

RESEARCH CENTRE

Paris

IN PARTNERSHIP WITH:

CNRS, Ecole normale supérieure de Paris

2020

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

# Contents

<b>Project-Team SIERRA</b>	<b>1</b>
<b>1 Team members, visitors, external collaborators</b>	<b>2</b>
<b>2 Overall objectives</b>	<b>3</b>
2.1 Statement . . . . .	3
<b>3 Research program</b>	<b>4</b>
3.1 Supervised Learning . . . . .	4
3.2 Unsupervised Learning . . . . .	4
3.3 Parsimony . . . . .	4
3.4 Optimization . . . . .	4
<b>4 Application domains</b>	<b>4</b>
4.1 Applications for Machine Learning . . . . .	4
<b>5 Highlights of the year</b>	<b>5</b>
<b>6 New results</b>	<b>5</b>
6.1 Implicit Bias of Gradient Descent for Wide Two-layer Neural Networks Trained with the Logistic Loss . . . . .	5
6.2 Learning with Differentiable Perturbed Optimizers . . . . .	5
6.3 Statistically Preconditioned Accelerated Gradient Method for Distributed Optimization . . . . .	5
6.4 Batch Normalization Provably Avoids Rank Collapse for Randomly Initialised Deep Networks . . . . .	6
6.5 Tight Nonparametric Convergence Rates for Stochastic Gradient Descent under the Noiseless Linear Model . . . . .	6
6.6 Consistent Structured Prediction with Max-Min Margin Markov Networks . . . . .	6
6.7 Fast and Robust Stability Region Estimation for Nonlinear Dynamical Systems . . . . .	6
6.8 Breaking the curse of dimensionality of Global Optimization of Non-convex functions . . . . .	7
6.9 Efficient improper learning for online logistic regression . . . . .	7
6.10 Improved Sleeping Bandits with Stochastic Actions Sets and Adversarial Rewards . . . . .	7
<b>7 Bilateral contracts and grants with industry</b>	<b>7</b>
7.1 Bilateral contracts with industry . . . . .	7
7.2 Bilateral grants with industry . . . . .	8
<b>8 Partnerships and cooperations</b>	<b>8</b>
8.1 International initiatives . . . . .	8
8.1.1 Inria International Labs . . . . .	8
8.2 European initiatives . . . . .	9
8.2.1 FP7 & H2020 Projects . . . . .	9
8.3 National initiatives . . . . .	9
<b>9 Dissemination</b>	<b>9</b>
9.1 Promoting scientific activities . . . . .	9
9.1.1 Scientific events: selection . . . . .	9
9.1.2 Journal . . . . .	9
9.1.3 Invited talks . . . . .	10
9.2 Teaching - Supervision - Juries . . . . .	11
9.2.1 Teaching . . . . .	11
9.2.2 Supervision . . . . .	11
9.2.3 Juries . . . . .	12

<b>10 Scientific production</b>	<b>12</b>
10.1 Major publications	12
10.2 Publications of the year	12

## **Project-Team SIERRA**

*Creation of the Team: 2011 January 01, updated into 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]
- Pierre Gaillard [Inria, Researcher, until Aug 2020]
- Alessandro Rudi [Inria, Researcher]
- Umut Simsekli [Inria, Researcher, from Nov 2020]
- Adrien Taylor [Inria, Starting Research Position]
- Alexandre d'Aspremont [CNRS, Senior Researcher]

## Post-Doctoral Fellows

- Martin Arjovsky [Inria]
- Alberto Bietti [Inria, until Aug 2020]
- Seyed Daneshmand [Inria, from Aug 2020]
- Remy Degenne [Inria, until Sep 2020]
- Ziad Kobeissi [Institut Louis Bachelier, from Oct 2020]
- Pierre-Yves Masse [Université technique de Prague - Tchéquie, until Mar 2020]
- Boris Muzellec [Inria, from Nov 2020]
- Yifan Sun [École Normale Supérieure de Paris, until Aug 2020]

## PhD Students

- Mathieu Barre [École Normale Supérieure de Paris]
- Eloise Berthier [DGA]
- Raphael Berthier [Inria]
- Margaux Bregere [EDE, until Oct 2020]
- Vivien Cabannes [Inria]
- Alexandre Defossez [Facebook, until Jun 2020]
- Radu Alexandru Dragomir [École polytechnique, co-directed with Jérôme Bolte]
- Gautier Izacard [CNRS, from Feb 2020]
- Remi Jezequel [École Normale Supérieure de Paris]
- Thomas Kerdreux [École polytechnique, PhD completed in Sept. 2020]
- Hans Kersting [Inria, from Oct 2020]
- Marc Lambert [DGA, from Sep 2020]
- Ulysse Marteau-Ferey [Inria]
- Gregoire Mialon [Inria, Co directed with Julie nMairal]
- Alex Nowak Vila [École Normale Supérieure de Paris]

- Loucas Pillaud Vivien [Ministère de l'Ecologie, de l'Energie, du Développement durable et de la Mer, until Aug 2020]
- Manon Romain [CNRS, from Sep 2020]

### **Technical Staff**

- Loïc Estève [Inria, Engineer, until Feb 2020]
- Gautier Izacard [CNRS, Engineer, until Jan 2020]

### **Interns and Apprentices**

- Stanislas Bénéteau [Ecole normale supérieure Paris-Saclay, from Apr 2020 until Aug 2020]
- Celine Moucer [École polytechnique, from Apr 2020 until Aug 2020]
- Quentin Rebjock [Inria, until Mar 2020]

### **Administrative Assistants**

- Helene Bessin Rousseau [Inria, until Jun 2020]
- Helene Milome [Inria]
- Scheherazade Rouag [Inria, from Nov 2020]

### **Visiting Scientists**

- Anant Raj [Institut Max-Planck, until Mar 2020]
- Manon Romain [CNRS, from Jun 2020 until Aug 2020]
- Aadirupa Saha [Institut Indien des Sciences, until Jan 2020]

## **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, text processing and neuro-imaging.

Machine learning is now a vast field of research and the team focuses on the following aspects: supervised learning (kernel methods, calibration), unsupervised learning (matrix factorization, statistical tests), parsimony (structured sparsity, theory and algorithms), and optimization (convex optimization, bandit learning). These four research axes are strongly interdependent, and the interplay between them is key to successful practical applications.

## 3 Research program

### 3.1 Supervised Learning

This part of our research focuses on methods where, given a set of examples of input/output pairs, the goal is to predict the output for a new input, with research on kernel methods, calibration methods, and multi-task learning.

### 3.2 Unsupervised Learning

We focus here on methods where no output is given and the goal is to find structure of certain known types (e.g., discrete or low-dimensional) in the data, with a focus on matrix factorization, statistical tests, dimension reduction, and semi-supervised learning.

### 3.3 Parsimony

The concept of parsimony is central to many areas of science. In the context of statistical machine learning, this takes the form of variable or feature selection. The team focuses primarily on structured sparsity, with theoretical and algorithmic contributions.

### 3.4 Optimization

Optimization in all its forms is central to machine learning, as many of its theoretical frameworks are based at least in part on empirical risk minimization. The team focuses primarily on convex and bandit optimization, with a particular focus on large-scale optimization.

## 4 Application domains

### 4.1 Applications for Machine Learning

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. In collaboration with Institut Curie.
- Text processing: document collection modeling, language models.
- Audio processing: source separation, speech/music processing.
- Neuro-imaging: brain-computer interface (fMRI, EEG, MEG).

## 5 Highlights of the year

- A. Rudi: Recipient of an ERC starting grant
- F. Bach: Election at the French Academy of Sciences
- F.P. Paty, A. d’Aspremont, M. Cuturi: AISTATS 2020 notable paper award

## 6 New results

### 6.1 Implicit Bias of Gradient Descent for Wide Two-layer Neural Networks Trained with the Logistic Loss

Neural networks trained to minimize the logistic (a.k.a. cross-entropy) loss with gradient-based methods are observed to perform well in many supervised classification tasks. Towards understanding this phenomenon, we analyze the training and generalization behavior of infinitely wide two-layer neural networks with homogeneous activations. We show that the limits of the gradient flow on exponentially tailed losses can be fully characterized as a max-margin classifier in a certain nonHilbertian space of functions. In presence of hidden low-dimensional structures, the resulting margin is independent of the ambient dimension, which leads to strong generalization bounds. In contrast, training only the output layer implicitly solves a kernel support vector machine, which a priori does not enjoy such an adaptivity. Our analysis of training is non-quantitative in terms of running time but we prove computational guarantees in simplified settings by showing equivalences with online mirror descent. Finally, numerical experiments suggest that our analysis describes well the practical behavior of two-layer neural networks with ReLU activations and confirm the statistical benefits of this implicit bias

### 6.2 Learning with Differentiable Perturbed Optimizers

Machine learning pipelines often rely on optimization procedures to make discrete decisions (e.g., sorting, picking closest neighbors, or shortest paths). Although these discrete decisions are easily computed, they break the back-propagation of computational graphs. In order to expand the scope of learning problems that can be solved in an end-to-end fashion, we propose a systematic method to transform optimizers into operations that are differentiable and never locally constant. Our approach relies on stochastically perturbed optimizers, and can be used readily together with existing solvers. Their derivatives can be evaluated efficiently, and smoothness tuned via the chosen noise amplitude. We also show how this framework can be connected to a family of losses developed in structured prediction, and give theoretical guarantees for their use in learning tasks. We demonstrate experimentally the performance of our approach on various tasks.

### 6.3 Statistically Preconditioned Accelerated Gradient Method for Distributed Optimization

We consider the setting of distributed empirical risk minimization where multiple machines compute the gradients in parallel and a centralized server updates the model parameters. In order to reduce the number of communications required to reach a given accuracy, we propose a preconditioned accelerated gradient method where the preconditioning is done by solving a local optimization problem over a subsampled dataset at the server. The convergence rate of the method depends on the square root of the relative condition number between the global and local loss functions. We estimate the relative condition number for linear prediction models by studying uniform concentration of the Hessians over a bounded domain, which allows us to derive improved convergence rates for existing preconditioned gradient methods and our accelerated method. Experiments on real-world datasets illustrate the benefits of acceleration in the ill-conditioned regime.



## 6.4 Batch Normalization Provably Avoids Rank Collapse for Randomly Initialised Deep Networks

Randomly initialized neural networks are known to become harder to train with increasing depth, unless architectural enhancements like residual connections and batch normalization are used. We here investigate this phenomenon by revisiting the connection between random initialization in deep networks and spectral instabilities in products of random matrices. Given the rich literature on random matrices, it is not surprising to find that the rank of the intermediate representations in unnormalized networks collapses quickly with depth. In this work we highlight the fact that batch normalization is an effective strategy to avoid rank collapse for both linear and ReLU networks. Leveraging tools from Markov chain theory, we derive a meaningful lower rank bound in deep linear networks. Empirically, we also demonstrate that this rank robustness generalizes to ReLU nets. Finally, we conduct an extensive set of experiments on real-world data sets, which confirm that rank stability is indeed a crucial condition for training modern-day deep neural architectures.

## 6.5 Tight Nonparametric Convergence Rates for Stochastic Gradient Descent under the Noiseless Linear Model

In the context of statistical supervised learning, the noiseless linear model assumes that there exists a deterministic linear relation  $Y = \langle \theta_*, \Phi(U) \rangle$  between the random output  $Y$  and the random feature vector  $\Phi(U)$ , a potentially non-linear transformation of the inputs  $U$ . We analyze the convergence of single-pass, fixed step-size stochastic gradient descent on the least-square risk under this model. The convergence of the iterates to the optimum  $\theta_*$  and the decay of the generalization error follow polynomial convergence rates with exponents that both depend on the regularities of the optimum  $\theta_*$  and of the feature vectors  $\Phi(U)$ . We interpret our result in the reproducing kernel Hilbert space framework. As a special case, we analyze an online algorithm for estimating a real function on the unit hypercube from the noiseless observation of its value at randomly sampled points; the convergence depends on the Sobolev smoothness of the function and of a chosen kernel. Finally, we apply our analysis beyond the supervised learning setting to obtain convergence rates for the averaging process (a.k.a. gossip algorithm) on a graph depending on its spectral dimension.

## 6.6 Consistent Structured Prediction with Max-Min Margin Markov Networks

Max-margin methods for binary classification such as the support vector machine (SVM) have been extended to the structured prediction setting under the name of max-margin Markov networks (M3N), or more generally structural SVMs. Unfortunately, these methods are statistically inconsistent when the relationship between inputs and labels is far from deterministic. We overcome such limitations by defining the learning problem in terms of a “max-min” margin formulation, naming the resulting method max-min margin Markov networks (M4N). We prove consistency and finite sample generalization bounds for M4N and provide an explicit algorithm to compute the estimator. The algorithm achieves a generalization error of  $O(1/\sqrt{n})$  for a total cost of  $O(n)$  projection-oracle calls (which have at most the same cost as the max-oracle from M3N). Experiments on multi-class classification, ordinal regression, sequence prediction and ranking demonstrate the effectiveness of the proposed method.

## 6.7 Fast and Robust Stability Region Estimation for Nonlinear Dynamical Systems

A linear quadratic regulator can stabilize a nonlinear dynamical system with a local feedback controller around a linearization point, while minimizing a given performance criteria. An important practical problem is to estimate the region of attraction of such a controller, that is, the region around this point where the controller is certified to be valid. This is especially important in the context of highly nonlinear dynamical systems. In this paper, we propose two stability certificates that are fast to compute and robust when the first, or second derivatives of the system dynamics are bounded. Associated with an efficient oracle to compute these bounds, this provides a simple stability region estimation algorithm compared to classic approaches of the state of the art. We experimentally validate that it can be applied to both

polynomial and non-polynomial systems of various dimensions, including standard robotic systems, for estimating region of attractions around equilibrium points, as well as for trajectory tracking.

## 6.8 Breaking the curse of dimensionality of Global Optimization of Non-convex functions

We consider the global minimization of smooth functions based solely on function evaluations. Algorithms that achieve the optimal number of function evaluations for a given precision level typically rely on explicitly constructing an approximation of the function which is then minimized with algorithms that have exponential running-time complexity. In this project, we consider an approach that jointly models the function to approximate and finds a global minimum. This is done by using infinite sums of square smooth functions and has strong links with polynomial sum-of-squares hierarchies. Leveraging recent representation properties of reproducing kernel Hilbert spaces, the infinite-dimensional optimization problem can be solved by subsampling in time polynomial in the number of function evaluations, and with theoretical guarantees on the obtained minimum.

Given  $n$  samples, the computational cost is  $O(n^{3.5})$  in time,  $O(n^2)$  in space, and we achieve a convergence rate to the global optimum that is  $O(n^{-m/d+1/2+3/d})$  where  $m$  is the degree of differentiability of the function and  $d$  the number of dimensions. The rate is nearly optimal in the case of Sobolev functions and more generally makes the proposed method particularly suitable for functions that have a large number of derivatives. Indeed, when  $m$  is in the order of  $d$ , the convergence rate to the global optimum does not suffer from the curse of dimensionality, which affects only the worst-case constants (that we track explicitly through the paper).

## 6.9 Efficient improper learning for online logistic regression

We considered the setting of online logistic regression with the objective of minimizing the regret with respect to the  $\ell_2$ -ball of radius  $B$ . It was known (see [Hazan et al., 2014]) that any proper algorithm which had logarithmic regret in the number of samples (denoted  $n$ ) necessarily suffered an exponential multiplicative constant in  $B$ . In this work, we designed an efficient improper algorithm that avoids this exponential constant while preserving a logarithmic regret. Indeed, [Foster et al., 2018] showed that the lower bound does not apply to improper algorithms and proposed a strategy based on exponential weights with prohibitive computational complexity. Our new algorithm based on regularized empirical risk minimization with surrogate losses satisfies a regret scaling as  $O(B \log(Bn))$  with a per-round time-complexity of order  $O(d^2)$ .

## 6.10 Improved Sleeping Bandits with Stochastic Actions Sets and Adversarial Rewards

We considered the problem of sleeping bandits with stochastic action sets and adversarial rewards. In this setting, in contrast to most work in bandits, the actions may not be available at all times. For instance, some products might be out of stock in item recommendation. The best existing efficient (i.e., polynomial-time) algorithms for this problem only guarantee an  $O(T^{2/3})$  upper-bound on the regret. Yet, inefficient algorithms based on EXP4 can achieve  $O(\sqrt{T})$ . In this work, we provided a new computationally efficient algorithm inspired by EXP3 satisfying a regret of order  $O(\sqrt{T})$  when the availabilities of each action  $i \in \mathcal{A}$  are independent. We then studied the most general version of the problem where at each round available sets are generated from some unknown arbitrary distribution (i.e., without the independence assumption) and proposed an efficient algorithm with  $O(2^K \sqrt{T})$  regret guarantee. Our theoretical results were corroborated with experimental evaluations.

# 7 Bilateral contracts and grants with industry

## 7.1 Bilateral contracts with industry

Microsoft Research: “Structured Large-Scale Machine Learning”. Machine learning is now ubiquitous in industry, science, engineering, and personal life. While early successes were obtained by applying

off-the-shelf techniques, there are two main challenges faced by machine learning in the “big data” era: structure and scale. The project proposes to explore three axes, from theoretical, algorithmic and practical perspectives: (1) large-scale convex optimization, (2) large-scale combinatorial optimization and (3) sequential decision making for structured data. The project involves two Inria sites (Paris and Grenoble) and four MSR sites (Cambridge, New England, Redmond, New York). Project website: <http://www.msr-inria.fr/projects/structured-large-scale-machine-learning/>.

## 7.2 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.

# 8 Partnerships and cooperations

## 8.1 International initiatives

### 8.1.1 Inria International Labs

#### 4TUNE

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

**Duration:** 2020 - 2022

**Coordinator:** Francis Bach

**Partners:**

- Machine Learning group, CWI (Netherlands)

**Inria contact:** Francis Bach

**Website:** <http://pierre.gaillard.me/4tune/>

**Summary:** 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.

#### FOAM

**Title:** First-Order Accelerated Methods for machine learning

**Duration:** 2020 - 2022

**Coordinator:** Alexandre d’Aspremont

**Partners:**

- Mathematical and Computational Engineering, Pontificia Universidad Católica de Chile (Chile)

**Inria contact:** Alexandre d’Aspremont

**Website:** <https://sites.google.com/view/cguzman/talks-and-events/foam-associate-team>

**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.

## 8.2 European initiatives

### 8.2.1 FP7 & H2020 Projects

- European Research Council: SEQUOIA project (grant number 724063), 2017-2022 (F Bach), “Robust algorithms for learning from modern data”.

### 8.3 National initiatives

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

## 9 Dissemination

### 9.1 Promoting scientific activities

#### 9.1.1 Scientific events: selection

##### Member of the conference program committees

- Pierre Gaillard, member of the program committee for the Conference on Learning Theory (COLT), 2020

##### Reviewer

- Adrien Taylor, reviewer for International Conference on Machine Learning (ICML), 2020 (top reviewer award).
- Adrien Taylor, reviewer for International Conference on Neural Information Processing Systems (Neurips), 2020 (top reviewer award).
- Adrien Taylor, reviewer for Conference on Decision and Control (CDC), 2020.
- Pierre Gaillard, reviewer for the International Conference on Artificial Intelligence and Statistics (Aistats), 2020

#### 9.1.2 Journal

##### Member of the editorial boards

- Francis Bach, co-editor-in-chief, Journal of Machine Learning Research
- Francis Bach, associate Editor, Mathematical Programming
- Francis Bach, associate editor, Foundations of Computational Mathematics (FoCM)

**Reviewer - reviewing activities**

- Adrien Taylor, reviewer for Automatica.
- Adrien Taylor, reviewer for Journal of Machine Learning Research (JMLR).
- Adrien Taylor, reviewer for Mathematical Programming (MAPR).
- Adrien Taylor, reviewer for SIAM Journal on Optimization (SIOPT).
- Adrien Taylor, reviewer for Computational Optimization and Applications (COAP).
- Adrien Taylor, reviewer for Journal of Optimization Theory and Applications (JOTA).
- Pierre Gaillard, reviewer for Mathematics of Operations Research (MOR).

**9.1.3 Invited talks**

- Adrien Taylor, invited talk University of Cambridge (CCIMI seminars), February 2020, United Kingdom.
- Adrien Taylor, invited talk at Université catholique de Louvain (Mathematical engineering seminars), February 2020, Belgium.
- Adrien Taylor, invited talk at Pontificia Universidad Católica de Chile, April 2020, Online.
- Adrien Taylor, invited talk at One World Optimization seminars, June 2020, Online.
- Adrien Taylor, invited talk at CWI-INRIA workshop, September 2020, Online.
- Pierre Gaillard, invited talk at the Valpred workshop, March 2020
- Pierre Gaillard, invited talk at the Potsdamer research seminar, June 2020, online.
- Pierre Gaillard, invited talk at the seminar of the Statify research team, Inria Grenoble, September 2020
- Alessandro Rudi, invited talk at University College of London, Gatsby unit, London October 2020.
- Francis Bach, invited virtual talk at Optimization for machine learning, CIRM, Luminy, March 2020.
- Francis Bach, invited talk at MIT, September 2020
- Francis Bach, invited virtual talk at the University of Texas, Austin, October 2020
- Francis Bach, invited virtual talk at the Symposium on the Mathematical Foundations of Data Science, Johns Hopkins University, October 2020
- Francis Bach, invited virtual talk at Harvard University, November 2020
- Francis Bach, invited virtual talk at CIMAT, Centro de Investigación en Matemáticas, Mexico, November 2020

## 9.2 Teaching - Supervision - Juries

### 9.2.1 Teaching

- Master: Alexandre d'Aspremont, Optimisation Combinatoire et Convexe, avec Zhentao Li, (2015-Present) cours magistraux 30h, Master M1, ENS Paris.
- 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, Machine Learning, 20h, Master ICFP (Physique), Université PSL.
- Master: Pierre Gaillard, Alessandro Rudi, Introduction to Machine Learning, 52h, L3, ENS, Paris.
- Master: Pierre Gaillard, Sequential learning, 20h, Master M2 MVA, ENS PS.
- Hausdorff school on MCMC: Francis Bach, 6 hours.

### 9.2.2 Supervision

- PhD in progress : Raphaël Berthier, started September 2017, supervised by Francis Bach and Pierre Gaillard.
- PhD in progress : Radu - Dragomir Alexandru, Bregman Gradient Methods, 2018, Alexandre d'Aspremont (joint with Jérôme Bolte)
- PhD in progress : Mathieu Barré, Accelerated Polyak Methods, 2018, Alexandre d'Aspremont
- PhD in progress : Grégoire Mialon, Sample Selection Methods, 2018, Alexandre d'Aspremont (joint with Julien Mairal)
- PhD in progress : Manon Romain, Causal Inference Algorithms, 2020, Alexandre d'Aspremont
- PhD in progress: Alex Nowak-Vila, supervised by Francis Bach and Alessandro Rudi.
- PhD in progress: Ulysse Marteau Ferey, supervised by Francis Bach and Alessandro Rudi.
- PhD in progress: Vivien Cabannes, supervised by Francis Bach and Alessandro Rudi.
- PhD in progress: Eloise Berthier, supervised by Francis Bach.
- PhD in progress: Theo Ryffel, supervised by Francis Bach and David Pointcheval.
- PhD in progress: Rémi Jezequel, supervised by Pierre Gaillard and Alessandro Rudi.
- PhD in progress: Antoine Bambade, supervised by Jean-Ponce (Willow), Justin Carpentier (Willow), and Adrien Taylor.
- PhD in progress: Marc Lambert, supervised by Francis Bach and Silvére Bonnabel.
- PhD in progress: Ivan Lerner, co-advised with Anita Burgun et Antoine Neuraz.
- PhD defended: Alexandre Défossez, supervised by Francis Bach and Léon Bottou (Facebook AI Research), defended in July 2020
- PhD defended: Loucas Pillaud-Vivien, supervised by Francis Bach and Alessandro Rudi, defended October 30 2020
- PhD defended: Margaux Brégère, supervised by Pierre Gaillard and Gilles Stoltz (Université Paris-Sud), defended in December 2020
- PhD defended : Thomas Kerdreux, New Complexity Bounds for Frank Wolfe, 2017, Alexandre d'Aspremont

### 9.2.3 Juries

- HdR Pierre Weiss, IMT Toulouse, September 2019 (Alexandre d’Aspremont).
- HDR Rémi Flamary, Université de Nice, November 2019 (Francis Bach).

## 10 Scientific production

### 10.1 Major publications

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

### 10.2 Publications of the year

#### International journals

- [3] E. Berthier and F. Bach. ‘Max-Plus Linear Approximations for Deterministic Continuous-State Markov Decision Processes’. In: *IEEE Control Systems Letters* 4.3 (July 2020), pp. 767–772. DOI: [10.1109/LCSYS.2020.2973199](https://doi.org/10.1109/LCSYS.2020.2973199). URL: <https://hal.archives-ouvertes.fr/hal-02617479>.
- [4] A. D’Aspremont, M. Cucuringu and H. Tyagi. ‘Ranking and synchronization from pairwise measurements via SVD’. In: *Journal of Machine Learning Research* 22.19 (11th Feb. 2021), pp. 1–63. URL: <https://hal.archives-ouvertes.fr/hal-02340372>.
- [5] E. De Klerk, F. Glineur and A. Taylor. ‘Worst-Case Convergence Analysis of Inexact Gradient and Newton Methods Through Semidefinite Programming Performance Estimation’. In: *SIAM Journal on Optimization* 30.3 (Jan. 2020), pp. 2053–2082. DOI: [10.1137/19M1281368](https://doi.org/10.1137/19M1281368). URL: <https://hal.inria.fr/hal-02956367>.
- [6] Y. Drori and A. B. Taylor. ‘Efficient First-order Methods for Convex Minimization: a Constructive Approach’. In: *Mathematical Programming, Series A* 184 (2020), pp. 183–220. DOI: [10.1007/s10107-019-01410-2](https://doi.org/10.1007/s10107-019-01410-2). URL: <https://hal.inria.fr/hal-01902048>.
- [7] V. Roulet and A. D’Aspremont. ‘Sharpness, Restart and Acceleration’. In: *SIAM Journal on Optimization* 30.1 (29th Oct. 2020), pp. 262–289. DOI: [10.1137/18M1224568](https://doi.org/10.1137/18M1224568). URL: <https://hal.archives-ouvertes.fr/hal-02983236>.
- [8] E. Ryu, A. Taylor, C. Bergeling and P. Giselsson. ‘Operator Splitting Performance Estimation: Tight Contraction Factors and Optimal Parameter Selection’. In: *SIAM Journal on Optimization* 30.3 (Jan. 2020), pp. 2251–2271. DOI: [10.1137/19M1304854](https://doi.org/10.1137/19M1304854). URL: <https://hal.inria.fr/hal-02956361>.

#### International peer-reviewed conferences

- [9] A. Bellet, R. Guerraoui and H. Hendrikx. ‘Who started this rumor? Quantifying the natural differential privacy guarantees of gossip protocols’. In: DISC 2020 - 34th International Symposium on Distributed Computing. Freiburg / Virtual, Germany, 2020. URL: <https://hal.inria.fr/hal-02166432>.
- [10] L. Chizat and F. Bach. ‘Implicit Bias of Gradient Descent for Wide Two-layer Neural Networks Trained with the Logistic Loss’. In: COLT 2020 - 33rd Annual Conference on Learning Theory. Vol. PMLR. Proceedings of Thirty Third Conference on Learning Theory 125. Graz / Virtual, Austria, 9th July 2020, pp. 1305–1338. URL: <https://hal.archives-ouvertes.fr/hal-02473847>.
- [11] R. Degenne, P. Ménard, X. Shang and M. Valko. ‘Gamification of pure exploration for linear bandits’. In: ICML 2020 - 37th International Conference on Machine Learning. Vienna / Virtual, Austria, 12th July 2020. URL: <https://hal.archives-ouvertes.fr/hal-02884330>.

- [12] C. Fernandez, C. Shue Chen, P. Gaillard and A. Silva. ‘Experimental Comparison of Semi-parametric, Parametric, and Machine Learning Models for Time-to-Event Analysis Through the Concordance Index’. In: JDS 2020 - 52nd Statistics Days of the French Statistical Society (SfDS). Nice, France, 25th May 2020. URL: <https://hal.inria.fr/hal-02894974>.
- [13] H. Hendriks, F. Bach and L. Massoulié. ‘Dual-Free Stochastic Decentralized Optimization with Variance Reduction’. In: NeurIPS 2020 - 34th Conference on Neural Information Processing Systems. Advances in Neural Information Processing Systems Proceedings. Vancouver / Virtual, Canada, 2020. URL: <https://hal.archives-ouvertes.fr/hal-02974237>.
- [14] H. Hendriks, L. Xiao, S. Bubeck, F. Bach and L. Massoulié. ‘Statistically Preconditioned Accelerated Gradient Method for Distributed Optimization’. In: ICML 2020 - Thirty-seventh International Conference on Machine Learning. Proceedings of Machine Learning Research. Vienna / Virtual, Austria, 2020. URL: <https://hal.archives-ouvertes.fr/hal-02974232>.
- [15] N. Keriven, A. Bietti and S. Vaïter. ‘Convergence and Stability of Graph Convolutional Networks on Large Random Graphs’. In: NeurIPS 2020 - 34th Conference on Neural Information Processing Systems. Vancouver (virtual), Canada: <https://nips.cc/>, 6th Dec. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02976711>.
- [16] G. Mialon, D. Chen, A. D’Aspremont and J. Mairal. ‘A Trainable Optimal Transport Embedding for Feature Aggregation and its Relationship to Attention’. In: ICLR 2021 - The Ninth International Conference on Learning Representations. Virtual, France, 4th May 2021. URL: <https://hal.archives-ouvertes.fr/hal-02883436>.
- [17] G. Mialon, A. D’Aspremont and J. Mairal. ‘Screening Data Points in Empirical Risk Minimization via Ellipsoidal Regions and Safe Loss Functions’. In: AISTATS 2020 - 23rd International Conference on Artificial Intelligence and Statistics. Palermo / Virtual, Italy, 3rd June 2020. URL: <https://hal.archives-ouvertes.fr/hal-02395624>.
- [18] L. Pillaud-Vivien, F. Bach, T. Lelièvre, A. Rudi and G. Stoltz. ‘Statistical Estimation of the Poincaré constant and Application to Sampling Multimodal Distributions’. In: *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*. AISTATS 2020 : 23rd International Conference on Artificial Intelligence and Statistics. Palermo / Virtual, Italy, Aug. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02327453>.
- [19] A. Saha, P. Gaillard and M. Valko. ‘Improved sleeping bandits with stochastic action sets and adversarial rewards’. In: ICML 2020 - 37th International Conference on Machine Learning. Vienna / Virtual, Austria, 12th July 2020. URL: <https://hal.inria.fr/hal-02950106>.

#### Conferences without proceedings

- [20] A. Askari, A. D’Aspremont and L. E. Ghaoui. ‘Naive Feature Selection: Sparsity in Naive Bayes’. In: AISTATS 2020 - 23rd International Conference on Artificial Intelligence and Statistics. Palermo / Virtual, Italy, 3rd June 2020. URL: <https://hal.archives-ouvertes.fr/hal-02340374>.
- [21] M. Barré, A. Taylor and A. D’Aspremont. ‘Complexity Guarantees for Polyak Steps with Momentum’. In: COLT 2020 - 33rd Annual Conference on Learning Theory. Graz / Virtual, Austria, 9th July 2020. URL: <https://hal.archives-ouvertes.fr/hal-02470997>.
- [22] R. Berthier, F. Bach and P. Gaillard. ‘Tight Nonparametric Convergence Rates for Stochastic Gradient Descent under the Noiseless Linear Model’. In: NeurIPS ’20 - 34th International Conference on Neural Information Processing Systems. Vancouver, Canada, 7th Dec. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02866755>.
- [23] V. Cabannes, A. Rudi and F. Bach. ‘Structured Prediction with Partial Labelling through the Infimum Loss’. In: *Proceedings of the 37th International Conference on Machine Learning, Proceeding of Machine Learning Research*. ICML 2020 - 37th International Conference on Machine Learning. Vol. 119. Proceedings of the 37th International Conference on Machine Learning. Online, United States, 18th July 2020, pp. 1230–1239. URL: <https://hal.archives-ouvertes.fr/hal-03127110>.



- [24] O. Cifka, A. Ozerov, U. Simsekli and G. Richard. ‘Self-Supervised VQ-VAE for One-Shot Music Style Transfer’. In: ICASSP 2021 - IEEE International Conference on Acoustics, Speech and Signal Processing. Toronto / Virtual, Canada, 6th June 2021. URL: <https://hal.telecom-paris.fr/hal-03132940>.
- [25] R. Jézéquel, P. Gaillard and A. Rudi. ‘Efficient improper learning for online logistic regression’. In: COLT 2020 - 33rd Annual Conference on Learning Theory. Graz / Virtual, Austria, 9th July 2020. URL: <https://hal.archives-ouvertes.fr/hal-02510505>.
- [26] F.-P. Paty, A. D’Aspremont and M. Cuturi. ‘Regularity as Regularization: Smooth and Strongly Convex Brenier Potentials in Optimal Transport’. In: AISTATS 2020 - 23rd International Conference on Artificial Intelligence and Statistics. Palermo / Virtual, Italy, 3rd June 2020. URL: <https://hal.archives-ouvertes.fr/hal-02340371>.

#### Doctoral dissertations and habilitation theses

- [27] M. Brégère. ‘Stochastic bandit algorithms for demand side management’. Université Paris-Saclay, 10th Dec. 2020. URL: <https://hal.archives-ouvertes.fr/tel-03059605>.
- [28] T. Kerdreux. ‘Accelerating conditional gradient methods’. Université Paris sciences et lettres, 30th June 2020. URL: <https://tel.archives-ouvertes.fr/tel-03052834>.

#### Reports & preprints

- [29] A. Askari, Q. Rebjock, A. D’Aspremont and L. E. Ghaoui. *FANOK: Knockoffs in Linear Time*. 29th Oct. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02983262>.
- [30] F. Bach. *On the Effectiveness of Richardson Extrapolation in Machine Learning*. 9th July 2020. URL: <https://hal.archives-ouvertes.fr/hal-02470950>.
- [31] M. Barré, A. Taylor and F. Bach. *Principled Analyses and Design of First-Order Methods with Inexact Proximal Operators*. 23rd Sept. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02947160>.
- [32] M. Barré, A. Taylor and A. D’Aspremont. *Convergence of Constrained Anderson Acceleration*. 7th Dec. 2020. URL: <https://hal.archives-ouvertes.fr/hal-03044520>.
- [33] E. Berthier, J. Carpentier and F. Bach. *Fast and Robust Stability Region Estimation for Nonlinear Dynamical Systems*. 30th Oct. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02984348>.
- [34] R. Berthier, F. Bach, N. Flammarion, P. Gaillard and A. Taylor. *A Continuized View on Nesterov Acceleration*. 11th Feb. 2021. URL: <https://hal.archives-ouvertes.fr/hal-03138823>.
- [35] A. Bietti and F. Bach. *Deep Equals Shallow for ReLU Networks in Kernel Regimes*. 9th Oct. 2020. URL: <https://hal.inria.fr/hal-02963250>.
- [36] A. D’Aspremont and M. Pilanci. *Global Convergence of Frank Wolfe on One Hidden Layer Networks*. 29th Oct. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02983259>.
- [37] C. Fernandez, C. S. Chen, P. Gaillard and A. Silva. *Experimental Comparison of Semi-parametric, Parametric, and Machine Learning Models for Time-to-Event Analysis Through the Concordance Index*. 12th Mar. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02507132>.
- [38] T. Kerdreux, I. Colin and A. D’Aspremont. *An Approximate Shapley-Folkman Theorem*. 29th Oct. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02983258>.
- [39] M. Lambert, S. Bonnabel and F. Bach. *The recursive variational Gaussian approximation (R-VGA)*. 30th Dec. 2020. URL: <https://hal.inria.fr/hal-03086627>.
- [40] U. Marteau-Ferey, F. Bach and A. Rudi. *Non-parametric Models for Non-negative Functions*. 7th July 2020. URL: <https://hal.inria.fr/hal-02891640>.
- [41] D. M. Ostrovskii and F. Bach. *Finite-sample analysis of M-estimators using self-concordance*. 30th Nov. 2020. URL: <https://hal.archives-ouvertes.fr/hal-01895127>.

- 
- [42] A. Raj, P. Gaillard and C. Saad. *Non-stationary Online Regression*. 10th Nov. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02998781>.
  - [43] A. Rudi, U. Marteau-Ferey and F. Bach. *Finding Global Minima via Kernel Approximations*. 23rd Dec. 2020. URL: <https://hal.inria.fr/hal-03081675>.
  - [44] T. Ryffel, D. Pointcheval and F. Bach. *ARIANN: Low-Interaction Privacy-Preserving Deep Learning via Function Secret Sharing*. 10th July 2020. URL: <https://hal.inria.fr/hal-02896127>.
  - [45] H. Zenati, A. Bietti, M. Martin, E. Diemert and J. Mairal. *Counterfactual Learning of Continuous Stochastic Policies*. 29th June 2020. URL: <https://hal.archives-ouvertes.fr/hal-02883423>.