

2025 Activity Report

RESEARCH CENTRE: Inria Paris Centre

IN PARTNERSHIP WITH: CNRS, École normale supérieure - PSL

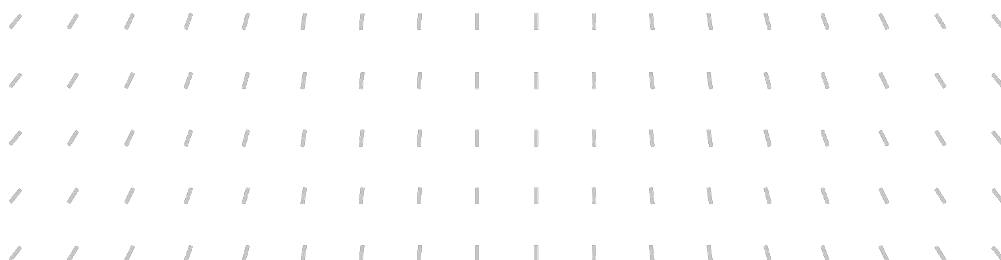

Project-Team

SIERRA

Machine Learning and Optimisation



In collaboration with Département d'Informatique de l'École Normale Supérieure



Project-Team SIERRA

Creation of the Project-Team: 2025 January 01

Each year, Inria research teams publish an Activity Report presenting their work and results over the reporting period. These reports follow a common structure, with some optional sections depending on the specific team. They typically begin by outlining the overall objectives and research programme, including the main research themes, goals, and methodological approaches. They also describe the application domains targeted by the team, highlighting the scientific or societal contexts in which their work is situated. The reports then present the highlights of the year, covering major scientific achievements, software developments, or teaching contributions. When relevant, they include sections on software, platforms, and open data, detailing the tools developed and how they are shared. A substantial part is dedicated to new results, where scientific contributions are described in detail, often with subsections specifying participants and associated keywords. Finally, the Activity Report addresses funding, contracts, partnerships, and collaborations at various levels, from industrial agreements to international cooperations. It also covers dissemination and teaching activities, such as participation in scientific events, outreach, and supervision. The document concludes with a presentation of scientific production, including major publications and those produced during the year.

Keywords

Computer sciences and digital sciences

A3.4. – Machine learning and statistics

A6.2. – Scientific computing, Numerical Analysis & Optimization

A7.1. – Algorithms

A8.2. – Optimization

A9.2. – Machine learning

A9.12. – Computer vision

Other research topics and application domains

B9.5.6. – Data science

Contents

Project-Team SIERRA	1
1 Team members, visitors, external collaborators	5
2 Overall objectives	7
2.1 Statement	7
3 Research program	7
4 Application domains	7
5 Social and environmental responsibility	8
6 Highlights of the year	8
6.1 Awards	8
6.2 Invited talks	8
7 Latest software developments, platforms, open data	8
7.1 Latest software developments	8
7.1.1 PEPit	8
7.2 Open data	9
8 New results	9
8.1 A PAC-Bayesian Link Between Generalisation and Flat Minima	9
8.2 Heavy-Tailed Diffusion with Denoising Lévy Probabilistic Models	9
8.3 Don't Be Greedy, Just Relax! Pruning LLMs via Frank-Wolfe	9
8.4 Algorithm- and Data-Dependent Generalization Bounds for Score-Based Generative Models	10
8.5 The surprising agreement between convex optimization theory and learning-rate scheduling for large model training	10
8.6 Augmented Lagrangian methods for infeasible convex optimization problems and diverging proximal-point algorithms	10
8.7 A constructive approach to strengthen algebraic descriptions of function and operator classes	10
8.8 Optimized projection-free algorithms for online learning: construction and worst-case analysis	11
8.9 Large Stepsizes Accelerate Gradient Descent for Regularized Logistic Regression	11
8.10 Statistical Advantage of Softmax Attention: Insights from Single-Location Regression	11
8.11 Phase Diagram of Dropout for Two-Layer Neural Networks in the Mean-Field Regime	12
8.12 Convergence of Shallow ReLU Networks on Weakly Interacting Data	12
8.13 Convergence of Deterministic and Stochastic Diffusion-Model Samplers: A Simple Analysis in Wasserstein Distance	12
8.14 Adaptive Coverage Policies in Conformal Prediction	12
8.15 Fast kernel methods: Sobolev, physics-informed, and additive models	12
8.16 On the Effectiveness of the z-Transform Method in Quadratic Optimization	13
8.17 Rethinking Early Stopping: Refine, Then Calibrate	13
8.18 Conditional Coverage Diagnostics for Conformal Prediction	13
8.19 Functional protein mining with conformal guarantees	14
8.20 Gradient equilibrium in online learning: Theory and applications	14
8.21 Universal log-optimality for general classes of e-processes and sequential hypothesis tests	14
8.22 The statistical fairness-accuracy frontier	14
9 Bilateral contracts and grants with industry	15
9.1 Bilateral grants with industry	15

10 Partnerships and cooperations	15
10.1 International initiatives	15
10.2 European initiatives	16
10.2.1 Horizon Europe	16
10.2.2 H2020 projects	17
10.3 National initiatives	18
10.4 Regional initiatives	18
11 Dissemination	18
11.1 Promoting scientific activities	18
11.1.1 Scientific events: organisation	18
11.1.2 Scientific events: selection	19
11.1.3 Journal	19
11.1.4 Invited talks	19
11.1.5 Leadership within the scientific community	21
11.1.6 Scientific expertise	21
11.1.7 Research administration	21
11.2 Teaching - Supervision - Juries - Educational and pedagogical outreach	21
11.2.1 Supervision	21
11.2.2 Juries	22
11.2.3 Educational and pedagogical outreach	22
11.3 Popularization	22
11.3.1 Participation in Live events	22
12 Scientific production	23
12.1 Major publications	23
12.2 Publications of the year	25
12.3 Cited publications	28

1 Team members, visitors, external collaborators

Research Scientists

- Francis Bach [Team leader, INRIA, HDR]
- Michael Jordan [Fondation Inria]
- Pierre Marion [INRIA, Researcher, from Sep 2025]
- Umut Simsekli [INRIA, Researcher]
- Adrien Taylor [INRIA, Researcher, HDR]
- Alexandre d'Aspremont [CNRS, Senior Researcher, HDR]

Post-Doctoral Fellows

- Luc Brogat-Motte [CENTRALESUPELEC, Post-Doctoral Fellow, until Mar 2025]
- Yurong Chen [INRIA, Post-Doctoral Fellow]
- Fajwel Fogel [ENS PARIS, Post-Doctoral Fellow, until Aug 2025]
- Armand Gissler [INRIA, Post-Doctoral Fellow, from Feb 2025]
- Maxime Haddouche [INRIA, Post-Doctoral Fellow]
- David Holzmüller [INRIA, Post-Doctoral Fellow, until Sep 2025]
- Frederik Kunstner [INRIA, Post-Doctoral Fellow]
- Fabian Schaipp [INRIA, Post-Doctoral Fellow]
- Corbinian Schlosser [INRIA, Post-Doctoral Fellow, until Jun 2025]
- Yang Su [ENS PARIS, Post-Doctoral Fellow, until Oct 2025]
- Manu Upadhyaya [INRIA, from Sep 2025]
- Julien Weibel [INRIA, Post-Doctoral Fellow]

PhD Students

- Roland Andrews [INRIA]
- Andrea Basteri [INRIA]
- Axel Benyamine [Ecole Polytechnique, from Sep 2025]
- Daniel Berg Thomsen [INRIA, from Nov 2025]
- Eugene Berta [INRIA]
- Eliot Beyler [INRIA, from Sep 2025]
- Pierre Boudart [INRIA]
- Nabil Boukir [INRIA]
- Sacha Braun [INRIA]
- Sarah Brood [ENS Paris]
- Arthur Calvi [CNRS, until Oct 2025]

- Aymeric Capitaine [Ecole Polytechnique]
- Léo Dana [ENS PARIS, from Oct 2025]
- Juliette Decugis [Meta, CIFRE]
- Benjamin Dupuis [INRIA]
- Alexandre Francois [INRIA, until May 2025]
- Etienne Gauthier [INRIA]
- Mahmoud Hegazy [Ecole Polytechnique]
- Clément Lezane [University of Twente, until Aug 2025]
- Simon Martin [ENS Paris]
- Gaëtan Narozniak [Meta, CIFRE, from Dec 2025]
- Antoine Scheid [Ecole Polytechnique]
- Dario Shariatian [INRIA]
- Lawrence Stewart [INRIA, until Mar 2025]
- Mario Tuci [INRIA, from Oct 2025]
- Weijia Wang [Sorbonne University]

Interns and Apprentices

- Theo Goix [ENS PARIS, Intern, from Jun 2025 until Jul 2025]
- Noah Liniger [ETH Zurich, Intern, from Sep 2025]
- Ayoub Melliti [INRIA, Intern, from Mar 2025 until Aug 2025]
- Si Yi Meng [INRIA, from Feb 2025]

Administrative Assistants

- Marina Kovacic [INRIA]
- Abigail Palma [INRIA]

Visiting Scientists

- Baptiste Abélès [Universitat Pompeu Fabra, from Oct 2025]
- Ioan-Liviu Aolaritei [U.C. Berkeley, until Jan 2025]
- Manish Krishan Lal [Technical University of Munich, from Oct 2025]

External Collaborator

- Marc Lambert [DGA, from Mar 2025]

2 Overall objectives

2.1 Statement

Machine learning is a recent scientific domain, positioned between applied mathematics, statistics and computer science. Its goals are the optimization, control, and modeling of complex systems from examples. It applies to data from numerous engineering and scientific fields (e.g., vision, bioinformatics, neuroscience, audio processing, text processing, economy, finance, etc.), the ultimate goal being to derive general theories and algorithms allowing advances in each of these domains. Machine learning is characterized by the high quality and quantity of the exchanges between theory, algorithms and applications: interesting theoretical problems almost always emerge from applications, while theoretical analysis allows the understanding of why and when popular or successful algorithms do or do not work, and leads to proposing significant improvements.

Our academic positioning is exactly at the intersection between these three aspects—algorithms, theory and applications—and our main research goal is to make the link between theory and algorithms, and between algorithms and high-impact applications in various engineering and scientific fields.

3 Research program

Machine learning has emerged as its own scientific domain in the last 30 years, providing a good abstraction of many problems and allowing exchanges of best practices between data oriented scientific fields. Among its main research areas, there are currently probabilistic models, supervised learning (including neural networks), unsupervised learning, reinforcement learning, and statistical learning theory. All of these are represented in the SIERRA team, but the main goals of the team are mostly related to supervised learning and optimization, and their mutual interactions, as well as with interdisciplinary collaborations. One particularity of the team is the strong focus on optimization (in particular convex optimization, but with more works in the non-convex world recently), leading to contributions in optimization which go beyond the machine learning context. Moreover, we interact more and more with other disciplines of applied mathematics (e.g., numerical analysis, control), and economics.

We have divided our research effort in four axes.

1. Optimization
2. Statistical machine learning
3. Machine learning in interaction
4. Incentives and machine learning

4 Application domains

Machine learning research can be conducted from two main perspectives: the first one, which has been dominant in the last 30 years, is to design learning algorithms and theories which are as generic as possible, the goal being to make as few assumptions as possible regarding the problems to be solved and to let data speak for themselves. This has led to many interesting methodological developments and successful applications. However, we believe that this strategy has reached its limit for many application domains, such as computer vision, bioinformatics, neuro-imaging, text and audio processing, which leads to the second perspective our team is built on: Research in machine learning theory and algorithms should be driven by interdisciplinary collaborations, so that specific prior knowledge may be properly introduced into the learning process, in particular with the following fields:

- Computer vision: object recognition, object detection, image segmentation, image/video processing, computational photography. In collaboration with the Willow project-team.
- Bioinformatics: cancer diagnosis, protein function prediction, virtual screening.

- Text processing: document collection modeling, language models.
- Audio processing: source separation, speech/music processing.
- Climate science (satellite imaging).
- AI for mathematical proofs and reasoning.

5 Social and environmental responsibility

As one domain within applied mathematics and computer science, machine learning and artificial intelligence may contribute positively to the environment for example by measuring climate change effect or reducing the carbon footprint of other sciences and activities. But it may also contribute negatively, notably by the ever-increasing sizes of machine learning models. Within the team, we work on these two aspects through our work on climate science and on frugal algorithms.

- Francis Bach: Member of the Comité consultatif national d'éthique du numérique (CCNEN).

6 Highlights of the year

6.1 Awards

- Election of Michael Jordan at the Chinese Academy of Sciences
- PhD award for Baptiste Goujaud: [2025 PhD award from department of mathematics, Ecole Polytechnique](#).
- PhD award for Antoine Bambade: [2025 Paul Caseau PhD award](#) (from EDF R&D).

6.2 Invited talks

- Plenary talk at ICCOPT 2025 for Alexandre d'Aspremont
- Plenary talk at COLT 2025 for Francis Bach
- Plenary talk at the France AI summit for Michael Jordan

7 Latest software developments, platforms, open data

7.1 Latest software developments

7.1.1 PEPit

Name: PEPit

Keyword: Optimisation

Functional Description: PEPit is a Python package aiming at simplifying the access to worst-case analyses of a large family of first-order optimization methods possibly involving gradient, projection, proximal, or linear optimization oracles, along with their approximate, or Bregman variants. In short, PEPit is a package enabling computer-assisted worst-case analyses of first-order optimization methods. The key underlying idea is to cast the problem of performing a worst-case analysis, often referred to as a performance estimation problem (PEP), as a semidefinite program (SDP) which can be solved numerically. For doing that, the package users are only required to write first-order methods nearly as they would have implemented them. The package then takes care of the SDP modelling parts, and the worst-case analysis is performed numerically via a standard solver.

This software is primarily based on the works on performance estimation problems by Adrien Taylor. Compared to other scientific software, its maintenance is relatively low cost (we can do it ourselves, together with students involved in using those techniques). We plan to continue updating this software by incorporating recent advances of the community, and with the clear long term idea of making it a tool for teaching first-order optimization.

URL: <https://pepit.readthedocs.io/en/0.2.0/>

Contact: Adrien Taylor

7.2 Open data

8 New results

8.1 A PAC-Bayesian Link Between Generalisation and Flat Minima

Modern machine learning usually involves predictors in the overparametrised setting (number of trained parameters greater than dataset size), and their training yield not only good performances on training data, but also good generalisation capacity. This phenomenon challenges many theoretical results, and remains an open problem. To reach a better understanding, in [14] we provide novel generalisation bounds involving gradient terms. To do so, we combine the PAC-Bayes toolbox with Poincaré and Log-Sobolev inequalities, avoiding an explicit dependency on dimension of the predictor space. Our results highlight the positive influence of flat minima (being minima with a neighbourhood nearly minimising the learning problem as well) on generalisation performances, involving directly the benefits of the optimisation phase.

8.2 Heavy-Tailed Diffusion with Denoising Lévy Probabilistic Models

Exploring noise distributions beyond Gaussian in diffusion models remains an open challenge. While Gaussian-based models succeed within a unified SDE framework, recent studies suggest that heavy-tailed noise distributions, like α -stable distributions, may better handle mode collapse and effectively manage datasets exhibiting class imbalance, heavy tails, or prominent outliers. Recently, Yoon et al. (NeurIPS 2023), presented the Lévy-Itô model (LIM), directly extending the SDE-based framework to a class of heavy-tailed SDEs, where the injected noise followed an α -stable distribution, a rich class of heavy-tailed distributions. However, the LIM framework relies on highly involved mathematical techniques with limited flexibility, potentially hindering broader adoption and further development. In [30], instead of starting from the SDE formulation, we extend the denoising diffusion probabilistic model (DDPM) by replacing the Gaussian noise with α -stable noise. By using only elementary proof techniques, the proposed approach, Denoising Lévy Probabilistic Models (DLPM), boils down to vanilla DDPM with minor modifications. As opposed to the Gaussian case, DLPM and LIM yield different training algorithms and different backward processes, leading to distinct sampling algorithms. These fundamental differences translate favorably for DLPM as compared to LIM: our experiments show improvements in coverage of data distribution tails, better robustness to unbalanced datasets, and improved computation times requiring smaller number of backward steps.

8.3 Don't Be Greedy, Just Relax! Pruning LLMs via Frank-Wolfe

Pruning is a common technique to reduce the compute and storage requirements of Neural Networks. While conventional approaches typically retrain the model to recover pruning-induced performance degradation, state-of-the-art Large Language Model (LLM) pruning methods operate layer-wise, minimizing the per-layer pruning error on a small calibration dataset to avoid full retraining, which is considered computationally prohibitive for LLMs. However, finding the optimal pruning mask is a hard combinatorial problem and solving it to optimality is intractable. Existing methods hence rely on greedy heuristics that ignore the weight interactions in the pruning objective. In [74], we instead consider the convex relaxation of these combinatorial constraints and solve the resulting problem using the Frank-Wolfe (FW) algorithm. Our method drastically reduces the per-layer pruning error, outperforms strong baselines on state-of-the-art GPT architectures, and remains memory-efficient. We provide theoretical justification by showing that, combined

with the convergence guarantees of the FW algorithm, we obtain an approximate solution to the original combinatorial problem upon rounding the relaxed solution to integrality.

8.4 Algorithm- and Data-Dependent Generalization Bounds for Score-Based Generative Models

Score-based generative models (SGMs) have emerged as one of the most popular classes of generative models. A substantial body of work now exists on the analysis of SGMs, focusing either on discretization aspects or on their statistical performance. In the latter case, bounds have been derived, under various metrics, between the true data distribution and the distribution induced by the SGM, often demonstrating polynomial convergence rates with respect to the number of training samples. However, these approaches adopt a largely approximation theory viewpoint, which tends to be overly pessimistic and relatively coarse. In particular, they fail to fully explain the empirical success of SGMs or capture the role of the optimization algorithm used in practice to train the score network. To support this observation, in [10], we first present simple experiments illustrating the concrete impact of optimization hyperparameters on the generalization ability of the generated distribution. Then, this paper aims to bridge this theoretical gap by providing the first algorithmic- and data-dependent generalization analysis for SGMs. In particular, we establish bounds that explicitly account for the optimization dynamics of the learning algorithm, offering new insights into the generalization behavior of SGMs. Our theoretical findings are supported by empirical results on several datasets.

8.5 The surprising agreement between convex optimization theory and learning-rate scheduling for large model training

In [28], we show that learning-rate schedules for large model training behave surprisingly similar to a performance bound from non-smooth convex optimization theory. We provide a bound for the constant schedule with linear cooldown; in particular, the practical benefit of cooldown is reflected in the bound due to the absence of logarithmic terms. Further, we show that this surprisingly close match between optimization theory and practice can be exploited for learning-rate tuning: we achieve noticeable improvements for training 124M and 210M Llama-type models by (i) extending the schedule for continued training with optimal learning-rate, and (ii) transferring the optimal learning-rate across schedules.

8.6 Augmented Lagrangian methods for infeasible convex optimization problems and diverging proximal-point algorithms

In [2], we investigate the convergence behavior of augmented Lagrangian methods (ALMs) when applied to convex optimization problems that may be infeasible. ALMs are a popular class of algorithms for solving constrained optimization problems. We establish progressively stronger convergence results, ranging from basic sequence convergence to precise convergence rates, under a hierarchy of assumptions.

In particular, we demonstrate that, under mild assumptions, the sequences of iterates generated by ALMs converge to solutions of the “closest feasible problem”. This study leverages the classical relationship between ALMs and the proximal-point algorithm applied to the dual problem. A key technical contribution is a set of concise results on the behavior of the proximal-point algorithm when applied to functions that may not have minimizers. These results pertain to its convergence in terms of its subgradients and of the values of the convex conjugate.

8.7 A constructive approach to strengthen algebraic descriptions of function and operator classes

It is well known that functions (resp. operators) satisfying a property p on a subset $Q \subset \mathbb{R}^d$ cannot necessarily be extended to a function (resp. operator) satisfying p on the whole of \mathbb{R}^d . Given $Q \subseteq \mathbb{R}^d$, this work considers the problem of obtaining necessary and ideally sufficient conditions to be satisfied by a function (resp. operator) on Q , ensuring the existence of an extension of this function (resp. operator) satisfying p on \mathbb{R}^d .

More precisely, given some property p , we present in [26] a refinement procedure to obtain stronger necessary conditions to be imposed on Q . This procedure can be applied iteratively until the stronger conditions are also sufficient. We illustrate the procedure on a few examples, including the strengthening of existing descriptions for the classes of smooth functions satisfying a Łojasiewicz condition, convex blockwise smooth functions, Lipschitz monotone operators, strongly monotone cocoercive operators, and uniformly convex functions.

In most cases, these strengthened descriptions can be represented, or relaxed, to semi-definite constraints, which can be used to formulate tractable optimization problems on functions (resp. operators) within those classes.

8.8 Optimized projection-free algorithms for online learning: construction and worst-case analysis

In [33], we study and develop projection-free algorithms for online learning with linear optimization oracles (a.k.a. Frank–Wolfe) for handling the constraint set. More precisely, this work (i) shows how to exploit semidefinite programming to jointly design and analyze online Frank–Wolfe-type algorithms numerically in a variety of settings, (ii) leverage those design techniques to propose an improved (optimized) variant of an online Frank–Wolfe algorithm along with its conceptually simple potential-based proof, and (iii) its anytime version which benefits from similar $O(T^{3/4})$ regret rate without requiring to know the time horizon T in advance. We are not aware of other direct regret guarantees for anytime version of online Frank–Wolfe without using the classical doubling trick.

Based on the semidefinite technique, we conclude with strong numerical evidence suggesting that no pure online Frank–Wolfe algorithm within our model class can have a regret guarantee better than $O(T^{3/4})$ without additional assumptions, that the current algorithms do not have optimal constants, and that multiple linear optimization rounds do not generally help to obtain better regre

8.9 Large Stepsizes Accelerate Gradient Descent for Regularized Logistic Regression

In [35], we investigate the convergence dynamics of gradient descent (GD) with constant stepsizes for l_2 -regularized logistic regression on linearly separable data. While classical optimization theory prescribes small stepsizes to ensure monotonic objective reduction, yielding a convergence rate linear in the condition number κ , this study demonstrates that large stepsizes can accelerate this rate to $\tilde{O}(\sqrt{\kappa})$. This acceleration leverages the "Edge of Stability" regime, where the objective evolves non-monotonically, effectively matching the optimal rates of Nesterov's momentum without explicit acceleration terms. We extend prior analyses from unregularized convex settings to the strongly convex case with finite minimizers. Furthermore, the study establishes that these benefits extend to generalization bounds, improving the best-known bounds for minimizing population risk under separable distribution. Finally, the work provides a sharp characterization of the maximum stepsize threshold for local convergence.

8.10 Statistical Advantage of Softmax Attention: Insights from Single-Location Regression

In [11], we provide a theoretical grounding for the prevalence of softmax attention over linear alternatives in Large Language Models. Focusing on the "Single-Location Regression" task, where the output depends on a single token at a random position, we employ statistical physics techniques to analyze the learning dynamics in the high-dimensional limit. We prove that softmax attention achieves the optimal Bayes risk, whereas linear attention fundamentally falls short due to inherent approximation limitations.

In particular, the study characterizes generalization performance through a small set of order parameters, demonstrating that both the exponential nonlinearity and the normalization scheme are critical for this optimality. We further derive self-consistent equations to describe the regularized empirical risk minimizer and extend their analysis to the finite-sample regime. In this regime, while softmax is no longer strictly Bayes-optimal, it is shown to consistently outperform linear attention, offering robust statistical evidence for its practical dominance.

8.11 Phase Diagram of Dropout for Two-Layer Neural Networks in the Mean-Field Regime

In [6], we investigate the training dynamics of two-layer neural networks trained with dropout in the large-width mean-field regime. We derive a rich asymptotic phase diagram comprising five distinct nondegenerate phases, determined by the relative scalings of width, learning rate, and dropout rate. A key finding is that the conventional “penalty” interpretation of dropout as an implicit regularizer only persists for impractically small learning rates of order $O(1/\text{width})$. In the more practical regime of larger learning rates, the study demonstrates that dropout acts instead as a “random geometry” modification, equivalent to a random block-coordinate descent. In this limit, the dynamics are described by mean-field jump processes driven by Poisson or Bernoulli clocks. The analysis employs a combination of coupling techniques for mean-field particle systems and martingale methods to establish convergence in both path and distribution spaces.

8.12 Convergence of Shallow ReLU Networks on Weakly Interacting Data

We analyse in [50] the convergence of one-hidden-layer ReLU networks trained by gradient flow on n data points. Our main contribution leverages the high dimensionality of the ambient space, which implies low correlation of the input samples, to demonstrate that a network with width of order $\log(n)$ neurons suffices for global convergence with high probability. Our analysis uses a Polyak–Łojasiewicz viewpoint along the gradient-flow trajectory, which provides an exponential rate of convergence of $1/n$. When the data are exactly orthogonal, we give further refined characterizations of the convergence speed, proving its asymptotic behavior lies between the orders $\frac{1}{n}$ and $1/\sqrt{n}$, and exhibiting a phase-transition phenomenon in the convergence rate, during which it evolves from the lower bound to the upper, and in a relative time of order $1/\log(n)$.

8.13 Convergence of Deterministic and Stochastic Diffusion-Model Samplers: A Simple Analysis in Wasserstein Distance

We provide in [54] new convergence guarantees in Wasserstein distance for diffusion-based generative models, covering both stochastic (DDPM-like) and deterministic (DDIM-like) sampling methods. We introduce a simple framework to analyze discretization, initialization, and score estimation errors. Notably, we derive the first Wasserstein convergence bound for the Heun sampler and improve existing results for the Euler sampler of the probability flow ODE. Our analysis emphasizes the importance of spatial regularity of the learned score function and argues for controlling the score error with respect to the true reverse process, in line with denoising score matching. We also incorporate recent results on smoothed Wasserstein distances to sharpen initialization error bounds.

8.14 Adaptive Coverage Policies in Conformal Prediction

Traditional conformal prediction methods construct prediction sets such that the true label falls within the set with a user-specified coverage level. However, poorly chosen coverage levels can result in uninformative predictions, either producing overly conservative sets when the coverage level is too high, or empty sets when it is too low. Moreover, the fixed coverage level cannot adapt to the specific characteristics of each individual example, limiting the flexibility and efficiency of these methods. In this work, we leverage recent advances in e-values and post-hoc conformal inference, which allow the use of data-dependent coverage levels while maintaining valid statistical guarantees. We propose in [66] to optimize an adaptive coverage policy by training a neural network using a leave-one-out procedure on the calibration set, allowing the coverage level and the resulting prediction set size to vary with the difficulty of each individual example. We support our approach with theoretical coverage guarantees and demonstrate its practical benefits through a series of experiments.

8.15 Fast kernel methods: Sobolev, physics-informed, and additive models

Physics-informed machine learning typically integrates physical priors into the learning process by minimizing a loss function that includes both a data-driven term and a partial differential equation (PDE) regularization.

Building on the formulation of the problem as a kernel regression task, we use in [62] Fourier methods to approximate the associated kernel, and propose a tractable estimator that minimizes the physics-informed risk function. We refer to this approach as physics-informed kernel learning (PIKL). This framework provides theoretical guarantees, enabling the quantification of the physical prior’s impact on convergence speed. We demonstrate the numerical performance of the PIKL estimator through simulations, both in the context of hybrid modeling and in solving PDEs. In particular, we show that PIKL can outperform physics-informed neural networks in terms of both accuracy and computation time. Additionally, we identify cases where PIKL surpasses traditional PDE solvers, particularly in scenarios with noisy boundary conditions.

8.16 On the Effectiveness of the z-Transform Method in Quadratic Optimization

The z-transform of a sequence is a classical tool used within signal processing, control theory, computer science, and electrical engineering. It allows for studying sequences from their generating functions, with many operations that can be equivalently defined on the original sequence and its z-transform. In particular, the z-transform method focuses on asymptotic behaviors and allows the use of Taylor expansions. We present a sequence of results of increasing significance and difficulty for linear models and optimization algorithms, demonstrating the effectiveness and versatility of the z-transform method in deriving new asymptotic results. Starting from the simplest gradient descent iterations in an infinite-dimensional Hilbert space, we show in [51] how the spectral dimension characterizes the convergence behavior. We then extend the analysis to Nesterov acceleration, averaging techniques, and stochastic gradient descent.

8.17 Rethinking Early Stopping: Refine, Then Calibrate

Machine learning classifiers often produce probabilistic predictions that are critical for accurate and interpretable decision-making in various domains. The quality of these predictions is generally evaluated with proper losses like cross-entropy, which decompose into two components: calibration error assesses general under/overconfidence, while refinement error measures the ability to distinguish different classes. In [52], we provide theoretical and empirical evidence that these two errors are not minimized simultaneously during training. Selecting the best training epoch based on validation loss thus leads to a compromise point that is suboptimal for both calibration error and, most importantly, refinement error. To address this, we introduce a new metric for early stopping and hyperparameter tuning that makes it possible to minimize refinement error during training. The calibration error is minimized after training, using standard techniques. Our method integrates seamlessly with any architecture and consistently improves performance across diverse classification tasks.

8.18 Conditional Coverage Diagnostics for Conformal Prediction

Evaluating conditional coverage remains one of the most persistent challenges in assessing the reliability of predictive systems. Although conformal methods can give guarantees on marginal coverage, no method can guarantee to produce sets with correct conditional coverage, leaving practitioners without a clear way to interpret local deviations. To overcome sample-inefficiency and overfitting issues of existing metrics, we cast in [58] conditional coverage estimation as a classification problem. Conditional coverage is violated if and only if any classifier can achieve lower risk than the target coverage. Through the choice of a (proper) loss function, the resulting risk difference gives a conservative estimate of natural miscoverage measures such as L1 and L2 distance, and can even separate the effects of over- and under-coverage, and non-constant target coverages. We call the resulting family of metrics excess risk of the target coverage (ERT). We show experimentally that the use of modern classifiers provides much higher statistical power than simple classifiers underlying established metrics like CovGap. Additionally, we use our metric to benchmark different conformal prediction methods. Finally, we release an open-source package for ERT as well as previous conditional coverage metrics. Together, these contributions provide a new lens for understanding, diagnosing, and improving the conditional reliability of predictive systems.

8.19 Functional protein mining with conformal guarantees

Molecular structure prediction and homology detection offer promising paths to discovering protein function and evolutionary relationships. However, current approaches lack statistical reliability assurances, limiting their practical utility for selecting proteins for further experimental and in-silico characterization. To address this challenge, we **introduce** a statistically principled approach to protein search leveraging principles from conformal prediction, offering a framework that ensures statistical guarantees with user-specified risk and provides calibrated probabilities (rather than raw ML scores) for any protein search model. Our method (1) lets users select many biologically-relevant loss metrics (i.e. false discovery rate) and assigns reliable functional probabilities for annotating genes of unknown function; (2) achieves state-of-the-art performance in enzyme classification without training new models; and (3) robustly and rapidly pre-filters proteins for computationally intensive structural alignment algorithms. Our framework enhances the reliability of protein homology detection and enables the discovery of uncharacterized proteins with likely desirable functional properties.

8.20 Gradient equilibrium in online learning: Theory and applications

We **present** a new perspective on online learning that we refer to as gradient equilibrium: a sequence of iterates achieves gradient equilibrium if the average of gradients of losses along the sequence converges to zero. In general, this condition is not implied by, nor implies, sublinear regret. It turns out that gradient equilibrium is achievable by standard online learning methods such as gradient descent and mirror descent with constant step sizes (rather than decaying step sizes, as is usually required for no regret). Further, as we show through examples, gradient equilibrium translates into an interpretable and meaningful property in online prediction problems spanning regression, classification, quantile estimation, and others. Notably, we show that the gradient equilibrium framework can be used to develop a debiasing scheme for black-box predictions under arbitrary distribution shift, based on simple post hoc online descent updates. We also show that post hoc gradient updates can be used to calibrate predicted quantiles under distribution shift, and that the framework leads to unbiased Elo scores for pairwise preference prediction.

8.21 Universal log-optimality for general classes of e-processes and sequential hypothesis tests

We **consider** the problem of sequential hypothesis testing by betting. For a general class of composite testing problems – which include bounded mean testing, equal mean testing for bounded random tuples, and some key ingredients of two-sample and independence testing as special cases – we show that any e-process satisfying a certain sublinear regret bound is adaptively, asymptotically, and almost surely log-optimal for a composite alternative. This is a strong notion of optimality that has not previously been established for the aforementioned problems and we provide explicit test supermartingales and e-processes satisfying this notion in the more general case. Furthermore, we derive matching lower and upper bounds on the expected rejection time for the resulting sequential tests in all of these cases. The proofs of these results make weak, algorithm-agnostic moment assumptions and rely on a general-purpose proof technique involving the aforementioned regret and a family of numeraire portfolios. Finally, we discuss how all of these theorems hold in a distribution-uniform sense, a notion of log-optimality that is stronger still and seems to be new to the literature.

8.22 The statistical fairness-accuracy frontier

Machine learning models must balance accuracy and fairness, but these goals often conflict, particularly when data come from multiple demographic groups. A useful tool for understanding this trade-off is the fairness-accuracy (FA) frontier, which characterizes the set of models that cannot be simultaneously improved in both fairness and accuracy. Prior analyses of the FA frontier provide a full characterization under the assumption of complete knowledge of population distributions – an unrealistic ideal. We study the FA frontier in the finite-sample regime, showing how it deviates from its population counterpart and quantifying the worst-case gap between them. In particular, we **derive** minimax-optimal estimators that depend on the designer’s knowledge of the covariate distribution. For each estimator, we characterize how finite-sample

effects asymmetrically impact each group’s risk, and identify optimal sample allocation strategies. Our results transform the FA frontier from a theoretical construct into a practical tool for policymakers and practitioners who must often design algorithms with limited data.

9 Bilateral contracts and grants with industry

9.1 Bilateral grants with industry

- Chaire “Marchés et Apprentissage”, portée par Michael Jordan au sein de la Fondation Inria, et lancée en Juillet 2024. En partenariat avec Air Liquide, BNP Paribas Asset Management Europe, EDF, Orange et la SNCF.
- Francis Bach: Co-advised PhD student with Meta.
- Pierre Marion: Co-advised PhD student with Meta.
- Pierre Marion: Gift from Google.org.

10 Partnerships and cooperations

10.1 International initiatives

GHOST

Title: Generative modeling, Heavy tails, Outliers, Sparse Training

Duration: 2025 to 2028

Partners:

- INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET AUTOMATIQUE (INRIA), France
- University of Calgary, Canada

Inria contact: Umut Simsekli

Coordinator: Umut Simsekli

Summary: Generative Artificial Intelligence (GAI) models are expensive, with massive energy requirements for both training and inference (use in applications). As GAI models are increasingly adopted to solve problems across industry, significant changes in how we train and use these models are required both to realize carbon emission goals, and democratize access to GAI models and research. State-of-the-art approaches for compressing neural networks are of limited efficacy when used with GAI models. While in most neural networks 85-95% of the weights can be pruned while maintaining performance, GAI cannot be pruned beyond 70% sparsity without significant degradation in performance. Empirically it has been observed that GAI models have different training dynamics that are likely responsible for affecting their compressibility: (a) trained GAI models have outlier weights/activations that appear to be important, and render conventional pruning and quantization less effective, (b) it appears that lower-magnitude weights carry more importance in GAI models than other deep learning models. Both of these empirical observations are currently poorly understood. Recently, we have illustrated that such outliers in optimization may occur due to the emergence of “heavy tails”, and heavy-tailed distributions have tight links with compressibility. In this proposal, our main objective is to develop a theoretically sound algorithmic framework for achieving state-of-the-art compression techniques for GAI. We will first explore the connections between heavy-tails and the behavior of the outliers observed in GAI, and understand how the training dynamics of GAI differ from other deep learning models. By exploiting this connection, we will then develop efficient algorithms that will significantly reduce the computational complexity both in memory and run-time. We will produce open-source software and test their performance on applications on computer vision, audio/language processing.

10.2 European initiatives

10.2.1 Horizon Europe

DYNASTY [DYNASTY project on cordis.europa.eu](https://cordis.europa.eu/project/DYNASTY)

Title: Dynamics-Aware Theory of Deep Learning

Duration: From October 1, 2022 to September 30, 2027

Partners:

- INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET AUTOMATIQUE (INRIA), France

Inria contact: Umut Simsekli

Coordinator:

Summary: The recent advances in deep learning (DL) have transformed many scientific domains and have had major impacts on industry and society. Despite their success, DL methods do not obey most of the wisdoms of statistical learning theory, and the vast majority of the current DL techniques mainly stand as poorly understood black-box algorithms.

Even though DL theory has been a very active research field in the past few years, there is a significant gap between the current theory and practice: (i) the current theory often becomes vacuous for models with large number of parameters (which is typical in DL), and (ii) it cannot capture the interaction between data, architecture, training algorithm and its hyper-parameters, which can have drastic effects on the overall performance. Due to this lack of theoretical understanding, designing new DL systems has been dominantly performed by ad-hoc, 'trial-and-error' approaches.

The main objective of this proposal is to develop a mathematically sound and practically relevant theory for DL, which will ultimately serve as the basis of a software library that provides practical tools for DL practitioners. In particular, (i) we will develop error bounds that closely reflect the true empirical performance, by explicitly incorporating the dynamics aspect of training, (ii) we will develop new model selection, training, and compression algorithms with reduced time/memory/storage complexity, by exploiting the developed theory.

To achieve the expected breakthroughs, we will develop a novel theoretical framework, which will enable tight analysis of learning algorithms in the lens of dynamical systems theory. The outcomes will help relieve DL from being a black-box system and avoid the heuristic design process. We will produce comprehensive open-source software tools adapted to all popular DL libraries, and test the developed algorithms on a wide range of real applications arising in computer vision, audio/music/natural language processing.

CASPER [CASPER project on cordis.europa.eu](https://cordis.europa.eu/project/CASPER)

Title: Systematic and computer-aided performance certification for numerical optimization

Duration: From November 1, 2024 to October 31, 2029

Partners:

- INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET AUTOMATIQUE (INRIA), France

Inria contact: Adrien Taylor

Coordinator:

Summary: Numerical optimization is a fundamental tool with a growing impact in many disciplines from science to industry. Many of its successes are due to theoretical advances, which are key to developing trust in numerical algorithms. While trust is non-negotiable in many applications, the complexity level of modern and future problems makes it very hard for theory to keep up with efficient proposals. Arguably worse, while both theory and experimental practice are key to the field, their respective recommendations often conflict with each other and the gap between theory and practice gets embarrassingly large.

The main objective of this proposal is to push forward the theoretical foundations of algorithmic optimization to drastically reduce the gap between fundamental theoretical understanding and practical scenarios. To achieve this, we will develop principled and systematic approaches to algorithmic analyses, as well as computer-aided performance certification tools. Whereas my recent works show that such techniques already allow going far beyond the surprisingly few classical templates for algorithmic analysis, they have currently very limited applicability beyond simple scenarios. We will largely broaden the techniques to develop and study modern algorithms with working guarantees that can (i) scale to unprecedented problem and data sizes, (ii) adapt to common problem structures, and (iii) be deployed on modern massively parallel computing environments. On the way, this project will allow for simplified certification and validation of existing theory, an absolute necessity in this era of massive scientific production.

Outcomes of CASPER will include symbolical and numerical algorithmic certification and development tools, as well as algorithms with unprecedented working guarantees. The tools will be released as open-source libraries and algorithms validated on key benchmarks that include challenging machine learning and robotic tasks.

10.2.2 H2020 projects

REAL [REAL project on cordis.europa.eu](http://real.project.on.cordis.europa.eu)

Title: Reliable and cost-effective large scale machine learning

Duration: From April 1, 2021 to March 31, 2026

Partners:

- INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET AUTOMATIQUE (INRIA), France
- UNIVERSITA COMMERCIALE LUIGI BOCCONI (UB), Italy

Inria contact: Alessandro Rudi

Coordinator:

Summary: In the last decade, machine learning (ML) has become a fundamental tool with a growing impact in many disciplines, from science to industry. However, nowadays, the scenario is changing: data are exponentially growing compared to the computational resources (post Moore's law era), and ML algorithms are becoming crucial building blocks in complex systems for decision making, engineering, science. Current machine learning is not suitable for the new scenario, both from a theoretical and a practical viewpoint: (a) the lack of cost-effectiveness of the algorithms impacts directly the economic/energetic costs of large scale ML, making it barely affordable by universities or research institutes; (b) the lack of reliability of the predictions affects critically the safety of the systems where ML is employed. To deal with the challenges posed by the new scenario, REAL will lay the foundations of a solid theoretical and algorithmic framework for reliable and cost-effective large scale machine learning on modern computational architectures. In particular, REAL will extend the classical ML framework to provide algorithms with two additional guarantees: (a) the predictions will be reliable, i.e., endowed with explicit bounds on their uncertainty guaranteed by the theory; (b) the algorithms will be cost-effective, i.e., they will be naturally adaptive to the new architectures and will provably achieve the desired reliability and accuracy level, by using minimum possible computational resources. The algorithms resulting from REAL will be released as open-source libraries for distributed

and multi-GPU settings, and their effectiveness will be extensively tested on key benchmarks from computer vision, natural language processing, audio processing, and bioinformatics. The methods and the techniques developed in this project will help machine learning to take the next step and become a safe, effective, and fundamental tool in science and engineering for large scale data problems.

10.3 National initiatives

- Alexandre d’Aspremont, Francis Bach, Michael Jordan: Chairs from the PRAIRIE-PSAI Cluster.

10.4 Regional initiatives

- Pierre Marion: Tremplin Chair from the PRAIRIE-PSAI Cluster.

Title: Mathematical Foundations of Modern Deep Learning

Duration: From September 1, 2025 to September 30, 2029

Summary: Recent years have witnessed breakthroughs across many fields of artificial intelligence (AI), largely driven by rapid advances in deep learning techniques. At the same time, modern AI models also present fundamental flaws: hallucinations, copyright infringements, biases, brittleness to adversarial attacks, economic and ecological cost. On the theoretical side, many fundamental questions regarding the striking effectiveness of deep learning remain open. While general theories of deep learning have provided valuable insights, they do not always capture the wide variety of settings encompassed in practice. My overarching research goal is to **address some of these challenges, by leveraging a mathematically-grounded approach to understand and improve modern AI techniques**. My research proposal is structured around three complementary axes towards advancing this goal: (i) Theoretical insights on generative models. The first axis explores core methodologies underpinning modern generative AI, particularly denoising diffusion models and Transformers, which form the backbone of large language models (LLMs). I seek to analyze how specific architectural choices and training procedures impact performance, robustness, and efficiency. (ii) Deep learning optimization. The remarkable effectiveness of stochastic gradient descent at finding good solutions in deep learning settings—large non-convex optimization problems—remains only partially understood. My research in this axis focuses on the role of regularization, especially through the lens of optimization dynamics. (iii) LLMs for formal mathematical reasoning. AI-assisted formal reasoning is a rapidly emerging field, recently achieving successes at the level of Olympiad mathematics. These advances bring us closer to AI-assisted theorem proving, with the potential to revolutionize the practice of mathematical research, while also serving as a testbed for the reasoning abilities of LLMs. A particularly promising route involves using LLMs to generate proofs in a formal language such as Lean or Rocq. This raises crucial methodological questions that for now have been little investigated. For instance: Is it advantageous to represent proofs as trees rather than unstructured sequences of text? If so, how can we guide LLMs in exploring the proof tree efficiently? And can reinforcement learning be used to train LLMs in this context, despite the absence of a standard two-player game framework (as in chess or Go)?

11 Dissemination

11.1 Promoting scientific activities

11.1.1 Scientific events: organisation

Member of the organizing committees

- Adrien Taylor: Cluster chair at EUROPT 2025
- Pierre Marion: Affinity and Inclusion Chair at EurIPS in 2025.
- Pierre Marion: Organizer of the Workshop on Principles of Generative Modeling at EurIPS in 2025.

- Francis Bach, Maxime Haddouche: Organizers of NeurIPS in Paris 2025.

11.1.2 Scientific events: selection

Member of the conference program committees

- Umut Simsekli: area chair for Conference on Learning Theory
- Umut Simsekli: area chair for Advances in Neural Processing Systems
- Pierre Marion: reviewer for International Conference on Learning Representations (ICLR 2026).
- Umut Simsekli: reviewer for Conference on Learning Theory

11.1.3 Journal

Member of the editorial boards

- Adrien Taylor & Alexandre d’Aspremont: invited editors, Mathematical Programming series B (“Systematic and computer-aided analyses of optimization algorithms”) with Aymeric Dieuleveut (Ecole Polytechnique) and Laurent Lessard (Northeastern University, US).
- Alexandre d’Aspremont: SIAM Journal on the Mathematics of Data Science.

Reviewer - reviewing activities

- Adrien Taylor: reviewer for Foundations of Computational Mathematics (FOCM).
- Adrien Taylor: reviewer for Automatica.
- Adrien Taylor: reviewer for Journal of Optimization Theory and Applications (JOTA).
- Adrien Taylor: reviewer for SIAM Journal on Optimization (SIOPT).
- Adrien Taylor: reviewer for Mathematical Programming (MPA) – **Service award**.
- Pierre Marion: reviewer for SIAM Journal on Mathematics of Data Science (SIMODS).
- Pierre Marion: reviewer for SIAM Journal on Optimization (SIOPT).
- Pierre Marion: reviewer for Neurocomputing.
- Pierre Marion: reviewer for Journal on Machine Learning Research (JMLR).
- Pierre Marion: reviewer for Bernoulli.
- Umut Simsekli: reviewer for JMLR
- Umut Simsekli: reviewer for Bernoulli

11.1.4 Invited talks

- Adrien Taylor: invited talks at **Probabilistic perspectives in neural network-based machine learning** workshop (10/2025, Oberwolfach).
- Adrien Taylor: invited talk at **Conference on advances in continuous optimization** (09/2025, Southampton).
- Adrien Taylor: invited talk at **Rice in Paris: large-scale learning and optimization** (06/2025, Paris).
- Adrien Taylor: invited talk at **MALGA seminar** (06/2025, Genova).

- Adrien Taylor: invited talk at [Séminaire images optimisation et probabilités](#) (04/2025, Bordeaux).
- Adrien Taylor (declined [ecological reasons]): invited talk at International Conference on Continuous Optimization (ICCOPT) (07/2025, Los Angeles).
- Pierre Marion: invited talk at the [19th International Joint Conference Computational and Financial Econometrics-Computational and Methodological Statistics](#) (12/2025, London).
- Pierre Marion: invited seminar at Centre de Sciences des Données, DI ENS (12/2025, Paris).
- Pierre Marion: invited seminar at ENSAE-CREST (09/2025, Palaiseau).
- Pierre Marion (declined [ecological reasons]): invited talk at the [2025 Canadian Mathematical Society Winter Meeting](#) (12/2025, Toronto).
- Pierre Marion (declined [ecological reasons]): invited talk at the [Workshop Recent Advances in Optimization, Control and AI](#) (11/2025, Shanghai).
- Umut Simsekli: invited talk at Istanbul-Ankara Stochastic days
- Umut Simsekli: invited talk at Lab. Math. de Versaille
- Umut Simsekli: invited talk at Geometry and Machine Learning workshop
- Michael Jordan: Keynote Speaker, AI, Science, and Society, Paris, France, 2/6/25
- Michael Jordan: Keynote Speaker, Next Generation AI and Economic Applications, Morocco, 2/24/25
- Michael Jordan: Keynote Speaker, Workshop on Generative Models and Uncertainty Quantification, Copenhagen, 9/17/25
- Michael Jordan: Invited Speaker, Lawrence Brown Memorial Lecture Series, University of Pennsylvania, 9/29/25-10/2/25
- Michael Jordan: Keynote Speaker, Conference on Croissance, IA et Bien Commun, Paris, 9/25/25
- Michael Jordan: Keynote Speaker, Workshop on AI and Economics, Paris School of Economics, Paris, 10/7/25
- Michael Jordan: Keynote Speaker, Conference on Games and AI for Security, Athens, 10/14/25
- Michael Jordan: Invited Speaker, Collège de France, Colloque de Rentrée, 10/16/25
- Michael Jordan: Keynote Speaker, EurIPS Conference, Copenhagen, 12/4/25
- Francis Bach: invited talk at Workshop on Overparametrization, Regularization, Identifiability and Uncertainty in Machine Learning, Oberwolfach, January 2025
- Francis Bach: invited talk, AI summit, February 2025
- Francis Bach: Aisenstadt Chair invited talks, Montreal, May 2025
- Francis Bach: keynote speaker, International Conference on Stochastic Programming, Paris, July 2025
- Francis Bach: invited speaker, Graduate Summer School on Mathematical Aspects of Data Science, EPFL, September 2025
- Francis Bach: keynote speaker, Conference on Mathematics of Machine Learning, Hamburg, September 2025
- Francis Bach: invited talk, Symposium "60 years FIM", ETH Zurich, June 2025
- Francis Bach: Keynote Speaker, Conference on Learning Theory, July 2025

- Francis Bach: Keynote speaker at workshop on Learned methods for operations research, CWI, November 2025
- Francis Bach: Keynote Speaker, IMS International Conference on Statistics and Data Science (ICSIDS), December 15-18, 2025, Seville, Spain
- Alexandre d’Aspremont: Keynote speaker, ICCOPT 2025, Los Angeles.
- Alexandre d’Aspremont: Centre de recherches mathématiques, Université de Montréal, May 2025.

11.1.5 Leadership within the scientific community

- Francis Bach: member of the ICML board.

11.1.6 Scientific expertise

- Pierre Marion: grant external assessor for NSERC.
- Francis Bach: member of the scientific council of Ile-de-France region.

11.1.7 Research administration

- Adrien Taylor: comité de suivant des doctorants.

11.2 Teaching - Supervision - Juries - Educational and pedagogical outreach

- Adrien Taylor: Convex Optimization (M1, ENS; 21h)
- Adrien Taylor: Convex Optimization (MVA; 3h)
- Adrien Taylor: Optimization & deep learning (M1, X/HEC; 30h)
- Alexandre d’Aspremont: Convex Optimization (MVA; 21h)
- Umut Simsekli: Introduction to Machine Learning (ENS, L3; 12h)
- Francis Bach: Learning Theory from First Principles (M2 IASD; 27h)

11.2.1 Supervision

- Adrien Taylor
 - New PhD student: Daniel Berg Thomsen
 - PhD in progress: Roland Andrews
 - PhD in progress: Weijia Wang
- Pierre Marion
 - new PhD student (started 12/2025): Gaëtan Narozniak.
- Umut Simsekli
 - new Phd student (Mario Tuci, 10/2025)
 - PhD in progress: Benjamin Dupuis
 - PhD in progress: Dario Shariatian
- Alexandre d’Aspremont
 - PhD in progress: Sarah Brood
 - PhD in progress: Arthur Calvi

- PhD in progress: Pierre Boudart (co-advised with Alessandro Rudi)
- PhD in progress: Alvin Opler (co-advised with Philippe Ciais)
- Francis Bach
 - new PhD student: Eliot Beyler
 - new PhD student: Leo Dana
 - PhD in progress: Simon Martin, co-advised with Giulio Biroli (ENS)
 - PhD in progress: Juliette Decugis, co-advised with Gabriel Synnaeve and Taco Cohen (Meta)
 - PhD in progress: Eugène Berta (co-advised with Michael Jordan)
 - PhD in progress: Sacha Braun (co-advised with Michael Jordan)
 - PhD defended: Lawrence Stewart [80]
- Michael Jordan
 - PhD in progress: Nabil Boukir (co-advised with Francis Bach)
 - PhD in progress: Etienne Gauthier (co-advised with Francis Bach)
 - PhD in progress: Antoine scheid
 - PhD in progress: Mahmoud Hegazy
 - PhD in progress: Aymeric Capitaine

11.2.2 Juries

- Adrien Taylor: PhD Jury of Teodor Rotaru (KULeuven, Belgium). November 2025.
- Adrien Taylor: PhD Jury of Joao Vitor Cavalcanti Vilela (MIT, US). August 2025.
- Adrien Taylor: PhD Jury of Nizar Bousselmi (UCLouvain, Belgium). June 2025.
- Umut Simsekli: Phd jury of Aël Queleunenec (Telecom Paris)
- Francis Bach: Phd jury of Sybille Marcotte (ENS Paris)
- Francis Bach: PhD jury of Lorenzo Noci (ETH Zurich)
- Francis Bach: HDR jury of Sebastien Gerchinovitz (Université de Toulouse)
- Alexandre d’Aspremont: HDR jury of Clément Royer (Université de Paris Dauphine)
- Alexandre d’Aspremont: PhD jury of Charles Guille-Escuret, Université de Montréal.

11.2.3 Educational and pedagogical outreach

- Umut Simsekli: Co-organizer of CIMPA summer school on probability and analysis (Istanbul)

11.3 Popularization

11.3.1 Participation in Live events

- Permanent & non-permanent researchers participated in “fête de la science 2025” (Jussieu) (Andrea Basteri, Marc Lambert, Pierre Marion, Adrien Taylor, Julien Weibel).
- Adrien Taylor: demi-heure de la science (Inria Paris).
- Pierre Marion: RJMI Speed meeting.

12 Scientific production

12.1 Major publications

- [1] R. Andreeva, B. Dupuis, R. Sarkar, T. Birdal and U. Şimşekli. ‘Topological Generalization Bounds for Discrete-Time Stochastic Optimization Algorithms’. In: *PMLR. Advances in Neural Information Processing Systems*. Vancouver, Canada, 2024. URL: <https://hal.science/hal-04872158>.
- [2] R. Andrews, J. Carpentier and A. Taylor. *Augmented Lagrangian methods for infeasible convex optimization problems and diverging proximal-point algorithms*. 27th June 2025. URL: <https://hal.science/hal-05461095> (cit. on p. 10).
- [3] A. Askari, A. d’Aspremont and L. E. Ghaoui. ‘Approximation Bounds for Sparse Programs’. In: *SIAM Journal on Mathematics of Data Science* 4.2 (1st June 2022), pp. 514–530. DOI: [10.1137/21M1398677](https://doi.org/10.1137/21M1398677). URL: <https://hal.archives-ouvertes.fr/hal-03165622>.
- [4] A. Bertazzi, D. Shariatian, U. Simsekli, E. Moulines and A. Durmus. ‘Piecewise deterministic generative models’. In: *PMLR. Advances in Neural Information Processing Systems*. Vancouver, Canada, 2024. URL: <https://hal.science/hal-04872155>.
- [5] T. Cantelobre, C. Ciliberto, B. Guedj and A. Rudi. *Measuring dissimilarity with diffeomorphism invariance*. 24th Feb. 2022. DOI: [10.48550/arXiv.2202.05614](https://doi.org/10.48550/arXiv.2202.05614). URL: <https://hal.inria.fr/hal-03573479>.
- [6] L. Chizat, P. Marion and Y. Yesbay. *Phase Diagram of Dropout for Two-Layer Neural Networks in the Mean-Field Regime*. 10th Oct. 2025. URL: <https://hal.science/hal-05307271> (cit. on p. 12).
- [7] R.-A. Dragomir, A. Taylor, A. d’Aspremont and J. Bolte. ‘Optimal Complexity and Certification of Bregman First-Order Methods’. In: *Mathematical Programming* 194.1 (1st July 2022), pp. 41–83. DOI: [10.1007/s10107-021-01618-1](https://doi.org/10.1007/s10107-021-01618-1). URL: <https://hal.inria.fr/hal-02384167>.
- [8] B. Dupuis, G. Deligiannidis and U. Şimşekli. ‘Generalization Bounds using Data-Dependent Fractal Dimensions’. In: *Proceedings of Machine Learning Research*. International Conference on Machine Learning (ICML 2023). Honolulu, United States, 10th July 2023. URL: <https://inria.hal.science/hal-04438550>.
- [9] B. Dupuis, D. Shariatian, M. Haddouche, A. Durmus and U. Simsekli. ‘Algorithm- and Data-Dependent Generalization Bounds for Score-Based Generative Models’. In: *Advances in Neural Information Processing Systems*. San Diego, United States, 2025. URL: <https://hal.science/hal-05466290>.
- [10] B. Dupuis and U. Şimşekli. ‘Generalization Bounds for Heavy-Tailed SDEs through the Fractional Fokker-Planck Equation’. In: *PMLR. International Conference on Machine Learning*. Vienna, Austria, 2024. URL: <https://hal.science/hal-04872149> (cit. on p. 10).
- [11] O. Duranthon, P. Marion, C. Boyer, B. Loureiro and L. Zdeborová. *Statistical Advantage of Softmax Attention: Insights from Single-Location Regression*. 3rd Oct. 2025. URL: <https://hal.science/hal-05295736> (cit. on p. 11).
- [12] M. Gurbuzbalaban, Y. Hu, U. Simsekli, K. Yuan and L. Zhu. ‘Heavy-Tail Phenomenon in Decentralized SGD’. In: *IJSE Transactions* (2024). URL: <https://hal.science/hal-04872162>.
- [13] M. Gürbüzbalaban, Y. Hu, U. Şimşekli and L. Zhu. ‘Cyclic and Randomized Stepsizes Invoke Heavier Tails in SGD than Constant Stepsize’. In: *Transactions on Machine Learning Research Journal* (2023). URL: <https://inria.hal.science/hal-04478948>.
- [14] M. Haddouche, P. Viallard, U. Şimşekli and B. Guedj. ‘A PAC-Bayesian Link Between Generalisation and Flat Minima’. In: *ALT 2025 - 36th International Conference on Algorithmic Learning Theory*. Milan, Italy, 2025, pp. 1–31. URL: <https://hal.science/hal-04455639> (cit. on p. 9).
- [15] L. Hodgkinson, U. Şimşekli, R. Khanna and M. W. Mahoney. ‘Generalization Bounds using Lower Tail Exponents in Stochastic Optimizers’. In: *International Conference on Machine Learning*. Baltimore, United States, 2022. URL: <https://hal.inria.fr/hal-03935798>.

- [16] S. Kolouri, K. Nadjahi, S. Shahrapour and U. Simsekli. ‘Generalized Sliced Probability Metrics’. In: ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Singapore, Singapore: IEEE, 23rd May 2022, pp. 4513–4517. DOI: [10.1109/ICASSP43922.2022.9746016](https://doi.org/10.1109/ICASSP43922.2022.9746016). URL: <https://hal.inria.fr/hal-03935833>.
- [17] T. Lauvaux, C. Giron, M. Mazzolini, A. d’Aspremont, R. Duren, D. Cusworth, D. Shindell and P. Ciaï. ‘Global assessment of oil and gas methane ultra-emitters’. In: *Science* 375.6580 (4th Feb. 2022), pp. 557–561. DOI: [10.1126/science.abj4351](https://doi.org/10.1126/science.abj4351). URL: <https://hal.archives-ouvertes.fr/hal-03565371>.
- [18] S. H. Lim, Y. Wan and U. Şimşekli. ‘Chaotic Regularization and Heavy-Tailed Limits for Deterministic Gradient Descent’. In: *Advances in Neural Processing Systems*. New Orleans, United States, 2022. URL: <https://hal.inria.fr/hal-03935819>.
- [19] U. Marteau-Ferey, F. Bach and A. Rudi. ‘Non-parametric Models for Non-negative Functions’. working paper or preprint. July 2020. URL: <https://hal.inria.fr/hal-02891640>.
- [20] S. Park, U. Şimşekli and M. A. Erdogdu. ‘Generalization Bounds for Stochastic Gradient Descent via Localized ε -Covers’. In: *Advances in Neural Processing Systems*. Baltimore, United States, 19th Sept. 2022. URL: <https://hal.inria.fr/hal-03935856>.
- [21] K. L. Pavasovic, A. Durmus and U. Simsekli, eds. *Approximate Heavy Tails in Offline (Multi-Pass) Stochastic Gradient Descent*. *Advances in Neural Information Processing Systems*. 27th Oct. 2023. URL: <https://inria.hal.science/hal-04478942>.
- [22] A. Raj, M. Barsbey, M. Gürbüzbalaban, L. Zhu and U. Şimşekli. ‘Algorithmic Stability of Heavy-Tailed Stochastic Gradient Descent on Least Squares’. In: *Algorithmic Learning Theory*. Singapore, Singapore, 2023. URL: <https://inria.hal.science/hal-04478947>.
- [23] A. Raj, U. Şimşekli and A. Rudi, eds. *Efficient Sampling of Stochastic Differential Equations with Positive Semi-Definite Models*. *Advances in Neural Information Processing Systems*. 2023. URL: <https://inria.hal.science/hal-04478943>.
- [24] A. Raj, L. Zhu, M. Gürbüzbalaban and U. Şimşekli. ‘Algorithmic Stability of Heavy-Tailed SGD with General Loss Functions’. In: *International Conference on Machine Learning*. Honolulu, United States, 2023. URL: <https://inria.hal.science/hal-04478946>.
- [25] V. Roulet and A. D’Aspremont. ‘Sharpness, Restart and Acceleration’. In: *SIAM Journal on Optimization* 30.1 (Oct. 2020), pp. 262–289. DOI: [10.1137/18M1224568](https://doi.org/10.1137/18M1224568). URL: <https://hal.archives-ouvertes.fr/hal-02983236>.
- [26] A. Rubbens, J. M. Hendrickx and A. Taylor. *A constructive approach to strengthen algebraic descriptions of function and operator classes*. 18th Sept. 2025. URL: <https://hal.science/hal-05461102> (cit. on p. 11).
- [27] S. Sachs, T. van Erven, L. Hodgkinson, R. Khanna and U. Simsekli. ‘Generalization Guarantees via Algorithm-dependent Rademacher Complexity’. In: *Conference on Learning Theory*. Bangalore (Virtual event), India, 4th July 2023. URL: <https://inria.hal.science/hal-04478945>.
- [28] F. Schaipp, A. Hägele, A. Taylor, U. Simsekli and F. Bach. ‘The Surprising Agreement Between Convex Optimization Theory and Learning-Rate Scheduling for Large Model Training’. In: *ICML 2025 - 42nd International Conference on Machine Learning*. Vancouver (BC), Canada, 13th July 2025. URL: <https://hal.science/hal-05461118> (cit. on p. 10).
- [29] M. Sefidgaran, A. Gohari, G. Richard and U. Şimşekli. ‘Rate-Distortion Theoretic Generalization Bounds for Stochastic Learning Algorithms’. In: *COLT 2022 - 35th Annual Conference on Learning Theory*. Vol. 178. *Proceedings of Machine Learning Research*. London, United Kingdom, 2nd July 2022. URL: <https://hal.telecom-paris.fr/hal-03759597>.
- [30] D. Shariatian, U. Simsekli and A. Durmus. ‘Heavy-Tailed Diffusion with Denoising Lévy Probabilistic Models’. In: *International Conference on Learning Representations*. Singapore, Singapore, 2025. URL: <https://hal.science/hal-05466297> (cit. on p. 9).

- [31] P. Viillard, M. Haddouche, U. Şimşekli and B. Guedj. ‘Learning via Wasserstein-Based High Probability Generalisation Bounds’. In: *NeurIPS 2023 - Thirty-seventh Conference on Neural Information Processing Systems*. New Orleans, United States, 7th June 2023. DOI: [10.48550/arXiv.2306.04375](https://arxiv.org/abs/2306.04375). URL: <https://hal.science/hal-04121624>.
- [32] Y. Wan, M. Barsbey, A. Zaidi and U. Simsekli. ‘Implicit Compressibility of Overparametrized Neural Networks Trained with Heavy-Tailed SGD’. In: *PMLR. International Conference on Machine Learning*. Vienna, Austria, 2024. URL: <https://hal.science/hal-04872152>.
- [33] J. Weibel, P. Gaillard, W. M. Koolen and A. Taylor. *Optimized projection-free algorithms for online learning: construction and worst-case analysis*. 3rd June 2025. URL: <https://hal.science/hal-05097004> (cit. on p. 11).
- [34] B. Woodworth, F. Bach and A. Rudi. *Non-Convex Optimization with Certificates and Fast Rates Through Kernel Sums of Squares*. 8th Apr. 2022. DOI: [10.48550/arXiv.2204.04970](https://arxiv.org/abs/2204.04970). URL: <https://hal.inria.fr/hal-03635236>.
- [35] J. Wu, P. Marion and P. L. Bartlett. ‘Large Stepsizes Accelerate Gradient Descent for Regularized Logistic Regression’. In: *NeurIPS 2025 - 39th Annual Conference on Neural Information Processing Systems*. Vol. *Advances in Neural Information Processing Systems*. 38. San Diego (CA), United States, 2nd Dec. 2025. URL: <https://hal.science/hal-05345727> (cit. on p. 11).
- [36] L. Zhu, M. Gurbuzbalaban, A. Raj and U. Simsekli, eds. *Uniform-in-Time Wasserstein Stability Bounds for (Noisy) Stochastic Gradient Descent*. *Advances in Neural Information Processing Systems*. 2023. URL: <https://inria.hal.science/hal-04478941>.

12.2 Publications of the year

International journals

- [37] A. Bambade, F. Schramm, S. E. Kazdadi, S. Caron, A. Taylor and J. Carpentier. ‘PROXQP: an Efficient and Versatile Quadratic Programming Solver for Real-Time Robotics Applications and Beyond’. In: *IEEE Transactions on Robotics* (6th June 2025). URL: <https://inria.hal.science/hal-04198663>.
- [38] R. Bonalli and A. Rudi. ‘Non-Parametric Learning of Stochastic Differential Equations with Fast Rates of Convergence’. In: *Foundations of Computational Mathematics* (6th Mar. 2025). URL: <https://hal.science/hal-04381810>.
- [39] N. Doumèche, F. Bach, G. Biau and C. Boyer. ‘Physics-informed kernel learning’. In: *Journal of Machine Learning Research* 26.124 (1st Sept. 2025), pp. 1–39. URL: <https://hal.science/hal-04701052>.
- [40] B. Dubois-Taine, R. Akiki and A. D’aspremont. ‘Iteratively Reweighted Least Squares for Phase Unwrapping’. In: *Optimization Methods and Software* (10th Sept. 2025), pp. 1–41. DOI: [10.1080/10556788.2025.252348](https://arxiv.org/abs/2509.25234). URL: <https://hal.science/hal-04710547>.
- [41] B. Dubois-Taine and A. D’aspremont. ‘Frank-Wolfe meets Shapley-Folkman: a systematic approach for solving nonconvex separable problems with linear constraints’. In: *Mathematical Programming* (12th Aug. 2025). DOI: [10.1007/s10107-025-02271-8](https://arxiv.org/abs/2508.1007). URL: <https://hal.science/hal-04715357>.
- [42] C. Schlosser, M. Tacchi and A. Lazarev. ‘Convergence rates for the moment-SoS hierarchy’. In: *Numerical Algebra, Control and Optimization* 16 (May 2025), pp. 105–156. DOI: [10.3934/naco.2025011](https://arxiv.org/abs/2505.25011). URL: <https://hal.univ-grenoble-alpes.fr/hal-04430952>.
- [43] H. Zenati, A. Bietti, M. Martin, E. Diemert, P. Gaillard and J. Mairal. ‘Counterfactual Learning of Stochastic Policies with Continuous Actions’. In: *Transactions on Machine Learning Research Journal* (Mar. 2025). URL: <https://hal.science/hal-02883423>.

Invited conferences

- [44] M. Lambert. ‘The LQR-Schrödinger Bridge’. In: *session "optimal transportation methods for estimation and control"*. 64th IEEE Conference on Decision and Control. Rio de Janeiro, France, 10th Dec. 2025. URL: <https://inria.hal.science/hal-05108441>.

International peer-reviewed conferences

- [45] F. Fogel, Y. Perron, N. Besic, L. Saint-André, A. Pellissier-Tanon, M. Schwartz, T. Boudras, I. Fayad, A. d’Aspremont, L. Landrieu and P. Ciais. ‘Open-Canopy: Towards Very High Resolution Forest Monitoring’. In: 2025 Conference on Computer Vision and Pattern Recognition. Nashville, United States, 2025. DOI: [10.1109/CVPR52734.2025.00138](https://doi.org/10.1109/CVPR52734.2025.00138). URL: <https://hal.science/hal-04710775>.
- [46] M. Haddouche, P. Viallard, U. Şimşekli and B. Guedj. ‘A PAC-Bayesian Link Between Generalisation and Flat Minima’. In: ALT 2025 - 36th International Conference on Algorithmic Learning Theory. Milan, Italy, 2025, pp. 1–31. URL: <https://hal.science/hal-04455639>.
- [47] S. Naldi, M. Safey El Din, A. Taylor and W. Wang. ‘Solving generic parametric linear matrix inequalities’. In: *ISSAC ’25: Proceedings of the 2025 International Symposium on Symbolic and Algebraic Computation*. ISSAC ’25: International Symposium on Symbolic and Algebraic Computation. ISSAC ’25. Guanajuato, Mexico: Association for Computing Machinery, 10th Nov. 2025, pp. 267–276. DOI: [10.1145/3747199.3747570](https://doi.org/10.1145/3747199.3747570). URL: <https://hal.sorbonne-universite.fr/hal-04973502>.
- [48] J. Qu, D. Holzmüller, G. Varoquaux and M. Le Morvan. ‘TabICL: A Tabular Foundation Model for In-Context Learning on Large Data’. In: ICML 2025 - 42nd International Conference on Machine Learning. Vancouver, Canada, 13th July 2025. URL: <https://hal.science/hal-05100518>.
- [49] J. Wu, P. Marion and P. L. Bartlett. ‘Large Stepsizes Accelerate Gradient Descent for Regularized Logistic Regression’. In: *NeurIPS 2025 - 39th Annual Conference on Neural Information Processing Systems*. Vol. *Advances in Neural Information Processing Systems*. 38. San Diego (CA), United States, 2nd Dec. 2025. URL: <https://hal.science/hal-05345727>.

Conferences without proceedings

- [50] L. Dana, L. Pillaud-Vivien and F. Bach. ‘Convergence of Shallow ReLU Networks on Weakly Interacting Data’. In: *Neural Information Processing Systems 2025*. San Diego (CA), United States, 2nd Dec. 2025. URL: <https://hal.science/hal-04959124> (cit. on p. 12).

Reports & preprints

- [51] F. Bach. *On the Effectiveness of the z-Transform Method in Quadratic Optimization*. 14th July 2025. URL: <https://hal.science/hal-05143189> (cit. on p. 13).
- [52] E. Berta, D. Holzmüller, M. I. Jordan and F. Bach. *Rethinking Early Stopping: Refine, Then Calibrate*. 31st Jan. 2025. URL: <https://hal.science/hal-04929578> (cit. on p. 13).
- [53] R. Berthier and M. B. Li. *Acceleration of Gossip Algorithms through the Euler-Poisson-Darboux Equation*. 17th July 2025. URL: <https://hal.science/hal-03582008>.
- [54] E. Beyler and F. Bach. *Convergence of Deterministic and Stochastic Diffusion-Model Samplers: A Simple Analysis in Wasserstein Distance*. 12th Nov. 2025. URL: <https://hal.science/hal-05198651> (cit. on p. 12).
- [55] E. Beyler and F. Bach. *Optimal Denoising in Score-Based Generative Models: The Role of Data Regularity*. 30th Sept. 2025. URL: <https://hal.science/hal-04991557>.
- [56] E. Beyler and F. Bach. *Variational Inference on the Boolean Hypercube with the Quantum Entropy*. 13th Feb. 2025. URL: <https://hal.science/hal-04753816>.
- [57] P. Boudart, P. Gaillard and A. Rudi. *Enjoying Non-linearity in Multinomial Logistic Bandits*. 4th July 2025. URL: <https://hal.science/hal-05145114>.

- [58] S. Braun, D. Holzmüller, M. I. Jordan and F. Bach. *Conditional Coverage Diagnostics for Conformal Prediction*. 12th Dec. 2025. URL: <https://hal.science/hal-05419143> (cit. on p. 13).
- [59] L. Chizat, P. Marion and Y. Yesbay. *Phase Diagram of Dropout for Two-Layer Neural Networks in the Mean-Field Regime*. 10th Oct. 2025. URL: <https://hal.science/hal-05307271>.
- [60] A. Dieuleveut, G. Fort, M. Hegazy and H.-T. Wai. *Federated Majorize-Minimization: Beyond Parameter Aggregation*. 23rd July 2025. URL: <https://hal.science/hal-05189632>.
- [61] N. Doumèche, F. Bach, É. Bedek, G. Biau, C. Boyer and Y. Goude. *Forecasting time series with constraints*. 13th Feb. 2025. URL: <https://hal.science/hal-04942529>.
- [62] N. Doumèche, F. Bach, G. Biau and C. Boyer. *Fast kernel methods: Sobolev, physics-informed, and additive models*. 1st Sept. 2025. URL: <https://hal.science/hal-05224318> (cit. on p. 13).
- [63] O. Duranthon, P. Marion, C. Boyer, B. Loureiro, L. Zdeborová and L. Zdeborová. *Statistical Advantage of Softmax Attention: Insights from Single-Location Regression*. 3rd Oct. 2025. URL: <https://hal.science/hal-05295736>.
- [64] N. Erickson, L. Purucker, A. Tschalzev, D. Holzmüller, P. M. Desai, D. Salinas and F. Hutter. *TabArena: A Living Benchmark for Machine Learning on Tabular Data*. 2025. URL: <https://hal.science/hal-05136945>.
- [65] I. Fayad, M. Zimmer, M. Schwartz, F. Gieseke, P. Ciais, G. Belouze, S. Brood, A. de Truchis and A. d’Aspremont. *DUNIA: Pixel-Sized Embeddings via Cross-Modal Alignment for Earth Observation Applications*. 2025. URL: <https://hal.science/hal-05347039>.
- [66] E. Gauthier, F. Bach and M. I. Jordan. *Adaptive Coverage Policies in Conformal Prediction*. 7th Oct. 2025. URL: <https://hal.science/hal-05305315> (cit. on p. 12).
- [67] E. Gauthier, F. Bach and M. I. Jordan. *E-Values Expand the Scope of Conformal Prediction*. 19th Mar. 2025. URL: <https://hal.science/hal-05003357>.
- [68] E. Gauthier, F. Bach and M. I. Jordan. *Backward Conformal Prediction*. 26th May 2025. URL: <https://hal.science/hal-05085303>.
- [69] E. Gauthier, F. Bach and M. I. Jordan. *Statistical Collusion by Collectives on Learning Platforms*. 7th Feb. 2025. URL: <https://hal.science/hal-04941041>.
- [70] D. Holzmüller and M. Schöpple. *Beyond ReLU: How Activations Affect Neural Kernels and Random Wide Networks*. 27th June 2025. DOI: [10.48550/arXiv.2506.22429](https://doi.org/10.48550/arXiv.2506.22429). URL: <https://hal.science/hal-05136950>.
- [71] M. Kalimuthu, D. Holzmüller and M. Niepert. *LOGLO-FNO: Efficient Learning of Local and Global Features in Fourier Neural Operators*. 2025. DOI: [10.48550/arXiv.2504.04260](https://doi.org/10.48550/arXiv.2504.04260). URL: <https://hal.science/hal-05136953>.
- [72] M. Lambert, F. Bach and S. Bonnabel. *Entropy Regularized Variational Dynamic Programming for Stochastic Optimal Control*. 22nd Mar. 2025. URL: <https://inria.hal.science/hal-05016406>.
- [73] L.-T.-N. Pham, D. Shariatian, A. Ocello, G. Conforti and A. Durmus. *Discrete Markov Probabilistic Models*. 17th Feb. 2025. URL: <https://hal.science/hal-04951334>.
- [74] C. Roux, M. Zimmer, A. d’Aspremont and S. Pokutta. *Don’t Be Greedy, Just Relax! Pruning LLMs via Frank-Wolfe*. 2025. DOI: [10.48550/arXiv.2510.13713](https://doi.org/10.48550/arXiv.2510.13713). URL: <https://hal.science/hal-05347023> (cit. on p. 9).
- [75] L. Stewart, F. Bach and Q. Berthet. *Building Bridges between Regression, Clustering, and Classification*. 16th Feb. 2025. URL: <https://hal.science/hal-04929718>.
- [76] Y. Su, N. Besic, X. Zhang, Y. Xu, S. Francini, G. d’Amico, G. Chirici, M. Schwartz, I. Fayad, S. Brood, A. Pellissier-Tanon, K. Yu, H. Chen, S. Chen, A. d’Aspremont and P. Ciais. *A fused canopy height map of Italy (2004–2024) from spaceborne and airborne LiDAR, and Landsat via deep learning and Bayesian averaging*. 8th Sept. 2025. DOI: [10.5194/essd-2025-378](https://doi.org/10.5194/essd-2025-378). URL: <https://hal.science/hal-05347031>.
- [77] D. B. Thomsen, A. Taylor and A. Dieuleveut. *Tight analyses of first-order methods with error feedback*. 2025. DOI: [10.48550/arXiv.2506.05271](https://doi.org/10.48550/arXiv.2506.05271). URL: <https://hal.science/hal-05346805>.

- [78] J. Weibel, P. Gaillard, W. M. Koolen and A. Taylor. *Optimized projection-free algorithms for online learning: construction and worst-case analysis*. 3rd June 2025. URL: <https://hal.science/hal-05097004>.
- [79] Y.-H. Wu, Q. Berthet, G. Biau, C. Boyer, E. Romuald and P. Marion. *Optimal stopping in latent diffusion models*. 9th Oct. 2025. URL: <https://hal.science/hal-05306599>.

12.3 Cited publications

- [80] L. Stewart. ‘Understanding and Formulating Training Objectives: Key Insights for Deep Learning’. Theses. INRIA, June 2025. URL: <https://theses.hal.science/tel-05218461> (cit. on p. 22).