LOL 2026 – Learning and Optimization in Luminy Workshop CIRM

Organizers

Elsa Cazelles (CNRS, Université de Toulouse)
Hadrien Hendrikx (INRIA Grenoble)
Mathurin Massias (INRIA Lyon)
Kimia Nadjahi (CNRS, ENS Paris)
Lorenzo Rosasco (University of Genova & MIT)

SCHEDULE

Monday	
09:30 – 09:45	Welcome talk
09:45 – 10:30	Bruno Galerne
10:30 – 11:00	**Coffee break**
11:00 – 11:45	Zebang Shen
11:45 – 12:30	Alexander Tong
12:30 – 14:00	**Lunch**
14:00 – 14:45	Christian Wald
14:45 – 15:30	Viktor Stein
15:30 – 16:00	**Coffee break**
16:00 – 16:45	Caroline Moosmueller
16:45 – 17:30	Lightning talks
19:30 – 20:30	**Dinner**

Tuesday	
09:00 – 09:45	Claire Boyer
09:45 – 10:30	Francis Bach
10:30 – 11:00	**Coffee break**
11:00 – 11:45	Gabriel Clara
11:45 – 12:30	Elvis Dohmatob
12:30 – 14:00	**Lunch**
14:00 – 14:45	Eloi Tanguy
14:45 – 15:30	Laetitia Chapel
15:30 – 16:00	**Coffee break**
16:00 – 16:45	Giacomo Meanti
17:00 – 19:15	**Poster**
19:30 – 20:30	**Dinner**

Wednesday	
09:00 – 09:45	Aymeric Dieuleveut
09:45 – 10:30	Clément Bonet
10:30 – 11:00	**Coffee break**
11:00 – 11:45	Lê Nguyên Hoang
11:45 – 12:30	George Kevrekidis
12:30 – 14:00	**Lunch**
14:00 – 19:30	**Free afternoon**
19:30 – 20:30	**Dinner**

Thursday	
09:00 – 09:45	Joey Bose
09:45 – 10:30	Arthur Leclaire
10:30 – 11:00	**Coffee break**
11:00 – 11:45	Antonio Sclocchi
11:45 – 12:30	Zahra Kadkhodaie
12:30 – 13:45	**Lunch**
13:45 – 14:30	Lorenzo Rosasco
14:30 – 15:15	Audrey Repetti
15:15 – 15:30	**Coffee break**
15:30 – 16:15	Samuel Hurault
19:30 – 20:30	**Dinner**

Friday	
09:00 – 09:45	Antonio Silveti-Falls
09:45 – 10:30	Nirupam Gupta
10:30 – 11:00	**Coffee break**
11:00 – 11:45	Edwige Cyffers
11:45 – 12:30	Martin Van Waerebeke
12:30 – 14:00	**Lunch**
14:00 – 15:00	**Goodbye**

MONDAY

09:30 – 09:45 — Welcome talk

09:45 – 10:30 — Bruno Galerne

Title : Diffusion models for Gaussian distributions: Exact solutions and Wasserstein errors

Abstract : Diffusion or score-based models recently showed high performance in image generation. They rely on a forward and a backward stochastic differential equations (SDE). The sampling of a data distribution is achieved by numerically solving the backward SDE or its associated flow ODE. Studying the convergence of these models necessitates to control four different types of error: the initialization error, the truncation error, the discretization error and the score approximation. In this work, we theoretically study the behavior of diffusion models and their numerical implementation when the data distribution is Gaussian. Our first contribution is to derive the analytical solutions of the backward SDE and the probability flow ODE and to prove that these solutions and their discretizations are all Gaussian processes. Our second contribution is to compute the exact Wasserstein errors between the target and the numerically sampled distributions for any numerical scheme. This allows us to monitor convergence directly in the data space, while experimental works limit their empirical analysis to Inception features. If time allows, extension of our approach for solving linear inverse problems will be discussed.
Joint work with Emile Pierret https://arxiv.org/abs/2405.14250

10:30 – 11:00 — Coffee break

11:00 – 11:45 — Zebang Shen

Title : Understanding diffusion model generalization through manifold coverage

Abstract : Diffusion models often generate novel samples even when the learned score is only coarse, a phenomenon not fully explained by the standard view of diffusion training as density estimation. Under the data manifold hypothesis, we propose a geometric explanation based on a localized generalization of measure domination. Specifically, for two measures μ and μ_data supported near a data manifold $\mathcal{M}$, we say that μ (δ, c)-covers μ_data if, for every geodesic ball of radius δ on $\mathcal{M}$, the mass assigned by μ is uniformly bounded below by the corresponding mass under μ_data, up to a positive constant c depending only on μ_data. This criterion captures whether a generative model covers all non-negligible regions of the data manifold at resolution δ, without requiring full density recovery. Our main result shows that, even with only coarsely learned scores, diffusion model can achieve coverage of μ_data at resolution $\delta_{DM} = \tilde{O}(N^{-\beta/4k})$, where β denotes the smoothness of the data manifold and k its intrinsic dimension. In contrast, the empirical measure cannot achieve coverage at any resolution finer than the empirical spacing scale $\delta_{emp} = \tilde{O}(N^{-1/k})$. Thus, when the manifold is sufficiently smooth, diffusion models can generalize beyond memorization by learning the geometry of the data support before learning the full data distribution.

11:45 – 12:30 — Alexander Tong

Title : On Couplings for Generative Models

Abstract : The practical success of diffusion and continuous normalizing flow models has been driven by the ability to optimize them at scale using simple conditional losses. By marginalizing over random pairings of noise and data, we effectively learn powerful generative priors. Yet, an intriguing phenomenon occurs during this process: the resulting marginal couplings naturally align closely with Euclidean optimal transport couplings, especially in high-dimensional spaces. In this talk, I will analyze the optimization and statistical learning implications of this alignment. I will outline strategies for exploiting these emergent optimal transport properties to improve model efficiency and sample quality, particularly within the highly constrained few-step and one-step generative domains.

12:30 – 14:00 — Lunch

14:00 – 14:45 — Christian Wald

Title : Self-aware Markov models and jump diffusion in embedding space for discrete generative modelling

Abstract : Discrete generative modeling aims to learn how to sample from a probability distribution on a discrete state space. In text generation, for example, this involves generating sequences of length L from a vocabulary of tokens. A standard approach in discrete flow matching starts with a special masked token and unmasks positions in the sequence, approximating a stochastic process. A major disadvantage of naive implementations of this method is that once a position flips from the masked token to a target token, it can never be corrected during the sampling procedure. We address this issue and explore two frameworks: a self-aware Markov kernel that can correct wrongly predicted tokens, and jump-diffusion processes in a continuous embedding space that combine discrete token updates with a continuous diffusion process.

14:45 – 15:30 — Viktor Stein

Title : Gradient flows on probability measures and their inertial variants: from Wasserstein geometry to transformers

Abstract : Gradient flows on spaces of probability measures (so-called density manifolds) provide a geometric framework for deriving and analyzing nonlinear evolution equations from an energy functional and a metric structure. I will begin by explaining the Wasserstein gradient flow, its kernelized variants, and their particle approximation. I will then detail how one can pass from first-order dissipative dynamics to inertial, damped Hamiltonian dynamics on density manifolds, where particles carry both positions and velocities. This geometric viewpoint has recently appeared in the analysis of Transformer architectures: attention layers can be interpreted as interacting particle systems, with mean-field limits given by gradient flows of interaction energies on spaces of probability measures. In our "SympFormer" work (joint with Wuchen Li and Gabriele Steidl), we use this interpretation to construct accelerated attention blocks from inertial dynamics on density spaces. The resulting architecture can be viewed as a structure-preserving particle discretization of an accelerated probability-measure flow. The talk is intended to be accessible without prior knowledge of Transformers. The main theme is that ideas from gradient flows, Hamiltonian mechanics, and infinite-

dimensional geometry can be used not only to analyze PDEs, but also to design neural network architectures.

15:30 – 16:00 — Coffee break

16:00 – 16:45 — Caroline Moosmueller

Title : Learning in the space of probability measures

Abstract : Many datasets in modern applications, from cell gene expression and images to shapes and text documents, are naturally interpreted as probability measures, distributions, histograms, or point clouds. This perspective motivates the development of learning algorithms that operate directly in the space of probability measures. However, this space presents unique challenges: it is nonlinear and infinite-dimensional. Fortunately, it possesses a natural Riemannian-type geometry which enables meaningful learning algorithms. This talk will provide an introduction to the space of probability measures and present approaches to unsupervised, supervised, and manifold learning within this framework. We will examine temporal evolutions on this space, including flows involving stochastic gradient descent and trajectory inference, with applications to analyzing gene expression in single cells. The proposed algorithms are furthermore demonstrated in pattern recognition tasks in imaging and medical applications.

16:45 – 17:30 — Lightning talks

19:30 – 20:30 — Dinner

TUESDAY

09:00 – 09:45 — Claire Boyer

Title : Tricks and insights in attention-based mechanisms

Abstract : We first show that, in large-prompt regimes, softmax attention effectively linearizes, enabling precise control of its outputs and training dynamics via concentration arguments. This perspective allows optimization analyses developed for linear attention to transfer directly to softmax attention when prompts are sufficiently long, revealing that large-prompt softmax attention inherits the analytical structure of its linear counterpart. As a result, we obtain a principled and broadly applicable toolkit for studying both the optimization and statistical behavior of softmax attention in long-context regimes. Building on this framework, we then show that attention trained on Gaussian data under unsupervised objectives aligns with the leading eigenvectors of the covariance matrix—thereby implicitly performing PCA. This establishes a direct theoretical link between attention mechanisms and classical spectral methods, positioning attention as an optimization-driven, implicit analogue of PCA.

09:45 – 10:30 — Francis Bach

Title : A spectral framework for closed-form relative density estimation

Abstract : Estimating relative densities and information-theoretic divergences from samples is a central problem in statistics and machine learning, but standard variational approaches to Kullback-Leibler (KL) divergence estimation often require nonlinear optimization and may suffer from numerical instability because of exponential terms. This talk presents a closed-form spectral framework for relative log-density estimation in linearly parameterized probabilistic models, including unnormalized and conditional models. The key idea is to express the KL divergence as an integral of weighted chi-squared divergences. This converts divergence estimation into a family of least-squares problems and yields explicit spectral formulas depending only on first- and second-order feature moments. The framework extends naturally to kernel methods and learned representations, including neural networks, with convergence guarantees and efficient learning algorithms in both settings.

10:30 – 11:00 — Coffee break

11:00 – 11:45 — Gabriel Clara

Title : Training Diagonal Linear Networks with Stochastic Sharpness-Aware Minimization

Abstract : We analyze the landscape and training dynamics of diagonal linear networks in a linear regression task, with the network parameters being perturbed by isotropic normal noise during training. The addition of such noise may be interpreted as a stochastic form of sharpness-aware minimization (SAM) and we prove several results that relate its action on the underlying landscape and training dynamics to the sharpness of the loss. In particular, the noise induces a weighted mixture of fractional norm penalties on the network parameters, which forces the individual layers to balance at a fast rate and changes the underlying landscape to favor solutions that result from a shrinkage-thresholding operator applied to the

true parameter. We show that balancing the layers equates to minimizing the average sharpness, as well as the trace of the Hessian matrix, among all possible factorizations of the same linear predictor. Further, we characterize how the noise level of the normal perturbations acts as a regularization parameter, with exact descriptions of its effect on shrinkage, thresholding, and balancing speed. Based on joint work with Johannes Schmidt-Hieber and Sophie Langer.

11:45 – 12:30 — Elvis Dohmatob

Title : Throw away (almost) all the data: a modern theory of data-selection

Abstract : A growing body of empirical evidence, including results from DeepSeek and others, has shown that in the very large data regimes characteristic of LLM training, it is often beneficial for generalization, to train on only a strategically selected, tiny fraction of the available data. This seems like magic, and a satisfactory theoretical explanation of the phenomenon is still lacking. I will present recent results, obtained with my collaborators, in which we compute the optimal data-selection strategy as a function of the relevant "cosmological" parameters, such as sample size and input dimension. Working in the solvable setting of high-dimensional regression with Gaussian data, we obtain a precise phase-transition picture. In the very large data regime, where the sample size is much larger than the input dimension, the optimal strategy is to keep a vanishing fraction of the data, consisting only of the hardest examples. By contrast, in the low-data regime, the optimal strategy is to use as much of the available data as possible, while prioritizing the easiest examples.
The main tools are random matrix theory and the replica method from statistical physics, which I will endeavor to introduce to the audience.

12:30 – 14:00 — Lunch

14:00 – 14:45 — Eloi Tanguy

Title : Computing Optimal Transport Barycentres

Abstract : Wasserstein barycentres represent average distributions between multiple probability measures for the Wasserstein distance. The numerical computation of Wasserstein barycentres is notoriously challenging. A common approach is to use Sinkhorn iterations, where an entropic regularisation term is introduced to make the problem more manageable. Another approach involves using fixed-point methods, akin to those employed for computing Fréchet means on manifolds. The convergence of such methods for 2-Wasserstein barycentres, specifically with a quadratic cost function and absolutely continuous measures, was studied by Alvarez-Esteban et al.. In this paper, we delve into the main ideas behind this fixed-point method and explore how it can be generalised to accommodate more diverse transport costs and generic probability measures, thereby extending its applicability to a broader range of problems. We show convergence results for this approach and illustrate its numerical behaviour on several barycentre problems.

14:45 – 15:30 — Laetitia Chapel

Title : Sliced Wasserstein Plans

Abstract : Optimal Transport (OT) has emerged as a foundational tool in modern machine learning, primarily due to its capacity to provide meaningful comparisons between probability distributions. One of the key advantages of OT lies in its dual nature: it introduces a

mathematically rigorous framework that defines Wasserstein distances but also constructs an optimal coupling (or transport plan) between distributions. This coupling reveals explicit correspondences between samples, enabling a wide range of applications. Despite the many successes of optimal transport in machine learning, and despite many tools to approximate the Wasserstein distance, computing OT plans remains a computationally challenging problem. In this talk, I will present a new methodology to efficiently compute sliced transport plans. The formulation can be recast as a bilevel optimization problem, and I will propose a differentiable generalized approximation that can be further adapted to data residing on manifolds. Finally, I will demonstrate the practical value of this approach by introducing a novel sliced OT-based conditional flow matching method for image generation, an application where fast computation of transport plans is crucial.

15:30 – 16:00 — Coffee break

16:00 – 16:45 — Giacomo Meanti

Title : Enhancing PnP Image Restoration with ProxiMAP

Abstract : Plug-and-Play methods have become standard tools for solving imaging inverse problems by replacing the intractable maximum a posteriori denoiser with the MMSE one. While this mismatch has been widely treated as unavoidable, we have shown that — under convexity assumptions — the true MAP can be targeted using the log-gradient of the probability distribution (the score). We discovered that using the score learned by diffusion models is problematic in practice: learned scores do not match true ones and MAP-targeting iterations converge to cartoon-like images rather than realistic ones. Better results are obtained by stopping short of convergence. We turn this observation into a design principle and introduce ProxiMAP, an approximation to the theoretical iteration which matches the denoiser's training noise schedule. This keeps the denoiser in-distribution where its score is reliable, and yields implicit early stopping that avoids the failure mode above. ProxiMAP is a modular drop-in replacement for MMSE denoisers in standard PnP algorithms and consistently sharpens reconstructions across deblurring, inpainting, super-resolution, and phase retrieval.

17:00 – 19:15 — Poster

19:30 – 20:30 — Dinner

WEDNESDAY

09:00 – 09:45 — Aymeric Dieuleveut

Title : Computer-aided proofs in first-order optimization, with applications to error feedback.

Abstract : First-order methods are widely used in optimization and machine learning, and their behavior is often analyzed through the spectrum of worst case convergence rates. Obtaining such guarantees is often difficult and both time consuming and error-prone. Starting with the work of Drori and Teboulle (2014), novel techniques have been used to gain numerical insights, leading to the release of various performance estimation (PE) software. In this talk, I will show how various computer-aided techniques can be used to study first-order optimization methods in a systematic way. From performance estimation problems with automated Lyapunov discovery, to symbolic regression and computer algebra systems, novel tools completely reshape the way we approach theory of optimization. As a main example, I will focus on error feedback methods used with compressed communication in distributed optimization. While error feedback has been widely studied, existing theory often provides untight (thus unreliable) bounds. I will present tight analyses with matching lower bounds that allow a fair comparison between error feedback schemes and standard compressed gradient descent, and help explain when error feedback is useful and when it is not. Overall, the talk aims to show how various computer-aided proofs can lead to clearer and more reliable insights into first-order optimization methods.
Based on :

- A Tight Theory of Error Feedback Algorithms in Distributed Optimization, DB Thomsen, A Taylor, A Dieuleveut International Conference on Machine Learning (ICML 2026)
- Tight analyses of first-order methods with error feedback, DB Thomsen, A Taylor, A Dieuleveut Advances in Neural Information Processing Systems (NeurIPS) (2025). PEPit: computer-assisted worst-case analyses of first-order optimization methods in Python B Goujaud, C Moucer, F Glineur, JM Hendrickx, AB Taylor, A Dieuleveut Mathematical Programming Computation 16 (3), 337-367 (2024)
- (eventually - as a detour) Provable non-accelerations of the heavy-ball method B Goujaud, A Taylor, A Dieuleveut, Mathematical Programming, 1-59 (2025)

09:45 – 10:30 — Clément Bonet

Title : Difference of convex programming in the Wasserstein Space with applications to MMD optimization

Abstract : Optimizing functionals over the space of probability measures is now ubiquitous in machine learning. A widely used approach is to perform the optimization directly over the Wasserstein space, but many objective functionals ofpractical interest are non-convex along Wasserstein geodesics, making the analysis of standard first-order methods challenging. In this work, we study a class of objectives over the Wasserstein space that admit a difference-of-convex (DC) decomposition and we lift the classical convex-concave procedure (CCCP) to

this setting. Under smoothness and strong convexity assumptions on the convex components of the decomposition, we prove almost stationarity along the iterates of the resulting algorithm. Our main focus is on the Maximum Mean Discrepancy (MMD) and the Energy Distance (ED) functionals, for which we develop explicit Wasserstein DC decompositions, and establish local convergence of the scheme under mild assumptions. Empirically, we show that well-chosen DC decompositions yield faster and more stable convergence than Wasserstein gradient descent on these MMD objectives.

10:30 – 11:00 — Coffee break

11:00 – 11:45 — Lê Nguyên Hoang

Title : Towards social and democratic recommendation systems

Abstract : Proprietary engagement-maximizing recommendation systems have recently become the targets of historical judicial sentences, both in US and EU. However, the current fallback is the chronological feed, which has very poor mathematical and empirical properties. In this talk, I will discuss a radically different framework for their design, which is rooted in social choice theory, and which my collaborators and I deployed on Tournesol.app and BlueSky. I will also discuss the philosophical foundations of our approach, centered around Tournesol's four digital democracy pillars, and the challenges to formalize and implement them mathematically and algorithmically.

11:45 – 12:30 — George Kevrekidis

Title : When rates are geometric: Certificate-preserving contact splittings for discrete-time acceleration

Abstract : The behavior of discrete optimization algorithms is often analyzed asymptotically via continuous-time limiting ODEs. However, the translation of convergence rates from continuous-time to discrete-time is not always straightforward, and may not hold in general. In this note, we use contact Hamiltonian systems as a framework for both optimizer design and rate-certificate transfer. We show that a contact Hamiltonian carries an intrinsic conformal decay identity, and that on compact regions where this Hamiltonian controls the objective gap it yields a continuous-time Lyapunov certificate. For sufficiently smooth contact splittings, backward error analysis shows that over the backward-error timescale the modified conformal factor matches the continuous decay rate up to O(hr) perturbations; the remaining BEA defect is algebraic for finite smoothness and improves to exponentially small under analyticity. Within this certificate-preserving framework, we construct contact splitting algorithms with state-dependent damping and demonstrate competitive behavior on ill-conditioned deterministic benchmarks and deep-learning tasks.

12:30 – 14:00 — Lunch

14:00 – 19:00 — Free afternoon

19:30 – 20:30 — Dinner

THURSDAY

09:00 – 09:45 — Joey Bose

Title : Normalizing Flow Maps

Abstract : Scalable approaches to learning dynamic measure transport maps have shaped the forefront of generative modeling, leading to paradigm-defining methods for simulation-free training such as diffusion models and flow matching. The natural evolution of this trajectory is to demand the same modeling power at a fraction of the inference cost---that is, to directly learn the *flow-map* operator, or integrator, associated with a generative dynamical system. In this paper, we model such flow-map operators using invertible neural networks, forming a new class of exactly invertible models termed Normalizing Flow Maps (NFM). NFMs generalize classical Normalizing Flows (NFs) by learning time-indexed transport maps between arbitrary pairs of times, while retaining efficient exact likelihood evaluation through change of variables. In contrast to standard flow-map approaches such as Consistency Models and Mean Flow models, NFMs provide both fast few-step sampling and tractable density estimation, making them suitable for domains where likelihoods are essential, including Boltzmann Generators.
To train NFMs scalably, we develop a family of flow-map consistency objectives that operate at both the pointwise and distributional levels. Beyond regression-based Lagrangian, Eulerian, and self-distillation losses, we introduce likelihood-compatible distributional consistency objectives, including forward- and reverse-KL formulations over intermediate marginals and interpolant conditionals. These objectives exploit the exact change-of-variables structure of NFM to regularize the learned flow maps as density-preserving transports, bridging regression-based flow-map training with maximum-likelihood principles. Empirically, we demonstrate that NFM achieve strong performance on image generation, yielding state-of-the-art sample quality among normalizing-flow models while enabling fast sampling in either direction via the time-inverse structure of flow maps. We further show that NFMs outperform prior normalizing-flow baselines, including RegFlow-style methods, for equilibrium sampling of small peptide systems in Cartesian coordinates. Together, these results establish normalizing flow maps as a practical framework for exact-likelihood generative modeling with fast, consistent, and bidirectional sampling.

09:45 – 10:30 — Arthur Leclaire

Title : Stochastic Plug-and-Play Methods

Abstract : Plug-and-Play methods constitute a class of iterative algorithms for ill-posed imaging inverse problems where regularization is performed by an off-the-shelf denoiser. In this talk, we will give an overview of several variants of plug-and-play algorithms. We will focus especially on stochastic plug-and-play algorithms where the input to the denoiser is renoised beforehand, as is usually done in diffusion-based algorithms, which helps to avoid a distribution shift in the denoiser use. We will provide convergence bounds showing that these stochastic plug-and-play algorithms may maximize or sample the posterior distribution up to some error, depending on the normalization of the added noise.

10:30 – 11:00 — Coffee break

11:00 – 11:45 — Antonio Sclocchi

Title : Learning the hidden grammar of data with diffusion models

Abstract : Why can deep generative models produce novel, coherent data from far fewer examples than the ambient dimension would suggest? I argue that the key lies in the hierarchical and compositional structure of natural data—its hidden "grammar"—and that diffusion models provide a natural probe of this structure. Using probabilistic context-free grammars as tractable models of hierarchical data, I present three results. First, the denoising dynamics undergoes a phase transition: as noise decreases, high-level features are reconstructed at critical noise scales, providing a dynamical signature of hierarchical organization. Second, during training, diffusion models learn compositional rules sequentially across levels of abstraction, with the sample complexity at each level controlled by the corresponding contextual correlations, yielding polynomial rather than exponential learnability. Third, hierarchical generalization and memorization appear as competing dynamical phases, with a crossover time that grows with dataset size. Taken together, these results outline a statistical physics approach to generative models, in which simple models illuminate emergent phenomena in modern AI.

11:45 – 12:30 — Zahra Kadkhodaie

Title : Blind denoising diffusion models and adaptive sampling algorithms

Abstract : Denoising diffusion models (DDMs) are state-of-the-art methods for learning densities from data across numerous domains, yet many aspects of the training and sampling pipeline remain poorly understood. In particular, noise conditioning requires practitioners to incorporate contrived unprincipled noise embeddings into neural network architectures and to use ad hoc noise schedules for sampling. To address these drawbacks, we provide a complete theory for blind denoising diffusion models (BDDMs): a variant of DDMs where the noise amplitude is not passed into the neural network during training or sampling, obviating the need for the aforementioned design choices. We justify the correctness of BDDMs as a sampling algorithm under an assumption of low intrinsic dimensionality of the underlying data distribution relative to the ambient dimension. This assumption arises through the introduction of the Bayesian problem of estimating noise levels from a single noisy sample, which might be of independent interest. We empirically compare the performance of BDDMs to standard DDMs, showcasing the benefits of an adaptive scheme which is rigorously justified by our analysis.

12:30 – 13:45 — Lunch

13:45 – 14:30 — Lorenzo Rosasco

Title : Learning ergodic dynamical systems

Abstract : We consider the problem of learning ergodic dynamical systems from a single finite trajectory. We derive learning guarantees for a least-squares estimator and contrast them with classical supervised learning guarantees for independent and identically distributed data. We further extend the analysis to higher-order systems, finite-state systems, and Koopman operator learning. Our analysis combines tools from statistical learning theory and Markov processes with a suitable concentration result for Hilbert-space-valued Markov processes.

14:30 – 15:15 — Audrey Repetti

Title : Analysis and synthesis approximated denoisers for forward-backward plug-and-play algorithms

Abstract : Markov Chain Monte Carlo (MCMC) algorithms are standard approaches to solve imaging inverse problems and quantify estimation uncertainties, a key requirement in absence of ground-truth data. To improve estimation quality, Plug-and-Play MCMC algorithms, such as PnP-ULA, have been recently developed to accommodate priors encoded by a denoising neural network. Designing scalable samplers for high-dimensional imaging inverse problems remains a challenge: drawing and storing high-dimensional samples can be prohibitive, especially for high-resolution images. To address this issue, this work proposes a distributed sampler based on approximate data augmentation and PnP-ULA to solve very large problems. The proposed sampler uses lightweight denoising convolutional neural network, to efficiently exploit multiple GPUs on a Single Program Multiple Data architecture. We illustrate the behaviour of the proposed approach in terms of reconstruction performance and scalability on several imaging problems, investigating communication and computation overheads due to the denoiser.

15:15 – 15:30 — Coffee break

15:30 – 16:15 — Samuel Hurault

Title : Geometric-aware discretization error of diffusion models

Abstract : Practical diffusion sampling is a numerical approximation problem: under a fixed inference budget, one must simulate a reverse-time ODE or SDE using only a limited number of denoising steps, so discretization error is often the dominant source of error. Existing non-asymptotic analyses provide convergence guarantees, but are typically too loose and too insensitive to diffusion parameters to guide practical design: broad families of schedules receive the same rates, which depend on coarse worst-case quantities such as the dimension or the drift Lipschitz constant. We take a less ambitious but more informative route. In the exact-score setting, we derive first-order asymptotic expansions of the Euler-Maruyama weak and Fréchet discretization errors. These formulas hold for general smooth reverse diffusions and become fully explicit under Gaussian data. They show how discretization error adapts to the geometry of the data through the covariance spectrum, and how this geometry interacts with key diffusion parameters, including the diffusion schedules and the diffusion-term coefficient. This yields tractable objectives for geometry-aware parameter optimization. Finally, we show that the qualitative predictions of the Gaussian formulas remain robust across diffusion sampling problems with different geometries, including image generation on different datasets and image posterior sampling.

19:30 – 20:30 — Dinner

FRIDAY

09:00 – 09:45 — Antonio Silveti-Falls

Title : Training Neural Networks at Any Scale with Scion

09:45 – 10:30 — Nirupam Gupta

Title : Learning in untrusted environments

Abstract : Machine learning systems are increasingly deployed in messy, real-world conditions where data can be noisy, corrupted, or even adversarial. From federated learning across untrusted devices to training on web-scale data of uncertain quality, robustness is not a luxury but a necessity. In this talk, I will discuss how to design learning algorithms that stay reliable under such conditions. A central question is feasibility: when can we actually learn robustly, and how does the heterogeneity of data across sources affect what is possible? I will present recent results on this question, show that heterogeneity plays a more subtle role than we previously thought, and draw out key differences between robust ML and classic robust statistics. I will close with open problems and directions for future work.

10:30 – 11:00 — Coffee break

11:00 – 11:45 — Edwige Cyffers

Title : Optimizing Performative Learning

Abstract : Machine learning systems are widely deployed, including for decision making processes that affect people's lives, such as loan approval or hiring. Gaming the system by modifying one's features therefore creates an iterative feedback loop, where the data distribution changes under the performative effect of the model. Performative learning tackles this setting and aims to optimize not a single time step model for a fixed distribution, but to find an optimal long term solution. In this talk, we will introduce performative learning and present two results for optimizing performative risk.
Refs: https://arxiv.org/abs/2411.02023 (NeurIPS 2024) and https://arxiv.org/abs/2510.12249 (ICML 2026)

11:45 – 12:30 — Martin Van Waerebeke

Title : When to forget ? Complexity trade-offs in machine unlearning

Abstract : Machine Unlearning (MU) aims at removing the influence of specific data points from a trained model, striving to achieve this at a fraction of the cost of full model retraining. We analyze the efficiency of unlearning methods and establish the first upper and lower bounds on minimax computation times for this problem, characterizing the performance of the most efficient algorithm against the most difficult objective function. Specifically, for strongly convex objective functions, we describe when unlearning can outperform retraining, depending on whether the forget set is assumed to be accessible.
Based on https://arxiv.org/abs/2502.17323 and https://arxiv.org/pdf/2602.14938.

12:30 – 14:00 — Lunch