LOL 2024 – Learning and Optimization in Luminy Workshop CIRM

Organizers

- Aymeric Dieuleveut
- Elsa Cazelles
- Thomas Moreau
- Mathurin Massias
- Lorenzo Rosasco

Schedule

Monday

09:00-09:45	Aryan Mokhtari
09:45-10:30	Saverio Salzo
10:30-11:00	**Coffee break**
11:00-11:45	Flavien Léger
11:45-12:30	Clarice Poon
12:30-14:00	**Lunch**
14:00-14:45	Emmanuel Soubies
14:45-15:30	Silvia Villa
15:30-16:00	**Coffee break**
16:00-16:45	Daniel McKenzie
16:45-17:30	Lightning Talks
19:30-12:30	**Diner**

Tuesday

09:00-09:45	Claire Boyer
09:45-10:30	Atsushi Nitanda
10:30-11:00	**Coffee break**
11:00-11:45	Taiji Suzuki
11:45-12:30	Aurélie Boisbunon
12:30-14:00	**Lunch**
14:00-14:45	Marco Cuturi
14:45-15:30	Kimia Nadjahi
15:30-16:00	**Break**
16:00-16:45	Gabriel Peyré
17:00-19:15	**Poster Session**
19:30-12:30	**Diner**

Wednesday

09:00-09:45	Anastasia Koloskova
09:45-10:30	Cyril Letrouit
10:30-11:00	**Coffee break**
11:00-11:45	Mathieu Even
11:45-12:30	Mikael Johansson
12:30-14:00	**Lunch**
14:00-19:00	**Free Afternoon**
19:30-12:30	**Diner**

Thursday

09:00-09:45	Nicolas Courty
09:45-10:30	Julio Backhoff
10:30-11:00	**Coffee break**
11:00-11:45	Rémi Flamary
11:45-12:30	Aram-Alexander Pooladian
12:30-14:00	**Lunch**
14:00-14:45	Sinho Chewi
14:45-15:30	Rémi Gribonval
15:30-16:00	**Coffee break**
16:00-16:45	Audrey Repetti
19:30-21:30	**Diner**

Friday

09:00-09:45	Barbara Pascal
09:45-10:30	Julien Mairal
10:30-11:00	**Coffee break**
11:00-11:45	Malgorzata Bogdan
11:45-12:30	Hadrien Hendrikx
12:30-14:00	**Lunch**
14:00-15:00	**Goodbye :)**

Monday

09:00 - 09:45 – Aryan Mokhtari

Title: Online Learning Guided Quasi-Newton Methods: Improved Global Non-asymptotic Guarantees

Abstract: Quasi-Newton (QN) methods are popular iterative algorithms known for their superior practical performance compared to Gradient Descent (GD)-type methods. However, the existing theoretical results for this class of algorithms do not sufficiently justify their advantage over GD-type methods. In this talk, we discuss our recent efforts to address this issue. Specifically, in the strongly convex setting, we propose the first "globally" convergent QN method that achieves an explicit "non-asymptotic superlinear" rate. We show that the rate presented for our method is provably faster than GD after at most $O(d)$ iterations, where d is the problem dimension. Additionally, in the convex setting, we present an accelerated variant of our proposed method that provably outperforms the accelerated gradient method and converges at a rate of $O(\min\{1/k^2, \sqrt{d \log k}/k^{2.5}\})$, where k is the number of iterations. To attain these results, we diverge from conventional approaches and construct our QN methods based on the Hybrid Proximal Extragradient (HPE) framework and its accelerated variants. Furthermore, a pivotal algorithmic concept underpinning our methodologies is an online learning framework for updating the Hessian approximation matrices. Specifically, we relate our method's convergence rate to the regret of a specific online convex optimization problem in the matrix space and choose the sequence of Hessian approximation matrices to minimize its overall regret.

09:45 - 10:30 – Saverio Salzo

Title: Nonsmooth Implicit Differentiation: Deterministic and Stochastic Convergence Rates

Abstract: I will address the problem of efficiently computing the derivative of the fixed-point of a parametric nondifferentiable contraction map. This problem has wide applications in machine learning, including hyperparameter optimization, meta-learning and data poisoning attacks. Two popular approaches are analyzed: iterative differentiation (ITD) and approximate implicit differentiation (AID). A key challenge behind the nonsmooth setting is that the chain rule does not hold anymore. Building upon the recent work by Bolte et al. (2022), who proved linear convergence of nondifferentiable ITD, I will show an improved linear rate for ITD and a slightly better rate for AID, both in the deterministic case. I will also introduce NSID, a new stochastic method to compute the implicit derivative when the fixed point is defined as the composition of an outer map and an inner map which is accessible only through a stochastic unbiased estimator. Rates for such stochastic method rates will be presented.

10:30 - 11:00 – Coffee break

11:00 - 11:45 – Flavien Léger

Title: A nonsmooth geometry for alternating minimization

Abstract: Given two arbitrary sets and an objective function defined on their product, I will introduce a nonsmooth version of a geometric condition known as nonnegative cross-curvature (NNCC). Then I will show that NNCC provides Alternating Minimization with an intrinsic geometry, leading to convergence rates results depending on intrinsic convexity conditions. Applications include new convergence theories for proximal schemes on infinite-dimensional spaces of measures, metric measure spaces, general gradient-descent like methods and a nonsmooth definition of mirror descent.

Joint works with Pierre-Cyril Aubin, Gabriele Todeschi, François-Xavier Vialard.

11:45 - 12:30 – Clarice Poon

Title: Sparse recovery guarantees for inverse optimal transport

Abstract: Optimal Transport is a useful metric to compare probability distributions and to compute a pairing given a ground cost. Its entropic regularization variant (eOT) is crucial to have fast algorithms and reflect fuzzy/noisy matchings. This work focuses on Inverse Optimal Transport (iOT), the problem of inferring the ground cost from samples drawn from a coupling that solves an eOT problem. It is a relevant problem that can be used to infer unobserved/missing links, and to obtain meaningful information about the structure of the ground cost yielding the pairing. On one side, iOT benefits from convexity, but on the other side, being ill-posed, it requires regularization to handle the sampling noise. This work presents a study of l1 regularization to model for instance Euclidean costs with sparse interactions between features. Specifically, we derive a sufficient condition for the robust recovery of the sparsity of the ground cost that can be seen as a generalization of the Lasso's celebrated "Irrepresentability Condition". To provide additional insight into this condition, we consider the Gaussian case. We show that as the entropic penalty varies, the iOT problem interpolates between a graphical Lasso and a classical Lasso, thereby establishing a connection between iOT and graph estimation. This is joint work with Francisco Andrade and Gabriel Peyré.

12:30 - 14:00 – Lunch

14:00 - 14:45 – Emmanuel Soubies

Title: Exact Continuous Relaxations of L0-Regularized Criteria

Abstract: Sparse generalized linear models are widely used in fields such as statistics, computer vision, signal/image processing and machine learning. The natural sparsity promoting regularizer is the l0 pseudo-norm which is discontinuous and non-convex. In this talk, we will present the l0-Bregman relaxation (B-Rex), a general framework to compute exact continuous relaxations of such l0-regularized criteria. Although in general still non-convex, these continuous relaxations are qualified as exact in the sense that they let unchanged the set of global minimizer while enjoying a better optimization landscape. In particular, we will show that some local minimizers of the initial functional are eliminated by these relaxations. Finally, these properties will be illustrated on both sparse Kullback-Leibler regression and sparse logistic regression problems.

14:45 - 15:30 – Silvia Villa

Title: Zeroth order optimization with structured directions

Abstract: In this talk I will present new algorithms for zero-order optimization via finite difference structured gradient estimation. I will describe their convergence properties under various assumptions, such as smoothness or convexity. I will also discuss empirical results that support the theoretical findings.

15:30 - 16:00 – Coffee break

16:00 - 16:45 – Daniel McKenzie

Title: Reducing the complexity of derivative-free optimization using compressed sensing.

Abstract: Derivative-Free Optimization (DFO) is concerned with minimizing a function whose gradients are inaccessible. While DFO has been classically applied to problems with 10^3 variables or fewer, emerging applications in machine learning require solving DFO problems with 10^5 variables or more. In this talk I'll survey some of my recent work in extending DFO to this high-dimensional regime, primarily by exploiting sparsity in gradients to construct good gradient approximations cheaply.

16:45 - 17:30 – Lightning Talks

19:30 - 12:30 – Diner

Tuesday

09:00 - 09:45 – Claire Boyer

Title: A primer on physics-informed learning

Abstract: Physics-informed machine learning combines the expressiveness of data-based approaches with the interpretability of physical models. In this context, we consider a general regression problem where the empirical risk is regularized by a partial differential equation that quantifies the physical inconsistency. We prove that for linear differential priors, the problem can be formulated as a kernel regression task, giving a rigorous framework to analyze physics-informed ML. In particular, the physical prior can help in boosting the estimator convergence.

The direct implementation of physics-informed kernel estimators can be tedious, and practitioners often resort to physics-informed neural networks (PINNs) instead. We offer some food for thought and statistical insight into the proper use of PINNs.

09:45 - 10:30 – Atsushi Nitanda

Title: Improved Particle Approximation Error for Mean Field Neural Networks

Abstract: Mean-field Langevin dynamics (MFLD) minimizes an entropy-regularized nonlinear convex functional defined over the space of probability distributions. MFLD has gained attention due to its connection with noisy gradient descent for mean-field two-layer neural networks. Unlike standard Langevin dynamics, the nonlinearity of the objective functional induces particle interactions, necessitating multiple particles to approximate the dynamics in a finite-particle setting. Recent works have demonstrated the uniform-in-time propagation of chaos for MFLD, showing that the gap between the particle system and its mean-field limit uniformly shrinks over time as the number of particles increases. In this work, we improve the dependence on logarithmic Sobolev inequality (LSI) constants in their bounds, which can exponentially deteriorate with the regularization coefficient, in the particle approximation error. Specifically, we establish an LSI-constant-free particle approximation error concerning the objective gap by leveraging the problem structure in risk minimization. As the application, we demonstrate improved convergence of MFLD, sampling guarantee for the mean-field stationary distribution, and uniform-in-time Wasserstein propagation of chaos in terms of particle complexity.

10:30 - 11:00 – Coffee break

11:00 - 11:45 – Taiji Suzuki

Title: Nonlinear feature learning of neural networks with gradient descent: Information theoretic optimality and in-context learning

Abstract: In this presentation, we consider a optimization guarantee of nonlinear feature learning in training neural networks and its application to in-context learning by Transformer architectures. In the first part, we study the problem of gradient descent learning of a single-index target function under isotropic Gaussian data, where the link function is an unknown degree-q polynomial with information exponent p. We prove that a two-layer neural network optimized by an SGD-based algorithm learns the true function with arbitrary polynomial link using $\widetilde{O}(d)$ time and sample complexity, regardless of information exponent; this matches the information theoretic limit up to polylogarithmic factors. Core to our analysis is the reuse of minibatch in the gradient computation, which gives rise to higher-order information beyond correlational queries. Second, we consider in-context learning (ICL) by Transformers. We study ICL of a nonlinear function class via transformer with nonlinear MLP layer. We show that a nonlinear transformer optimized by gradient descent learns the target function with a prompt length that only depends on the intrinsic dimension of the target function class; in contrast, an algorithm that directly trains only on the test prompt requires a statistical complexity scaling with the ambient dimension d. In addition to that, if we have time, we also considers a convergence guarantee of mean field gradient flow (MFGF) for training Transformers to obtain nonlinear features in the pretraining procedure of in-context learning, and show that the objective is strict-saddle and thus the MFGF is not captured by a critical point almost surely.

11:45 - 12:30 – Aurélie Boisbunon

Title: Deep Unsupervised Domain Adaptation for Time Series Classification: a Benchmark

Abstract: Unsupervised Domain Adaptation (UDA) aims to harness labeled source data to train models for unlabeled target data. While research in UDA is extensive in domains like computer vision and natural language processing, it remains underexplored for time series data despite its widespread real-world applications. Our paper addresses this gap by introducing a comprehensive benchmark for evaluating UDA techniques for time series classification, with a focus on deep learning methods. We provide a fair and standardized UDA method assessments with state of the art neural network backbones (e.g. Inception) for time series data, as well as seven additional datasets besides the most used ones. This benchmark offers insights into the strengths and limitations of the evaluated approaches while preserving the unsupervised nature of DA, making it directly applicable to practical problems. It serves as a beneficial resource for researchers and practitioners, fostering innovation in this critical field.

12:30 - 14:00 – Lunch

14:00 - 14:45 – Marco Cuturi

Title: Elastic Costs and Monge Transport Maps

Abstract: I will present in this short talk recent developments on choosing carefully the cost function when computing optimal transport maps to ”shape” Monge map estimators. In particular, I will present a pipeline to learn, within a family of elastic costs, a suitable set of parameters that can help when running OT estimation.

14:45 - 15:30 – Kimia Nadjahi

Title: Slicing Mutual Information Generalization Bounds for Neural Networks

Abstract: The ability of machine learning (ML) algorithms to generalize well to unseen data has been studied through the lens of information theory, by bounding the generalization error with the input-output mutual information (MI), i.e., the MI between the training data and the learned hypothesis. Yet, these bounds have limited practicality for modern ML applications (e.g., deep learning), due to the difficulty of evaluating MI in high dimensions. Motivated by recent findings on the compressibility of neural networks, we consider algorithms that operate by ”slicing” the parameter space, i.e., trained on random lower-dimensional subspaces. We introduce new, tighter information-theoretic generalization bounds tailored for such algorithms, demonstrating that slicing improves generalization. Our bounds offer significant computational and statistical advantages over standard MI bounds, as they rely on scalable alternative measures of dependence, i.e., disintegrated mutual information and k-sliced mutual information. Then, we extend our analysis to algorithms whose parameters do not need to exactly lie on random subspaces, by leveraging rate-distortion theory. This strategy yields generalization bounds that incorporate a distortion term measuring model compressibility under slicing, thereby tightening existing bounds without compromising performance or requiring model compression. Building on this, we propose a regularization scheme enabling practitioners to control generalization through compressibility. Finally, we empirically validate our results and achieve the computation of non-vacuous information-theoretic generalization bounds for neural networks, a task that was previously out of reach.

15:30 - 16:00 – Break

16:00 - 16:45 – Gabriel Peyré

Title: A Survey of Wasserstein Flow in Neural Network Training Analysis

Abstract: In this talk, I will first introduce the concept of Wasserstein gradient flow, an optimization process over the space of measures. This approach provides a unified framework for describing the gradient descent method applied to particle positions and can handle an arbitrary, possibly infinite, number of particles. Additionally, it enables the modeling of diffusion phenomena, which are not easily described by particle systems, and can be beneficial for sampling problems. A significant recent application of this method is in studying the convergence of gradient training in shallow neural networks, where particles represent neuron weights. I will conclude by discussing its application in deep learning, particularly in training ResNet architectures, where optimal transport is applied independently to each residual connection. This final part is based on joint work with Raphaël Barboni and François-Xavier Vialard.

17:00 - 19:15 – Poster Session

19:30 - 12:30 – Diner

Wednesday

09:00 - 09:45 – Anastasia Koloskova

Title: Gradient Descent with Linearly Correlated Noise: Theory and Applications to Differential Privacy

Abstract: We study gradient descent under linearly correlated noise. Our work is motivated by recent practical methods for optimization with differential privacy (DP), such as DP-FTRL, which achieve strong performance in settings where privacy amplification techniques are infeasible (such as in federated learning). These methods inject privacy noise through a matrix factorization mechanism, making the noise linearly correlated over iterations. We propose a simplified setting that distills key facets of these methods and isolates the impact of linearly correlated noise. We analyze the behavior of gradient descent in this setting, for both convex and non-convex functions. Our analysis is demonstrably tighter than prior work and recovers multiple important special cases exactly (including anticorrelated perturbed gradient descent). We use our results to develop new, effective matrix factorizations for differentially private optimization, and highlight the benefits of these factorizations theoretically and empirically.

09:45 - 10:30 – Cyril Letrouit

Title: Some insights on training two-layers transformers

Abstract: TBA

10:30 - 11:00 – Coffee break

11:00 - 11:45 – Mathieu Even

Title: Asynchronous speedups in centralized and decentralized distributed optimization

Abstract: In this talk, we will first introduce a distributed optimization setting, in which an optimizer aims at minimizing a loss function with respect to data distributed amongst many agents. We differentiate scenarii in which (i) a central entity orchestrates the training (centralized setting) and (ii) the decentralized setting, in which there is no central orchestrator and agents collaborate to reach a model consensus. In both cases, we will study the effect of asynchrony on the optimization procedure (mainly focusing on the convex optimization setting): how is the training time — in terms of wall-clock running time — affected by eventual delays, and when can optimizers gain from asynchrony? We will quantify precisely this gain with respect to baseline algorithms in terms of wall-clock running time (coined ”asynchronous speedup”), as functions of communication delays between agents and local compute times, first in the centralized setting, and then see that it can be generalized to the decentralized setting.

11:45 - 12:30 – Mikael Johansson

Title: Better models for asynchronous optimization

Abstract: A key challenge in asynchronous optimization is that gradients provide local descent directions. This makes it challenging to combine gradients computed at different iterates into a meaningful search direction that ensures descent. In contrast, gradients (together with function values and convexity) provide global lower bounds on the objective function, making it easier to combine gradients evaluated at widely different points into a valid lower bound of the objective function.

In this talk, I will describe a parallel and asynchronous optimization algorithm that uses this idea to decouple gradient evaluations at the workers from the decision vector updates at the coordinating master. The resulting method becomes insensitive to the amount of asynchrony in the system: it converges for all bounded delays and can be implemented without knowledge of the maximum delay. Numerical experiments confirm that the proposed algorithm converges quickly in practice with essentially no tuning.

12:30 - 14:00 – Lunch

14:00 - 19:00 – Free Afternoon

19:30 - 12:30 – Diner

Thursday

09:00 - 09:45 – Nicolas Courty

Title: Unbalancing Sliced Wasserstein

Abstract: Optimal transport (OT) has emerged as a powerful framework to compare probability measures, a fundamental task in many statistical and machine learning problems. Substantial advances have been made over the last decade in designing OT variants which are either computationally and statistically more efficient, or more robust to the measures and datasets to compare. Among them, sliced OT distances have been extensively used to mitigate optimal transport's cubic algorithmic complexity and curse of dimensionality. In parallel, unbalanced OT was designed to allow comparisons of more general positive measures, while being more robust to outliers. In this talk, I will discuss how to combine these two concepts, namely slicing and unbalanced OT, to develop a general framework for efficiently comparing positive measures. We propose two new loss functions based on the idea of slicing unbalanced OT, and study their induced topology and statistical properties. We then develop a fast Frank-Wolfe-type algorithm to compute these loss functions, and show that the resulting methodology is modular as it encompasses and extends prior related work. We finally conduct an empirical analysis of our loss functions and methodology on both synthetic and real datasets, to illustrate their relevance and applicability.

09:45 - 10:30 – Julio Backhoff

Title: Martingale Benamou-Brenier: duality and gradient flow

Abstract: A central quest in optimal transport is to determine the (distribution of) paths which, while remaining close to a constant velocity reference path, interpolate between an intial and terminal law. In mathematical finance a closely related question has been recently posed: what is the martingale which interpolates between an intial and terminal law and stays as close as possible to a constant volatility reference martingale?

In this talk we characterize the solution to this question, emphasizing duality techniques. We also present a new numerical method for this problem, based on a gradient flow formulation.

10:30 - 11:00 – Coffee break

11:00 - 11:45 – Rémi Flamary

Title: Any2Graph: Deep End-To-End Supervised Graph Prediction With An Optimal Transport Loss

Abstract: We present Any2graph, a generic framework for end-to-end Supervised Graph Prediction (SGP) i.e. a deep learning model that predicts an entire graph for any kind of input. The framework is built on a novel Optimal Transport loss, the Partially-Masked Fused Gromov-Wasserstein, that exhibits all necessary properties (permutation invariance, differentiability and scalability) and is designed to handle any-sized graphs. Numerical experiments showcase the versatility of the approach that outperform existing competitors on a novel challenging synthetic dataset and a variety of real-world tasks such as map construction from satellite image (Sat2Graph) or molecule prediction from fingerprint (Fingerprint2Graph).

11:45 - 12:30 – Aram-Alexander Pooladian

Title: Algorithms for mean-field variational inference via polyhedral optimization in the Wasserstein space

Abstract: We develop a theory of finite-dimensional polyhedral subsets over the Wasserstein space and optimization of functionals over them via first-order methods. Our main application is to the problem of mean-field variational inference, which seeks to approximate a distribution, called the posterior, by the closest product measure in the sense of Kullback–Leibler divergence, called the mean-field approximation. We propose a novel optimization procedure for computing the mean-field approximation, where we are able to provide concrete algorithmic guarantees under the standard assumption that the posterior is strongly log-concave and log-smooth. Joint work with Yiheng Jiang and Sinho Chewi.

12:30 - 14:00 – Lunch

14:00 - 14:45 – Sinho Chewi

Title: Log-concave sampling

Abstract: I will provide a survey on log-concave sampling from the perspective of optimization via the theory of optimal transport.

14:45 - 15:30 – Rémi Gribonval

Title: Conservation laws for gradient flows

Abstract: Understanding the geometric properties of gradient descent dynamics is a key ingredient in deciphering the recent success of very large machine learning models. A striking observation is that trained over-parameterized models retain some properties of the optimization initialization. This “implicit bias” is believed to be responsible for some favorable properties of the trained models and could explain their good generalization properties. In this work, we expose the definitions and properties of ”conservation laws”, that define quantities conserved during gradient flows of a given machine learning model, such as a ReLU network, with any training data and any loss. After explaining how to find the maximal number of independent conservation laws via Lie algebra computations, we provide algorithms to compute a family of polynomial laws, as well as to compute the number of (not necessarily polynomial) conservation laws. We obtain that on a number of architecture there are no more laws than the known ones, and we identify new laws for certain flows with momentum and/or non-Euclidean geometries. (Joint work with Sibylle Marcotte and Gabriel Peyré)

15:30 - 16:00 – Coffee break

16:00 - 16:45 – Audrey Repetti

Title: An optimisation view of learning robust and flexible denoisers for inverse imaging problems

Abstract: Proximal methods have been extensively used for solving inverse imaging problems. In this context, they are used to find an estimate of an unknown image from degraded measurements, by solving a regularised variational problem. Recently, proximal methods have been mixed with learning approaches to further improve the reconstruction quality. In particular, several works have proposed to replace the operator related to the regularisation by a more sophisticated denoiser, leading to plug-and-play (PnP) methods. In this presentation, we will discuss how optimisation theory can be leveraged to build robust and flexible denoisers. We will show that the resulting PnP methods are competitive with state-of-the-art methods for solving inverse imaging problems.

19:30 - 21:30 – Diner

Friday

09:00 - 09:45 – Barbara Pascal

Title: Bilevel optimization for automated data-driven inverse problem resolution

Abstract: Most inverse problems in signal and image processing are ill-posed. To remove the ambiguity about the solution and design noise-robust estimators, a priori properties, e.g., smoothness or sparsity, can be imposed to the solution through regularization. The main bottleneck to use the obtained variational regularized estimators in practice, i.e., without access to ground truth, is that the quality of the estimates strongly depends on the fine-tuning of the level of regularization. A classical approach to automated and data-driven selection of regularization parameter consists in designing a data-dependent unbiased estimator of the error, the minimization of which provides an approximate of the optimal parameters. The resulting overall procedure can be formulated as a bilevel optimization problem, the inner loop computing the variational regularized estimator and the outer loop selecting hyperparameters. The design of a fully automated data-driven procedure adapted to inverse problems corrupted with highly correlated noise will be described in detail and exemplified on a texture segmentation problem. Its applicability to other inverse problems will be demonstrated through numerical simulations on both synthetic and real-world data.

09:45 - 10:30 – Julien Mairal

Title: Functional Bilevel Optimization for Machine Learning

Abstract: In this talk, we introduce a new functional point of view on bilevel optimization problems for machine learning, where the inner objective is minimized over a function space. These types of problems are most often solved by using methods developed in the parametric setting, where the inner objective is strongly convex with respect to the parameters of the prediction function. The functional point of view does not rely on this assumption and notably allows using over-parameterized neural networks as the inner prediction function. We propose scalable and efficient algorithms for the functional bilevel optimization problem and illustrate the benefits of our approach on instrumental regression and reinforcement learning tasks. This is a joint work with Ieva Petrulionyte and Michael Arbel.

10:30 - 11:00 – Coffee break

11:00 - 11:45 – Malgorzata Bogdan

Title: Unveiling low-dimensional patterns induced by convex non-differentiable regularizers

Abstract: Popular regularizers with non-differentiable penalties, such as Lasso, Elastic Net, Generalized Lasso, or SLOPE, reduce the dimension of the parameter space by inducing sparsity or clustering in the estimators' coordinates. In this paper, we focus on linear regression and explore the asymptotic distributions of the resulting low-dimensional patterns when the number of regressors p is fixed, the number of observations n goes to infinity, and the penalty function increases at the rate of sqrt(n). While the asymptotic distribution of the rescaled estimation error can be derived by relatively standard arguments, the convergence of the pattern does not simply follow from the convergence in distribution, and requires a careful and separate treatment. For this purpose, we use the Hausdorff distance as a suitable mode of convergence for subdifferentials, resulting in the desired pattern convergence. Furthermore, we derive the exact limiting probability of recovering the true model pattern. This probability goes to 1 if and only if the penalty scaling constant diverges to infinity and the regularizer-specific asymptotic irrepresentability condition is satisfied. We then propose simple two-step procedures that asymptotically recover the model patterns, irrespective whether the irrepresentability condition holds.

Interestingly, our theory shows that Fused Lasso cannot reliably recover its own clustering pattern, even for independent regressors. It also demonstrates how this problem can be resolved by "concavifying" the Fused Lasso penalty coefficients. Additionally, sampling from the asymptotic error distribution facilitates comparisons between different regularizers. We provide short simulation studies showcasing an illustrative comparison between the asymptotic properties of Lasso, Fused Lasso, and SLOPE. If time permits we will present extensions to robust and quantile regression and the asymptotic results on the FDR control by SLOPE. This is a joint work with Ivan Hejny, Jonas Wallin and Michal Kos.

11:45 - 12:30 – Hadrien Hendrikx

Title: Investigating Variance Definitions for Stochastic Mirror Descent with Relative Smoothness

Abstract: This talk will be a gentle introduction to (Stochastic) Mirror Descent, which should (hopefully) still be relevant for people that are very familiar with it. We will discuss basic assumptions, convergence results, and how to adapt known objects to the relative setting. We will also try and give a new point of view on several objects, and reflect on how to handle stochasticity in non-Euclidean environment. There will be many few-lines proofs, backed with extensive hand waving about the meaning of these derivations. A more conventional abstract can be found below.

Mirror Descent is a popular algorithm, that extends Gradients Descent (GD) beyond the Euclidean geometry. One of its benefits is to enable strong convergence guarantees through smooth-like analyses, even for objectives with exploding or vanishing curvature. This is achieved through the introduction of the notion of relative smoothness, which holds in many of the common use-cases of Mirror descent. While basic deterministic results extend well to the relative setting, most existing stochastic analyses require additional assumptions on the mirror, such as strong convexity (in the usual sense), to ensure bounded variance. In this work, we revisit Stochastic Mirror Descent (SMD) proofs in the (relatively-strongly-) convex and relatively-smooth setting, and introduce a new (less restrictive) definition of variance which can generally be bounded (globally) under mild regularity assumptions. We then investigate this notion in more details, and show that it naturally leads to strong convergence guarantees for stochastic mirror descent. Finally, we leverage this new analysis to obtain convergence guarantees for the Maximum Likelihood Estimator of a Gaussian with unknown mean and variance.

12:30 - 14:00 – Lunch

14:00 - 15:00 – Goodbye :)