

RESEARCH CENTRE

Inria Paris Center

IN PARTNERSHIP WITH:

Ecole normale supérieure de Paris, CNRS

2022

ACTIVITY REPORT

Project-Team

SIERRA

Statistical Machine Learning and Parsimony

IN COLLABORATION WITH: Département d'Informatique de l'Ecole
Normale Supérieure

DOMAIN

**Applied Mathematics, Computation and
Simulation**

THEME

**Optimization, machine learning and
statistical methods**

The Inria logo is a stylized, cursive script in red, positioned in the bottom right corner of the page.

Contents

Project-Team SIERRA	1
1 Team members, visitors, external collaborators	2
2 Overall objectives	3
2.1 Statement	3
3 Research program	3
4 Application domains	4
5 Highlights of the year	4
5.1 Awards	4
6 New software and platforms	4
6.1 New software	4
6.1.1 PEPit	4
7 New results	5
7.1 Information theory through kernel methods	5
7.2 Optimal Algorithms for Stochastic Complementary Composite Minimization	5
7.3 Super-Acceleration with Cyclical Step-sizes	5
7.4 An optimal gradient method for smooth strongly convex minimization	5
7.5 PEPit: computer-assisted worst-case analyses of first-order optimization methods in Python	6
7.6 Last-Iterate Convergence of Optimistic Gradient Method for Monotone Variational Inequalities	6
7.7 Fast Stochastic Composite Minimization and an Accelerated Frank-Wolfe Algorithm under Parallelization	6
7.8 Vector-Valued Least-Squares Regression under Output Regularity Assumptions.	6
7.9 On the Benefits of Large Learning Rates for Kernel Methods	7
7.10 Active Labeling: Streaming Stochastic Gradients	7
7.11 Sampling from Arbitrary Functions via PSD Models	7
7.12 On the Consistency of Max-Margin Losses	7
7.13 Exponential convergence of sum-of-squares hierarchies for trigonometric polynomials	7
7.14 Measuring dissimilarity with diffeomorphism invariance	8
7.15 Non-Convex Optimization with Certificates and Fast Rates Through Kernel Sums of Squares	8
7.16 Generalization Bounds using Lower Tail Exponents in Stochastic Optimizers	8
7.17 Rate-Distortion Theoretic Generalization Bounds for Stochastic Learning Algorithms	8
7.18 Chaotic Regularization and Heavy-Tailed Limits for Deterministic Gradient Descent	9
7.19 Generalized Sliced Probability Metrics	9
7.20 Generalization Bounds for Stochastic Gradient Descent via Localized epsilon-Covers	9
8 Bilateral contracts and grants with industry	10
8.1 Bilateral grants with industry	10
9 Partnerships and cooperations	10
9.1 International initiatives	10
9.1.1 Associate Teams in the framework of an Inria International Lab or in the framework of an Inria International Program	10
9.2 International research visitors	11
9.2.1 Visits of international scientists	11
9.3 European initiatives	11
9.3.1 Horizon Europe	11
9.3.2 H2020 projects	11
9.4 National initiatives	13

10 Dissemination	13
10.1 Promoting scientific activities	13
10.1.1 Scientific events: organisation	13
10.1.2 Journal	14
10.1.3 Invited talks	14
10.1.4 Leadership within the scientific community	15
10.1.5 Scientific expertise	15
10.2 Teaching - Supervision - Juries	15
10.2.1 Teaching	15
10.2.2 Supervision	15
10.2.3 Juries	16
10.3 Popularization	16
10.3.1 Interventions	16
11 Scientific production	16
11.1 Major publications	16
11.2 Publications of the year	17

Project-Team SIERRA

Creation of the Project-Team: 2012 January 01

Keywords

Computer sciences and digital sciences

A3.4. – Machine learning and statistics

A5.4. – Computer vision

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

A7.1. – Algorithms

A8.2. – Optimization

A9.2. – Machine learning

Other research topics and application domains

B9.5.6. – Data science

1 Team members, visitors, external collaborators

Research Scientists

- Francis Bach [Team leader, INRIA, Senior Researcher, HDR]
- Alessandro Rudi [INRIA, Researcher]
- Umut Simsekli [INRIA, Researcher]
- Alexandre d'Aspremont [CNRS, Senior Researcher]

Post-Doctoral Fellows

- Pierre-Cyril Aubin [INRIA]
- Hans Kersting [INRIA]
- Ziad Kobeissi [ILB]
- Clément Mantoux [INRIA]
- Anant Raj [INRIA - Marie Curie Fellowship]
- Blake Woodworth [DI-ENS]

PhD Students

- Andrea Basteri [Inria, from Oct 2022]
- Eloise Berthier [DGA]
- Gaspard Beugnot [INRIA]
- Theophile Cantelobre [INRIA]
- Bertille Follain [ENS PARIS]
- Gautier Izacard [CNRS]
- Remi Jezequel [ENS PARIS]
- Marc Lambert [DGA]
- Clément Lezane [UNIV TWENTE]
- Ulysse Marteau-Ferey [INRIA]
- Gregoire Mialon [Inria, until Jan 2022]
- Céline Moucer [ENS PARIS-SACLAY]
- Benjamin Paul-Dubois-Taine [UNIV PARIS SACLAY]
- Lawrence Stewart [INRIA]

Technical Staff

- Loic Esteve [INRIA]

Interns and Apprentices

- Benjamin Dupuis [Inria, from Sep 2022]
- David Holzmüller [UNIV TU BERLIN, from Mar 2022 until Sep 2022]
- Krunoslav Lehman Pavasovic [Inria, from Oct 2022]
- Sarah Sachs [University of Amsterdam, from Nov 2022]

Administrative Assistants

- Helene Bessin Rousseau [INRIA]
- Helene Milome [INRIA]

Visiting Scientists

- Silvere Bonnabel [UNIV NOUVELLE-CALEDONIE, from Sep 2022]
- Steffen Grunewalder [UNIV NEWCASTLE, from Sep 2022]
- Cristobal Guzman [INRIA, from Sep 2022]

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 modelisation 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, in particular computer vision, bioinformatics, audio processing, and text processing.

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.

We have thus divided our research effort in three parts:

1. Convex optimization

2. Non-convex optimization
3. 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).

5 Highlights of the year

5.1 Awards

- Starting investigator ERC grant Dynasty (U. Simsekli)
- ICASSP 2022 best paper award (U. Simsekli)
- Médaille des Assises des Mathématiques (F. Bach)

6 New software and platforms

6.1 New software

6.1.1 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 ourself, 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 New results

7.1 Information theory through kernel methods

We consider the analysis of probability distributions through their associated covariance operators from reproducing kernel Hilbert spaces. We show that the von Neumann entropy and relative entropy of these operators are intimately related to the usual notions of Shannon entropy and relative entropy, and share many of their properties. They come together with efficient estimation algorithms from various oracles on the probability distributions. We also consider product spaces and show that for tensor product kernels, we can define notions of mutual information and joint entropies, which can then characterize independence perfectly, but only partially conditional independence. We finally show how these new notions of relative entropy lead to new upper-bounds on log partition functions, that can be used together with convex optimization within variational inference methods, providing a new family of probabilistic inference methods. [52]

7.2 Optimal Algorithms for Stochastic Complementary Composite Minimization

Inspired by regularization techniques in statistics and machine learning, we study complementary composite minimization in the stochastic setting. This problem corresponds to the minimization of the sum of a (weakly) smooth function endowed with a stochastic first-order oracle, and a structured uniformly convex (possibly nonsmooth and non-Lipschitz) regularization term. Despite intensive work on closely related settings, prior to our work no complexity bounds for this problem were known. We close this gap by providing novel excess risk bounds, both in expectation and with high probability. Our algorithms are nearly optimal, which we prove via novel lower complexity bounds for this class of problems. We conclude by providing numerical results comparing our methods to the state of the art.

7.3 Super-Acceleration with Cyclical Step-sizes

We develop a convergence-rate analysis of momentum with cyclical step-sizes. We show that under some assumption on the spectral gap of Hessians in machine learning, cyclical step-sizes are provably faster than constant step-sizes. More precisely, we develop a convergence rate analysis for quadratic objectives that provides optimal parameters and shows that cyclical learning rates can improve upon traditional lower complexity bounds. We further propose a systematic approach to design optimal first order methods for quadratic minimization with a given spectral structure. Finally, we provide a local convergence rate analysis beyond quadratic minimization for the proposed methods and illustrate our findings through benchmarks on least squares and logistic regression problems.

7.4 An optimal gradient method for smooth strongly convex minimization

We present an optimal gradient method for smooth strongly convex optimization. The method is optimal in the sense that its worst-case bound on the distance to an optimal point exactly matches the lower bound on the oracle complexity for the class of problems, meaning that no black-box first-order method can have a better worst-case guarantee without further assumptions on the class of problems at hand. In addition, we provide a constructive recipe for obtaining the algorithmic parameters of the method and illustrate that it can be used for deriving methods for other optimality criteria as well.

7.5 PEPit: computer-assisted worst-case analyses of first-order optimization methods in Python

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.

7.6 Last-Iterate Convergence of Optimistic Gradient Method for Monotone Variational Inequalities

The Past Extragradient (PEG) [Popov, 1980] method, also known as the Optimistic Gradient method, has known a recent gain in interest in the optimization community with the emergence of variational inequality formulations for machine learning. Recently, in the unconstrained case, Golowich et al. [2020] proved that a $O(1/N)$ last-iterate convergence rate in terms of the squared norm of the operator can be achieved for Lipschitz and monotone operators with a Lipschitz Jacobian. In this work, by introducing a novel analysis through potential functions, we show that (i) this $O(1/N)$ last-iterate convergence can be achieved without any assumption on the Jacobian of the operator, and (ii) it can be extended to the constrained case, which was not derived before even under Lipschitzness of the Jacobian. The proof is significantly different from the one known from Golowich et al. [2020], and its discovery was computer-aided. Those results close the open question of the last iterate convergence of PEG for monotone variational inequalities.

7.7 Fast Stochastic Composite Minimization and an Accelerated Frank-Wolfe Algorithm under Parallelization

We consider the problem of minimizing the sum of two convex functions. One of those functions has Lipschitz-continuous gradients, and can be accessed via stochastic oracles, whereas the other is “simple”. We provide a Bregman-type algorithm with accelerated convergence in function values to a ball containing the minimum. The radius of this ball depends on problem-dependent constants, including the variance of the stochastic oracle. We further show that this algorithmic setup naturally leads to a variant of Frank-Wolfe achieving acceleration under parallelization. More precisely, when minimizing a smooth convex function on a bounded domain, we show that one can achieve an ϵ primal-dual gap (in expectation) in $\tilde{O}(1/\sqrt{\epsilon})$ iterations, by only accessing gradients of the original function and a linear maximization oracle with $O(1/\sqrt{\epsilon})$ computing units in parallel. We illustrate this fast convergence on synthetic numerical experiments.

7.8 Vector-Valued Least-Squares Regression under Output Regularity Assumptions.

In [19] we propose and analyse a reduced-rank method for solving least-squares regression problems with infinite dimensional output. We derive learning bounds for our method, and study under which setting statistical performance is improved in comparison to full-rank method. Our analysis extends the interest of reduced-rank regression beyond the standard low-rank setting to more general output regularity assumptions. We illustrate our theoretical insights on synthetic least-squares problems. Then, we propose a surrogate structured prediction method derived from this reduced-rank method. We assess its benefits on three different problems: image reconstruction, multi-label classification, and metabolite identification.

7.9 On the Benefits of Large Learning Rates for Kernel Methods

In [30], we study an intriguing phenomenon related to the good generalization performance of estimators obtained by using large learning rates within gradient descent algorithms. First observed in the deep learning literature, we show that such a phenomenon can be precisely characterized in the context of kernel methods, even though the resulting optimization problem is convex. Specifically, we consider the minimization of a quadratic objective in a separable Hilbert space, and show that with early stopping, the choice of learning rate influences the spectral decomposition of the obtained solution on the Hessian's eigenvectors. This extends an intuition described by Nakkiran (2020) on a two-dimensional toy problem to realistic learning scenarios such as kernel ridge regression. While large learning rates may be proven beneficial as soon as there is a mismatch between the train and test objectives, we further explain why it already occurs in classification tasks without assuming any particular mismatch between train and test data distributions.

7.10 Active Labeling: Streaming Stochastic Gradients

The workhorse of machine learning is stochastic gradient descent. To access stochastic gradients, it is common to consider iteratively input/output pairs of a training dataset. Interestingly, it appears that one does not need full supervision to access stochastic gradients, which is the main motivation of this paper. After formalizing the "active labeling" problem, which focuses on active learning with partial supervision, we provide a streaming technique that provably minimizes the ratio of generalization error over the number of samples. We illustrate our technique in depth for robust regression. The results are presented in [31]

7.11 Sampling from Arbitrary Functions via PSD Models

In many areas of applied statistics and machine learning, generating an arbitrary number of independent and identically distributed (i.i.d.) samples from a given distribution is a key task. When the distribution is known only through evaluations of the density, current methods either scale badly with the dimension or require very involved implementations. Instead, we take a two-step approach by first modeling the probability distribution and then sampling from that model. We use the recently introduced class of positive semi-definite (PSD) models, which have been shown to be efficient for approximating probability densities. We show that these models can approximate a large class of densities concisely using few evaluations, and present a simple algorithm to effectively sample from these models. We also present preliminary empirical results to illustrate our assertions. The results are presented in [37].

7.12 On the Consistency of Max-Margin Losses

The foundational concept of Max-Margin in machine learning is ill-posed for output spaces with more than two labels such as in structured prediction. In this paper, we show that the Max-Margin loss can only be consistent to the classification task under highly restrictive assumptions on the discrete loss measuring the error between outputs. These conditions are satisfied by distances defined in tree graphs, for which we prove consistency, thus being the first losses shown to be consistent for Max-Margin beyond the binary setting. We finally address these limitations by correcting the concept of Max-Margin and introducing the Restricted-Max-Margin, where the maximization of the loss-augmented scores is maintained, but performed over a subset of the original domain. The resulting loss is also a generalization of the binary support vector machine and it is consistent under milder conditions on the discrete loss. The results are presented in the paper [45].

7.13 Exponential convergence of sum-of-squares hierarchies for trigonometric polynomials

In [55] we consider the unconstrained optimization of multivariate trigonometric polynomials by the sum-of-squares hierarchy of lower bounds. We first show a convergence rate of $O(1/s^2)$ for the relaxation with degree s without any assumption on the trigonometric polynomial to minimize. Second, when the

polynomial has a finite number of global minimizers with invertible Hessians at these minimizers, we show an exponential convergence rate with explicit constants. Our results also apply to the minimization of regular multivariate polynomials on the hypercube.

7.14 Measuring dissimilarity with diffeomorphism invariance

Measures of similarity (or dissimilarity) are a key ingredient to many machine learning algorithms. We introduce DID, a pairwise dissimilarity measure applicable to a wide range of data spaces, which leverages the data's internal structure to be invariant to diffeomorphisms. We prove that DID enjoys properties which make it relevant for theoretical study and practical use. By representing each datum as a function, DID is defined as the solution to an optimization problem in a Reproducing Kernel Hilbert Space and can be expressed in closed-form. In practice, it can be efficiently approximated via Nyström sampling. Empirical experiments support the merits of DID. The results are presented in the paper [58].

7.15 Non-Convex Optimization with Certificates and Fast Rates Through Kernel Sums of Squares

In [41] We consider potentially non-convex optimization problems, for which optimal rates of approximation depend on the dimension of the parameter space and the smoothness of the function to be optimized. In this paper, we propose an algorithm that achieves close to optimal a priori computational guarantees, while also providing a posteriori certificates of optimality. Our general formulation builds on infinite-dimensional sums-of-squares and Fourier analysis, and is instantiated on the minimization of multivariate periodic functions.

7.16 Generalization Bounds using Lower Tail Exponents in Stochastic Optimizers

Despite the ubiquitous use of stochastic optimization algorithms in machine learning, the precise impact of these algorithms and their dynamics on generalization performance in realistic non-convex settings is still poorly understood. While recent work has revealed connections between generalization and heavy-tailed behavior in stochastic optimization, this work mainly relied on continuous-time approximations; and a rigorous treatment for the original discrete-time iterations is yet to be performed. To bridge this gap, we present novel bounds linking generalization to the lower tail exponent of the transition kernel associated with the optimizer around a local minimum, in both discrete- and continuous-time settings. To achieve this, we first prove a data- and algorithm-dependent generalization bound in terms of the celebrated Fernique-Talagrand functional applied to the trajectory of the optimizer. Then, we specialize this result by exploiting the Markovian structure of stochastic optimizers, and derive bounds in terms of their (data-dependent) transition kernels. We support our theory with empirical results from a variety of neural networks, showing correlations between generalization error and lower tail exponents. The results are presented in the paper [32].

7.17 Rate-Distortion Theoretic Generalization Bounds for Stochastic Learning Algorithms

Understanding generalization in modern machine learning settings has been one of the major challenges in statistical learning theory. In this context, recent years have witnessed the development of various generalization bounds suggesting different complexity notions such as the mutual information between the data sample and the algorithm output, compressibility of the hypothesis space, and the fractal dimension of the hypothesis space. While these bounds have illuminated the problem at hand from different angles, their suggested complexity notions might appear seemingly unrelated, thereby restricting their high-level impact. In this study, we prove novel generalization bounds through the lens of rate-distortion theory, and explicitly relate the concepts of mutual information, compressibility, and fractal dimensions in a single mathematical framework. Our approach consists of (i) defining a generalized notion of compressibility by using source coding concepts, and (ii) showing that the 'compression error rate' can be linked to the generalization error both in expectation and with high probability. We show that in the 'lossless compression' setting, we recover and improve existing mutual information-based bounds,

whereas a ‘lossy compression’ scheme allows us to link generalization to the rate-distortion dimension—a particular notion of fractal dimension. Our results bring a more unified perspective on generalization and open up several future research directions. The results are presented in the paper [46].

7.18 Chaotic Regularization and Heavy-Tailed Limits for Deterministic Gradient Descent

Recent studies have shown that gradient descent (GD) can achieve improved generalization when its dynamics exhibits a chaotic behavior. However, to obtain the desired effect, the step-size should be chosen sufficiently large, a task which is problem dependent and can be difficult in practice. In this study, we incorporate a chaotic component to GD in a controlled manner, and introduce multiscale perturbed GD (MPGD), a novel optimization framework where the GD recursion is augmented with chaotic perturbations that evolve via an independent dynamical system. We analyze MPGD from three different angles: (i) By building up on recent advances in rough paths theory, we show that, under appropriate assumptions, as the step-size decreases, the MPGD recursion converges weakly to a stochastic differential equation (SDE) driven by a heavy-tailed Lévy-stable process. (ii) By making connections to recently developed generalization bounds for heavy-tailed processes, we derive a generalization bound for the limiting SDE and relate the worst-case generalization error over the trajectories of the process to the parameters of MPGD. (iii) We analyze the implicit regularization effect brought by the dynamical regularization and show that, in the weak perturbation regime, MPGD introduces terms that penalize the Hessian of the loss function. Empirical results are provided to demonstrate the advantages of MPGD. The results are presented in the paper [36].

7.19 Generalized Sliced Probability Metrics

Sliced probability metrics have become increasingly popular in machine learning, and they play a quintessential role in various applications, including statistical hypothesis testing and generative modeling. However, in a practical setting, the convergence behavior of the algorithms built upon these distances have not been well established, except for a few specific cases. In this paper, we introduce a new family of sliced probability metrics, namely Generalized Sliced Probability Metrics (GSPMs), based on the idea of slicing high-dimensional distributions into a set of their one-dimensional marginals. We show that GSPMs are true metrics, and they are related to the Maximum Mean Discrepancy (MMD). Exploiting this relationship, we consider GSPM-based gradient flows and show that, under mild assumptions, the gradient flow converges to the global optimum. Finally, we demonstrate that various choices of GSPMs lead to new positive definite kernels that could be used in the MMD formulation while providing a unique integral geometric interpretation. We illustrate the application of GSPMs in gradient flows. The results are presented in the paper [33].

7.20 Generalization Bounds for Stochastic Gradient Descent via Localized ϵ -Covers

In [40], we propose a new covering technique localized for the trajectories of SGD. This localization provides an algorithm-specific complexity measured by the covering number, which can have dimension-independent cardinality in contrast to standard uniform covering arguments that result in exponential dimension dependency. Based on this localized construction, we show that if the objective function is a finite perturbation of a piecewise strongly convex and smooth function with P pieces, i.e. non-convex and non-smooth in general, the generalization error can be upper bounded by $O(\sqrt{(\log n \log(nP))/n})$, where n is the number of data samples. In particular, this rate is independent of dimension and does not require early stopping and decaying step size. Finally, we employ these results in various contexts and derive generalization bounds for multi-index linear models, multi-class support vector machines, and K -means clustering for both hard and soft label setups, improving the known state-of-the-art rates.

8 Bilateral contracts and grants with industry

8.1 Bilateral grants with industry

- Alexandre d'Aspremont, Francis Bach, Martin Jaggi (EPFL): Google Focused award.
- Francis Bach: Gift from Facebook AI Research.
- Alexandre d'Aspremont: fondation AXA, "Mécénat scientifique", optimisation & machine learning.

9 Partnerships and cooperations

9.1 International initiatives

9.1.1 Associate Teams in the framework of an Inria International Lab or in the framework of an Inria International Program

FOAM

Title: First-Order Accelerated Methods for Machine Learning.

Duration: 2020 ->

Coordinator: Cristobal Guzman (crguzmanp@mat.uc.cl)

Partners:

- Pontificia Universidad Católica de Chile Santiago (Chili)

Inria contact: Alexandre d'Aspremont

Summary: Our main interest is to investigate novel and improved convergence results for first-order iterative methods for saddle-points, variational inequalities and fixed points, under the lens of PEP. Our interest in improving first-order methods is also deeply related with applications in machine learning. Particularly in sparsity-oriented inverse problems, optimization methods are the workhorse for state of the art results. On some of these problems, a set of new hypothesis and theoretical results shows improved complexity bounds for problems with good recovery guarantees and we plan to extend these new performance bounds to the variational framework.

4TUNE

Title: Adaptive, Efficient, Provable and Flexible Tuning for Machine Learning

Duration: 2020 ->

Coordinator: Peter Grünwald (pdg@cwi.nl)

Partners: CWI

Inria contact: Pierre Gaillard

The long-term goal of 4TUNE is to push adaptive machine learning to the next level. We aim to develop refined methods, going beyond traditional worst-case analysis, for exploiting structure in the learning problem at hand. We will develop new theory and design sophisticated algorithms for the core tasks of statistical learning and individual sequence prediction. We are especially interested in understanding the connections between these tasks and developing unified methods for both. We will also investigate adaptivity to non-standard patterns encountered in embedded learning tasks, in particular in iterative equilibrium computations.

9.2 International research visitors

9.2.1 Visits of international scientists

- Sarah Sachs - PhD Student visiting from University of Amsterdam
- Mert Gurbuzbalaban - Associate Professor visiting from Rutgers University

9.3 European initiatives

9.3.1 Horizon Europe

DYNASTY [DYNASTY project on cordis.europa.eu](#)

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.

9.3.2 H2020 projects

SEQUOIA [SEQUOIA project on cordis.europa.eu](#)

Title: Robust algorithms for learning from modern data

Duration: From September 1, 2017 to August 31, 2023

Partners:

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

Inria contact: Francis BACH

Coordinator:

Summary: Machine learning is needed and used everywhere, from science to industry, with a growing impact on many disciplines. While first successes were due at least in part to simple supervised learning algorithms used primarily as black boxes on medium-scale problems, modern data pose new challenges. Scalability is an important issue of course: with large amounts of data, many current problems far exceed the capabilities of existing algorithms despite sophisticated computing architectures. But beyond this, the core classical model of supervised machine learning, with the usual assumptions of independent and identically distributed data, or well-defined features, outputs and loss functions, has reached its theoretical and practical limits.

Given this new setting, existing optimization-based algorithms are not adapted. The main objective of this proposal is to push the frontiers of supervised machine learning, in terms of (a) scalability to data with massive numbers of observations, features, and tasks, (b) adaptability to modern computing environments, in particular for parallel and distributed processing, (c) provable adaptivity and robustness to problem and hardware specifications, and (d) robustness to non-convexities inherent in machine learning problems.

To achieve the expected breakthroughs, we will design a novel generation of learning algorithms amenable to a tight convergence analysis with realistic assumptions and efficient implementations. They will help transition machine learning algorithms towards the same wide-spread robust use as numerical linear algebra libraries. Outcomes of the research described in this proposal will include algorithms that come with strong convergence guarantees and are well-tested on real-life benchmarks coming from computer vision, bioinformatics, audio processing and natural language processing. For both distributed and non-distributed settings, we will release open-source software, adapted to widely available computing platforms.

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

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

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.

NN-OVEROPT [NN-OVEROPT project on cordis.europa.eu](https://cordis.europa.eu/project/nn-overopt)

Title: Neural Network : An Overparametrization Perspective

Duration: From November 1, 2021 to October 31, 2024

Partners:

- INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET AUTOMATIQUE (INRIA), France
- THE BOARD OF TRUSTEES OF THE UNIVERSITY OF ILLINOIS (UNIVERSITY OF ILLINOIS), United States

Inria contact: Francis Bach

Coordinator:

Summary: In recent times, overparametrized models where the number of model parameters far exceeds the number of training samples available are the methods of choice for learning problems and neural networks are amongst the most popular overparametrized methods used heavily in practice. It has been discovered recently that overparametrization surprisingly improves the optimization landscape of a complex non-convex problem, i.e., the training of neural networks, and also has positive effects on the generalization performance. Despite improved empirical performance of overparametrized models like neural networks, the theoretical understanding of these models is quite limited which hinders the progress of the field in the right direction. Any progress in the understanding of the optimization as well as generalization aspects for these complex models especially neural networks will lead to big technical advancement in the field of machine learning and artificial intelligence. During the Marie Skłodowska-Curie Actions Individual Fellowship-Global Fellowship (MSCA-IF-GF), I plan to study the optimization problem arising while training overparametrized neural networks and generalization in overparametrized neural networks. The end goal for this project is to provide better theoretical understanding of the optimization landscape while training overparametrized models as a result of which to provide better optimization algorithms for training as well as to study the universal approximation guarantees of overparametrized models. We also aim to study the implicit bias induced by optimization algorithms while training overparametrized complex models. To achieve the objective discussed above, I will be using tools from traditional optimization theory, statistical learning theory, gradient flows, as well as from statistical physics.

9.4 National initiatives

- Alexandre d’Aspremont: IRIS, PSL “Science des données, données de la science”.

10 Dissemination

10.1 Promoting scientific activities

10.1.1 Scientific events: organisation

Member of the organizing committees

- F. Bach: Workshop on on Uncertainty Quantification, Erwin Schrödinger International Institute for Mathematics and Physics (ESI), co-organizer, June 2022.

Member of the conference program committees

- A. Rudi: area chair, ICML 2022.
- U. Simsekli: area chair, ICML and NeurIPS 2022.

10.1.2 Journal

Member of the editorial boards

- A. d’Aspremont, Associate Editor, SIAM Journal on Optimization.
- A. d’Aspremont, Section Editor, SIAM Journal on the Mathematics of Data Science.
- F. Bach, co-editor-in-chief, Journal of Machine Learning Research
- F. Bach: Series Editor, Adaptive Computation and Machine Learning, MIT Press, since 2016.

Reviewer - reviewing activities

- Adrien Taylor: reviewer for “SIAM Journal on Optimization”.
- Adrien Taylor: reviewer for “Journal of Optimization Theory and Applications”.
- Adrien Taylor: reviewer for “IMA Journal on Numerical Analysis”.
- Adrien Taylor: reviewer for “Conference on Learning Theory 2022”.
- Adrien Taylor: reviewer for “Journal of Machine Learning Research”.
- A. Rudi: reviewer for “Journal of Machine Learning Research”.
- U. Simsekli: reviewer for “Journal of Applied Probability”.

10.1.3 Invited talks

- A. d’Aspremont, ISMP 2022, semi-plenary.
- Adrien Taylor: invited talk at Trade-OPT Workshop on Algorithmic and Continuous Optimization, Louvain-la-Neuve.
- Adrien Taylor: invited talk at Workshop on Conic Linear Optimization for Computer-assisted Proofs, Oberwolfach.
- Adrien Taylor: invited talk at Rice University (ECE speaker seminar series, online).
- A. Rudi: Czech-French Workshop in AI, Prague, Sept 2022
- A. Rudi: ELLIS Theory Workshop, Arenzano, Italy, June 2022
- A. Rudi: Colloquium PRAIRIE, Paris, May 2022
- A. Rudi: Séminaire Parisien de Statistique — Institut Henri Poincaré ‘, Paris, April 2022
- A. Rudi: Workshop MAS-MODE, SMAI, Paris, March 2022
- A. Rudi: Seminars of IA and Mathematics by Centro Nazionale di Ricerche, March 2022
- A. Rudi: Séminaire Parisien d’Optimisation — Institut Henri Poincaré ‘, Paris, March 2022
- A. Rudi: S-DCE Seminar Series - Alan Turing Institute, London, March 2022

- A. Rudi: Unione Matematica Italiana, Colloquium Prisma, Feb. 2022
- U. Simsekli: The Statistics and Machine Learning Thematic Seminar, University of Amsterdam, Nov. 2022
- U. Simsekli: On Future Trends and Opportunities for Monte Carlo methods, Warsaw, Dec. 2022
- F. Bach: Workshop on optimization, CIRM, Luminy, May 2022
- F. Bach: Workshop on probabilistic programming, Collège de France, July 2022.
- F. Bach: International Congress of Mathematicians, online, July 2022, invited speaker.
- F. Bach: ECML-PKDD, Grenoble, invited speaker, September 2022.
- F. Bach: Workshop on optimization, CIRM, Luminy, October 2022

10.1.4 Leadership within the scientific community

- F. Bach: president of the board of ICML
- F. Bach: Member of the Scientific Council of the Société Informatique de France, since 2022.

10.1.5 Scientific expertise

- F. Bach: Deputy Scientific director, Inria Paris

10.2 Teaching - Supervision - Juries

10.2.1 Teaching

- Master: Alexandre d'Aspremont, Optimisation convexe: modélisation, algorithmes et applications cours magistraux 21h (2011-Present), Master M2 MVA, ENS PS.
- Master : Francis Bach, Optimisation et apprentissage statistique, 20h, Master M2 (Mathématiques de l'aléatoire), Université Paris-Sud, France.
- Master : Francis Bach, Learning theory from first principles, 27h, Master M2 MASH, Université Paris Dauphine PSL, France.
- Master: Alessandro Rudi, Umut Simsekli. Introduction to Machine Learning, 52h, L3, ENS, Paris.

10.2.2 Supervision

- PhD in progress: Antoine Bambade supervised by Jean Ponce (WILLOW), Justin Carpentier (WILLOW), and Adrien Taylor.
- PhD in progress: Baptiste Goujaud supervised by Eric Moulines (École Polytechnique), Aymeric Dieuleveut (École Polytechnique), and Adrien Taylor.
- PhD in progress: Céline Moucer supervised by Francis Bach and Adrien Taylor.
- PhD in progress: Theophile Cantelobre supervised by Benjamin Guedji (INRIA), Carlo Ciliberto (UCL), A. Rudi.
- PhD in progress: Gaspard Beugnot supervised by Julien Mairal (INRIA) and A. Rudi.
- PhD in progress: Andrea Basteri supervised by A. Rudi.
- PhD in progress: Bertille Follain supervised by F. Bach and U. Simsekli.
- PhD in progress: Marc Lambert supervised by F. Bach and S. Bonnabel.

- PhD in progress: Ivan Lerner, co-advised with Anita Burgun et Antoine Neuraz.
- PhD in progress: Lawrence Stewart, co-advised by Francis Bach and Jean-Philippe Vert.
- PhD in progress: Gautier Izacard, co-advised by Alexandre d'Aspremont and Edouard Grave (Meta).
- PhD in progress: Rémi Jezequel, co-advised by Alessandro Rudi and Pierre Gaillard (Inria Grenoble)
- PhD in progress: Cle'ment Lezane, co-advised by Alexandre d'Aspremont and Cristobal Guzman.
- PhD defended: Vivien Cabannes, supervised by Francis Bach and Alessandro Rudi.
- PhD defended: Eloise Berthier, supervised by Francis Bach.
- PhD defended: Ulysse Marteau Ferey, supervised by Alessandro Rudi and Francis Bach.
- PhD defended: Grégoire Mialon, Sample Selection Methods, 2018, Alexandre d'Aspremont (joint with Julien Mairal)
- PHD defended: Theo Ryffel, supervised by Francis Bach and David Pointcheval.

10.2.3 Juries

- A. d'Aspremont. PhD exam, David Martinez, Oxford.
- A. d'Aspremont. PhD exam, Guillaume Braun, U. Lille.
- A. d'Aspremont. HdR, Andrea Simonetto, ENSTA.
- A. Rudi. PhD exam, Jonas Wacker, Eurecom.
- A. Rudi. PhD exam, Alex Lambert, Telecom Paris.
- A. Rudi, PhD exam, Belhal Karimi, Ecole Polytechnique.
- U. Simsekli, PhD exam, Thibault Séjourné, ENS Paris
- U. Simsekli, PhD exam, Merve Gürel, ETH Zurich
- F. Bach, PhD exam, El Mehdi Achour, Université de Toulouse
- F. Bach, PhD exam, Daniel LeJeune, Rice University

10.3 Popularization

10.3.1 Interventions

- A. d'Aspremont. Vivatech
- A. d'Aspremont. AI for Finance workshop, palais Brongniart.

11 Scientific production

11.1 Major publications

- [1] 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>.
- [2] 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>.

- [3] 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>.
- [4] 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>.
- [5] S. Kolouri, K. Nadjahi, S. Shahrampour 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>.
- [6] T. Lauvaux, C. Giron, M. Mazzolini, A. d'Aspremont, R. Duren, D. Cusworth, D. Shindell and P. Ciais. '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>.
- [7] 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>.
- [8] 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>.
- [9] 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>.
- [10] 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>.
- [11] 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>.
- [12] 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://doi.org/10.48550/arXiv.2204.04970). URL: <https://hal.inria.fr/hal-03635236>.

11.2 Publications of the year

International journals

- [13] 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.science/hal-03165622>.
- [14] P.-C. Aubin-Frankowski and A. Bensoussan. 'The reproducing kernel Hilbert spaces underlying linear SDE Estimation, Kalman filtering and their relation to optimal control'. In: *Pure and Applied Functional Analysis* (2022). URL: <https://hal.science/hal-03830472>.
- [15] M. Barré, A. Taylor and F. Bach. 'A note on approximate accelerated forward-backward methods with absolute and relative errors, and possibly strongly convex objectives'. In: *Open Journal of Mathematical Optimization* (11th Jan. 2022). DOI: [10.5802/ojmo.12](https://doi.org/10.5802/ojmo.12). URL: <https://hal.inria.fr/hal-03377374>.
- [16] M. Barré, A. Taylor and A. d'Aspremont. 'Convergence of a Constrained Vector Extrapolation Scheme'. In: *SIAM Journal on Mathematics of Data Science* (11th July 2022). DOI: [10.1137/21M1428030](https://doi.org/10.1137/21M1428030). URL: <https://hal.science/hal-03044520>.

- [17] R. Bollapragada, D. Scieur and A. d’Aspremont. ‘Nonlinear Acceleration of Momentum and Primal-Dual Algorithms’. In: *Mathematical Programming, Series A* (9th Feb. 2022), pp. 1–38. DOI: [10.1007/s10107-022-01775-x](https://doi.org/10.1007/s10107-022-01775-x). URL: <https://hal.science/hal-01893921>.
- [18] M. Brégère and M. Huard. ‘Online hierarchical forecasting for power consumption data’. In: *International Journal of Forecasting* 38.1 (Jan. 2022), pp. 339–351. DOI: [10.1016/j.ijforecast.2021.05.011](https://doi.org/10.1016/j.ijforecast.2021.05.011). URL: <https://hal.science/hal-03884826>.
- [19] L. Brogat-Motte, A. Rudi, C. Brouard, J. Rousu and F. d’Alché-Buc. ‘Vector-Valued Least-Squares Regression under Output Regularity Assumptions’. In: *Journal of Machine Learning Research* (2022). URL: <https://hal.telecom-paris.fr/hal-03888760>.
- [20] 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>.
- [21] Y. Drori and A. Taylor. ‘On the oracle complexity of smooth strongly convex minimization’. In: *Journal of Complexity* (1st Feb. 2022). DOI: [10.1016/j.jco.2021.101590](https://doi.org/10.1016/j.jco.2021.101590). URL: <https://hal.inria.fr/hal-03154582>.
- [22] A. Fallah, M. Gurbuzbalaban, A. Ozdaglar, U. Şimşekli and L. Zhu. ‘Robust Distributed Accelerated Stochastic Gradient Methods for Multi-Agent Networks’. In: *Journal of Machine Learning Research* (2022). URL: <https://hal.inria.fr/hal-03857867>.
- [23] T. Kerdreux, I. Colin and A. d’Aspremont. ‘Stable Bounds on the Duality Gap of Separable Nonconvex Optimization Problems’. In: *Mathematics of Operations Research* (26th July 2022). DOI: [10.1287/moor.2022.1291](https://doi.org/10.1287/moor.2022.1291). URL: <https://hal.science/hal-03837876>.
- [24] T. Kerdreux, A. d’Aspremont and S. Pokutta. ‘Restarting Frank-Wolfe’. In: *Journal of Optimization Theory and Applications* 192.3 (1st Mar. 2022), pp. 799–829. URL: <https://hal.science/hal-01893922>.
- [25] M. Lambert, S. Bonnabel and F. Bach. ‘The limited-memory recursive variational Gaussian approximation (L-RVGA)’. In: *Statistics and Computing* (2022). URL: <https://hal.inria.fr/hal-03501920>.
- [26] 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.science/hal-03565371>.
- [27] A. Taylor and Y. Drori. ‘An optimal gradient method for smooth convex minimization’. In: *Mathematical Programming, Series A* (2022). DOI: [10.1007/s10107-022-01839-y](https://doi.org/10.1007/s10107-022-01839-y). URL: <https://hal.inria.fr/hal-03154583>.

International peer-reviewed conferences

- [28] P.-C. Aubin-Frankowski and A. Bensoussan. ‘Operator-valued Kernels and Control of Infinite dimensional Dynamic Systems’. In: CDC 2022 - 61st IEEE Conference on Decision and Control. Cancun, Mexico, 2022. URL: <https://hal.science/hal-03811584>.
- [29] E. Berthier, J. Carpentier, A. Rudi and F. Bach. ‘Infinite-Dimensional Sums-of-Squares for Optimal Control’. In: 61st IEEE Conference on Decision and Control. Cancun, Mexico, 6th Dec. 2022. URL: <https://hal.science/hal-03377120>.
- [30] G. Beugnot, A. Rudi and J. Mairal. ‘On the Benefits of Large Learning Rates for Kernel Methods’. In: COLT 2022 - 35th Annual Conference on Learning Theory. Vol. 178. Proceedings of Thirty Fifth Conference on Learning Theory. London, United Kingdom, 5th July 2022, pp. 254–282. URL: <https://hal.inria.fr/hal-03878527>.
- [31] V. Cabannes, F. S. Bach, V. Perchet and A. Rudi. ‘Active Labeling: Streaming Stochastic Gradients’. In: NeurIPS 2022 - 36th Conference on Neural Information Processing Systems. New Orleans, United States, 28th Nov. 2022. URL: <https://hal.inria.fr/hal-03806666>.

- [32] L. Hodgkinson, U. Şimşekli, R. Khanna and M. W. Mahoney. ‘Generalization Bounds using Lower Tail Exponents in Stochastic Optimizers’. In: ICML 2022 - 39th International Conference on Machine Learning. Baltimore, United States, 2022. URL: <https://hal.inria.fr/hal-03935798>.
- [33] S. Kolouri, K. Nadjahi, S. Shahrampour and U. Simsekli. ‘Generalized Sliced Probability Metrics’. In: ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing. 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>.
- [34] M. Lambert, S. Bonnabel and F. Bach. ‘The continuous-discrete variational Kalman filter (CD-VKF)’. In: 61st IEEE Conference on Decision and Control. Cancun, Mexico, 6th Dec. 2022. URL: <https://hal.inria.fr/hal-03665666>.
- [35] M. Lambert, S. Chewi, F. S. Bach, S. Bonnabel and P. Rigollet. ‘Variational inference via Wasserstein gradient flows’. In: NeurIPS 2022 - Thirty-sixth Conference on Neural Information Processing Systems. Nouvelle Orléans (Louisiane), United States, 28th Nov. 2022. DOI: [10.48550/arXiv.2205.15902](https://doi.org/10.48550/arXiv.2205.15902). URL: <https://hal.inria.fr/hal-03953336>.
- [36] S. H. Lim, Y. Wan and U. Şimşekli. ‘Chaotic Regularization and Heavy-Tailed Limits for Deterministic Gradient Descent’. In: NeurIPS 2022 - Thirty-sixth Conference on Neural Information Processing Systems. New Orleans, United States, 28th Nov. 2022. URL: <https://hal.inria.fr/hal-03935819>.
- [37] U. Marteau-Ferey, F. Bach and A. Rudi. ‘Sampling from Arbitrary Functions via PSD Models’. In: AISTATS 2022 - 25th International Conference on Artificial Intelligence and Statistics. Valencia (virtual), Spain, 28th Mar. 2022. URL: <https://hal.inria.fr/hal-03386544>.
- [38] K. Mishchenko, F. Bach, M. Even and B. Woodworth. ‘Asynchronous SGD Beats Minibatch SGD Under Arbitrary Delays’. In: NeurIPS 2022 - Thirty-sixth Conference on Neural Information Processing Systems. Advances in Neural Information Processing Systems. New Orleans, United States, 28th Nov. 2022. URL: <https://hal.science/hal-03867190>.
- [39] A. Orvieto, H. Kersting, F. Proske, F. Bach and A. Lucchi. ‘Anticorrelated Noise Injection for Improved Generalization’. In: ICML 2022 - 39th International Conference on Machine Learning. Baltimore, United States, 17th July 2022. URL: <https://hal.science/hal-03883872>.
- [40] S. Park, U. Şimşekli and M. A. Erdogdu. ‘Generalization Bounds for Stochastic Gradient Descent via Localized ε -Covers’. In: NeurIPS 2022 - Thirty-sixth Conference on Neural Information Processing Systems. Advances in Neural Processing Systems. Baltimore, United States, 19th Sept. 2022. URL: <https://hal.inria.fr/hal-03935856>.
- [41] B. Woodworth, F. Bach and A. Rudi. ‘Non-Convex Optimization with Certificates and Fast Rates Through Kernel Sums of Squares’. In: *Conference on Learning Theory*. Conference on Learning Theory 2022. Vol. 178. London, United Kingdom: PMLR, 2nd July 2022, pp. 4620–4642. DOI: [10.48550/arXiv.2204.04970](https://doi.org/10.48550/arXiv.2204.04970). URL: <https://hal.inria.fr/hal-03635236>.

Conferences without proceedings

- [42] F. Bach and L. Chizat. ‘Gradient Descent on Infinitely Wide Neural Networks: Global Convergence and Generalization’. In: International Congress of Mathematicians. Saint-Petersbourg, Russia, 2022. URL: <https://hal.science/hal-03379011>.
- [43] A. Bambade, S. El-Kazdadi, A. Taylor and J. Carpentier. ‘PROX-QP: Yet another Quadratic Programming Solver for Robotics and beyond’. In: RSS 2022 - Robotics: Science and Systems. New York, United States, June 2022. URL: <https://hal.inria.fr/hal-03683733>.
- [44] E. Gorbunov, A. Taylor and G. Gidel. ‘Last-Iterate Convergence of Optimistic Gradient Method for Monotone Variational Inequalities’. In: NeurIPS 2022 - Thirty-sixth Conference on Neural Information Processing Systems. Advances in Neural Information Processing Systems 35. New Orleans, United States, 28th Nov. 2022. URL: <https://hal.science/hal-03780329>.
- [45] A. Nowak-Vila, A. Rudi and F. Bach. ‘On the Consistency of Max-Margin Losses’. In: AISTATS 2022 - 25th International Conference on Artificial Intelligence and Statistics. València, Spain, 28th Mar. 2022. URL: <https://hal.science/hal-03615096>.

- [46] 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>.

Doctoral dissertations and habilitation theses

- [47] E. Berthier. ‘Efficient Algorithms for Control and Reinforcement Learning’. Université PSL (Paris Sciences & Lettres), 27th Oct. 2022. URL: <https://theses.hal.science/tel-03850657>.
- [48] V. Cabannes. ‘From weakly supervised learning to active labeling’. Ecole Normale Supérieure (ENS), 18th July 2022. URL: <https://hal.inria.fr/tel-03806675>.
- [49] U. Marteau-Ferey. ‘Modelling functions with kernels: from logistic regression to global optimisation’. PSL, 7th Sept. 2022. URL: <https://hal.inria.fr/tel-03989753>.
- [50] G. Mialon. ‘On Inductive Biases for Machine Learning in Data Constrained Settings’. Ecole Normale Supérieure (ENS), Paris, FRA., 19th Jan. 2022. URL: <https://hal.science/tel-03999078>.
- [51] T. Ryffel. ‘Cryptography for Privacy-Preserving Machine Learning’. PSL, 23rd June 2022. URL: <https://hal.science/tel-03998109>.

Reports & preprints

- [52] F. Bach. *Information Theory with Kernel Methods*. 26th Aug. 2022. URL: <https://hal.science/hal-03577992>.
- [53] F. Bach. *On the relationship between multivariate splines and infinitely-wide neural networks*. 6th Feb. 2023. URL: <https://hal.science/hal-03974669>.
- [54] F. Bach. *Sum-of-Squares Relaxations for Information Theory and Variational Inference*. 26th Aug. 2022. URL: <https://hal.science/hal-03703475>.
- [55] F. Bach and A. Rudi. *Exponential convergence of sum-of-squares hierarchies for trigonometric polynomials*. 30th Jan. 2023. URL: <https://hal.science/hal-03843458>.
- [56] E. Berthier, Z. Kobeissi and F. Bach. *A Non-asymptotic Analysis of Non-parametric Temporal-Difference Learning*. 23rd May 2022. URL: <https://hal.science/hal-03672958>.
- [57] R. Berthier and M. B. Li. *Acceleration of Gossip Algorithms through the Euler-Poisson-Darboux Equation*. 21st Feb. 2022. URL: <https://hal.science/hal-03582008>.
- [58] 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>.
- [59] A. Défossez, C. Caucheteux, J. Rapin, O. Kabeli and J.-R. King. *Decoding speech from non-invasive brain recordings*. 10th Oct. 2022. URL: <https://hal.science/hal-03808317>.
- [60] B. Dubois-Taine, F. Bach, Q. Berthet and A. Taylor. *Fast Stochastic Composite Minimization and an Accelerated Frank-Wolfe Algorithm under Parallelization*. 5th Sept. 2022. URL: <https://hal.science/hal-03769647>.
- [61] B. Goujaud, C. Moucer, F. Glineur, J. Hendrickx, A. Taylor and A. Dieuleveut. *PEPit: computer-assisted worst-case analyses of first-order optimization methods in Python*. 19th Sept. 2022. URL: <https://hal.science/hal-03780353>.
- [62] B. Goujaud, A. Taylor and A. Dieuleveut. *Optimal first-order methods for convex functions with a quadratic upper bound*. 19th Sept. 2022. URL: <https://hal.science/hal-03780321>.
- [63] R. Jézéquel, D. M. Ostrovskii and P. Gaillard. *Efficient and Near-Optimal Online Portfolio Selection*. 27th Sept. 2022. URL: <https://hal.science/hal-03787674>.
- [64] Z. Kobeissi and F. Bach. *On a Variance-Reduction Correction for the Temporal-Difference Learning in the Stochastic Continuous Setting*. 10th June 2022. URL: <https://hal.inria.fr/hal-03574645>.

-
- [65] L. Montaut, Q. Le Lidec, A. Bambade, V. Petrik, J. Sivic and J. Carpentier. *Differentiable Collision Detection: a Randomized Smoothing Approach*. 29th Sept. 2022. URL: <https://hal.science/hal-03780482>.
- [66] C. Moucer, A. Taylor and F. Bach. *A systematic approach to Lyapunov analyses of continuous-time models in convex optimization*. 25th May 2022. URL: <https://hal.inria.fr/hal-03677528>.
- [67] A. El-Nouby, G. Izacard, H. Touvron, I. Laptev, H. Jégou and E. Grave. *Are Large-scale Datasets Necessary for Self-Supervised Pre-training?* 14th Feb. 2022. URL: <https://hal.science/hal-03572721>.
- [68] S. Peng, C. Giron, G. Liu, A. d’Aspremont, A. Benoit, T. Lauvaux, X. Lin, H. D. Rodrigues, M. Saunois and P. Ciaï. *High resolution assessment of coal mining methane emissions by satellite in Shanxi, China*. 3rd Nov. 2022. DOI: [10.31223/X5493W](https://doi.org/10.31223/X5493W). URL: <https://hal.science/hal-03837844>.
- [69] T. Ryffel, F. Bach and D. Pointcheval. *Differential Privacy Guarantees for Stochastic Gradient Langevin Dynamics*. 5th Feb. 2022. URL: <https://hal.inria.fr/hal-03547726>.
- [70] L. Stewart, F. Bach, Q. Berthet and J.-P. Vert. *Regression as Classification: Influence of Task Formulation on Neural Network Features*. 10th Nov. 2022. URL: <https://hal.science/hal-03846706>.
- [71] B. Woodworth, K. Mishchenko and F. Bach. *Two Losses Are Better Than One: Faster Optimization Using a Cheaper Proxy*. Feb. 2023. URL: <https://hal.inria.fr/hal-03977083>.