, and Patrizio Frosini *Generalized Permutants and Graph GENEOs*, Machine Learning and Knowledge Extraction 5.4 (2023), pp. 1905-1920. DOI: 10.3390/make5040092

[4] Mattia G. Bergomi, Patrizio Frosini, Daniela Giorgi, et al. *Towards a topological–geometrical theory of group equivariant non-expansive operators for data analysis and machine learning*, Nature Machine Intelligence 1.9 (2019), pp. 423-433. ISSN: 2522-5839. DOI: 10.1038/s42256-019-0087-3

[5] Ryoma Sato *A Survey on The Expressive Power of Graph Neural Networks*, Preprint at arXiv (2020). DOI:10.48550/arXiv.2003.04078.

Mitigating the adverse effects of data scarcity through pre-trained physics-informed DL-ROMs

Simone Brivio†, Stefania Fresca, Andrea Manzoni

MOX, Dept. of Mathematics, Politecnico di Milano, P.zza Leonardo da Vinci 32, Milano, I-20133, Italy

simone.brivio@polimi.it, stefania.fresca@polimi.it, andrea1.manzoni@polimi.it

Deep learning-based reduced order models (DL-ROMs) provide a comprehensive paradigm for nonlinear model order reduction enabling the construction of fast and efficient surrogate models for the simulation of nonlinear parametrized PDEs [3]. Experimental evidence and theoretical results have recently demonstrated that the prediction accuracy of data-driven DL-ROMs is often unsatisfactory when only an insufficient amount of labeled data is available at the training stage [1]. Unfortunately, data scarcity is common in Scientific Machine Learning (SciML) applications. Indeed, data are usually generated through synthetic solvers, which provide highly accurate and reliable simulations, but generally demand excessive computational resources. For this reason, we are normally only able to generate only a handful of labeled data, which are often not representative of the entire parametric space.

To compensate for the accuracy shortfall brought about by data scarcity, we build on the fact that the governing equations convey the same information as the data synthetically generated through numerical solvers. Consequently, it is sound to minimize the residual of the governing equation in the regions of the parametric space that are not properly covered by labeled training data. The resulting *physics-informed* approach is unsupervised by nature and does not need additional input-output pairs.

However, especially as the problem complexity increases, such physics-informed architecture requires a significant amount of computational resources to be suitably trained, and its optimization phase is prone to convergence failure. To avoid these side effects, by further intertwining data and physics, we devise a novel two-step training strategy, consisting of *(i)* a fast and efficient *pre-training* stage that enables the optimizer to quickly and stably approach the minimum in the loss landscape, and *(ii)* a *fine-tuning* phase that further enhances the prediction accuracy.

Ultimately, we showcase the potential of the resulting paradigm, termed Pre-Trained Physics-Informed DL-ROM (PTPI-DL-ROM), by assessing its performance in terms of prediction accuracy and training efficiency [2]. To this end, we consider a series of numerical experiments involving parametrized PDEs stemming from computational fluid dynamics and mathematical biology.

References

[1] Brivio, S., Fresca, S., Franco, N. & Manzoni, A. Error estimates for POD-DL-ROMs: a deep learning framework for reduced order modeling of nonlinear parametrized PDEs enhanced by proper orthogonal decomposition. *Adv. Comput. Math.*. **50** (2024)

[2] Brivio, S., Fresca, S. & Manzoni, A. PTPI-DL-ROMs: Pre-trained physics-informed deep learning-based reduced order models for nonlinear parametrized PDEs. *Computer Methods In Applied Mechanics And Engineering*. **432** pp. 117404 (2024)

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

Majorization-Minimization for multiclass classification in a big data scenario

Filippo Camellini

Department of Physics, Informatics and Mathematics, Via Campi 213/B, 41125, Modena, Italy

filippo.camellini@unimore.it

Emilie Chouzenoux, Jean–Christophe Pesquet

CVN, Inria, CentraleSupélec, University Paris Saclay, 9 rue Joliot Curie, 91190, Gif-sur-Yvette, France

emilie.chouzenoux@inria.fr, jean-christophe@pesquet.eu

Giorgia Franchini, Federica Porta

Department of Physics, Informatics and Mathematics, Via Campi 213/B, 41125, Modena, Italy

giorgia.franchini@unimore.it, federica.porta@unimore.it

Majorization-minimization (MM) algorithms are numerical optimization methods that simplify the original problem by iteratively replacing the objective function with more tractable approximations. Through an iterative process, at each step the minimization problem associated with a surrogate function is solved. This surrogate function, also called the tangent majorant, approximates the objective function and exhibits good properties, typically convexity.

The MM algorithms just described are widely used in the field of supervised learning. For instance, in [1], an MM approach with a quadratic surrogate function is exploited to train a binary SVM-based linear model with a squared hinge loss as the data fidelity term and a smooth regularization term that induces sparsity.

The main objective and original contribution of our work is to propose a highly scalable MM algorithm for training a linear multiclass classification model that leverages a data-fidelity function and a regularizer term which are L-smooth. Specifically, when extending the MM approach presented in [1] to the multiclass classification case, two challenges arise. The first involves formulating a differentiable objective function, which leads to the use of the Weston-Watkins formulation [2]. The second challenge relates to the case where the size of the training set makes gradient computation unfeasible. In this big data context, it becomes necessary to introduce techniques that use gradient approximation while still allowing the exploitation of MM methods.

The Incremental MM algorithm we propose is inspired by a classical incremental gradient scheme [3] in which the descent direction is rescaled using a symmetric positive definite matrix. The use of a scaling matrix derived from the MM Quadratic approach presented in [1] allows us to leverage a second-order approximation of the function, rather than relying solely on the gradient. Unlike a classic second-order method that involves high computational costs at each iteration, the structure of our proposed matrix enables more efficient computation. A square matrix of order equal to the gradient dimension is calculated in a warm-up phase that precedes the optimization process and can be easily parallelized. Subsequently, for each epoch, this matrix remains constant and is updated by computing only a diagonal matrix.

The experiments presented in our work highlight the advantage of using the second-order information introduced by the scaling matrix. In particular, MM Incremental exhibits better performance compared to other algorithms commonly used in big data contexts, such as Incremental or Stochastic Gradient Descent.

References

[1] Alessandro Benfenati, Emilie Chouzenoux, Giorgia Franchini, Salla Latva-Äijö, Dominik Narnhofer, Jean–Christophe Pesquet, Sebastian J. Scott, Mahsa Yousefi *Majoration-Minimization for Sparse*

SVMs, Advanced Techniques in Optimization for Machine Learning and Imaging, Springer Nature Singapore, 2024

[2] Yutong Wang, Clayton Scott *Weston-Watkins Hinge Loss and Ordered Partitions*, Advances in Neural Information Processing Systems, 2020

[3] Dimitri P. Bertsekas, John N. Tsitsiklis *Gradient Convergence in Gradient methods with Errors*, SIAM Journal on Optimization, 2000

Implicit Neural Field Reconstruction on Complex Shapes from Scattered Data

Davide Carrara[a], Marc Hirschvogel[a], Francesco Regazzoni[a], Simone Pezzuto[b], Stefano Pagani[a]

[a] MOX, Dipartimento di Matematica, Politecnico di Milano, Milan, Italy

[a] {davide.carrara, marc.hirschvogel, francesco.regazzoni, stefano.pagani}@polimi.it

[b] Dipartimento di Matematica, Università di Trento, Trento, Italy

[b] simone.pezzuto@unitn.it

In many engineering and medical applications, reconstructing physical fields and domain geometries from noisy, scattered data collected by local sensors is a critical task. Both the statistical reconstruction of distributed quantities and the simulation of physical processes (typically modeled by means of partial differential equations) depend heavily on accurate geometry reconstruction.

Meshless approaches, such as using Multi-Layer Perceptrons to represent Signed or Unsigned Distance Function (S/U-DF) from target geometries, have been effective in tackling this challenge, but often require intense preprocessing of the data and are not suited for sparse datasets. We propose two novel approaches for geometry reconstruction, tailored to scenarios of low and high data numerosity, that require only point cloud representations and do not need mesh or point correspondences for input. We present applications of each method for reconstructing cardiac geometries.

For cases with high-quality data, we propose a supervised reconstruction pipeline using the *DeepSDF* architecture [1]. This method combines an embedding model and a regression network to learn and reconstruct the shapes of multiple objects using a shared network. Each geometry is associated with a latent code that encodes shape information, enabling the generation of realistic new synthetic shapes by sampling the latent space. We demonstrate the application of this method for solving nonlinear PDEs on reconstructed geometries, where the latent code is used for network conditioning [2]. For scenarios with limited or noisy data where SDF computation is not feasible, we introduce a novel method [3] that reconstructs the geometry from surface-level point measurements. Our approach employs a tailored loss function combining fit and regularization terms, including a differential term based on the eikonal equation to enhance model generalization. The reconstructed shape model is then used to predict distributed quantities on the surface, taking into account its geometry. High accuracy and geometric fidelity are ensured through supervised training and validation against derived surface properties such as gradients, which are computed using automatic differentiation. We validate this method on both synthetic datasets and an atrial cardiac geometry.

This project has been partially funded from the project PRIN2022, MUR, Italy, 2023–2025, P2022N5ZNP "SIDDMs: shape-informed data-driven models for parametrized PDEs, with application to computational cardiology".

References

[1] Park, J. J., Florence, P., Straub, J., Newcombe, R., and Lovegrove, S. *DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation.* , 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 165–174.

[2] Regazzoni F., Pagani S., Quarteroni A. *Universal Solution Manifold Net works (USM-Nets): Non-Intrusive Mesh-Free Surrogate Models for Problems in Variable Domains* , Journal of Biomechanical Engineering, 144(12), 2022

[3] Carrara D., Regazzoni F., Pagani S. *Implicit neural field reconstruction on complex shapes from scattered and noisy data* , Mox Report 40/2024

Operator Learning Techniques in Computational Cardiology

Edoardo Centofanti

Università di Pavia, Via Ferrata 5, 27100, Pavia, Italy edoardo.centofanti01@universitadipavia.it

Operator Learning methods are gaining significant attention in biomathematics and computational cardiology due to their ability to efficiently approximate complex dynamical systems. These methods offer new opportunities for reducing computational costs while maintaining accuracy, making them particularly suited for addressing challenges in cardiac modeling. In this talk, we will explore the application of Operator Learning techniques to tackle two key challenges in cardiac electrophysiology. First, we examine their capability of learning ionic models [1], which play a critical role in describing cellular excitability and action potential generation, but present a challenging nonlinear and stiff dynamics. Second, we focus on applying the Fourier Neural Operator (FNO) to learn activation and repolarization times [2], as evaluated through the monodomain cardiac model. By comparing these approaches to traditional numerical solvers, we will highlight their potential for accurately reconstructing electrophysiological dynamics while significantly improving computational efficiency.

References

[1] E. Centofanti, M. Ghiotto, L.F. Pavarino *Learning the Hodgkin-Huxley Model with Operator Learning Techniques*, Computer Methods in Applied Mechanics and Engineering 432, Part A (2024): 117381.

[2] Joint work with G. Ziarelli, S. Scacchi, *in preparation*

A unified framework for equivariant neural network

Francesco Conti

Address francesco.conti@phd.unipi.it

Equivariant neural networks are proving effective in many real-world scenrios [1]. For example, Convolutional Neural Networks are the state-of-the-art in computer vision tasks and Topological Data Analysis [2] (TDA) is achieving great accomplishment with noisy datasets. In this talk, we are going to present a unified mathematical framework for equivariant neural networks and show that both CNNs and TDA can be expressed using this framework that we call Group Equivariant Non-Expansive Operators (GENEOs) [3].

References

[1] Gerken, Jan E and Aronsson, Jimmy and Carlsson, Oscar and Linander, Hampus and Ohlsson, Fredrik and Petersson, Christoffer and Persson, Daniel *Geometric deep learning and equivariant neural network* , Springer Artificial Intelligence Review

[2] Wasserman, Larry *Topological data analysis* , Annual Review of Statistics and Its Application

[3] Bergomi, Mattia G and Frosini, Patrizio and Giorgi, Daniela and Quercioli, Nicola *Towards a topological–geometrical theory of group equivariant non-expansive operators for data analysis and machine learning* , Nature Machine Intelligence

Integrating Molecular Dynamics and Machine Learning Algorithms to Predict the Functional Profile of Kinase Ligands

Ivan Cucchi

University of Pavia - Dept. of Mathematics ivan.cucchi01@universitadipavia.it

Elena Frasnetti

University of Pavia - Dept. of Chemistry

The modulation of protein function via designed small molecules is providing new opportunities in chemical biology and medicinal chemistry. While drugs have traditionally been developed to block enzymatic activities through active site occupation, a growing number of strategies now aim to control protein functions in an allosteric fashion, allowing for the tuning of a target's activation or deactivation via the modulation of the populations of conformational ensembles that underlie its function. In the context of the discovery of new active leads, it would be very useful to generate hypotheses for the functional impact of new ligands. Since the discovery and design of allosteric modulators (inhibitors/activators) is still a challenging and often serendipitous target, the development of a rapid and robust approach to predict the functional profile of a new ligand would significantly speed up candidate selection. Herein, we present different machine learning (ML) classifiers to distinguish between potential orthosteric and allosteric binders. Our approach integrates information on the chemical fingerprints of the ligands with descriptors that recapitulate ligand effects on protein functional motions. The latter are derived from molecular dynamics (MD) simulations of the target protein in complex with orthosteric or allosteric ligands. In this framework, we train and test different ML architectures, which are initially probed on the classification of orthosteric versus allosteric ligands for cyclin-dependent kinases (CDKs). The results demonstrate that different ML methods can successfully partition allosteric versus orthosteric effectors (although to different degrees). Next, we further test the models with FDA-approved CDK drugs, not included in the original dataset, as well as ligands that target other kinases, to test the range of applicability of these models outside of the domain on which they were developed. Overall, the results show that enriching the training dataset with chemical physics-based information on the protein–ligand dynamic cross-talk can significantly expand the reach and applicability of approaches for the prediction and classification of the mode of action of small molecules.

References

[1] E. Frasnetti, I. Cucchi, S. Pavoni, F. Frigerio, F. Cinquini, S. A. Serapian, L. F. Pavarino, G. Colombo *Integrating Molecular Dynamics and Machine Learning Algorithms to Predict the Functional Profile of Kinase Ligands* , Journal of Chemical Theory and Computation, Vol. 20, Issue 20

GANs through the Lens of Topological Data Analysis

Ben Cullen

University of Pisa ben.cullen@phd.unipi.it

B. T. Corradini[a], C. Gallegati[b], S. Marziali[b], G. A. D'Inverno[c], M. Bianchini[b], F. Scarselli[b]

[a] University of Florence, [b] University of Siena, [c] SISSA, Trieste

Generative Adversarial Networks (GANs) [1] aim to produce realistic samples by mapping a low-dimensional latent space to a high-dimensional data space by exploiting an adversarial training mechanism. Despite achieving state-of-the-art results, GAN training faces significant challenges such as mode collapse, vanishing gradients, and inefficiencies in hyperparameter tuning, relying on computationally expensive trial-and-error methods. In addition, GANs lack a clear early stopping criterion, often leading to resource-intensive training processes.
This work investigates GANs using Topological Data Analysis (TDA) tools [3] to gain deeper insights into their training dynamics and generative capabilities. By employing persistent homology, we examine the evolution of topological features during training, focusing on the convergence of the generated manifold to that of real data. Through various experiments on MNIST and CIFAR-10 datasets with different GAN models, we analyze the interplay between model architecture, training stability, and performance, as well as characterise common issues in GANs. In particular, we show that the Wasserstein distance between persistence diagrams, which summarise the topological features of manifolds, is a robust tool for quantifying similarities between generated and real data, offering a novel perspective on evaluating samples beyond conventional metrics like the Frechet Inception Distance (FID) [2]. Indeed, the FID score is shown to be insufficient in assessing the quality of generated images, neither alone nor in combination with the Intrinsic Dimension estimation [4]. Our results suggest that homological features provide a suitable characterisation of the generative process that can be valuable for uncovering insights about the structural transformations occurring during the training of a GAN. This study lays the foundation for integrating topology-based approaches into the optimization and assessment of generative models, potentially enabling the formulation of an early stopping criterion.

References

[1] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. & Bengio, Y. Generative adversarial nets. *Advances In Neural Information Processing Systems.* **27** (2014)

[2] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B. & Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances In Neural Information Processing Systems.* **30** (2017)

[3] Chazal, F. & Michel, B. An introduction to topological data analysis: fundamental and practical aspects for data scientists. *Frontiers In Artificial Intelligence.* **4** pp. 667963 (2021)

[4] Pope, P., Zhu, C., Abdelkader, A., Goldblum, M. & Goldstein, T. The intrinsic dimension of images and its impact on learning. *ArXiv Preprint ArXiv:2104.08894.* (2021)

Approximation properties of neural ODEs

Arturo De Marinis

GSSI - Gran Sasso Science Institute, L'Aquila, Italy arturo.demarinis[at]gssi.it

Davide Murari[1], Elena Celledoni[2], Nicola Guglielmi[3], Brynjulf Owren[4], Francesco Tudisco[5]

[1]University of Cambridge, Cambridge, United Kingdom dm2011[at]cam.ac.uk

[2]NTNU - Norges teknisk-naturvitenskapelige universitet, Trondheim, Norway elena.celledoni[at]ntnu.no

[3]GSSI - Gran Sasso Science Institute, L'Aquila, Italy nicola.guglielmi[at]gssi.it

[4]NTNU - Norges teknisk-naturvitenskapelige universitet, Trondheim, Norway brynjulf.owren[at]ntnu.no

[5]University of Edinburgh, Edinburgh, United Kingdom f.tudisco[at]ed.ac.uk

We study the universal approximation property (UAP) of shallow neural networks whose activation function is defined as the flow of a neural ODE. We prove the UAP for the space of such shallow neural networks in the space of continuous functions. In particular, we also prove the UAP with the weight matrices constrained to have unit norm.

Furthermore, in [1] we are able to bound from above the Lipschitz constant of the flow of the neural ODE, that tells us how much a perturbation in input is amplified or shrunk in output. If the upper bound is large, then so it may be the Lipschitz constant, leading to the undesirable situation where certain small perturbations in input cause large changes in output. Therefore, in [2] we compute a perturbation to the weight matrix of the neural ODE such that the flow of the perturbed neural ODE has Lipschitz constant bounded from above as we desire. This leads to a stable flow and so to a stable shallow neural network.

However, the stabilized shallow neural network with unit norm weight matrices does not satisfy the universal approximation property anymore. Nevertheless, we are able to prove approximation bounds that tell us how poorly and how accurately a continuous target function can be approximated by the stabilized shallow neural network.

The results presented during this talk are being collected in [3].

References

[1] N. Guglielmi, A. De Marinis, A. Savostianov, and F. Tudisco. Contractivity of neural ODEs: an eigenvalue optimization problem. *arXiv preprint arXiv:2402.13092*, 2024.

[2] A. De Marinis, N. Guglielmi, S. Sicilia, and F. Tudisco. Stability of neural ODEs by a control over the expansivity of their flows. *Work in progress.*

[3] A. De Marinis, D. Murari, E. Celledoni, N. Guglielmi, B. Owren and F. Tudisco, Approximation properties of neural ODEs. *Work in progress.*

Learning Variably Scaled Kernels and Scaling Functions via Discontinuous Neural Networks

Francesco Della Santa
Dipartimento di Scienze Matematiche, Politecnico di Torino

francesco.dellasanta@gmail.com

Gianluca Audone, Emma Perracchione, Sandra Pieraccini
Dipartimento di Scienze Matematiche, Politecnico di Torino

gianluca.audone@polito.it, emma.perracchione@polito.it, sandra.pieraccini@polito.it

This presentation describes a novel methodology to improve the accuracy of interpolation techniques based on Variably Scaled Kernels (VSKs) by learning the scaling function directly from data. The importance of selecting an appropriate scaling function in VSK methods is well-documented, with studies suggesting that the function should mimic the key features of the target, such as its discontinuities [1]. However, theoretical explanations for these observations and practical methods for constructing such scaling functions are almost missing. This work addresses both challenges, offering a theoretical framework alongside a practical, automated solution for learning scaling functions using Discontinuous Neural Networks (δNNs) [2].

The theoretical results illustrated in this presentation justify the intuition that having scaling functions mimicking the behavior of the target function can significantly improve approximation accuracy. These theoretical findings also provide a robust foundation for understanding the role of scaling functions in interpolation. To bridge the gap between theory and application, we propose a novel approach for automatically learning scaling functions through δNNs, which are designed to effectively learn both continuous and discontinuous features of a target function [2]. By leveraging the properties of δNNs, our method constructs scaling functions with characteristics that resemble the target function's ones; this is observed both for continuous and discontinuous target functions.

The presentation includes numerical experiments that validate the theoretical claims, demonstrating the practical efficacy of our approach. These examples involve classical interpolation problems with both continuous and discontinuous target functions. Notably, the results highlight that δNN-based scaling functions enable VSK methods to achieve greater accuracy, particularly in challenging scenarios with discontinuities, outperforming conventional kernel interpolation techniques.

References

[1] S. De Marchi, W. Erb, F. Marchetti, E. Perracchione, M. Rossini *Shape-driven interpolation with discontinuous kernels: Error analysis, edge extraction, and applications in magnetic particle imaging*, SIAM J. Sci. Comput. 42 (2) (2020) B472–B491

[2] F. Della Santa, S. Pieraccini *Discontinuous neural networks and discontinuity learning*, J. Comput. Appl. Math. 419 (2023)

Spectral Complexity of Deep Neural Networks

Simmaco Di Lillo, Domenico Marinucci, Michele Salvi, Stefano Vigogna

RoMaDS - Department of Mathematics, University of Rome Tor Vergata, Rome, Italy

dilillo[at]mat.uniroma2.it

Understanding the spectral properties of neural networks is critical for unveiling their theoretical foundations and practical performance. Fully connected networks with random initialization are known to converge to isotropic Gaussian processes in the infinite-width limit. In this work, we propose a novel approach to characterize network complexity by leveraging the angular power spectrum of these limiting Gaussian fields. Specifically, we define sequences of random variables associated with the angular power spectrum and provide a comprehensive asymptotic characterization of their distribution as network depth grows.

This framework enables a new classification of neural networks into three categories: low-disorder, sparse, and high-disorder. Our analysis reveals distinct behaviors of common activation functions, with particular attention to the sparsity properties of ReLU networks. These theoretical insights are supported by extensive numerical simulations.

A Neural Preconditioner for the Numerical Solutions of Parametrised PDEs

Nunzio Dimola*

MOX, Department of Mathematics, Politecnico di Milano, Italy. nunzio.dimola@polimi.it

Nicola Rares Franco

MOX, Department of Mathematics, Politecnico di Milano, Italy. nicolarares.franco@polimi.it

Paolo Zunino

MOX, Department of Mathematics, Politecnico di Milano, Italy. paolo.zunino@polimi.it

Numerical solution of PDEs is widely recognized as fundamental in scientific and engineering applications; nonetheless, the algebraic structure of the resulting discretized system presents challenges in constructing efficient solution algorithms in many relevant applications. In addition, the computational effort increases as the problem needs to be solved multiple times, to address various instances of the parameter.

In this work, we propose a novel, matrix-free preconditioning strategy that leverages operator learning to efficiently address a class of parametrized 3D-1D mixed-dimensional PDEs [1]. The proposed preconditioner generalizes across varying shapes of the 1D manifold without requiring retraining procedure, making it robust to changes in graph topology. Theoretical ground for *preconditioner learning* approach is established, developing a fully unsupervised training procedure, removing the need for prior problem solution data.

A key contribution is the introduction of a problem-specific data augmentation set, tailored to the spectral properties of the 3D-1D coupling operator's kernel. This enhancement enables the preconditioner to smooth high-frequency error components associated with the coupling term, thereby removing the need for explicit pre- or post-smoothing stages, often required by other approaches (e.g. [1, 2]).

Numerical experiments demonstrate the competitiveness of the proposed approach against established preconditioners, particularly in accelerating convergence in iterative solvers. The preconditioner maintains robust performance across the parameter space without requiring a setup stage, showing scalability for large-scale problems.

This study establishes a foundation for extending machine learning-based preconditioning techniques to broader classes of coupled multi-physics systems, providing a powerful tool for overcoming complex computational challenges in scientific computing.

References

[1] Federica Laurino and Paolo Zunino *Derivation and analysis of coupled PDEs on manifolds with high dimensionality gap arising from topological model reduction.*, ESAIM: Mathematical Modelling and Numerical Analysis, 53(6), 2047-2080.

[2] Yael Azulay and Eran Treister *Multigrid-augmented deep learning preconditioners for the helmholtz equation.*, SIAM Journal on Scientific Computing, 45(3):S127–S151, 2022

[3] Alena Kopani**v**cáková and George Em Karniadakis *Deeponet based preconditioning strategies for solving parametric linear systems of equations.*, arXiv preprint arXiv:2401.02016, 2024.805

Optimizing patient admission in the emergency department with machine learning-based survival models

Davide Duma, Vittorio Meini

Dipartimento di Matematica, Università di Pavia davide.duma@unipv.it,

vittorio.meini01@universitadipavia.it

Roberto Aringhieri

Dipartimento di Informatica, Università degli Studi di Torino roberto.aringhieri@unito.it

In Emergency Department (ED) management, optimizing key performance indicators such as Door-To-Doctor Time (DTDT) and Emergency Department Length of Stay (EDLOS) is crucial for improving efficiency and care quality[1]. DTDT measures the time from patient arrival to doctor consultation, reflecting the speed of initial care. EDLOS tracks the total time a patient spends in the ED, from admission to discharge. Reducing both indicators helps minimize waiting times, prevent overcrowding, and ensure timely care delivery.

In addition, the rate of patients who Leave Without Being Seen (LWBS) is a critical performance metric for EDs. LWBS refers to patients who leave the ED before being seen by a doctor, often due to long wait times or overcrowding. These episodes pose a significant risk to patients, as delays or lack of medical attention can lead to a worsening of their health conditions. In contrast, prior optimization approaches have primarily focused on metrics such as DTDT and EDLOS, overlooking the direct minimization of the LWBS rate.

Despite the widespread use of descriptive analytics to study this issue in the medical literature [2], there have been limited efforts to explore predictive and prescriptive analytics. Previous Machine Learning (ML) studies have proposed classification models to identify categories of patients at risk of abandon (e.g., see [3]), but these models are not suitable to capture their behavior over time. However, to model LWBS as an optimization objective, it is essential to estimate the risk of abandonment as a function of waiting time.

This work introduces a dynamic physician-patient assignment framework based on an Integer Linear Programming (ILP) model informed by predictions about the LWBS risk and solved in a reactive way when the waiting list is updated. Using ML for survival analysis on censored data [4], patient behavior is modeled based on their information collected at the triage. This approach not only supports better decision-making in patient admission processes but also enables the simulation of scenarios with extended DTDT compared to the one observed in the electronic health records.

We consider a real case study of a medium-size Italian ED. A computational analysis evaluates the proposed framework's performance, analyzing the trade-off between DTDT, EDLOS, and LWBS rates, by considering both the aleatory and the epistemic uncertainty. Results demonstrate the potential of the proposed interplay beetween ML, ILP, and simulation for improving ED admissions.

References

[1] D Duma, R Aringhieri. *Real-time resource allocation in the emergency department: A case study*, Omega 117, 102844, 2023.

[2] M Johnson, S Myers, J Wineholt, M Pollack, AL Kusmiesz. *Patients Who Leave the Emergency Department Without Being Seen*, Journal of Emergency Nursing 35(2), 105-108, 2009.

[3] M Sheraton, C Gooch, R Kashyap. *Patients leaving without being seen from the emergency department: A prediction model using machine learning on a nationwide database*, Journal of the American College of Emergency Physicians Open 1(6), 1684-1690, 2020.

[4] P Wang, Y Li, CK Reddy. *Machine Learning for Survival Analysis: A Survey*, ACM Computing Surveys 15(6): 1-36, 2019.

Low-rank approximation methods for real data analysis and integration

Flavia Esposito

Dipartimento di Matematica, Università degli Studi di Bari Aldo Moro flavia.esposito@uniba.it

Over the years, low-rank approximation models have gained significant attention due to their effectiveness in analyzing real data.

The key idea is that real data has a structured form (such as vectors, matrices, or tensors) and admits a low-rank representation. A data matrix $X \in \mathbb{R}^{n\times m}$, with n samples and m features, can be represented as a product of two factors $W \in \mathbb{R}^{n\times r}$ and $H \in \mathbb{R}^{r\times m}$, with $r < \min(m, n)$, such that $X \approx WH$.

The problem of finding such a pair (W, H) can be mathematically formulated as a penalized optimization task:

$$\min_{W,H\in\mathcal{C}} Div(X, WH) + \mu_1 J_1(W) + \mu_2 J_2(H) + \mu_3 J_3(W, H) \tag{1}$$

where $Div(\cdot,\cdot)$ is a divergence function that evaluates the quality of the approximation, $\mathcal{C}$ is a feasible set that encodes structural or physical information about the data, J_i $(i = 1, 2, 3)$are the penalty functions that enforce additional properties on W and H, and μ_i are the penalty hyperparameters, balancing the bias-variance trade off in approximating X and satisfying factor properties.

In this talk, we review some theoretical and computational issues related to specific low-rank approximation models and numerical methods defined on the set $\mathcal{C}$ of nonnegative matrices. We address several mathematical challenges, including the selection of an appropriate divergence function tailored to the specific data domain, and the proper definition of J_i to integrate domain-specific prior knowledge. We also emphasize real-world applications, particularly in the biomedical and environmental fields. Morover, we also investigate how additional constraints encoded by the peculiar form of J_i can be advantageously handled using manifold optimization techniques.

References

[1] Gillis, N. Nonnegative matrix factorization. (SIAM, 2020)

[2] Boumal, N. An introduction to optimization on smooth manifolds. (Cambridge University Press, 2023)

A random matrix approach to Hopfield-like neural networks: addressing generalization and overfitting

Alberto Fachechi

Department of Mathematics, Sapienza University of Rome alberto.fachechi@uniroma1.it

Theoretical investigations in modern Artificial Intelligence focus on developing robust mathematical frameworks for understanding information processing in learning neural networks. A key challenge lies in characterizing the ability of these systems to extract hidden features from empirical data and utilize them for effective generalization. In this context, spin glass models with structural disorder (to use the words by Marc Mézard [1]) serve as essential toy models for exploring the functional regimes of artificial neural networks. In this talk, I will present recent findings on example-based Hopfield-like models, which provide an ideal theoretical framework for this purpose and naturally emerge from a statistical inference perspective. Specifically, we exploit [2, 3] the properties of random Hebbian-like coupling matrices in these models to explore the attractor landscape, thereby providing insights into the mechanisms underlying generalization and overfitting. Furthermore, we extend the Marchenko-Pastur theorem to this class of random matrices, using the resulting distribution to estimate crucial model characteristics, including the attractive capacity of hidden ground truths.

References

[1] M. Mézard, Spin glass theory and its new challenge: structured disorder. Indian J Phys 98, 3757–3768 (2024).

[2] E. Agliari et al., "Regularization, early-stopping and dreaming: a Hopfield-like setup to address generalization and overfitting." Neural Networks 177 (2024): 106389.

[3] E. Agliari, A. Fachechi, D. Luongo. "A spectral approach to Hebbian-like neural networks." Applied Mathematics and Computation 474 (2024): 128689.

On latent dynamics learning in nonlinear reduced order modeling

Nicola Farenga

MOX, Department of Mathematics, Politecnico di Milano, Milan, Italy

nicola.farenga@polimi.it

Stefania Fresca, Simone Brivio, Andrea Manzoni

MOX, Department of Mathematics, Politecnico di Milano, Milan, Italy

{stefania.fresca, simone.brivio, andrea1.manzoni} @polimi.it

In this work, we present the novel mathematical framework of *latent dynamics models* (LDMs) for reduced order modeling of parameterized nonlinear time-dependent PDEs. Our framework casts this latter task as a nonlinear dimensionality reduction problem, while constraining the latent state to evolve accordingly to an (unknown) dynamical system, namely a latent vector ordinary differential equation (ODE). A time-continuous setting is employed to derive error and stability estimates for the LDM approximation of the full order model (FOM) solution. We analyze the impact of using an explicit Runge-Kutta scheme in the time-discrete setting, resulting in the ΔLDM formulation, and further explore the learnable setting, ΔLDM_θ, where deep neural networks approximate the discrete LDM components, while providing a bounded approximation error with respect to the FOM. Moreover, we extend the concept of parameterized Neural ODE – recently proposed as a possible way to build data-driven dynamical systems with varying input parameters – to be a convolutional architecture, where the input parameters information is injected by means of an affine modulation mechanism, while designing a convolutional autoencoder neural network able to retain spatial-coherence, thus enhancing interpretability at the latent level. Numerical experiments, including the Burgers' and the advection-reaction-diffusion equations, demonstrate the framework's ability to obtain, in a multi-query context, a *time-continuous* approximation of the FOM solution, thus being able to query the LDM approximation at any given time instance while retaining a prescribed level of accuracy. Our findings highlight the remarkable potential of the proposed LDMs, representing a mathematically rigorous framework to enhance the accuracy and approximation capabilities of reduced order modeling for time-dependent parameterized PDEs.

References

[1] N. Farenga, S. Fresca, S. Brivio, A. Manzoni *On latent dynamics learning in nonlinear reduced order modeling*, https://arxiv.org/abs/2408.15183

Mathematical Transformations and Deep Learning Methodologies to enhance Tool Wear Monitoring using Audio Data

Stefania Ferrisi*, Rosita Guido, Giuseppina Ambrogio

Department of Mechanical, Energetic and Management Engineering, University of Calabria, Ponte P. Bucci, Rende, 87036, CS, Italy

*stefania.ferrisi@unical.it

The integration of deep learning methodologies with Internet of Things (IoT) sensor systems offers significant potential for real-time monitoring of tool conditions in milling processes. Tool condition monitoring systems provide critical insights into tool wear, allowing for timely replacement decisions, minimizing machine downtime, and preserving the quality of machined surfaces. These advancements contribute to the sustainability of manufacturing operations by reducing waste and optimizing resource utilization. Among various IoT sensors, microphones that capture audio signals during machining have emerged as a cost-effective and non-invasive approach.
This study investigates the use of mathematical transformations for audio signals to enhance the predictive accuracy of tool wear monitoring. This study examines two primary methods for processing audio data: numerical feature extraction and audio conversion into spectrograms using the Fast Fourier Transform (FFT). By decomposing complex audio waveforms into their frequency components, the FFT retains essential information that characterizes the progression of tool wear. The generated spectrograms, represented as high-resolution images, provide a detailed depiction of frequency and amplitude variations over time. When analyzed using convolutional neural networks, these spectrograms enable accurate classification of tool wear stages and estimation of the remaining useful life of cutting tools. This methodology highlights the effectiveness of combining rigorous mathematical signal processing techniques with artificial intelligence to address challenges in predictive maintenance.
The findings emphasize the potential of this approach to develop robust and scalable systems for real-time tool monitoring, aligning with the principles of modern manufacturing to improve efficiency, reduce operational costs, and support sustainable practices.

References

[1] Ferrisi Stefania, Ambrogio Giuseppina, Guido Rosita and Umbrello Domenico. *Artificial Intelligence techniques and Internet of things sensors for tool condition monitoring in milling: A review* , Materials Research Proceedings, Vol. 41, pp 2000-2010, 2024

[2] Stefania Ferrisi, Gabriele Zangara, David Rodríguez Izquierdo, Danilo Lofaro, Rosita Guido, Domenico Conforti and Giuseppina Ambrogio. *Tool Condition Monitoring for milling process using Convolutional Neural Networks* , Procedia Computer Science, Vol. 232, pp 1607-1616, 2024

Deep orthogonal decomposition: an adaptive basis approach to dimensionality reduction

Nicola Rares Franco[1]

Andrea Manzoni[1], Paolo Zunino[1], Jan S. Hesthaven[2]

[1]MOX, Department of Mathematics, Politecnico di Milano, 20133, Milan, Italy

[2]CMSS, École Polytechnique Fédérale de Lausanne, Station 8, 1015, Lausanne, Switzerland

nicolarares.franco@polimi.it

Linear dimensionality reduction methods like Principal Component Analysis (PCA) and Singular Value Decomposition (SVD) are ubiquitous in statistics, machine learning, and numerical analysis. Recently, several researchers have developed adaptive variants of these methods to address the challenge of integrating external sources of information —such as, e.g., contextual information or parameter dependency— within the dimensionality reduction process. We refer to these methods as to «algorithms for parameter-dependent low-rank approximation». Such approaches enable enhanced interpretability in statistical applications, such as extracting key patterns in data (e.g., ECG signals, images, or audio) conditioned on covariates like age or time, and improved performance in numerical applications, such as reduced-order modeling of PDEs with slowly decaying Kolmogorov n-widths.

Starting from here, we present a unified theoretical framework for parametric low-rank approximations and propose *Deep Orthogonal Decomposition* (DOD) as a novel approach for dimensionality reduction in the context of reduced-order modeling of parameterized PDEs. DOD utilizes deep neural networks to construct an adaptive local bases that can capture the structure of the solution manifold in a dynamical manner. By combining linear and nonlinear elements, DOD overcomes the limitations of global methods, such as POD and deep autoencoders, providing both interpretability and precise error control. We validate the effectiveness of the DOD through numerical experiments based on the Navier-Stokes and Eikonal equations, demonstrating its capability to address challenging scenarios, including nonlinear PDEs, intricate geometries, and large parameter spaces. In doing so, we also explore certain connections between the DOD and the *Grassmann manifold*, thanks to which we are able to develop specific diagnostic tools that can facilitate practical implementation and analysis.

Finally, we come back to the general framework, with the purpose of deepening our understanding through a more abstract mathematical analysis. Specifically, we shall present some novel theoretical results that show how the efficacy of parametric low-rank approximation algorithms —such as the DOD— relates to certain regularity properties, which, in turn, depend on how the eigenvalues of the covariance operator change with the problem parameters. In particular, branching phenomena (crossing of the eigenvalues) can significantly impact model performance and needs to be accounted for when designing and implementing these approaches.

References

[1] Franco, N. R., Manzoni, A., Zunino, P., Hesthaven, J. S. (2024). *Deep orthogonal decomposition: a continuously adaptive data-driven approach to model order reduction*, arXiv preprint arXiv:2404.18841

[2] Franco, N. R. (2024). *Measurability and continuity of parametric low-rank approximation in Hilbert spaces: linear operators and random variables*, arXiv preprint arXiv:2409.09102

[3] Gupta, A., Barbu, A. (2018). *Parameterized principal component analysis*, Pattern Recognition, 78, 215-227.

[4] Amsallem, D., Farhat, C. (2011). *An online method for interpolating linear parametric reduced-order models*, SIAM Journal on Scientific Computing, 33(5), 2169-2198.

Penalized Maximum Likelihood and Loss Minimization for Classification

Bharath Krishnan Girishkumar

PhD student, MIDA group, Department of Mathematics, University of Genova

bharathkrishnan.girishkumar@edu.unige.it

Federico Benvenuto

Associate Professor, MIDA group, Department of Mathematics, University of Genova

federico.benvenuto@unige.it

This talk explores the parallelism between empirical loss minimization and binary classification as a maximum likelihood problem with data drawn from a Bernoulli distribution. We demonstrate that empirical loss minimization corresponds to penalized maximum likelihood estimation, where the penalty depends on the specific loss function. Furthermore, we establish a one-to-one correspondence between solutions of different loss functions via generalized linear model link functions. Remarkably, the resulting binary classifiers remain identical across the considered loss functions. We also show that the classification problem can be solved numerically using linear equations. However, due to potential ill conditioning in the case of square systems, iterative algorithms are often more effective. Finally, we extend these concepts to multiclass classification and present supporting numerical experiments.

Learning Passive Left Ventricular Mechanics via Shape Encoding Neural Networks

Marc Hirschvogel[a1], Davide Carrara[a], Stefano Pagani[a], Simone Pezzuto[b], Francesco Regazzoni[a]

[a] MOX, Dipartimento di Matematica, Politecnico di Milano, Milan, Italy

[b] Dipartimento di Matematica, Università di Trento, Trento, Italy

We present a novel scientific machine learning approach to predict the solution of partial differential equations on unseen domains. The methodology consists of a two-step procedure: First, the *DeepSDF* [1] neural network architecture is used to learn a signed-distance function (SDF) that is representative of the object's shape. Second, a fully connected neural network is trained with PDE solutions on different geometries, leveraging a latent vector that encodes shape information from the prior SDF training step [2]. The approach, in general, only requires a point cloud representation of the geometry, hence neither meshes nor any type of point-to-point correspondence between domains is needed. We test our approach for inferring anisotropic passive mechanics on left ventricular patient-specific and synthetically generated geometries, investigating alternative shape encoding via principal component analysis or input feature enhancement by universal ventricular coordinates. Our results highlight the potential of shape codes for surrogating nonlinear PDEs on a diverse cohort of ventricles and pave the way for real-time predictions of multi-physics phenomena such as cardiac electromechanics on complex geometries.

The present research has been supported by the project PRIN2022, MUR, founded by the European Union (grant P2022N5ZNP).

References

[1] Park, Jeong Joon and Florence, Peter and Straub, Julian and Newcombe, Richard and Lovegrove, Steven *DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation*, 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019

[2] Regazzoni, Francesco and Pagani, Stefano and Quarteroni, Alfio *Universal Solution Manifold Networks (USM-Nets): Non-Intrusive Mesh-Free Surrogate Models for Problems in Variable Domains*, Journal of Biomechanical Engineering, 144(12), 2022

[1] marc.hirschvogel@polimi.it

Data-driven parameterization for adaptive spline model reconstruction

Sofia Imperatore

IMATI-CNR "Enrico Magenes", via Ferrata 5/A, 27100, Pavia, Italy sofia.imperatore@imati.cnr.it

Carlotta Giannelli, Angelos Mantzaflaris, Dominik Mokri**v**s, Felix Scholz

Dipartimento di Matematica e Informatica "Ulisse Dini", Università degli Studi di Firenze, Viale Morgani 67/A, Firenze, 50137, Italy carlotta.giannelli@unifi.it

Inria Centre at Université Côte d'Azur, 2004 Route des Lucioles, 06902 Sophia Antipolis, France angelos.mantzaflaris@inria.fr

MTU Aero Engines AG, Dachauer Strasse 665, Munich, 80995, Germany dominik.mokris@mtu.de, felix.scholz@mtu.de

In this talk, we combine Computer Aided Geometric Design (CAGD) methods with Deep Learning (DL) technologies. The final objective is the (re-)construction of highly accurate CAD models for the design of complex data-driven free-form adaptive spline geometries. In particular, we present two novel geometric deep learning techniques for parameterizing scattered point clouds in $\mathbb{R}^3$ on a planar parametric domain, by exploiting (graph) convolutional neural networks. Firstly, we introduce a data-driven parameterization model that builds upon existing meshless parameterization schemes and predicts the parametric values of the input point cloud from the proximity information of its 3D items and its dual line graph [1]. Secondly, we present an alternative learning model, that avoids line-graph computation, characterized by a new boundary informed message-passing input layer, that takes in input boundary conditions and propagates them into the new features of the interior points [2]. Finally, we show the effectiveness of these learning models for surface fitting with adaptive spline constructions and moving parameterization, thus merging CAGD methods with DL technologies [3].

References

[1] Giannelli, C., Imperatore, S., Mantzaflaris, A., and Scholz, F. *Learning meshless parameterization with graph convolutional neural networks*, In International conference on WorldS4 (pp. 375–387). Singapore: Springer Nature Singapore, 2023.

[2] Giannelli, C., Imperatore, S., Mantzaflaris, A., and Scholz, F. *BIDGCN: boundary-informed dynamic graph convolutional network for adaptive spline fitting of scattered data*, Neural Computing and Applications, 1–24, 2024

[3] Giannelli, C., Imperatore, S., Mantzaflaris, A., and Mokri**v**s, D. *Leveraging moving parameterization and adaptive THB-splines for CAD surface reconstruction of aircraft engine components* In Smart Tools and Applications in Graphics-Eurographics Italian Chapter Conference. The Eurographics Association, 2023

A new mathematical model to analyze the spread of misinformation on Social Media

Samira Iscaro, Dajana Conte, Giovanni Pagano and Beatrice Paternoster

Dept. of Mathematics, University of Salerno, Fisciano (SA), Italy

siscaro@unisa.it, dajconte@unisa.it, gpagano@unisa.it, beapat@unisa.it

In recent years, research activity on the mathematical analysis of evolutionary problems has been focused on a fundamental problem of our society: the spread out of fake information on Social Media. Since the outcomes of many political or social events have been influenced by fake news [5], researches have developed different approaches to describe and analyze the aformentioned phenomenon, from the use of machine learning based detectors for fake information to the use of mathematical epidemiological models to make an attempt of prediction of the evolution of news spread through time [1, 4]. In particular, in [2], it has been shown that starting from a dataset composed of real data extracted from X (Twitter) and using an epidemiological model of SIR type, called the Ignorant-Spreader-Recovered model, it is possible to compute optimized parameters to make predictions on the spread of a certain news in terms of both the total number of individuals who share news and the moment of the peak of interest towards it. The main aim of this talk is to focus on a new mathematical model to analyze the spread of fake news on Social Media. More specifically, we will focus on a class of mathematical models called the Ignorant-Spreader-Counter Spreader-Recovered (ISCR), initially proposed in [6], and introduced to accurately describe the scenario in which there is a debate regarding a certain topic. In general, this class of models is suited to describe the situation in which two different groups of spreders are present: the one of the individuals that share the fake news (the Spreader class) and the one of individuals who try to restore the truth (the Counter Spreader class). Mathematical methods to numerically solve this kind of problem will also be introduced. In fact, we will show that using an ISCR model it is possible to predict the evolution of certain fake news, spread on X in recent years. However, in order to reduce the computational cost of the fitting phase as well as to preserve the positivity of the model, starting from strategies described in [3], we will introduce a Nonstandard Finite Difference (NSFD) scheme for the considered model, showing that it is truly applicable to reality to make predictions, as confirmed by several numerical experiments.

References

[1] Cardone, A., Diaz de Alba, P., Paternoster, B. *Analytical Properties and Numerical Preservation of an Age-Group Susceptible-Infected-Recovered Model: Application to the Diffusion of Information.* J. of Comput.l and Nonl.r Dyn. 19.6. https://doi.org/10.1115/1.4065437 (2024).

[2] Castiello, M., Conte, D., Iscaro, S. *Using Epidemiological Models to Predict the Spread of Information on Twitter.* Algorithms. 16, 391. https://doi.org/10.3390/a16080391 (2023).

[3] Conte, D., Guarino, N., Pagano, G., Paternoster, B. *Positivity-preserving and elementary stable nonstandard method for a COVID-19 SIR model.* Dolomites Research Notes on Approximation, 15(DRNA Volume 15.5), 65-77. (2022).

[4] D'Ambrosio, R., Giordano, G., Mottola, S., Paternoster, B. *Stiffness analysis to predict the spread out of fake information.* Future Internet, 13(9), 222. (2021).

[5] Maleki, M., Mead, E., Arani, M., Agarwal, N. *Using an epidemiological model to study the spread of misinformation during the Black Lives Matter Movement.* arXiv. https://arxiv.org/abs/2103.12191 (2021).

[6] Muhlmeyer, M., Agarwal, S. *Information spread in a social media age. In Modelling and Control.* CRC Press: Boca Raton, FL, USA; Taylor and Francis Group: Boca Raton, FL, USA; London, UK; New York, NY, USA. (2021).

Constructing Interpretable Prediction Models with Semi-Orthogonal 1D DNNs: An Example in Irregular ECG Classification

Giacomo Lancia

Università di Roma "La Sapienza"; Dipartimento di Scienze di Base Applicate all'Ingegneria (SBAI);

Via Antonio Scarpa, 16; Roma giacomo.lancia@uniroma1.it

Cristian Spitoni

Universiteit Utrecht; Mathematics Department; Budapestlaan, 6; Utrecht c.spitoni@uu.nl

When considering deep learning (DL)–based applications in medicine, ensuring interpretability is crucial for assessing the quality and safety of clinical predictions. At the same time, the increasing availability of massive, complex, high-dimensional data has heightened the demand for computational models capable of extracting salient features effectively and accurately.

Artificial Neural Network (ANN) based methods have emerged as powerful tools for achieving highly accurate predictions across diverse medical domains , often outperforming traditional approaches. However, the inherent "black-box" nature of ANN models limits their interpretability, making it challenging to establish causal relationships between input covariates and predictions. As such, integrating ANN-based methods within interpretable frameworks is essential for ensuring model transparency and clinical reliability.

To address this challenge, we propose a novel methodology that enforces a mathematical structure within the neural network architecture. Specifically, we incorporate Semi-Orthogonal constraints on convolutional kernels to ensure the invertibility of the learned features, thereby enhancing the interpretability of the convolutional neural network's (CNN) actions. This constraint enables a more transparent understanding of the transformations applied to the data, bridging the gap between predictive accuracy and explainability.

Saliency Map (SM)-based methods and, more broadly, eXplainable Artificial Intelligence (XAI) algorithms have been proposed to visualize and interpret the areas of input data where an ANN focuses its attention during decision-making.These methods aim to uncover the connections between outputs and inputs by analyzing the network's propagation rules. While SM-based approaches, such as Vanilla Gradient , CAM , and Layer-Wise Relevance Propagation , have demonstrated utility, they often fail to fully reveal causal relationships between inputs and predictions. Instead, they primarily highlight regions of interest, which may lead to misleading conclusions in clinical settings.

Building upon these foundations, our methodology leverages Semi-Orthogonal convolutional weights to enhance the interpretability of 1-D Deep Neural Networks (1D DNN). Our approach allows for the efficient reconstruction of 1D CNN inputs, providing deeper insights into how features are extracted and contributing to more interpretable feature representations. These extracted features are then incorporated into a Logistic Regression (LR) model, a simple yet interpretable framework, to classify irregular ECG patterns.

The interpretability of this framework is further evaluated through feature extraction and alignment with clinical knowledge. The proposed methodology ensures that the predictions not only achieve high accuracy but also maintain a transparent decision-making process, crucial for identifying conditions such as Atrial Fibrillation (AF), Myocardial Infarction (MI), and Sinus Bradycardia (SBR).

Through this mathematically grounded approach, our research bridges the gap between advanced deep learning techniques and the need for interpretable, clinically reliable predictions in medical applications. We demonstrate that Semi-Orthogonal constraints significantly improve the explainability of 1D CNN-based predictions, setting a precedent for developing robust and interpretable diagnostic models. A pre-print of this work is available at https://arxiv.org/abs/2410.12059.

Hexagonal Grid-Based Reinforcement Learning Environments for Marine Biodiversity Monitoring

Giulia Lombardi

University of Trento, Via Sommarive 14, I-38123 Povo (Trento) giulia.lombardi@unitn.it

Monitoring marine biodiversity presents profound scientific and logistical challenges, necessitating the integration of advanced mathematical frameworks, computational methodologies, and interdisciplinary expertise. These challenges are further compounded by the vast spatial scales and ecological complexity of marine environments, necessitating a rigorous and multifaceted approach to data acquisition and environmental monitoring [1].

Contemporary research leverages an array of technological modalities, including satellite imagery, remote sensing, autonomous underwater vehicles (AUVs), remotely operated vehicles (ROVs), environmental DNA (eDNA) sampling, and traditional vessel-based expeditions [2, 3]. While these methodologies yield valuable insights, they are frequently constrained by logistical limitations, financial burdens, and operational inflexibility. Moreover, the intensifying impacts of climate change—manifested in dynamic ocean conditions and shifting species distributions—exacerbate the demand for continuous, high-resolution monitoring systems.

To address these pressing challenges, we propose a novel framework for local data collection and monitoring, specifically designed to optimize AUV navigation. Central to our approach is the development of a reinforcement learning (RL) hexagonal grid-world environment which enables the agent to navigate under conditions of partial observability. This model exploits the geometric properties of hexagonal tiling — notably uniform neighbor connectivity and minimized edge effects — to enhance spatial coverage and enable realistic navigation in complex marine settings. This design outperforms traditional Cartesian grid-based systems, which are susceptible to inefficiencies in irregular or dynamic environments. In fact, similar methodologies have demonstrated success in terrestrial [5] and space-based [5] applications, yet remain underexplored in ecological modeling and monitoring [6].

The proposed approach addresses diverse ecological and operational needs by allowing the customization of multi-objective reward functions to suit specific monitoring tasks. This adaptability allows researchers to leverage the same framework to pursue diverse objectives, such as accurately identifying biodiversity hotspots, monitoring rare or endangered species, and detecting plastic debris, ultimately offering a powerful AI-driven tool for advancing marine conservation and environmental research.

References

[1] N. Bax et al. *Seascape ecology: identifying research priorities for an emerging ocean sustainability science*, Marine Ecology Progress Series, 663, 1–29, 2019.

[2] A. Miller, J. I. Virmani. *Advanced marine technologies for ocean research*, Deep Sea Research Part II: Topical Studies in Oceanography, Volume 212, 105340, 2023, ISSN 0967-0645.

[3] A. Miller, J. I. Virmani. *Advanced marine technologies for ocean research*, Deep Sea Research Part II: Topical Studies in Oceanography, Volume 212, 105340, 2023, ISSN 0967-0645.

[4] Uber Technologies Inc. *Hexagonal Grids in Urban Mobility Optimization: Applications to Uber*, Available at: https://www.uber.com/en-GB/blog/h3/.

[5] W. Wang, H. Zhou, S. Zheng, G. L"u, and L. Zhou. *Ocean surface currents estimated from satellite remote sensing data based on a global hexagonal grid*, International Journal of Digital Earth, 16:1, 1073–1093, 2023.

[6] C.P.D. Birch, S.P. Oom, J.A. Beecham. *Rectangular and hexagonal grids used for observation, experiment and simulation in ecology*, Ecological Modelling, 206(2007), 347–359.

Multi-fidelity reduced-order surrogate modelling

Andrea Manzoni, Paolo Conti
MOX – Department of Mathematics, Politecnico di Milano, Italy
andrea1.manzoni@polimi.it, paolo.conti@polimi.it

Mengwu Guo
Centre for Mathematical Sciences, Lund University, Sweden
mengwu.guo@math.lu.se

Attilio Frangi
Department of Civil and Environmental Engineering, Politecnico di Milano
attilio.frangi@polimi.it

Steven L. Brunton, J. Nathan Kutz
University of Washington, Seattle, United States
sbrunton@uw.edu, kutz@uw.edu

High-fidelity numerical simulations of partial differential equations (PDEs) given a restricted computational budget can significantly limit the number of parameter configurations considered and/or time window evaluated. Multi-fidelity surrogate modeling aims to leverage less accurate, lower-fidelity models that are computationally inexpensive in order to enhance predictive accuracy when high-fidelity data are scarce [4]. However, low-fidelity models, while often displaying the qualitative solution behavior, fail to accurately capture fine spatio-temporal and dynamic features of high-fidelity models.

To address this shortcoming, we present a data-driven strategy that combines dimensionality reduction with multi- fidelity neural network surrogates [1]. The key idea is to generate a spatial basis by applying proper orthogonal decomposition (POD) to high-fidelity solution snapshots, and approximate the dynamics of the reduced states – time-parameter-dependent expansion coefficients of the POD basis – using a multi-fidelity long short-term memory network [2, 3]. By mapping low-fidelity reduced states to their high-fidelity counterpart, the proposed reduced-order surrogate model enables the efficient recovery of full solution fields over time and parameter variations in a non-intrusive manner. A further extension to the case of multiple data sources, with low-fidelity models of different type, is also considered, in the spirit of progressive learning from multiple sources.

The generality of the proposed approach is demonstrated by a collection of PDE problems where the low-fidelity model can be defined by coarser meshes and/or time stepping, as well as by misspecified physical features.

References

[1] P. Conti, M. Guo, A. Manzoni, A. Frangi, S. L. Brunton, and J. Nathan Kutz. Multi-fidelity reduced-order surrogate modelling. *Proceedings of the Royal Society A*, 480(2283):20230655, 2024.

[2] P. Conti, M. Guo, A. Manzoni, and J. S. Hesthaven. Multi-fidelity surrogate modeling using long short-term memory networks. *Computer methods in applied mechanics and engineering*, 404:115811, 2023.

[3] M. Guo, A. Manzoni, M. Amendt, P. Conti, and J. S. Hesthaven. Multi-fidelity regression using artificial neural networks: Efficient approximation of parameter-dependent output quantities. *Computer methods in applied mechanics and engineering*, 389:114378, 2022.

[4] M. Torzoni, A. Manzoni, and S. Mariani. A multi-fidelity surrogate model for structural health monitoring exploiting model order reduction and artificial neural networks. *Mechanical Systems and Signal Processing*, 197:110376, 2023.

Convergence of quantum neural networks at infinite width

Anderson Melchor Hernandez

Piazza di Porta S. Donato, 5, Bologna, BO anderson.melchor@unibo.it

Filippo Girardi

Piazza dei Cavalieri, 7, Pisa, PI filippo.girardi@sns.it

Giacomo De Palma

Piazza di Porta S. Donato, 5, Bologna, BO giacomo.depalma@unibo.it

Davide Pastorello

Piazza di Porta S. Donato, 5, Bologna, BO davide.pastorello3@unibo.it

Quantum neural networks constitute the quantum version of deep neural models. These new models are based on quantum circuits and generate functions given by the expectation values of a quantum observable measured on the output of a quantum circuit made by parametric one-qubit and two-qubit gates [4]. The parameters of the circuit encode both the input data and the parameters of the model itself. These parameters are typically optimized by gradient descent, which involves iterative adjustment to minimize a cost function and improve the performance of the quantum circuit in the processing and analysis of data [1]. Significant progress has been made in addressing the question of whether training can perfectly fit the training examples while simultaneously avoiding overfitting. A fundamental breakthrough has been the proof that, in the limit of infinite width, the probability distribution of the function generated by a deep neural network trained on a supervised learning problem converges to a Gaussian process [2]. This recent has inspired renewed interest in quantum machine learning, raising the question of whether quantum neural networks exhibit analogous properties. In this presentation, I will explore some of the recent advancements in this area, highlighting key insights and findings [3].

References

[1] F. Girardi, G. De Palma, *Trained quantum neural networks are Gaussian processes,* arXiv:2402.08726 (2024).

[2] H. Boris, *Which neural net architectures give rise to exploding and vanishing gradients?* J. Adv Neural Inf Process Syst 31 (2018).

[3] A. Melchor Hernandez, F. Girardi, G. De Palma, D. Pastorello, *Quantitative convergence of trained quantum neural networks to a Gaussian process,* Preprint.

[4] M. Schuld, I. Sinayskiy, F. Petruccione, *An introduction to quantum machine learning,* J. Contemporary Physics 56 (2015) no. 2.

An all-around perspective on hybrid coupled models and parameter calibration for collective cell dynamics

Marta Menci

Università Campus Bio-Medico di Roma, m.menci@unicampus.it

The study of collective dynamics has garnered significant interest across various scientific domains due to its potential to model self-organization in complex systems and its wide range of applications. In the biological and biomedical world, an increasing number of phenomena benefits from the mathematical and numerical approach, aiming at in-silico models to investigate the phenomena of interest. In this field, collective cell dynamics play a critical role in several biological processes characterizing the human body. The main feature of those kind of collective behaviors, that need to be taken into account in the mathematical models, is that cells not only interact mechanically, but are also driven by chemical signals which lead cells moving towards higher concentrations of chemicals. In real applications, parameter estimation can be exceptionally challenging due to the large number of parameters that need to be simultaneously estimated and the costs of performing experiments to collect experimental data. To this end, machine learning algorithms are currently investigated, allowing for faster and robust optimization procedures for solving inverse problems associated with parameter estimation.

The talk will explore a recent class of multiscale hybrid coupled models to simulate migrations of cells in different scenarios [1, 2]. Originally conceived to model embriogenesis processes, the particular structure combine discrete cellular dynamics with continuous chemical signaling, offering a multiscale framework to describe the complex interactions between cells and their environment.

Although hybrid models provide an accurate description of cell behaviors, they can be computationally expensive, especially when dealing with large numbers of cells in higher-dimensional settings. To address this challenge, a macroscopic pressureless Euler-type model with nonlocal chemotaxis has been rigorously derived from the microscopic scale, describing cellular dynamics in term of evolution of a cell density, hence on a macroscopic scale [3, 4].

Numerical simulations of the considered models at different scales will be presented, including 2D and 3D scenarios. In particular, the hybrid coupled model is validated against experimental data (positions and velocities of cells acquired at different times during the experiments), whereas the macroscopic-derived version makes use of synthetic data generated from original microscopic real-data.

This work is based on ongoing collaborations with Roberto Natalini (Instituto per le Applicazioni del Calcolo - CNR), Thierry Paul (LYSM - CNRS) and Tommaso Tenna (Université Côte D'azur).

References

[1] E. Di Costanzo, M. Menci, E. Messina, R. Natalini and A. Vecchio *A hybrid model of collective motion of discrete particles under alignment and continuum chemotaxis* , Discrete & Continuous Dynamical Systems-B, 25(1), 2020.

[2] G. Bretti, E. Campanile, M. Menci, R. Natalini *A scenario-based study on hybrid PDE-ODE model for Cancer-on-chip experiment* , In: Problems in Mathematical Biophysics: A Volume in Memory of Alberto Gandolfi. Cham: Springer Nature Switzerland, 2024.

[3] R. Natalini and T. Paul. *The mean-field limit for hybrid models of collective motions with chemotaxis*, SIAM Journal on Mathematical Analysis, 55(2), 2023.

[4] M. Menci, R. Natalini, T. Paul. *Microscopic, kinetic and hydrodynamic models of collective motions with chemotaxis: a numerical study* , Mathematics and Mechanics of Complex Systems, 12(1), 2024.

Step-by-Step Time-Discrete Physics Informed Neural Networks for PDEs models

Giovanni Pagano

Department of Mathematics, University of Salerno, Italy, gpagano@unisa.it

[1]C. Valentino, [2]D. Conte, [2]B. Paternoster, [1]F. Colace, [3]M. Casillo

[1] Department of Industrial Engineering, University of Salerno, Italy, {cvalentino,fcolace}@unisa.it

[2]Department of Mathematics, University of Salerno, Italy, {dajconte,beapat}@unisa.it

[3]Department of Cultural Heritage Sciences, University of Salerno, Italy, mcasillo@unisa.it

Models based on Partial Differential Equations (PDEs) originate from different phenomena, such as: life cycle of batteries [1], evolution of vegetation [3], corrosion of materials [5], production of renewable energy [1]. For the related numerical solution, in addition to standard well-known methods, several techniques based on Artificial Neural Networks (ANNs) have recently been proposed, see e.g. [4]. In this context, the so-called Physics-Informed Neural Networks (PINNs) are considered, i.e. ANNs generally constructed in such a way as to compute a time-continuous and space-continuous approximation of the exact solution of the analyzed PDE.

This talk focuses on the derivation of a new approach based on PINNs, namely Time-Discrete PINNs, for the solution of PDEs. They are called this way since provide a solution which is continuous in space and discrete in time. Existing Time-Discrete PINNs from the literature are based on the immersion of classical Runge-Kutta (RK) methods within ANNs. That is, given every point of the spatial domain, the neural network is constructed in such a way as to furnish, as output, approximations of the stages of the selected RK method at a fixed time step.

Here, we propose new Step-by-Step (SBS) Time-Discrete PINNs, based on the implicit Euler and Crank-Nicolson methods [1]. We construct these PINNs in such a way as to obtain, as output, an approximation of the solution by the above-mentioned methods at each time step (unlike RK-based PINNs). Furthermore, we establish connections between the existing RK-based and the new SBS PINNs, which allow to use the same workflow for both in implementation. Several numerical experiments, conducted on PDEs models related to sustainability [1] and life cycle of batteries [2], show the advantages of the new SBS PINN over the RK-based ones, and also over classical continuous-time and continuous-space PINNs.

Acknowledgements: this work has been supported by the PRIN PNRR 2022 project P20228C2PP "BAT-MEN".

References

[1] M. Frittelli, B. Bozzini, I. Sgura. *Turing patterns in a 3D morpho-chemical bulk-surface reaction-diffusion system for battery modeling.* MinE (Mathematics in Engineering) 6(2), 363-393 (2024).

[2] D. Conte, G. Pagano, B. Paternoster. *Nonstandard finite differences numerical methods for a vegetation reaction–diffusion model.* J. Comput. Appl. Math. 419 (2023).

[3] G. Frasca-Caccia, C. Valentino, F. Colace, D. Conte. *An overview of differential models for corrosion of cultural heritage artefacts.* Math. Model. Nat. Phenom. 18, 27 (2023).

[4] 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.* J. Comput. Phys. 378, 686–707 (2019).

[5] C. Valentino, G. Pagano, D. Conte, B. Paternoster, F. Colace, M. Casillo. *Step-by-step time discrete Physics Informed Neural Networks with application to a sustainability PDE model.* Math. Comput. Simul., doi.org/10.1016/j.matcom.2024.10.043 (2024).

[6] C. Valentino, G. Pagano, D. Conte, B. Paternoster, F. Colace. *Physics Informed Neural Networks for a Lithium-ion batteries model: a case of study.* Submitted.

Training a quantum GAN with classical data

Davide Pastorello*, Giacomo De Palma* and Tristan Klein†
*Dept. of Mathematics, University of Bologna, Piazza di Porta San Donato 5, 40126 Bologna, IT

† ENS de Lyon, Département Informatique, 15 parvis René Descartes 69342 Lyon Cedex 07, France

davide.pastorello3@unibo.it, giacomo.depalma@unibo.it , tristan.klein@ens-lyon.fr

Quantum neural networks (QNNs) are defined by parametric quantum circuits which can be trained by backpropagation in analogy to classical feedforward neural networks. Parametric circuits can be applied to construct generators and discriminators within the quantum version of generative adversarial networks (GANs). In quantum generative adversarial networks (QGANs), the generator can be implemented using a series of quantum gates that manipulate the quantum state of a set of qubits and it is designed to generate data resembling those from the training dataset. The discriminator is also implemented as a quantum circuit, this circuit evaluates the likelihood of the data generated by the generator, comparing it with the real data from the training set. The loss function used to train a QGAN is often defined using quantum concepts, such as quantum state overlap or quantum divergence, rather than traditional loss metrics like cross-entropy. During the training, the parameters of the generator and discriminator quantum circuits are optimized using variational algorithms within an adversarial framework. In the quantum architecture, the training set is made by quantum states, which may encode classical data, assumed to be stored in a quantum memory.

In [1], we considered the so-called shadow protocol that is a procedure to construct classical estimates of quantum states, called *classical shadows*, by means of measurements and quantum/classical processing. The classical shadow is computed classically and stored as classical information and used to efficiently estimates expectation values of observables [2]. Moreover, for any n-qubit quantum state ρ, the computation of a number of classical shadows that is logarithmic in n provides an accurate estimate of ρ w.r.t. the *local quantum Wasserstein distance of order 1* that is a notion from the quantum optimal mass transport [1]. This distance is a measure of distinguishability between quantum states of a n-qubit system and it can be used to evaluate the convergence of the shadow protocol.

The accuracy in estimating a quantum state with classical shadows in this metric has a remarkable consequence in the training of a QGAN [1]. Considering a QGAN where the discriminator generates a classical estimate of the true state constructed as the empirical mean of $O(\log n)$ classical shadows, as proved in [1], no more copies of the true state will be needed and the information contained in its classical shadow will be sufficient. The generator and the discriminator are trained against each other in the adversarial scenario, and the expectation value of the discriminator observable on the true state is estimated via its classical estimate without needing further copies of the true state. After enough iterations, the generated state will be close to the classical shadow of the true state in the local quantum Wasserstein distance of order 1. As a consequence, a QGAN can be equivalently trained over classical shadows in place of true quantum states, if no prior information about the state is available.

In this talk we introduce the notion of the local quantum Wasserstein distance of order 1 as a tool in quantum optimal mass transport, its role in quantifying the convergence of the shadow protocol and how a QGAN can be trained by classical data estimating the quantum states of the training set in terms of classical shadows.

References

[1] De Palma, G., Klein, T., Pastorello, D. *Classical shadows meet quantum optimal mass transport.* Journal of Mathematical Physics. 65, 092201 (2024)

[2] Huang, H. Y., Kueng, R.,Preskill, J. *Predicting many properties of a quantum system from very few measurements.* Nature Physics 16, 1050-1057 (2020).

Linesearch-Enhanced Forward-Backward Methods for Inexact Nonconvex Scenarios

Danilo Pezzi

danilo.pezzi@unimore.it

Silvia Bonettini

silvia.bonettini@unimore.it

Giorgia Franchini

giorgia.franchini@unimore.it

Marco Prato

marco.prato@unimore.it

Via Campi 213/B, Modena

In recent times, optimization techniques have been widely applied to imaging problems, leading to increasingly sophisticated variational models in current research. Significant advancement from previous state-of-the-art methods have been achieved by considering nonconvex settings and combining machine learning strategies with the classical variational techniques. In this talk we introduce a forward-backward framework aimed at the minimization of an objective function composed of a differentiable term and a convex, non differentiable one. The scheme is able to handle two different challenges that can be presented by the objective function. On one hand, even if the differentiable part of the function may be non-convex, the method is is able to achieve convergence to a stationary point. On the other hand, only partial knowledge of the function is required. Indeed, all the key steps of the method can be performed inexactly. As this is a general scheme, it can incoporate a variety of algorithms for different problems. Here we present an application in the realm of bilevel optimization for imaging problems, where the scope is to combine classical variational techniques with machine learning approaches to improve the quality of the reconstructed images. The numerical experience shows that the method is competitive with other existing approaches [1][2].

References

[1] Pedregosa, Fabian *Hyperparameter Optimization with Approximate Gradient*, International Conference on Machine Learning, 2016

[2] Suonperä, Ensio and Valkonen, Tuomo *Linearly convergent bilevel optimization with single-step inner methods*, Computational Optimization and Applications, 2023

The Neural Approximated Virtual Element Method on general polygons

Moreno Pintore

Laboratoire Jacques-Louis Lions, Sorbonne Université, INRIA, 4 place Jussieu, 75005 Paris, France

moreno.pintore@inria.fr

Stefano Berrone, Gioana Teora

Dipartimento di Scienze Matematiche "G. L. Lagrange", Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy stefano.berrone@polito.it, gioana.teora@polito.it

In the Scientific Machine Learning framework, numerous new methods to solve engineering problems have been proposed in the last few years. Such methods combine the accuracy and stability of classical numerical methods with the efficiency and adaptability of machine learning techniques. The Neural Approximated Virtual Element Method (NAVEM) perfectly fits in this context, since it is a method inspired by the Virtual Element Method (VEM) [1], with which shares some features, and that heavily relies on the nonlinear approximation properties of deep neural networks.

The VEM is a numerical method used to solve partial differential equations using meshes comprising very general elements and basis functions that are not known in closed form. The idea of the NAVEM is to use the same meshes and to explicitly approximate the VEM basis functions through one or more neural networks. This approximation leads to a completely different method, that does not include projection or stabilization operators, but that relies on an offline-online splitting.

The NAVEM has been firstly introduced in [2] and then extended in [3] to more general two-dimensional meshes. In this presentation we focus on this second formulation, characterized by an approximation of the VEM basis functions through a novel set of harmonic functions. This choice is crucial in order to accurately approximate the VEM basis functions while reducing spurious oscillations that may characterize the output of a standard neural network. We also present the architecture of the involved neural networks and we theoretically discuss their approximation properties. We propose several numerical results to illustrate the performances of the method on different meshes and on different problems.

References

[1] L. Beirão da Veiga, F. Brezzi, A. Cangiani, G. Manzini, and A. Russo *Basic principles of Virtual Element Methods*, Mathematical Models and Methods in Applied Sciences, vol. 23, no. 01, pp. 199–214, 2013.

[2] S. Berrone, D. Oberto, M. Pintore, and G. Teora *The lowest-order neural approximated virtual element method*, ENUMATH 2023, Accepted. [Online]. Available: https://arxiv.org/abs/2311.18534.

[3] S. Berrone, M. Pintore, and G. Teora *The lowest-order neural approximated virtual element method*, ArXiv preprint arXiv:2409.15917, 2024.

Grokking as an entanglement transition in tensor network machine learning

Domenico Pomarico

National Institute for Nuclear Physics, Bari domenico.pomarico@ba.infn.it

Alfonso Monaco, Giuseppe Magnifico, Antonio Lacalamita, Ester Pantaleo, Loredana Bellantuono, Sabina Tangaro, Tommaso Maggipinto, Marianna La Rocca, Nicola Amoroso, Sebastiano Stramaglia, Roberto Bellotti

National Institute for Nuclear Physics, Bari alfonso.monaco@ba.infn.it

Generalizability is a fundamental property for machine learning algorithms, detected by a grokking transition during training dynamics [1]. In the quantum-inspired machine learning framework we numerically prove that a quantum many-body system shows an entanglement transition corresponding to a performances improvement in binary classification of unseen data. Two datasets are considered as use case scenarios, namely fashion MNIST and genes expression communities of hepatocellular carcinoma. The measurement of qubits magnetization and correlations is included in the matrix product state (MPS) simulation [2], in order to define meaningful genes subcommunities, verified by means of enrichment procedures.

References

[1] Kenzo Clauw, Sebastiano Stramaglia and Daniele Marinazzo, Information-Theoretic Progress Measures reveal Grokking is an Emergent Phase Transition, arXiv:2408.08944 (2024)

[2] E. Miles Stoudenmire and David J. Schwab, Supervised Learning with Quantum-Inspired Tensor Networks, arXiv:1605.05775 (2017)

A CNN-LSTM approach for parameter estimation for lithium metal battery cycling model

Maria Grazia Quarta, Ivonne Sgura

Department of Mathematics and Physics "Ennio De Giorgi", University of Salento, via per Arnesano, 73100 Lecce (LE), Italy

mariagrazia.quarta@unisalento.it, ivonne.sgura@unisalento.it

Benedetto Bozzini

Department of Energy, Politecnico di Milano, Via Lambruschi 4, 20156 Milano, Italy

benedetto.bozzini@polimi.it

Raquel Barreira

Instituto Politécnico de Setúbal, Escola Superior de Tecnologia de Setúbal Campus do IPS Estefanilha, 2914-508 Setúbal, Portugal raquel.barreira@estbarreiro.ips.pt

Symmetric coin cell cycling is an important tool for the analysis of battery materials, enabling the study of electrode/electrolyte systems under realistic operating conditions. Moreover, understanding the behavior of metal anodes in batteries and accurately predicting their performance is a challenge due to the methodological gap between theoretical models and experimental observations. In order to address this challenge, a PDE model describing the voltage profiles behavior of symmetrical coin cells testing the Galvanostatic Discharge-Charge (GDC) protocol has been developed [1, 2]

In this talk, based on [3], we propose a hybrid architecture of Convolutional Neural Network and Long-Short Term Memory layers (CNN-LSTM) to estimate some relevant physico-chemical parameters in the PDE system that describe GDC cycling of Li/Li symmetric cells. Our results show the neural network ability to capture characteristics of voltage profiles, such as peak and valley, saddle points, and concavity variations [1], that other traditional methods, such as Least Squares (LS) fitting, may overlook. Moreover, our Deep Learning algorithm can successfully estimate parameters also for experimental discharge-charge time series data. These results highlight the robustness of our approach, which allows us to bridge the gap between theory and experiments.

References

[1] F. Rossi, L. Mancini, I. Sgura, M. Boniardi, A. Casaroli, A.P. Kao, B. Bozzini, *Insight into the Cycling Behaviour of Metal Anodes, Enabled by X-ray Tomography and Mathematical Modelling*, ChemElectroChem 9, 2022.

[2] B. Bozzini, E. Emanuele, J. Strada, I. Sgura, *Mathematical modelling and parameter classification enable understanding of dynamic shape-change issues adversely affecting high energy-density battery metal anodes*, Applications in Engineering Science 13, 100125, 2023.

[3] M.G. Quarta, I. Sgura, E. Emanuele, J. Strada, R. Barreira, B. Bozzini, *A deep-learning approach to parameter fitting for a lithium metal battery cycling model*, submitted.

On the complexity of infinite argumentation

Luca San Mauro

University of Bari lucafrancesco.sanmauro@uniba.it

Uri Andrews

University of Wisconsin andrews@math.wisc.edu

The theory of abstract argumentation frameworks (AFs), introduced in Dung's seminal work [3], has become a foundational topic in knowledge representation. AFs provide a versatile and powerful tool for modeling diverse reasoning problems, especially in scenarios requiring the resolution of conflicting arguments. To accommodate varying argumentative contexts, a wide range of semantics has been developed to determine which arguments or extensions (i.e., sets of arguments) are considered acceptable (for an in-depth overview, see the handbook [2]).

While research has extensively explored finite AFs, the study of *infinite* AFs remains underdeveloped, creating theoretical, conceptual, and practical gaps. Our work [1] addresses these gaps by systematically analyzing the algorithmic complexity of problems associated with infinite AFs. Leveraging concepts from computability theory, we define computable AFs as those where a Turing machine can determine, for any pair of arguments, whether one attacks the other. Our results reveal that, for several established semantics, determining whether an argument is (credulously or skeptically) accepted reaches maximal complexity, properly belonging to the so-called Σ^1_1 and Π^1_1 classes.

Moreover, we demonstrate that a single, carefully constructed infinite AF suffices to witness our hardness results, highlighting that argument acceptability remains highly undecidable for an individual, specific framework. Finally, we propose a way of using Turing degrees to calibrate, for a given infinite AF, the exact difficulty of computing an extension in a given semantics. This approach uncovers a rich and intricate landscape of complexities, significantly advancing our understanding of infinite AFs and their computational properties.

References

[1] U. Andrews and L. San Mauro, *On computational problems for infinite argumentation frameworks: The complexity of finding acceptable extensions*, in Proceedings of the 22nd International Workshop on Nonmonotonic Reasoning (NMR 2024), CEUR Workshop Proceedings, 3835: 3–13, 2024

[2] Pietro Baroni, Dov Gabbay, Massimilino Giacomin, and Leendert van der Torre (eds), *Handbook of Formal Argumentation*, College Publications, London, 2018

[3] P. M. Dung, *On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games*, Artificial intelligence, 77: 321–357, 1995

Trade-off Invariance Principle for regularized functionals

Alessandro Scagliotti

Technical University of Munich & Munich Center for Machine Learning (MCML) scag@ma.tum.de

Massimo Fornasier, Jona Klemenc

Technical University of Munich & Munich Center for Machine Learning (MCML)

massimo.fornasier@cit.tum.de, jona.klemenc@tum.de

In this talk, we consider functionals $H_\alpha : U \to \mathbb{R} \cup \{+\infty\}$ of the form $H_\alpha(u) = F(u) + \alpha G(u)$ with $\alpha \in [0, +\infty)$, and where $U \neq \emptyset$ is a set without further structure. Assuming that

$$H_\alpha^\star := \arg\min_U H_\alpha$$

is non-empty for every $\alpha \in [a, b] \subset [0, +\infty)$ (with $0 \leq a < b$), we first show that —excluding at most countably many exceptional values of $\alpha \in [a, b]$— we have the following:

$$\inf_{H_\alpha^\star} F = \sup_{H_\alpha^\star} F, \qquad \inf_{H_\alpha^\star} G = \sup_{H_\alpha^\star} G,$$

i.e., for every $u_1^\star, u_2^\star \in H_\alpha^\star$ the identities $F(u_1^\star) = F(u_2^\star)$ and $G(u_1^\star) = G(u_2^\star)$ hold true.
We further prove a stronger result, which asserts that for all but countable many $\alpha \in [0, +\infty)$, if $\inf_{u \in U} H_\alpha(u) > -\infty$, then there exists a value $G_\alpha \in [-\infty, +\infty]$ such that $G(u_i) \to G_\alpha$ for every sequence $(u_i)_{i \in \mathbb{N}}$ such that $H_\alpha(u_i) \to \inf_{u \in U} H_\alpha(u)$ as $i \to \infty$.
This fact in turn implies an unexpected consequence for functionals regularized with uniformly convex norms: excluding again at most countably many values of α, it turns out that for a minimizing sequence, convergence to a minimizer in the weak or strong sense is equivalent.

References

[1] M. Fornasier, J. Klemenc, A. Scagliotti *Trade-off Invariance Principle for minimizers of regularized functional*, arXiv:2411.11639 (preprint).

Quantum Optimization in Environmental Resource Management: A Focus on Irrigation Scheduling

Vincenzo Schiano Di Cola

Istituto di Ricerca sulle Acque, Consiglio Nazionale delle Ricerche; Quantum2Pi s.r.l.

vincenzo.schianodicola@ba.irsa.cnr.it

Dimitri Jordan Kenne, Gabriele Intoccia

Dipartimento di Matematica e Fisica, Università degli Studi della Campania "Luigi Vanvitelli"

Salvatore Cuomo

Dipartimento di Matematica e Applicazioni, Universit'a degli Studi di Napoli Federico II

salcuomo@unina.it

Effective resource management in agriculture is essential for sustainability, given the increasing demand on water resources. Traditional optimization methods for irrigation scheduling frequently encounter difficulties in reconciling complex constraints, such as temporal dependencies, resource availability, and environmental considerations. Date et al. [1] recently proposed the use of quantum computers to accelerate machine learning model training. They formulated three machine learning problems (linear regression, support vector machine, and balanced k-means clustering) as Quadratic Unconstrained Binary Optimization (QUBO) problems and proposed solving them using adiabatic quantum computing. In this context, Quantum Approximate Optimization Algorithm (QAOA) serves as an alternative approach for obtaining effective approximate solutions to these problems. This could lead to the use of QAOA in deep learning for neural network training and to boost novel research opportunities including non-Gaussian gates, exploring quantum advantages with decoherence, developing specialized Quantum Neural Networks (QNNs), and a more profound examination of fundamental concepts in quantum physics as they relate to QNNs [2].

This research examines the application of quantum algorithms, specifically the QAOA, to enhance resource management in agriculture. This research presents irrigation scheduling as a QUBO problem and explores various ansatz in the setting of a Variational Quantum Eigensolver (VQE) [3, 4]. This study emphasizes the potential of quantum optimization in addressing critical challenges in agricultural water management, offering a method for improved sustainability via enhanced resource allocation. The suggested approach illustrates the broader applicability of quantum approximation in solving complex optimization problems across diverse environmental and industrial contexts, extending beyond irrigation.

References

[1] Date, P., Arthur, D., & Pusey-Nazzaro, L. (2021). *QUBO formulations for training machine learning models.* , Scientific reports, 11(1), 10029.

[2] Blekos, K., Brand, D., Ceschini, A., Chou, C. H., Li, R. H., Pandya, K., & Summer, A. (2024). *A review on quantum approximate optimization algorithm and its variants.* , Physics Reports, 1068, 1-66.

[3] Muhamediyeva, Dilnoz & Niyozmatova, Nilufar & Yusupova, Dilfuza & Samijonov, Boymirzo. (2024). *Quantum optimization methods in water flow control.*, E3S Web of Conferences. 590. 02003. 10.1051/e3sconf/202459002003.

[4] Scherer, Wolfgang. (2019). *Mathematics of Quantum Computing: An Introduction.*, J10.1007/978-3-030-12358-1

A Framework Combining Machine Learning and Statistical Modeling for Detecting Extreme Events in High-Dimensional Data

Dhruv Singhvi

Bond Street, Norwich, UK dhruv.singhvi@whiteboxfinancial.com

The detection and identification of extreme events in complex, high-dimensional datasets present significant challenges across various fields, including finance, environmental science, and engineering. We introduce a framework that combines machine learning-driven dimensionality reduction with advanced statistical transformations to address this challenge. This approach simplifies high-dimensional data while preserving its inherent probabilistic and relational properties, enabling effective and interpretable linear analysis.

The framework begins with dimensionality reduction using algorithms such as **Uniform Manifold Approximation and Projection (UMAP)** [McInnes et al., 2018], which project high-dimensional data into a two-dimensional space. UMAP facilitates the creation of a visually interpretable representation, ensuring the retention of essential topological features for subsequent analysis.

Following dimensionality reduction, the data is transformed into a linear progression through a combination of **cumulative distribution alignment** and **monotonic mapping** [Wasserman, 2006]. This process involves calculating the empirical cumulative distribution functions (ECDFs) of the reduced dimensions to map the data to a uniform distribution. The uniform coordinates are then aligned to a linear configuration, such as $y = mx+c$, using transformations that preserve the original dataset's probabilistic structure and relative relationships.

Once aligned, linear regression models are employed to identify deviations from expected patterns, with residual analysis pinpointing outliers and anomalies. These outliers are interpreted as potential extreme events, characterized by their deviation from the linear progression. This linear framework simplifies the detection of rare or unexpected phenomena, allowing traditional statistical techniques such as hypothesis testing and confidence interval estimation to be applied effectively [Wasserman, 2006].

This methodology bridges the gap between machine learning and statistical modeling, providing a scalable, interpretable, and versatile solution for extreme event detection. In the working paper, we demonstrate the effectiveness of the framework through applications to financial datasets, highlighting its capability to identify anomalies. Future research will explore its adaptability to domains such as environmental systems, biological data, and urban modeling, expanding its applicability to a wide range of critical challenges.

References

[1] McInnes, L., Healy, J., & Melville, J. (2018) *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*, Journal of Open Source Software

[2] Wasserman, L. (2006) All of Statistics: A Concise Course in Statistical Inference

A Deep-QLP Decomposition Algorithm and Applications

Cristiano Tamborrino
cristiano.tamborrino@uniba.it
In collaboration with: Antonella Falini, Francesca Mazzia
Dipartimento di Informatica, Università degli Studi di Bari Aldo Moro, Italy

Abstract

Singular value decomposition (SVD) is a fundamental tool in data analysis and machine learning. Starting from the Stewart's QLP decomposition [1], we propose an innovative *Deep-QLP* decomposition algorithm for efficiently computing an approximate Singular Value Decomposition (SVD) based on the preliminary work in [2]. Given a specified tolerance τ, the algorithm automatically computes a positive integer f and a factorization $\mathcal{U}_f\mathcal{L}_f^D\mathcal{V}_f^T$, with $\mathcal{L}_f^D$ diagonal matrix, $\mathcal{U}_f, \mathcal{V}_f$ matrices of rank f with orthonormal columns such that

$$\|A - \mathcal{U}_f\mathcal{L}_f^D\mathcal{V}_f^T\|_2 \leq 3\tau\|A\|_2.$$

The *Deep-QLP* algorithm stands out for its ability to return an approximation of the largest singular values, based on a fixed tolerance, to achieve significant dimensionality reduction while simultaneously preserving essential information in the data. In addition, it can also be used to return an approximation of the smallest singular values that can be used in some applications.
The algorithm has been successfully integrated with the randomized SVD [3], making the *Deep-QLP* algorithm particularly effective for sparse matrices, which are prevalent in numerous applications such as text mining.
Several numerical experiments have been conducted, demonstrating the effectiveness of the proposed method.

Acknowledgements. Cristiano Tamborrino, Francesca Mazzia and Antonella Falini acknowledge the support of the PNRR project FAIR - Future AI Research (PE00000013), Spoke 6 - Symbiotic AI (CUP H97G22000210007) under the NRRP MUR program funded by the NextGenerationEU. The authors thank the GNCS for its valuable support under the INdAM-GNCS project CUP E55F22000270001.

References

[1] Gilbert W. Stewart. *The QLP Approximation to the Singular Value Decomposition.* SIAM J. Sci. Comput., 20:1336–1348, 1999. https://api.semanticscholar.org/CorpusID:15701097.

[2] Antonella Falini and Francesca Mazzia. *Approximated Iterative QLP for Change Detection in Hyperspectral Images.* AIP Conference Proceedings, 3094(1):370003, 2024. https://doi.org/10.1063/5.0210496.

[3] Nathan Halko, Per-Gunnar Martinsson, and Joel A. Tropp. *Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions.* SIAM Review, 53(2):217–288, 2011. SIAM.

Variable metric proximal stochastic gradient methods with additional sampling

Ilaria Trombini, Valeria Ruggiero

Dept. of Mathematics and Computer Science, University of Ferrara, Ferrara, 44121, Italy

trmlri@unife.it, rgv@unife.it

Natavsa Krklec Jerinkić

Faculty of Sciences, Department of Mathematics and Informatics, University of Novi Sad, Trg Dositeja Obradovića 4, Novi Sad, 21000, Serbia natasa.krklec@dmi.uns.ac.rs

Federica Porta

Dept. of Physics, Informatics and Mathematics, University of Modena and Reggio Emilia, Modena, 41125, Italy federica.porta@unimore.it

Regularized empirical risk minimization problems are prevalent across various domains, such as machine learning, signal processing, and image processing. Many optimization challenges within machine learning can be formulated as the minimization of a composite function: the first component typically represents the expected risk, which is often substituted by empirical risk in practice, while the second imposes prior information on the solution. Generally, the first term is differentiable and the second term is convex, making proximal gradient methods particularly well-suited for tackling these optimization problems.

However, in the context of large-scale machine learning, calculating the full gradient of the differentiable term can be computationally prohibitive, rendering standard proximal gradient algorithms impractical. Consequently, proximal stochastic gradient methods have been extensively explored in optimization research over recent decades [1, 2, 3].

This talk introduces a class of proximal stochastic gradient methods built on three foundational elements: a variable metric underlying the iterative process, a stochastic line search mechanism to control the descent properties, and an incremental mini-batch size technique based on additional sampling. Convergence results for the proposed algorithms are established under varying assumptions on the objective function. Notably, no assumption regarding the Lipschitz continuity of the gradient of the differentiable part of the objective function is required. Possible strategies for the automatic selection of parameters in the proposed framework are also discussed. Numerical experiments on binary classification problems demonstrate the effectiveness of this approach in comparison to other leading proximal stochastic gradient methods.

References

[1] Xiao, L and Zhang, T *A proximal Stochastic Gradient Method with Progressive Variance Reduction*, SIAM J. Optim. 24, 4, (2014), 2057–2075.

[2] Phamy. N H, and Nguyen, L M and Phan, D T and Tran–Dinh, Q *ProxSARAH: an efficient algorithmic framework for stochastic composite nonconvex optimization*, J. Mach. Learn. Res. 21, 1, Article 110 (2020), 1–48.

[3] Wang, Zhe and Ji, Kaiyi and Zhou, Yi and Liang, Yingbin and Tarokh, Vahid *SpiderBoost and Momentum: Faster Stochastic Variance Reduction Algorithms*, Proceedings of the 33rd International Conference on Neural Information Processing Systems, Curran Associates Inc. 216, (2019), 2406–2416.

Industry Talks

Pirelli pratical development of LLM application for risk prevention on the workplace

Mattia Beretta

Generative AI tech lead, Pirelli & C. S.p.A.

mattia.beretta@pirelli.com

Pirelli, a leader in tire manufacturing, strengthens its commitment to ensuring workplace safety. The "Health, Safety and Environment" department, with the support of GenAI, can now not only analyze thousands of textual reports from global facilities more efficiently each year but also implement preventive actions to mitigate risk situations. By leveraging the capabilities of natural language LLM, Pirelli is able to automate and optimize the risk assessment process by summarizing reports and highlighting critical points.

Data, Math, and Machine Learning: Revolutionizing Earth Observation Technologies

Nicolò Taggio

GeoAI team coordinator, Planetek Italia S.r.l, taggio@planetek.com

The advent of advanced machine learning (ML) algorithms and imaging technologies (such as hyperspectral and multispectral sensors), has significantly transformed Earth Observation (EO). The connection between data, mathematics, and ML will be explored to understand how they are driving this transformation, revolutionizing interpretation and use of EO data for various applications.

A foundational overview of hyperspectral and multispectral imaging, highlighting their key differences and advantages, will be presented. By diving into the feature space, mathematical operations and machine learning techniques can be applied to combine spectral bands, creating indexes that enhance the detection and classification of surface features. Furthermore, to illustrate the practical application of these concepts, will be highlighted a **Burned Area Detection Using Non-Negative Matrix Factorization (NMF)**. This unsupervised approach leverages spectral signatures to identify and map burned areas accurately, showcasing the power of data-driven feature extraction.

Finally, a service called **Rheticus Network alert** will be presented, which integrates ML algorithms with data and mathematical models to provide actionable insights for pipeline monitoring. This service emphasizes user interaction, showcasing the importance of tailoring EO solutions to meet end-user needs. Looking ahead, the focus will be on the potential of cognitive cloud computing to optimize complex satellite networks through cooperative swarming. This approach leverages multi-objective functions inspired by game theory, enabling autonomous self-organization of satellite assets to achieve tasks even in scenarios with incomplete information. This future-oriented perspective highlights how advances in distributed intelligence and autonomous decision making are reshaping the next generation of space-based technologies.

Posters

MaxCutPool: Differentiable Feature-Aware MAXCUT for Pooling in Graph Neural Networks

Carlo Abate

carlo.abate@iit.it

Filippo Maria Bianchi

filippo.m.bianchi@uit.no

We propose a novel approach to compute the MAXCUT in attributed graphs, i.e. graphs with features associated with nodes and edges, by exploiting heterophilic message passing to assign connected nodes to different partitions. The approach is fully differentiable, making it possible to find solutions that jointly optimize the MAXCUT along with other objectives. Based on the obtained MAXCUT partition, we implement MaxCutPool, a hierarchical graph pooling layer for graph neural networks. The layer is sparse, differentiable, and particularly suitable for downstream tasks on heterophilic graphs. Our key contributions include: (1) a novel MAXCUT computation method for attributed graphs, (2) a new hierarchical pooling layer especially effective for heterophilic graphs, (3) a general scheme for node-to-supernode assignment, and (4) the introduction of the first heterophilic dataset for graph classification. Experimental results demonstrate that MaxCutPool achieves state-of-the-art performance across various graph classification and node classification tasks, highlighted by perfect accuracy on expressiveness tests and significant improvements on heterophilic graph classification.

References

[1] C. Abate and F. M. Bianchi. Maxcutpool: differentiable feature-aware maxcut for pooling in graph neural networks, 2024, https://arxiv.org/abs/2409.05100

Distributional forecast approaches to stochastic optimization in healthcare appointment scheduling

Sara Cambiaghi

Department of Mathematics "F. Casorati", University of Pavia

sara.cambiaghi01@universitadipavia.it

Davide Duma

Department of Mathematics "F. Casorati", University of Pavia

davide.duma@unipv.it

Appointment scheduling in healthcare involves optimizing the allocation of appointment slots to balance patient access and resource utilization, with one of the primary sources of uncertainty being the service time. In this talk we propose a stochastic optimization approach that leverages distributional forecasts to identify patient groups with similar service times.

We explore two alternative strategies to achieve the desired distributional forecast. The first method utilizes random forest leaf embeddings, where the encoded data are clustered using the K-means algorithm. The second method employs a Decision Tree to generate an initial partition of patients and their corresponding distributions, which are subsequently clustered using the Wasserstein K-Means algorithm. Both strategies enable the derivation of probability distributions for each resulting cluster.

The appointment scheduling problem is formulated through a stochastic programming model that minimizes waiting times for outpatients, completion times for emergency patients and inpatients, and overtime. A genetic algorithm is used to solve the optimization problem by estimating the Pareto front within a reasonable timeframe.

As a case study, we examine the CT scan service at Policlinico San Matteo in Pavia, Italy, where three categories of patients–outpatients, inpatients, and emergency cases–use the same resources but have different priorities and needs. The considered multi-objective decision problem concerns the outpatient CT scan scheduling over a weekly planning horizon, avoiding conflicts with the real-time scheduling of inpatients and emergencies and ensuring doctors have enough time to complete both the scans and their reports.

A computational analysis is performed to evaluate the effectiveness of the proposed optimization and machine learning approach for improving the efficiency of the CT scan service. We also compare different point predictions with distributional forecasts to represent the uncertainty of service time durations and understand which method provides better results.

References

[1] J. Marcak, Y.L. Huang, *Radiology procedure time slot redesign to improve scheduling efficiency*, Proceedings of the 62nd IIE Annual Conference and Expo.

[2] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, *A fast and elitist multiobjective genetic algorithm: NSGA-II*, IEEE Transactions on Evolutionary Computation.

[3] P. Bhattacharjee, P.K. Ray, *Scheduling appointments for multiple classes of patients in presence of unscheduled arrivals: Case study of a CT department*, Proceedings of the IISE Transactions on Healthcare Systems Engineering.

[4] T. Gneiting, A.E. Raftery, *Strictly Prover Scoring Rules, Prediction, and Estimation*, Journal of the American Statistical Association.

[5] M. Bicego, *K–Random Forest: a K–Means style algorithm for Random Forest clustering*, International Joint Conference on Neural Networks (IJCNN).

[6] Y. Zhuang, X. Chen, Y. Yang, *Wasserstein K–means for clustering probability distributions*, Advances in Neural Information Processing Systems.

Anticlustering for Large Scale Clustering

Anna Livia Croella

Sapienza University of Rome, Rome, Italy annalivia.croella@uniroma1.it

Veronica Piccialli, Antonio Maria Sudoso

Sapienza University of Rome, Rome, Italy veronica.piccialli@uniroma1.it,

antonio.sudoso@uniroma1.it

This research develops innovative methodologies for integrating clustering and anticlustering techniques into large-scale data analysis within AI frameworks. The objective is to establish a mechanism that generates tighter lower bounds for the clustering problem, starting from a heuristic solution that minimizes the Within-group Sum of Squares (WSS). A key insight is that the minimum WSS of the union of disjoint subsets is always greater than or equal to the sum of the minimum WSS of the individual subsets [1]. This indicates that summing the minimum WSS values of disjoint subsets provides a valid lower bound for the optimal WSS of the entire dataset. To enhance this lower bound, we maximize the minimum WSS of each subset by creating groups of points with high dissimilarity, a process known as anticlustering [2]. Through anticlustering, we developed a certification process to validate clustering solutions obtained using the k-means algorithm. We tested this mechanism on large-scale datasets containing 2,000 to 10,000 data points and between 2 and 500 features. Our procedure consistently achieved gaps between the clustering solution and the lower bound ranging from 0.1% to 5%. Future work will focus on iterative improvements to the clustering solutions through feedback loops, as well as integrating the generation of lower bounds into a Branch & Bound algorithm [3].

References

[1] Diehr, G. *Evaluation of a Branch and Bound Algorithm for Clustering*, SIAM J. Sci. Stat. Comput. 6, 2, 268–284 (1985)

[2] Papenberg, M. *K-Plus anticlustering: An improved k-means criterion for maximizing between-group similarity*, British Journal of Mathematical and Statistical Psychology, 77(1), 80–102 (2023)

[3] Piccialli, V., Russo, A.R., and Sudoso, A.M. *An Exact Algorithm for Semi-supervised Minimum Sum-of-Squares Clustering*, Computers & Operations Res., 147, 105958 (2021)

ROI Image Identification via Topological Data Analysis: A Case Study of Brain Tumor MRI

Serena Grazia De Benedictis

University of Bari Aldo Moro, serena.debenedictis@uniba.it

Grazia Gargano, Gaetano Settembre

University of Bari Aldo Moro, grazia.gargano@uniba.it, gaetano.settembre@uniba.it

In the medical context, modern imaging methods such as magnetic resonance imaging (MRI) have completely changed how diseases are diagnosed and tracked. Advanced image processing algorithms are increasingly employed to automate the interpretation of medical images, facilitating faster and more accurate diagnosis. This work presents a novel ensemble of methods using MRI data for the detection and classification of common brain cancers. The proposed approach combines dimensionality reduction technique with machine learning (ML) algorithms, and then integrates ML prediction with topological data analysis (TDA)-based results [2]. A low-rank Tucker decomposition [3] is used to reduce data dimensionality while maintaining the key structures and properties of preprocessed MRI scans. Robust tumor classification models can be developed with supervised machine learning classifiers that are trained on the low-dimensional representations of the data. The MRI scans are also parallelly processed using persistent homology (PH) [4], an algebraic method for measuring topological features of data to explore the spatial relationships and patterns present in the pixel distribution and the geometry of the images. Indeed, by extracting the most persistent connected component of the MRI scan, we can precisely identify region of interest (ROI) that can suggest the existence or features of a possible tumor and require further investigation. The promising results obtained by applying the proposed framework to a brain tumor image dataset demonstrate the effectiveness of integrating low-rank approximation, ML and TDA techniques for tumor detection and classification. This comprehensive approach provides a robust strategy for future research and clinical application, potentially extendable to other solid tumors.

References

[1] S.G. De Benedicitis, G. Gargano, and G. Settembre *Enhanced MRI brain tumor detection and classification via topological data analysis and low-rank tensor decomposition*, Journal of Computational Mathematics and Data Science (2024), 13, 100103. doi:10.1016/j.jcmds.2024.100103

[2] Dey, Tamal Krishna and Wang, Yusu *Computational Topology for Data Analysis*, Cambridge University Press, 2022. doi:10.1017/9781009099950.

[3] Kolda, Tamara G. and Bader *Tensor Decompositions and Applications*, SIAM Review 51 (3) (2009) 455–500. doi:10.1137/07070111x.

[4] Schenck, Hal *Algebraic Foundations for Applied Topology and Data Analysis* , Springer International Publishing, 2022. doi:10.1007/978-3-031-06664-1.

Inferring Failure Processes via Causality Analysis: from Event Logs to Predictive Fault Trees

Roberta De Fazio

Dipartimento di Matematica e Fisica, Università degli Studi della Campania Luigi Vanvitelli, Italy

roberta.defazio@unicampania.it

Benoît Depaire

Faculty of Business Informatics, Hasselt University, Belgium

benoit.depaire@uhasselt.be

Stefano Marrone, Laura Verde

Dipartimento di Matematica e Fisica, Università degli Studi della Campania Luigi Vanvitelli, Italy

stefano.marrone,laura.verde@unicampania.it

In the current Artificial Intelligence era, the integration of the Industry 4.0 paradigm in real-world settings requires robust and scientific methods and tools. Two concrete aims are the exploitation of large datasets [1] and the guarantee of a proper level of explainability, demanded by critical systems and applications [2]. Focusing on the predictive maintenance problem, this work leverages causality analysis to elicit knowledge about system failure processes. The result is a model expressed according to a newly introduced formalism: the Predictive Fault Trees [3]. This model is enriched by causal relationships inferred from dependability-related event logs. The proposed approach considers both fault-error-failure chains between system components and the impact of environmental variables (e.g., temperature, pressure) on the health status of the components. A proof of concept shows the effectiveness of the methodology, leveraging an event-based simulator [4].

References

[1] R. De Fazio, A. Balzanella, S. Marrone, F. Marulli, L. Verde, V. Reccia, P. Valletta *CaseID Detection for Process Mining: A Heuristic-Based Methodology*, Process Mining Workshops, Springer Nature Switzerland

[2] S. Ramezani, L. Cummins, B. Killen, R. Carley, A. Amirlatifi, S. Rahimi, M. Seale, L. Bian *Scalability, Explainability and Performance of Data-Driven Algorithms in Predicting the Remaining Useful Life: A Comprehensive Review*, IEEE Access,Institute of Electrical and Electronics Engineers (IEEE)

[3] R. De Fazio, S. Marrone, L. Verde, V. Reccia, P. Valletta *Towards an extension of Fault Trees in the Predictive Maintenance Scenario*, 19th European Dependable Computing Conference, arXiv pre-print

[4] C. Abate, L. Campanile, S. Marrone *A flexible simulation-based framework for model-based/data-driven dependability evaluation*, Proceedings - 2020 IEEE 31st International Symposium on Software Reliability Engineering Workshops, ISSREW 2020

A line-search based SGD algorithm with Adaptive Importance Sampling

Anna De Magistris

Dipartimento di Matematica e Fisica Luigi Vanvitelli, anna.demagistris@unicampania.it

Filippo Camellini, Serena Crisci, Giorgia Franchini

Università degli Studi di Modena e Reggio Emilia, filippo.camellini@edu.unife.it

Dipartimento di Matematica e Fisica Luigi Vanvitelli, serena.crisci@unicampania.it

Università degli Studi di Modena e Reggio Emilia, giorgia.franchini@unimore.it

Stochastic Gradient Methods are essential for solving large-scale optimization problems, particularly when the objective function F is expressed as the sum of n functions f_i, each with an L_i-Lipschitz continuous gradient [1]. Stochastic Gradient Descent (SGD), which computes an approximate gradient by sampling a function f_{i_k} from a probability distribution p^k, is highly efficient and scalable. However, its asymptotic performance is limited; with a constant step size, it converges only to a neighborhood of the optimum even under strong convexity assumptions [2]. To address this, variance-reduction techniques like SVRG [3] and SAGA [4] combine stochastic gradients with partial updates of the full gradient. Another approach involves dynamic sampling to increase the batch size progressively, as in algorithms like LISA [5, 6]. Importance sampling is also explored, optimizing the sampling distribution p^k to reduce variance based on Lipschitz constants L [7]. Yet, estimating L remains challenging, especially in deep learning contexts. A notable advancement is the SGD-AIS algorithm, which approximates an optimal sampling distribution without relying on L and demonstrates superior performance compared to SGD with uniform sampling [8]. However, the decreasing step size employed in SGD-AIS can slow convergence and demands careful parameter tuning. To overcome these limitations, we propose an automatic step size selection method using a stochastic Armijo-type line-search procedure. This approach simplifies parameter tuning, accelerates convergence, and leverages the importance sampling distribution of SGD-AIS. Our contributions include extending SGD-AIS with a stochastic line-search strategy and introducing a variant for mini-batch stochastic gradients. Theoretical convergence results and experiments on ℓ_2-regularized logistic regression and smooth hinge loss confirm the effectiveness of the proposed methods.

References

[1] F. E. Curtis and K. Scheinberg, *Optimization Methods for Supervised Machine Learning: From Linear Models to Deep Learning*, arXiv preprint arXiv:1706.10207, 2017.

[2] L. Bottou, F. E. Curtis, and J. Nocedal, *Optimization Methods for Large-Scale Machine Learning*, SIAM Review, vol. 60, no. 2, pp. 223–311, 2018.

[3] S. J. Reddi, A. Hefny, S. Sra, B. Póczós, and A. Smola, *Stochastic Variance Reduction for Nonconvex Optimization*, Proceedings of ICML 2016, pp. 314–323.

[4] A. Defazio, F. Bach, and S. Lacoste-Julien, *SAGA: A Fast Incremental Gradient Method with Support for Non-Strongly Convex Composite Objectives*, Proceedings of NIPS 2014, pp. 1646–1654.

[5] G. Franchini, F. Porta, V. Ruggiero, and I. Trombini, *A Line Search Based Proximal Stochastic Gradient Algorithm with Dynamical Variance Reduction*, Journal of Scientific Computing, 2022.

[6] G. Franchini, F. Porta, V. Ruggiero, I. Trombini, and L. Zanni, *A Stochastic Gradient Method with Variance Control and Variable Learning Rate for Deep Learning*, Journal of Computational and Applied Mathematics, vol. 451, p. 116083, 2024.

[7] L. Xiao and T. Zhang, *A Proximal Stochastic Gradient Method with Progressive Variance Reduction*, SIAM Journal on Optimization, vol. 24, no. 4, pp. 2057–2075, 2014.

[8] H. Liu, X. Wang, J. Li, and A. M.-C. So, *Low-Cost Lipschitz-Independent Adaptive Importance Sampling of Stochastic Gradients*, Proceedings of ICPR 2020, pp. 2150–2157.

[9] S. Vaswani, A. Mishkin, I. Laradji, M. Schmidt, G. Gidel, and S. Lacoste-Julien, *Painless Stochastic Gradient: Interpolation, Line-Search, and Convergence Rates*, Proceedings of NIPS 2019.

[10] P. Zhao and T. Zhang, *Stochastic Optimization with Importance Sampling for Regularized Loss Minimization*, Proceedings of PMLR 2015, pp. 1–9.

[11] D. Bertsekas, *Convex Optimization Theory*, Athena Scientific, Belmont, Massachusetts, 2009.

[12] C. Tan, S. Ma, Y.-H. Dai, and Y. Qian, *Barzilai-Borwein Step Size for Stochastic Gradient Descent*, Advances in Neural Information Processing Systems, vol. 29, 2016.

[13] A. Johnson, *Example Book*, Example Publisher, Example City, 2023.

Adapting SAM2 for Few-Shot Multi-Class Semantic Segmentation

Bernardo Forni

University of Pavia bernardo.forni01@universitadipavia.it

Segment Anything Model 2 (SAM2) has shown outstanding performance in zero-shot image and video segmentation. We introduce a novel module to adapt SAM2 for the challenging and underexplored task of few-shot multi-class semantic segmentation. This task involves labeling each pixel within an image using a limited set of mask-annotated images from multiple classes. Our approach leverages a transformer architecture that aggregates the SAM2 features of different classes, accommodating any N-way K-shot configurations.

Furthermore, we employ a meta-learning strategy to efficiently fine-tune the entire model, thereby improving its generalization capabilities. Our work is motivated by the demands of industrial image segmentation, where precise segmentation is crucial for detecting semantic anomalies. We achieved remarkable results on internal datasets.

Based on joint work with: Gabriele Lombardi, Mirco Planamente and Federico Pozzi.

References

[1] Ravi, N., Gabeur, V., Hu, Y.T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L. and Mintun, E. *Sam 2: Segment anything in images and videos*, arXiv preprint arXiv:2408.00714

[2] De Marinis, P., Fanelli, N., Scaringi, R., Colonna, E., Fiameni, G., Vessio, G. and Castellano, G. *Label Anything: Multi-Class Few-Shot Semantic Segmentation with Visual Prompts*, arXiv preprint arXiv:2407.02075

GANs through the Lens of Topological Data Analysis

Caterina Gallegati

University of Siena, c.gallegati@student.unisi.it

B. T. Corradini[a], B. Cullen[b], S. Marziali[c], G. A. D'Inverno[d], M. Bianchini[c], F. Scarselli[c]

[a] University of Florence, [b] University of Pisa, [c] University of Siena, [d] SISSA, Trieste

Generative Adversarial Networks (GANs) [1] aim to produce realistic samples by mapping a low-dimensional latent space to a high-dimensional data space by exploiting an adversarial training mechanism. Despite achieving state-of-the-art results, GAN training faces significant challenges such as mode collapse, vanishing gradients, and inefficiencies in hyperparameter tuning, relying on computationally expensive trial-and-error methods. In addition, GANs lack a clear early stopping criterion, often leading to resource-intensive training processes.

This work investigates GANs using Topological Data Analysis (TDA) tools [3] to gain deeper insights into their training dynamics and generative capabilities. By employing persistent homology, we examine the evolution of topological features during training, focusing on the convergence of the generated manifold to that of real data. Through various experiments on MNIST and CIFAR-10 datasets with different GAN models, we analyze the interplay between model architecture, training stability, and performance, as well as characterise common issues in GANs. In particular, we show that the Wasserstein distance between persistence diagrams, which summarise the topological features of manifolds, is a robust tool for quantifying similarities between generated and real data, offering a novel perspective on evaluating samples beyond conventional metrics like the Frechet Inception Distance (FID) [2]. Indeed, the FID score is shown to be insufficient in assessing the quality of generated images, neither alone nor in combination with the Intrinsic Dimension estimation [4]. Our results suggest that homological features provide a suitable characterisation of the generative process that can be valuable for uncovering insights about the structural transformations occurring during the training of a GAN. This study lays the foundation for integrating topology-based approaches into the optimization and assessment of generative models, potentially enabling the formulation of an early stopping criterion.

References

[1] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. & Bengio, Y. Generative Adversarial Networks. (arXiv,2014), https://arxiv.org/abs/1406.2661

[2] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B. & Hochreiter, S. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. (arXiv,2017), https://arxiv.org/abs/1706.08500

[3] Chazal, F. & Michel, B. An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists. (arXiv,2017), https://arxiv.org/abs/1710.04019

[4] Pope, P., Zhu, C., Abdelkader, A., Goldblum, M. & Goldstein, T. The Intrinsic Dimension of Images and Its Impact on Learning. (arXiv,2021), https://arxiv.org/abs/2104.08894

CAP: Copyright Audit via Prompt generation

Daniela Gallo

ICAR-CNR and University of Salento, Italy, daniela.gallo@icar.cnr.it

Angelica Liguori

ICAR-CNR, Italy, angelica.liguori@icar.cnr.it

Ettore Ritacco

University of Udine, Italy, ettore.ritacco@uniud.it

Luca Caviglione

IMATI-CNR, Italy, luca.caviglione@ge.imati.cnr.it

Fabrizio Durante

University of Salento, Italy, fabrizio.durante@unisalento.it

Giuseppe Manco

ICAR-CNR, Italy, giuseppe.manco@icar.cnr.it

To achieve accurate and unbiased predictions, Machine Learning (ML) models rely on large, heterogeneous, and high-quality datasets. However, this could raise ethical and legal concerns regarding copyright and authorization aspects, especially when information is gathered from the Internet. Indeed, such data may be protected by intellectual property rights, and proper authorizations for its usage should be granted on a case-by-case basis [1]. With the rise of generative models, being able to track data has become of particular importance. Indeed, as they require large datasets for being trained, they often rely on data derived from different sources without being able to discriminate among public or "restricted" sources. Consequently, they may (un)intentionally replicate copyrighted contents [2]. To this aim, we propose **C**opyright **A**udit via **P**rompts generation (**CAP**), a framework for automatically checking if the training set used by an ML model contains unauthorized data. Testing whether data has been used to train an ML model is known as *membership inference* problem. However, different from classical Membership Inference Attacks [3] that directly check if a given slice of information has been used in the training phase, we cannot directly inspect the training set used by the model, as only the owner knows it. To address this issue, CAP generates suitable keys that induce the model to reveal copyrighted content. Additionally, training prompt generators, which rely on complex architectures like transformers, require large computational demands. For this reason, we introduce an optimization procedure aiming to speed up the learning process. By leveraging a generalized Pareto distribution [4], we filter out irrelevant data based on model error, applying an 80% threshold to exclude extreme outliers. This reduces the dataset size while preserving the most impactful samples. Extensive evaluations across four realistic IoT scenarios and synthetic datasets demonstrate the effectiveness of our framework in identifying unauthorized data with high accuracy. This work offers a robust and efficient solution for ensuring responsible and ethical use of generative artificial intelligence models.

References

[1] Meuris, B., Qadeer, S. & Stinis, P. Machine-learning custom-made basis functions for partial differential equations. (arXiv,2021), https://arxiv.org/abs/2111.05307

[2] Li, H., Deng, G., Liu, Y., Wang, K., Li, Y., Zhang, T., Liu, Y., Xu, G., Xu, G. & Wang, H. Digger: Detecting Copyright Content Mis-usage in Large Language Model Training. (arXiv,2024), https://arxiv.org/abs/2401.00676

[3] Shokri, R., Stronati, M., Song, C. & Shmatikov, V. Membership Inference Attacks against Machine Learning Models. (arXiv,2016), https://arxiv.org/abs/1610.05820

[4] Vignotto, E. & Engelke, S. Extreme value theory for anomaly detection – the GPD classifier. *Extremes*. **23**, 501-520 (2020,9), http://dx.doi.org/10.1007/s10687-020-00393-0

A Low-Rank Multi-Factor Approach to Identify Differentially Expressed Genes in Transcriptome Data

Grazia Gargano

Department of Mathematics, University of Bari Aldo Moro, Italy, grazia.gargano@uniba.it

RNA-sequencing (RNA-seq) technology provides a robust platform for transcriptome-wide analysis of gene expression, enabling the study of transcriptional changes associated with different biological conditions. A primary application of RNA-seq is the identification of differentially expressed genes (DEGs) between different biological states (e.g., disease vs. normal, treatment vs. control), cell populations, or time points. DEGs are defined as genes that show statistically significant differences in expression levels or read counts across experimental conditions, reflecting changes in gene activity associated with biological processes such as disease progression, therapeutic response or developmental pathways. Despite their critical role in transcriptomic research, further advances are needed to improve the accuracy and efficiency of DEGs identification.

We propose a novel mathematical framework based on three-factor nonnegative matrix factorization (tri-NMF) [1] to identify genes that exhibit differential expression under two or more distinct experimental conditions. We represent gene expression data by a matrix $\mathbf{X} \in {\mathbb{R}_+}^{n\times m}$, where n is the number of samples (*e.g.*, patient groups, tissues, experiments, or time points) and m is the number of genes. To compare experimental conditions, we introduce a tri-NMF-based approach, formulated as a constrained penalized optimization task:

$$\min_{\mathbf{U}\geq 0,\mathbf{S}\geq 0,\mathbf{V}\geq 0} Div(\mathbf{X}|\mathbf{U}\mathbf{S}\mathbf{V}^\top) + \lambda_{\mathbf{U}}\mathcal{P}_1(\mathbf{U}) + \lambda_{\mathbf{S}}\mathcal{P}_2(\mathbf{S}) + \lambda_{\mathbf{V}}\mathcal{P}_3(\mathbf{V})$$

where $Div(\cdot,\cdot) : \mathbb{R}_+^{n\times m} \times \mathbb{R}_+^{n\times m} \to \mathbb{R}_+$ denotes some divergence function, which evaluates the goodness of fitting; $\mathbf{U} \in {\mathbb{R}_+}^{n\times k}$, $\mathbf{S} \in {\mathbb{R}_+}^{k\times r}$ and $\mathbf{V} \in {\mathbb{R}_+}^{m\times r}$ are the nonnegative factors of low-rank data representation; $\mathcal{P}_1 : \mathbb{R}^{n\times k} \to \mathbb{R}$, $\mathcal{P}_2 : \mathbb{R}^{k\times r} \to \mathbb{R}$, $\mathcal{P}_3 : \mathbb{R}^{m\times r} \to \mathbb{R}$ codify regularization constraints to enforce specific properties on the factor matrices, while $\lambda_{\mathbf{U}}$, $\lambda_{\mathbf{S}}$ and $\lambda_{\mathbf{V}}$ are some positive regularization parameters.

For DEGs identification, we consider the generalized Kullback-Leibler divergence as the cost function and set $k = r$ $(k, r < min(n, m))$ equal to the number of different conditions we want to compare. The information about the sample labels is encoded in the structure of the factor $\mathbf{U}$. We impose $\mathbf{U}$ to be a binary matrix representing sample clusters, where $\mathbf{U}_{ij} \in \{0, 1\}$ and $\sum_{j=1}^{k} \mathbf{U}_{ij} = 1$. This ensures that each sample is assigned to exactly one cluster. Imposing sparsity and orthogonality constraints on the columns of $\mathbf{V}$ ensures that the extracted list of DEGs has minimal or no overlap of genes. The objective function is minimized by using an alternating scheme with an appropriate choice of the multiplicative update rules [2]. To compute DEGs, we define a gene score criterion based on the normalized entropy, which is computed from the coefficients of the matrix $\mathbf{V}$ obtained during the factorization. We validate our approach on synthetic data to assess its performance and robustness under controlled conditions. Synthetic datasets are generated to simulate realistic biological scenarios, allowing us to test the model's ability to accurately identify DEGs.

This is a joint work with Nicoletta Del Buono and Flavia Esposito (Department of Mathematics, University of Bari Aldo Moro, Bari, Italy).

References

[1] Nicolas Gillis, *Nonnegative Matrix Factorization*, SIAM, Philadelphia, 2020.

[2] Daniel Lee and Hyunjune Seung, *Algorithms for Non-negative Matrix Factorization*, Advances in Neural Information Processing Systems (NeurIPS), Volume 13, 2001.

Analysis of Decision-Making Styles and Personality Traits in Women Undergoing Voluntary Termination of Pregnancy: A Bayesian Network Approach Using bnstruct

Letizia Lorusso
School of Medical Statistics and Biometry, Interdisciplinary Department of Medicine,

University of Bari "Aldo Moro", Bari, Italy. letizia.lorusso@uniba.it

In this study, we explore the application of Bayesian Networks to analyze the relationships between the General Decision-Making Style (GDMS) test 1, the Big Five Questionnaire (BFQ) 2 and the Personality Inventory for DSM-5 (PID-5) 3, and socio-demographic characteristics of women who undergo voluntary termination of pregnancy (VTP). Using the *bnstruct* package 4 for building Bayesian Networks, our goal is to compare the results of different alghortim applied with three scoring fuction, to define significant patterns that can reveal the underlying dynamics of these choices, considering variables such as personality type and decision-making aspects related to this experience. The data used comes from a database containing socio-demographic information of 122 women, as well as their personality and decision-making test results, with a total of 27 variables.
To this end, we construct a Bayesian network representing the probabilistic dependencies among the variables and compare the performance of four algorithms for structure learning: Structural Expectation-Maximization (SEM), Max-Min Parents-and-Children (MMPC), Max-Min Hill-Climbing (MMHC), and Hill-Climbing (HC). Each algorithm employs a different approach to structure learning, and we assess their effectiveness in identifying the most accurate causal relationships in our data 5.
Additionally, we compare the performance of the model using three main scoring methods: Bayesian Dirichlet equivalent score (BDeu), Bayesian Information Criterion (BIC), and Akaike Information Criterion (AIC). These scoring functions are employed to evaluate the quality of the model and determine which approach provides the best representation of the data.
In general, BDeu is particularly well-suited for data with discrete variables. AIC and BIC penalizes complexity, choosing the variables in the simpler models with good predictive ability. We evaluate these scoring functions in the context of the four algorithms. At the first we use the MMPC alghoritm to explore the conditional dependencies among variables, and to define the skeleton of the network, without directly optimizing a global score. Then, we applied the HC and the MMHC alghoritm: the first refines the network structure by selecting the best local changes to maximize the scoring function; the second identifies the parent-child relationships. At the end we use the SEM alghoritm, optimizing both the model's structure and its parameters simultaneously, to choose the final model.
The results of this comparison provide valuable insights into which algorithm and scoring function best capture the relationships among personality, decision-making style, and socio-demographic factors in the context of VTP decisions.

References

[1] Di Fabio, A. (2007). *General Decision Making Style (GDMS): Un primo contributo alla validazione italiana.*, GIPO, Giornale Italiano di Psicologia dell'Orientamento, 8(3), 17-25

[2] Caprara, G. V., Barbaranelli, C., Borgogni, L., & Perugini, M. (1993). *The Big Five Questionnaire: A new questionnaire to assess the five-factor model.* , Personality and Individual Differences, 15, 281–288. http://dx.doi.org/10.1016/0191-8869(93)90218-R

[3] Fossati, A., Krueger, R. F., Markon, K. E., Borroni, S., & Maffei, C. (2013). *Reliability and validity of the Personality Inventory for DSM-5 (PID-5) predicting DSM-IV personality disorders and psychopathy in community-dwelling Italian adults.*, Assessment, 20, 689 –708

[4] Sambo, F., & Franzin, A. (2015). *bnstruct: Bayesian Network Structure Learning from Data with Missing Values (p. 1.0.15)*, https://doi.org/10.32614/CRAN.package.bnstruct

[5] Alberto Franzin, Francesco Sambo, Barbara di Camillo. (2017) *"bnstruct: an R package for Bayesian Network structure learning in the presence of missing data."*, Bioinformatics, 33 (8): 1250-1252; Oxford University Press.

COSMONET 2.0: An R Package for Survival Analysis Using Screening-Network Methods

Maura Mecchi

University of Basilicata, Potenza, Italy maura.mecchi@gmail.com

Co-author: Antonella Iuliano

University of Basilicata, Potenza, Italy antonella.iuliano@unibas.it

Network-based methods are becoming increasingly crucial in precision oncology and healthcare. The advent of high-throughput technologies, coupled with advancements in the quantitative analysis of biomolecular data, has created new opportunities to investigate the mechanisms driving the onset and progression of complex diseases.

However, in this high-dimensional setting, several challenges arise. These include data heterogeneity, limited samples relative to the number of variables, multicollinearity between variables, and the need to integrate a priori biological information into the analysis. Equally important are the interpretation and validation of the results, which are essential for ensuring the reliability and clinical relevance of the findings.

Innovative statistical approaches are being developed to address some of these challenges. These methods aim to improve the accuracy and robustness of data analysis, enabling more reliable insights into complex biological processes and disease mechanisms. Among these, COSMONET (COx Survival Methods based On NETworks), introduced in [1], is an R package that integrates both biologically driven and data-driven screening techniques within a network-penalized Cox regression model. This approach allows for more accurate identification of key biomarkers while accounting for the complex interdependencies in biological networks (see [2, 3]). Here, we present COSMONET 2.0, an extended version that provides a comprehensive workflow, covering the entire process from data preprocessing to gene signature selection and survival outcome prediction. This enhanced version incorporates additional features, such as clinical variables. It includes implementation improvements that support more robust analysis, enabling the practical application of network-based methods to multi-omics data in survival analysis. In addition, COSMONET 2.0 introduces new functions for data preprocessing, visualization, survival prediction, and gene enrichment analysis, making it a powerful tool for integrating omics data in cancer survival analysis. These enhancements enable a more comprehensive approach to understanding the molecular underpinnings of cancer and predicting patient outcomes with increased accuracy and reliability. Moreover, the new version of the software is significantly faster in terms of computational costs.

We illustrate the package's efficiency using several cancer datasets from the GDC data portal (https://portal.gdc.cancer.gov) to evaluate its prediction accuracy under a large set of conditions. Various performance measures, including the concordance index (C-index) and other relevant metrics, are applied to assess the package's ability to reliably predict survival outcomes.

References

[1] Iuliano, A., Occhipinti, A., Angelini, C., De Feis, I., & Lió, P. (2021). *Cosmonet: An R package for survival analysis using screening-network methods*, Mathematics, 9(24), 3262.

[2] Fan, J., Feng, Y., & Wu, Y. (2010). *High-dimensional variable selection for Cox's proportional hazards model*, In Borrowing strength: Theory powering applications–a Festschrift for Lawrence D. Brown (Vol. 6, pp. 70-87). Institute of Mathematical Statistics.

[3] Sun, H., Lin, W., Feng, R., & Li, H. (2014). *Network-regularized high-dimensional Cox regression for analysis of genomic data*, *Statistica Sinica, 24(3), 1433.*

Efficiency-driven 3D CNN architectures for hyperspectral classification

Giuseppina Monteverde

Department of Basic and Applied Sciences for Engineering - Sapienza University of Rome

giuseppina.monteverde@uniroma1.it

Vittoria Bruni, Domenico Vitulano

Department of Basic and Applied Sciences for Engineering - Sapienza University of Rome

vittoria.bruni@uniroma1.it, domenico.vitulano@uniroma1.it

Hyperspectral imaging enables the simultaneous capture of spatial and spectral information across multiple wavelengths, yielding high-dimensional data suitable for a wide range of applications. 3D Convolutional Neural Networks (CNNs) can completely exploit the hyperspectral data structure through 3D convolutional filters, which jointly extract spatial and spectral features. This process improves classification performance by increasing intraclass variation and reducing interclass variation [1]. On the other side, the high computational cost of deep CNN architectures — both in terms of resource consumption and training time — when processing such high-dimensional data necessitates optimization techniques. These can be approached through dimensionality reduction or more efficient network architectures [2]. The former reduces the input dimensionality by transforming the data into a lower-dimensional yet representative form, while the latter focuses on streamlining the network architectures.

Two distinct approaches for enhancing hyperspectral classification efficiency using 3D CNNs are proposed. The first method employs feature extraction, projecting the data in a proper domain and automatically selecting relevant components in the transformed space based on the entropic normalized information distance. This approach is an adaptive and automatic method where the number of features to be selected is not pre-defined but automatically given [3]. The second methodology focuses on determining the filters size setting of convolutional layers in a 3D CNN, guided by Heisenberg's uncertainty principle. This principle inspires a rule for relating the spatial and spectral dimensions of convolutional filters as the network depth increases, enabling the network to learn discriminative features that capture both fine spatial resolution and broad spectral characteristics [4]. The effectiveness of CNNs in the proposed approaches is assessed using both raw and transformed input data. Both the features selected by the entropy-based method and the architectures with Heisenberg-based cascaded filter setting demonstrate a significant reduction in training time while preserving high classification accuracy. These strategies provide solutions for processing hyperspectral data, aimed at enhancing operational efficiency.

References

[1] M. Ahmad, A. M. Khan, M. Mazzara, S. Distefano, M. Ali and M. S. Sarfraz, *A Fast and Compact 3DCNN for Hyperspectral Image Classification*, in IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, 2022

[2] H. Fırat, M. E. Asker, D. Hanbay, *Classification of hyperspectral remote sensing images using different dimension reduction methods with 3D/2D CNN*, Remote Sensing Applications: Society and Environment, vol. 25, 2022

[3] V. Bruni, G. Monteverde, D. Vitulano, *An Entropy Based Speed Up For Hyperspectral Data Classification Via CNNn*, in 2022 12th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2022

[4] V. Bruni, G. Monteverde, D. Vitulano, *Heisenberg principle-inspired filters and size setting in 3D CNN for hyperspectral data classification*, accepted for publication in 2024 14th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2024

Bi-level algorithm for optimizing hyperparameters in penalized NMF

Laura Selicato

Water Research Institute – National Research Council (IRSA-CNR), Viale Francesco de Blasio, 5,

Bari, Italy laura.selicato@uniba.it

Over the past decade, machine learning has emerged as one of the main innovation drivers. Its research community is expanding at an unprecedented speed, thanks to the growing need to build accurate, reliable, and interpretable models that respond to the multitude of data generated. All the learning algorithms require the configuration of hyperparameters (HPs), i.e., parameters that govern the learning approach. HPs tuning is a crucial problem in the field of the learning process since the selection of the HPs has an important impact on the final performance of the algorithm. The main goal of the hyperparameter optimization (HPO) problem is to automate the search process, thereby improving the generalization performance of the model and enabling a more flexible design of the underlying learning algorithms. A reliable approach is to transform the HPO into a bi-level optimization problem that can be solved by gradient descent techniques. The challenge is the estimation of the gradient with respect to the HPs. In this work, we present a new mathematical framework for solving the HPO in Nonnegative Matrix Factorization (NMF) based on bi-level techniques, focusing on penalty HPs, which turn out to be useful to emphasize intrinsic properties in the data, such as sparsity. We design a novel algorithm, named Alternating Bi-level (AltBi), which incorporates the HPO into the updates of NMF factors. Finally, we provide results of the existence and convergence of solutions with also numerical experiments.

This is a joint work with Nicoletta Del Buono and Flavia Esposito (Department of Mathematics, University of Bari Aldo Moro, Bari, Italy).

Hybrid knowledge and data-driven approaches for Diffuse Optical Tomography reconstruction

Alessandra Serianni

University of Milan, alessandra.serianni@unimi.it

Diffuse Optical Tomography (DOT) is a non-invasive medical imaging technique which employs Near-Infrared (NIR) light to recover the spatial distribution of optical coefficients in biological tissues. Due to the limited availability of boundary measurements and the intense light scattering, DOT reconstruction is a severely ill-posed problem [1]. Recently, the success of deep learning methods has shifted the focus of tomographic imaging from purely knowledge-driven to data-driven approaches.
In this contribution, we propose a hybrid approach that combines model-based and deep learning techniques. Our idea is to leverage Graph Neural Networks (GNNs), that -once trained- we use as a fast forward model that solves partial differential equations, into an iterative optimization-based method for solving the inverse problem. Due to the severe ill-conditioning of the reconstruction problem, we also learn a prior over the space of solutions using an autoencoder-type neural network which maps the latent code to the estimated physical parameter, that is passed to the GNN to obtain the prediction. The latent code is finally optimized to minimize the difference between the recorded and predicted data.
By optimizing the latent code, we constrain the solution space to the manifold learned by the generative model. In order to add greater structure and meaning to the latent space, we learn a compact and non-degenerate intrinsic manifold basis [2] and the rank of the covariance matrix of the latent space is implicitly minimized [3], while encouraging better reconstructions.

References

[1] A. Benfenati, G. Bisazza, P. Causin, *A Learned SVD approach for Inverse Problem Regularization in Diffuse Optical Tomography*, arXiv preprint arXiv:2111.13401, (2021)

[2] K. Flouris, E. Konukoglu, *Canonical normalizing flows for manifold learning* Proceedings of the 37th International Conference on Neural Information Processing Systems, (2024), pp. 27294 - 27314

[3] J Mounayer, S Rodriguez, C Ghnatios, C Farhat, F Chinesta, *Rank Reduction Autoencoders–Enhancing interpolation on nonlinear manifolds*, arXiv preprint arXiv:2405.13980, (2024)

Spatial Informed Hierarchical Clustering for Hyperspectral Imagery via Total Variation

Gaetano Settembre*,[a], Nicoletta Del Buono[a], Flavia Esposito[a]
Nicolas Gillis[b]

[a] Department of Mathematics, University of Bari Aldo Moro, Italy
[b] Faculty of Engineering, University of Mons, Belgium

{gaetano.settembre, nicoletta.delbuono, flavia.esposito}@uniba.it,
nicolas.gillis@umons.ac.be

Hierarchical clustering algorithms offer powerful tools for hyperspectral image analysis, reflecting the inherent hierarchical structure of materials within images. Despite their potential, existing models often neglect critical image properties, such as the spatial similarity and proximity of neighboring pixels. Building on the H2NMF algorithm proposed in [1], which employs a rank-two nonnegative matrix factorization for binary cluster splitting, we propose two key improvements to enhance clustering performance.

Firstly, we refine the estimation of the basis matrix W. While the original approach relies on the successive projection algorithm, we employ more robust and advanced variants such as the smoothed successive projection algorithm (SSPA) and the smoothed vertex component analysis (SVCA) [2]. These methods address the limitations of the pure pixel assumption by better identifying the vertices of the convex hull of the data, even in noisy conditions.

Secondly, we incorporate Total Variation (TV) regularization [3] into the objective function to improve the estimation of the coefficient matrix H. This regularization exploits the spatial structure within hyperspectral images, promoting smoother and spatially coherent solutions while preserving critical edge information. The new objective function is defined as:

$$\min_{H \geq 0} \|X - WH\|_F^2 + \lambda \sum_{\ell=1}^{r} \|SH(\ell, :)\|_1,$$

where X represents the original hyperspectral image, W is derived from the aforementioned methods. In our case $r = 2$ and $S \in \mathbb{R}^{K \times n}$ is a sparse matrix encoding pixel neighborhood relationships such that $S(k, i) = 1$ and $S(k, j) = -1$ for some k if pixels i and j are neighbors. We solve this new optimization problem using an iterative gradient-based approach.

Several experiments are conducted on different real remote sensing hyperspectral datasets (e.g., Cuprite, Urban, Samson, etc) to evaluate the convergence curve of the algorithm, and then the effectiveness of our new proposed clustering method.

References

[1] Gillis, N., Kuang, D. & Park, H. Hierarchical Clustering of Hyperspectral Images Using Rank-Two Nonnegative Matrix Factorization. *IEEE Transactions On Geoscience And Remote Sensing.* **53**, 2066-2078 (2015,4), DOI: 10.1109/TGRS.2014.2352857.

[2] Nadisic, N., Gillis, N. & Kervazo, C. Smoothed separable nonnegative matrix factorization. *Linear Algebra And Its Applications.* **676** pp. 174-204 (2023,11), DOI: 10.1016/j.laa.2023.07.013.

[3] Rudin, L., Osher, S. & Fatemi, E. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena.* **60**, 259-268 (1992,11), DOI: 10.1016/0167-2789(92)90242-F.

Empowering Clinicians with Explainable AI: Predicting Mortality Risk in MAFLD with Counterfactual Analysis

Paolo Sorino, Domenico Lofù, Rossella Donghia, Caterina Bonfiglio, Gianluigi Giannelli, and Tommaso Di Noia

Address paolo.sorino@poliba.it, domenico.lofu@poliba.it,

rossella.donghia@irccsdebellis.it, catia.bonfiglio@irccsdebellis.it,

gianluigi.giannelli@irccsdebellis.it, tommaso.dinoia@poliba.it

Metabolic Dysfunction Associated with Fatty Liver Disease (MAFLD) represents a paradigm shift in liver disease classification, moving from the concept of a *"non-condition"* to an inclusive diagnostic entity. Introduced in 2020, MAFLD diagnosis is based on the presence of hepatic steatosis along with one of three metabolic conditions: overweight or obesity *(Subtype 1)*, metabolic dysregulation in lean individuals *(Subtype 2)*, or diabetes mellitus *(Subtype 3)*. As MAFLD is increasingly recognized as a public health concern, there is an urgent need for innovative approaches to improve early detection and management. In this context, Machine Learning (ML) has emerged as a game-changing technology in modern clinical practice, offering the capability to extract actionable insights from complex, high-dimensional datasets. By leveraging sophisticated algorithms, ML enables clinicians to address critical challenges such as early disease diagnosis, accurate risk stratification, and the development of personalised treatment strategies, making ML an indispensable tool for tackling multifaceted health problems such as MAFLD. To address the early identification of high-risk patients, we developed **MORIX**, an artificial intelligence-based framework for predicting mortality risk in individuals with MAFLD. The study cohort consisted of $1,675$ subjects (543 females and $1,132$ males) aged > 30 years, diagnosed with MAFLD and recruited between May 2005 and January 2007 from the National Institute of Gastroenterology, IRCCS 'S. De Bellis' in Castellana Grotte (Italy). The cohort was observed until December 31, 2023. Using this dataset, which included anthropometric and biochemical parameters, we applied Recursive Feature Elimination (RFE) with a Random Forest (RF) model to select the most relevant features. These features were then used to train and evaluate five machine learning algorithms—Random Forest (RF), eXtreme Gradient Boosting (XGB), Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Light Gradient Boosting Machine (LGBM)—using a 5-fold cross-validation approach. Among the tested models, RF demonstrated the highest performance, achieving an accuracy of 83%, with a precision and recall of 83% for mortality prediction, and an F1 score of 0.83. The Area Under the ROC Curve (AUC) was 0.88, confirming the RF model's ability to effectively distinguish between high- and low-risk patients. In comparison, XGB and SVM achieved slightly lower accuracies of 82% and 80%, while MLP and LGBM showed weaker results overall.

In addition, explainability was a core component of the MORIX framework. Explainable Artificial Intelligence (XAI) techniques, specifically Shapley Additive exPlanations (SHAP), were applied to the RF model to make the decision-making process transparent. SHAP values revealed that **age** and **blood glucose** were the most critical predictors of mortality, providing clinicians with clear insights into the model's decision-making process.

Furthermore, MORIX includes a counterfactual analysis feature, enabling clinicians to simulate "$what - if''$ scenarios. For instance, modifying biochemical parameters, such as cholesterol or weight, allows users to observe how these changes influence the predicted mortality risk. This capability offers actionable insights, supporting targeted interventions to improve patient outcomes.

To ensure accessibility, we developed a user-friendly web application that integrates the trained RF model. This application enables healthcare professionals to input new patient data, receive real-time mortality risk predictions, and access detailed explanations of the model's decisions.

In conclusion, MORIX exemplifies how ML can bridge the gap between complex data and practical clinical applications. By combining robust predictive performance with explainable AI and counterfactual analysis, MORIX offers a valuable tool for clinicians to make informed, data-driven decisions. Its integration into clinical workflows has the potential to enhance patient care by identifying high-risk MAFLD patients early and providing actionable insights into improving outcomes. Future work will focus on expanding the dataset to include additional clinical variables and exploring the use of Deep Learning (DL) to further enhance model performance.