Variational Methods and Non-Smooth Geometric Structures with Applications

3-5 JUNE 2025

Bocconi University | Aula AS01 Via Röntgen 1, Milano

Luigi Ambrosio

Title: Sharp PDE estimates for random two-dimensional bipartite matching with power cost function

Abstract: In this joint work with F.Vitillaro and D.Trevisan we study how the PDE ansatz of Caracciolo, Parisi, Lucibello and Sicuro needs to modified in the case then the cost function is the p-th power of the distance, with p different from 2. This involves the replacement of Poisson's equation with the q-Laplace equation and raises an open question about the convergence of this nonlinear PDE with a random right-hand side.

Charles Bertucci

Title: Hamilton-Jacobi equations on the space of probability measures

Abstract: Hamilton-Jacobi equations on the space of probability measures naturally arise in the optimal control of the continuity of Fokker-Planck equation, or in study of large deviations of mean field systems. After recalling how we derive such equations, I will explain the main difficulties in the adaptation of the theory of viscosity solutions to such infinite dimensional equations, namely in the proof of a comparison principle. I will treat both the case of the first order continuity equation and the second order Fokker-Planck equation. We shall see that when trying to pass from one to the other, a particularly helpful tool will be a smooth approximation of the 2-Wasserstein distance that we developed with PL Lions.

Giulia Cavagnari

Title: Convergence of Stochastic Euler Schemes via Probability Vector Fields

Abstract: We develop a unified measure-theoretic framework to study the convergence of stochastic time-discretization schemes—such as stochastic gradient descent and explicit Euler schemes for interacting particle systems—in a separable Hilbert space X. These schemes approximate solutions to deterministic ODEs in X, driven by dissipative fields generated via stochastic superposition. Our approach views these dynamics as evolutions in the Wasserstein space of probability measures over X, modeled by Probability Vector Fields—a notion of distributed velocity field. Under suitable dissipativity and boundedness conditions, we prove that the laws of the interpolated explicit Euler trajectories converge to those of an implicit limit evolution governed by a maximal dissipative extension of the associated barycentric field. As a consequence, we recover the convergence of classical stochastic schemes to the unique solution of the underlying deterministic ODE. This is a joint work with Giuseppe Savaré (Bocconi University) and Giacomo Enrico Sodini (Universität Wien).

Organized by BIDSA Bocconi Institute for Data Science and Analytics, and Department of Decision Sciences
PRIN Grant 202244A7YL (2023-2025) "Gradient Flows and Non-Smooth Geometric Structures with Applications to Optimization and Machine Learning" - Funded by the European Union NextGeneration EU, missione 4, componente 2, investimento 1.1 - CUP J53D23003570001.

Lénaïc Chizat

Title: Analysis of Annealed Sinkhorn’s Algorithm for (Unbalanced) Optimal Transport

Abstract: The Annealed Sinkhorn algorithm—a variant of Sinkhorn’s iterations where the entropic regularization parameter decreases over iterations—is a popular heuristic for solving large-scale optimal transport (OT) problems. Despite its widespread use, its theoretical foundations remain scarce compared to those of standard Sinkhorn's iterations. In this talk, I will present an analysis of the Annealed Sinkhorn algorithm that fills this gap. Our results provide a precise characterization of annealing schedules that ensure convergence to optimal transport plans. We also identify the optimal schedules and introduce a debiasing trick that permits faster annealing than what the standard implementation allows. Furthermore, our analysis reveals that overly aggressive annealing leads the algorithm to solve a semi-relaxed OT problem instead. This observation motivates a natural question: can we leverage annealing to solve unbalanced (KL-relaxed) optimal transport problems? We answer this affirmatively by analyzing an alternating minimization scheme on the semi-coupling formulation of unbalanced OT.
This is based on https://arxiv.org/abs/2408.11620 and work in preparation

Ernesto De Vito

Title: Learning a bounded operator: misspecified setting

Abstract: We consider the problem of estimating a bounded operator A between two Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$ from a training set of noisy observations:

$$y_i = Ax_i + \varepsilon_i \quad \text{for } i = 1, \dots, n.$$

We establish a finite-sample bound for the estimator $\dot{W}_n$, defined as the minimizer of the regularized empirical risk:

$$\frac{1}{n}\sum_{i=1}^{n} \|Wx_i - y_i\|_{\mathcal{Y}}^2 + \lambda_n \|W\|_{\mathcal{S}}^2, \quad \lambda_n > 0,$$

over the Hilbert space $\mathcal{S}$ of Hilbert–Schmidt operators from $\mathcal{X}$ to $\mathcal{Y}$. Notably, our bound does not require the true operator A to be Hilbert–Schmidt (misspecified setting).

Virginie Ehrlacher

Title: Numerical solution of eigenvalue Schrödinger problems using infinite-width two-layer networks

Abstract: The aim of this talk is to present recent results concerning the analysis of numerical schemes using two-layer neural networks with infinite width for the resolution of high-dimensional Schrödinger eigenvalue problems. Using Barron's representation of the solution with a measure of probability, the Rayleigh quotient associated to the eigenvalue problem is minimized thanks to a constrained gradient curve dynamic on the 2-Wasserstein space of parameters defining the neural network. Inspired by the work from Bach and Chizat, we prove that if the gradient curve converges, then the represented function is a solution of the eigenvalue problems considered, but not necessarily the lowest one. At least up to our knowledge, this is the first theoretical result of this type concerning the minimization of non-convex functionals. Numerical experiments are given to illustrate the advantages and drawbacks of the method. Open questions related to the interest of such approaches for the resolution of many-body electronic Schrödinger equations will be discussed.

Massimo Fornasier

Title: A new look at distributional regression: Wassertein Sobolev functions and their numerical approximations

The talk presents a collection of results with Pascal Heid, Giuseppe Savaré, and Giacomo Sodini.

Abstract: We start the talk by presenting general results of strong density of sub-algebras of bounded Lipschitz functions in metric Sobolev spaces. We apply such results to show the density of smooth cylinder functions in Sobolev spaces of functions on the Wasserstein space $\mathcal P_2$ endowed with a finite positive Borel measure. As a byproduct, we obtain the infinitesimal Hilbertianity of Wassertein Sobolev spaces. By taking advantage of these results, we further address the challenging problem of the numerical approximation of Wassertein Sobolev functions defined on probability spaces. Our particular focus centers on the Wasserstein distance function, which serves as a relevant example. In contrast to the existing body of literature focused on approximating efficiently pointwise evaluations, we chart a new course to define functional approximants by adopting three machine learning-based approaches:

1. Solving a finite number of optimal transport problems and computing the corresponding Wasserstein potentials.
2. Employing empirical risk minimization with Tikhonov regularization in Wasserstein Sobolev spaces.
3. Addressing the problem through the saddle point formulation that characterizes the weak form of the Tikhonov functional's Euler-Lagrange equation.

As a theoretical contribution, we furnish explicit and quantitative bounds on generalization errors for each of these solutions. In the proofs, we leverage the theory of metric Sobolev spaces introduced above and we combine it with techniques of optimal transport, variational calculus, and large deviation bounds. In our numerical implementation, we harness appropriately designed neural networks to serve as basic functions. Consequently, our constructive solutions significantly enhance at equal accuracy the evaluation speed, surpassing that of state-of-the-art methods by several orders of magnitude.

Anna Korba

Title: Flowing Datasets with Wasserstein over Wasserstein Gradient Flows

Abstract: Many applications in machine learning involve data represented as probability distributions. The emergence of such data requires radically novel techniques to design tractable gradient flows on probability distributions over this type of (infinite-dimensional) objects. We endow this space with a metric structure from optimal transport, namely the Wasserstein over Wasserstein (WoW) distance, derive a differential structure on this space, and define WoW gradient flows. The latter enables to design dynamics over this space that decrease a given objective functional. We apply our framework to transfer learning and dataset distillation tasks, leveraging our gradient flow construction as well as novel ractable functionals that take the form of Maximum Mean Discrepancies with Sliced-Wasserstein based kernels between probability distributions.

Hugo Lavenant

Title: How quickly does the Gibbs sampler converge for log-concave distributions?

Abstract: The Gibbs sampler (a.k.a. Glauber dynamics and heat-bath algorithm) is a popular Markov Chain Monte Carlo algorithm that iteratively samples from the conditional distributions of the probability measure of interest. Under the assumption of log-concavity, for its random scan version, we provide a sharp bound on the speed of convergence in relative entropy. Assuming that evaluating conditionals is cheap compared to evaluating the joint density, our results imply that the number of full evaluations required for the Gibbs sampler to mix grows linearly with the condition number and is independent of the dimension. This contrasts with gradient-based methods such as overdamped Langevin or Hamiltonian Monte Carlo (HMC), whose mixing time typically increases with the dimension. This is joint work with Filippo Ascolani and Giacomo Zanella.

Edoardo Mainini

Title: Hypercontractivity in Wasserstein gradient flows

Abstract: We obtain regularizing effect and L^p hypercontractivity estimates for Wasserstein gradient flows, starting from the JKO scheme. Emphasis is given to aggregation-diffusion models. This is a joint work with Stefano Lisini.

Cesare Molinari

Title: Learning from data via overparameterization

Abstract: The goal of machine learning is to achieve a good prediction by exploiting training data and some a-priori information about the model. The most common methods to achieve the last objective are explicit and implicit regularization. In the first technique, a regularizer is explicitly introduced to find, among all the solutions, a good generalizing one. The second technique, i.e., implicit regularization, is based on the inductive bias intrinsically induced by the specific method used to optimize the parameters involved. Recently, the success of learning has been related to re- and over-parameterization, which are widely used - for instance - in neural network applications and the optimization method used. However, there is still an open question of how to find systematically what is the inductive bias hidden behind the model for a particular optimization scheme. The goal of this talk is to take a step in this direction by extensively studying many reparameterizations used in the state of the art and providing a common structure to analyze the problem in a unified way. We show that gradient descent on the empirical loss for many reparameterizations is equivalent, in the original problem, to a generalization of mirror descent. The mirror function depends on the reparameterization and introduces an inductive bias, which plays the role of the regularizer. Our theoretical results provide asymptotic behavior and convergence in the simplified setting of linear models.

Emanuele Naldi

Title: Inexact JKO and proximal-gradient algorithms in the Wasserstein space: links and differences from the Hilbert case

Abstract: In this talk, we explore the asymptotic convergence properties of inexact Jordan–Kinderlehrer–Otto (JKO) scheme and proximal-gradient algorithm in the Wasserstein space. While the classical JKO scheme assumes exact minimization at each step, practical implementations rely on approximate solutions due to computational constraints. We analyze two types of inexactness: errors in Wasserstein distance and errors in functional evaluations. We establish rigorous convergence guarantees under controlled error conditions. Beyond the inexact setting, we also extend the convergence results by considering varying stepsizes. Our analysis expands previous approaches, providing new insights into discrete Wasserstein gradient flows. We finish the talk with a comparison to the Hilbert space setting, where the proximity operator is nonexpansive, a property that plays a central role in many classical convergence results. In the Wasserstein setting, the nonexpansivity of the proximity operator generally fails, even for geodesically convex functionals. We discuss the class of functions for which this property still holds and highlight potential directions for future research. These results contribute to the broader understanding of approximate optimization in Wasserstein spaces, with potential applications in sampling, partial differential equations, and machine learning.

Filippo Riva

Title: Existence of gradient flows via trajectory-minimization in spaces of measures

Abstract: We present a novel global-in-time variational approach to gradient flows and doubly nonlinear equations in (reflexive) Banach spaces. It is based on the De Giorgi's principle, which states that solving a gradient flow is equivalent to being a null-minimizer of a suitable energy functional among all trajectories sharing the same initial position. As for the similar Brezis-Ekeland-Nayroles (BEN) principle (which applies only to a convex framework), finding a minimizer for such functional is not difficult in general, but proving that the minimum is zero poses a real challenge. In the BEN formulation, the task has been accomplished by Ghoussoub, resorting to the tool of self-dual Lagrangians.
Our approach allows to extend the analysis to nonconvex energies, directly dealing with the De Giorgi's functional, and it relies on a convexification of the problem in spaces of measures exploiting the so-called superposition principle. The validity of the null-minimization is then recovered by a careful application of the Von Neumann minimax theorem, and by employing the "backward boundedness" property of the dual Hamilton-Jacobi equation. The talk is based on a joint work with A. Pinzi and G. Savaré.

Lorenzo Rosasco

Title: Learning Multi-Index Models with Hyper-Kernel Ridge Regression

Abstract: Deep neural networks excel in high-dimensional problems, outperforming models such as kernel methods, which suffer from the curse of dimensionality. However, the theoretical foundations of this success remain poorly understood. We follow the idea that the compositional sparsity of the learning task is the key factor determining when deep networks outperform other approaches. Taking a step towards formalizing this idea, we consider a simple sparse compositional model, namely the multi-index model (MIM). In this context, we introduce and study hyper-kernel ridge regression (HKRR), an approach blending neural networks and kernel methods. Our main contribution is a sample complexity result demonstrating that HKRR can efficiently learn MIM, thus overcoming the curse of dimensionality.
Further, we note that the kernel nature of the estimator allows us to develop ad hoc optimization solutions. Based on this observation, we contrast alternating minimization and alternating gradient methods both theoretically and numerically. Towards this end, different initialization schemes are considered. These numerical results complement and reinforce our theoretical findings.

Alessandro Rudi

Title: Non-parametric Learning of Stochastic Differential Equations with Non-asymptotic Fast Rates of Convergence

Abstract: We propose a novel non-parametric learning paradigm for the identification of drift and diffusion coefficients of multi-dimensional non-linear stochastic differential equations, which relies upon discrete-time observations of the state. The key idea essentially consists of fitting a RKHS-based approximation of the corresponding Fokker–Planck equation to such observations, yielding theoretical estimates of non-asymptotic learning rates which, unlike previous works, become increasingly tighter when the regularity of the unknown drift and diffusion coefficients becomes higher. Our method being kernel-based, offline pre-processing may be profitably leveraged to enable efficient numerical implementation, offering excellent balance between precision and computational complexity.
Associated paper: https://link.springer.com/article/10.1007/s10208-025-09705-x

Silvia Villa

Title: A structured tour of optimization with finite differences

Abstract: Finite-difference methods are widely used for zeroth-order optimization in settings where gradient information is unavailable or expensive to compute. These procedures mimick first-order strategies by approximating gradients through function evaluations along a set of random directions.
In this talk, I will illustrate recent theoretical results under various assumptions on the objective function showing convergence rates for the case where a structure — such as orthogonality — is imposed on the random set of directions. I will also review review and extend several strategies for constructing structured direction matrices and empirically compare them with unstructured approaches in terms of computational cost, gradient approximation quality, and convergence behavior.
This is a joint work with: Marco Rando, Cesare Molinari, and Lorenzo Rosasco

Federico Vitillaro

Title: The superposition principle for local 1-dimensional currents

Abstract: In 1993, S. Smirnov proved that every one-dimensional Euclidean normal current admits a nice integral representation through currents associated to curves with finite length. This superposition principle was later extended to the metric setting (i.e. to Ambrosio-Kirchheim normal currents) by E. Paolini and E. Stepanov. In this joint work with L. Ambrosio and F. Renzi, we give a further generalization to the case of locally normal metric currents (intended in the sense of U. Lang and S. Wenger), providing a decomposition in (possibly unbounded) curves with locally finite length. Our result holds in Polish spaces, or more generally in complete metric spaces for 1-currents with tight support.