Partial Differential Equations, Inverse Problems and Machine Learning

Equipo organizador

- Carlos Esteve Yagüe (Universitat d'Alacant)
- Domènec Ruiz i Balet (Universitat de Barcelona)

Descripción

This minisymposium explores recent advances at the intersection of partial differential equations, control theory, and machine learning, with a particular focus on applications to neural architectures such as Transformers and generative models. Several talks address the mathematical structure of self-attention and its mean-field and gradient flow limits, unveiling phenomena such as metastability and clustering. Others examine the control of probability distributions via continuity equations, variational methods, and entropy-based metrics. Novel perspectives on numerical methods for partial differential equations and inverse problems further broaden the scope of this session. Altogether, the symposium highlights the deepening dialogue between classical analysis and modern data-driven modeling.

Palabras clave: Partial Differential Equations; Mean-field models; Inverse Problems; Neural Networks; Transformers.

Programa

JUEVES, 22 de enero

11:00 – 11:30	Gissell Estrada-Rodríguez (Universitat Politècnica de Catalunya) *Nonlocal interaction kernels inference in nonlinear gradient flow equations*
11:30 – 12:00	Johannes Schwab (University of Innsbruck) *Solving inverse problems with diffusion models*
12:00 – 12:30	Nadja Gruber (Medical University of Innsbruck) *Noisier2Inverse: Self-Supervised Learning for Image Reconstruction with Correlated Noise*
12:30 – 13:00	Jon Asier Bárcena Petisco (Universidad del País Vasco) *Resolution of parametric elliptic PDEs via Neural Networks*
15:30 – 16:00	Valérie Castin, (École Normale Supérieure de Paris) *The mean-field dynamics of deep Transformers*
16:00 – 16:30	Anna Shalova (TU Eindhoven) *Variational analysis of toy Transformers*
16:30 – 17:00	Hugo Koubbi (Université París Dauphine) *Dynamic metastability in the self attention model*
17:00 – 17:30	Yury Korolev (University of Bath) *Large-time dynamics in transformer architectures with layer normalisation*
18:00 – 18:30	Ponente (Universidad) *Título*
18:30 – 19:00	Ponente (Universidad) *Título*

VIERNES, 23 de enero

11:00 – 11:30	Francisco Periago (Universidad Politécnica de Cartagena) *Universal approximation of set-valued maps and application to control*
11:30 – 12:00	Carlos Castro (Universidad Politécnica de Madrid) *Numerical approximation of the 3-D Navier-Lamé system in an exterior domain using the fundamental solutions method*
12:00 – 12:30	Antonio Álvarez López (Universidad Autónoma de Madrid) *Entropy-driven control of the continuity equation for normalizing flows*

Nonlocal interaction kernels inference in nonlinear gradient flow equations

Gissell Estrada-Rodríguez

Department of Mathematics, Universitat Politècnica de Catalunya

gissell.estrada@upc.edu

Resumen. When applying nonlinear aggregation-diffusion equations to model real life phenomenon, a major challenge lies on the choice of the interaction potential. Previous numerical and theoretical studies typically required predetermination of terms and the goal is often to reproduce the observed dynamics qualitatively, not quantitatively. In this talk, we address the inverse problem of identifying nonlocal interaction potentials in nonlinear aggregation-diffusion equations from noisy discrete trajectory data. Our approach involves formulating and solving a regularised variational problem, which requires minimising a quadratic error functional across a set of hypothesis functions. A key theoretical contribution is our novel stability estimate for the PDE, validating the error functional ability in controlling the 2-Wasserstein distance between solutions generated using the true and estimated interaction potentials. We demonstrate the effectiveness of the methods through various 1D and 2D examples showcasing collective behaviours.

Solving inverse problems with diffusion models

Johannes Schwab

Department of Mathematics, University of Innsbruck

johannes.schwab@uibk.ac.at

Resumen. Inverse problems arise in many scientific domains where the goal is to recover signals from indirect or incomplete measurements. The main difficulty lies in the ill-posedness of the problem. Traditional approaches typically rely on handcrafted priors or regularization techniques, which can struggle with complex data distributions. Recently, diffusion models have emerged as powerful generative tools capable of learning intricate data priors from large-scale datasets.

In this talk, we explore how diffusion models can be leveraged to solve inverse problems in a principled way. We present a framework where diffusion models serve as implicit priors, enabling high-quality reconstructions while at the same time guarantee consistency with the measured data. Applications include image deblurring, super-resolution, and medical imaging, among others. We discuss the theoretical underpinnings, practical implementation strategies, and show initial results which are compared to classical and deep learning baselines.

Noisier2Inverse: Self-Supervised Learning for Image Reconstruction with Correlated Noise

Nadja Gruber

Faculty of AISCM, Medical University of Innsbruck

nadja.gruber@i-med.ac.at

Resumen. In this talk, I will present Noisier2Inverse, a novel self-supervised deep learning framework for solving general inverse problems—without the need for clean ground truth data. Unlike classical methods that rely on uncorrelated noise assumptions, Noisier2Inverse is specifically designed for statistically correlated noise, which is common in applications like computed tomography, microscopy, and seismic imaging.

Building on ideas from Noisier2Noise, our method leverages synthetic noise augmentation to train a reconstruction network. However, instead of learning to recover the original noisy data, we train directly in the measurement space to reconstruct an extrapolated image. This key innovation removes the need for a separate extrapolation step at inference time—avoiding the instability often introduced by ill-posedness.

We show that Noisier2Inverse significantly outperforms existing self-supervised approaches when it comes to handling complex, correlated noise structures, making it a robust solution for real-world imaging tasks where high-quality ground truth is not available.

Resolution of parametric elliptic PDEs via Neural Networks

Jon Asier Bárcena Petisco

Departamento de Matemáticas, Universidad del País Vasco

jonasier.barcena@ehu.eus

Resumen. In this talk we construct a Neural Network that approximates the matrix multiplication operator for any activation function such that there exists a Neural Network which can approximate the scalar multiplication function. In particular, we use the Strassen algorithm to bound the number of weights and layers needed for such Neural Networks. This allows us to define another Neural Network for approximating the inverse matrix operator. Also, by relying on the Galerkin method, we apply those Neural Networks to solve parametric elliptic PDEs for a whole set of parameters. Finally, we discuss improvements with respect to the prior results.

The mean-field dynamics of deep Transformers

Valérie Castin

Département de Mathématiques et Applications, Ecole Normale Supérierure de Paris

valerie.castin@ens.psl.eu

Resumen. Transformers, that underlie the recent successes of large language models, represent the data as sequences of vectors called tokens. This representation is leveraged by the attention function, which learns dependencies between tokens and is key to the success of Transformers. However, the dynamics induced by the iterative application of attention across layers remain to be fully understood. To analyze these dynamics, we identify each input sequence with a probability measure, thus handling input sequences of arbitrary length, and model its evolution as a Vlasov equation called Transformer PDE, whose velocity field is non-linear in the probability measure. For compactly supported initial data and several self-attention variants, we show the Transformer PDE is well-posed and is the mean-field limit of an interacting particle system. We also study the case of Gaussian initial data, which has the nice property of staying Gaussian across the dynamics. This allows us to identify typical behaviors theoretically and numerically, and to highlight a clustering phenomenon that parallels previous results in the discrete case.

Variational analysis of toy transformers

Anna Shalova

Departament of Mathematics and Computer Sciences, Technische Universiteit Eindhoven

a.shalova@tue.nl

Resumen. Transformers are, arguably, the state-of-the-art models in natural language processing and various other applications. While transformers are widely used in practice, the theoretical understanding of their behavior is very limited. However, recently, the clustering phenomena in toy transformers has been explained by Geshkovski et al. (2024). While being one of the key phenomena arising in transformers, clustering discourages variability of predictions and thus cannot be the only driving mechanism behind transformers. In this talk I will discuss two aggregation-diffusion models which can be interpreted as extensions of the toy transformers mentioned above. I will give a variational characterization of solutions to the underlying PDEs and show that the diffusive terms can be interpreted as the mechanisms promoting diversity of predictions in transformers. The talk is based on joint works with André Schlichting and Mark Peletier.

Dynamic metastability in the self attention model

Hugo Koubbi

CEREMADE, Université Paris Dauphine

hugo.koubbi@yale.edu

Resumen. We consider the self-attention model—an interacting particle system on the unit sphere, which serves as a toy model for Transformers, the deep neural network architecture behind the recent successes of large language models. We prove the appearance of dynamic metastability conjectured in [**?**]—although particles collapse to a single cluster in infinite time, they remain trapped near a configuration of several clusters for an exponentially long period of time. By leveraging a gradient flow interpretation of the system, we also connect our result to an overarching framework of slow motion of gradient flows proposed by Otto and Reznikoff [**?**] in the context of coarsening and the Allen-Cahn equation. We finally probe the dynamics beyond the exponentially long period of metastability, and illustrate that, under an appropriate time-rescaling, the energy reaches its global maximum in finite time and has a staircase profile, with trajectories manifesting saddle-to-saddle-like behavior, reminiscent of recent works in the analysis of training dynamics via gradient descent for two-layer neural networks.

Referencias

[1] B. Geshkovski, C. Letrouit, Y. Polyanskiy and Ph. Rigollet (2023). A mathematical perspective on transformers. *arXiv preprint arXiv:2312.10794*.

[2] F. Otto and M.G. Reznikoff (2007). Slow motion of gradient flows. *Journal of Differential Equations*, 237 **2**, 372–420.

Large-time dynamics in transformer architectures with layer normalisation

Yury Korolev

Department of Mathematical Sciences, University of Bath

ymk30@bath.ac.uk

Resumen. Transformers have become the backbone of many modern AI systems. A series of recent works have demonstrated that they can be understood mathematically as transformations of measures. We focus on a special case when the propagation of a measure through the transformer follows a gradient flow in the space of probability measures on the unit sphere under a variant of the Wasserstein metric with a non-local mobility term. This allows us to investigate the emergence of either clusters or absolutely continuous measures in the large-time limit and to characterise them as stationary points of an interaction energy. We further investigate how the stationary points depend on the parameters of the transformer, in particular on the eigenvalues and eigenvectors of the product of the key and query matrices. The rigorous framework for studying the gradient flow that we provide also suggests a possible metric geometry for studying the general case (i.e. one that is not described by a gradient flow).

Universal approximation of set-valued maps and application to control

Francisco Periago

Departamento de Matemática Aplicada y Estadística, Universidad Politécnica de Cartagena

f.periago@upct.es

Resumen. Set-valued maps appear in a number of situations in Control Theory. For instance, the operator that maps each initial condition of a dynamical system to a control that steers the state of the associated system to a given target state in a prescribed horizon is, in general, multi-valued. In order to support the use of neural networks for approximating such set-valued operators, a universal approximation theorem for set-valued maps is presented. Then, an approximation scheme is proposed for specific selections of the set-valued maps using DeepONets. Interestingly, it is shown that the proposed scheme breaks the curse of dimensionality for both approximation and estimation errors. Additionally, and as a first step towards the approximation of the whole set-valued map, Single Input Multiple Output Nets (SIMONets) are introduced and analysed. Numerical simulation results for the emblematic examples of the heat and wave equations will be shown as well.

Numerical approximation of the 3-D Navier-Lamé system in an exterior domain using the fundamental solutions method

Carlos Castro

Departamento de Matemática e Informática, Universidad Politécnica de Madrid

carlos.castro@upm.es

Resumen. We obtain a new addition theorem for the fundamental solution of the Navier-Lamé system in three dimensions, incorporating radiation conditions to simulate outgoing waves at infinity. This provides an expansion of this fundamental solution that requires only the evaluation of Bessel functions and scalar spherical harmonics. This is particularly useful in collocation numerical methods based on fundamental solutions, such as the boundary element method or the method of fundamental solutions. For the latter, we show its efficiency when approximating the Navier-Lamé system in exterior domains. We also give some applications to inverse problems.

Entropy-driven control of the continuity equation for normalizing flows

Antonio álvarez López

Departamento de Matemáticas, Universidad Autónoma de Madrid

antonio.alvarezl@uam.es

Resumen. Normalizing flows are generative models that transform a simple (typically Gaussian) reference probability distribution into a complex target distribution through a sequence of smooth, invertible maps. Within the continuous-time framework of neural ODEs, the objective can be recast as a control problem for the continuity equation: we seek a time-dependent vector field that drives the final-time distribution arbitrarily close to the target.
Under a mild tail-compatibility assumption, we establish approximate controllability of this system when the error is measured in relative entropy. The proof is constructive and combines a reverse Pinsker inequality with a piecewise-constant-in-time control scheme that yields explicit bounds on the required number of switches. The result sheds light on the reachable set of the continuity equation in relative entropy and links classical control theory with modern flow-based generative modeling.