



IN PARTNERSHIP WITH:
CNRS

**Ecole normale supérieure de
Paris**

Activity Report 2015

Project-Team SIERRA

Statistical Machine Learning and Parsimony

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

RESEARCH CENTER
Paris - Rocquencourt

THEME
**Optimization, machine learning and
statistical methods**

Table of contents

1. Members	1
2. Overall Objectives	2
3. Research Program	2
3.1. Supervised Learning	2
3.2. Unsupervised Learning	2
3.3. Parsimony	2
3.4. Optimization	3
4. Application Domains	3
5. Highlights of the Year	3
6. New Software and Platforms	3
6.1. DICA: Moment Matching for Latent Dirichlet Allocation (LDA) and Discrete Independent Component Analysis (DICA)	3
6.2. LinearFW: Implementation of linearly convergent versions of Frank-Wolfe	4
6.3. CNN-Head-Detection: Context-aware CNNs for person head detection	4
7. New Results	4
7.1. On the Global Linear Convergence of Frank-Wolfe Optimization Variants	4
7.2. Barrier Frank-Wolfe for Marginal Inference	5
7.3. Sequential Kernel Herding: Frank-Wolfe Optimization for Particle Filtering	5
7.4. Variance Reduced Stochastic Gradient Descent with Neighbors	5
7.5. Rethinking LDA: Moment Matching for Discrete ICA	5
7.6. Tensorizing Neural Networks	6
7.7. Context-Aware CNNs for Person Head Detection	6
7.8. Unsupervised Learning from Narrated Instruction Videos	6
7.9. On Pairwise Cost for Multi-Object Network Flow Tracking	7
7.10. Multi-utility Learning: Structured-Output Learning with Multiple Annotation-Specific Loss Functions	7
7.11. Convex Optimization for Parallel Energy Minimization	7
7.12. Active-set Methods for Submodular Optimization	8
7.13. Spectral Norm Regularization of Orthonormal Representations for Graph Transduction	8
7.14. On the Equivalence between Quadrature Rules and Random Features	8
7.15. Preconditioning of a Generalized Forward-Backward Splitting and Application to Optimization on Graphs	9
7.16. A Riemannian Low-Rank Method for Optimization over Semidefinite Matrices with Block-Diagonal Constraints	9
7.17. Tightness of the Maximum Likelihood Semidefinite Relaxation for Angular Synchronization	9
7.18. Coherent Diffractive Imaging Using Randomly Coded Masks	10
7.19. Renegar's Condition Number and Compressed Sensing Performance	10
7.20. Supervised Clustering in the Data Cube	10
7.21. Convex Relaxations for Permutation Problems	10
7.22. Phase Recovery, MaxCut and Complex Semidefinite Programming	11
7.23. Choice of V for V -Fold Cross-Validation in Least-Squares	11
7.24. Gains and Losses are Fundamentally Different in Regret Minimization: The Sparse Case	11
7.25. Batched Bandit Problems	11
7.26. Online Learning in Repeated Auctions	12
8. Bilateral Contracts and Grants with Industry	12
8.1. Bilateral Contracts with Industry	12
8.2. Bilateral Grants with Industry	12
9. Partnerships and Cooperations	12
9.1. National Initiatives	12

9.1.1.	ANR: Calibration	12
9.1.2.	CNRS: BeFast	13
9.2.	European Initiatives	13
9.2.1.1.	SIPA	13
9.2.1.2.	SpaRTaN	13
9.2.1.3.	MacSeNet	14
9.3.	International Research Visitors	14
10.	Dissemination	15
10.1.	Promoting Scientific Activities	15
10.1.1.	Scientific events organisation	15
10.1.1.1.	General chair, scientific chair	15
10.1.1.2.	Member of the organizing committees	15
10.1.2.	Scientific events selection	15
10.1.3.	Journal	15
10.1.4.	Others	15
10.1.5.	Invited talks	15
10.2.	Teaching - Supervision - Juries	16
10.2.1.	Teaching	16
10.2.2.	Supervision	16
10.2.3.	Juries	17
10.3.	Popularization	17
11.	Bibliography	17

Project-Team SIERRA

Creation of the Team: 2011 January 01, updated into Project-Team: 2012 January 01

Keywords:

Computer Science and Digital Science:

- 3.4. - Machine learning and statistics
- 6.2. - Scientific Computing, Numerical Analysis & Optimization
- 7.3. - Operations research, optimization, game theory
- 8.2. - Machine learning

Other Research Topics and Application Domains:

- 9.4.5. - Data science

1. Members

Research Scientists

Francis Bach [Team leader, Inria, HdR]
Sylvain Arlot [CNRS, Researcher, until Sept 2015]
Simon Lacoste-Julien [Inria, Starting Research position]
Alexandre d'Aspremont [CNRS, HdR]

Faculty Member

Vianney Perchet [Univ. Paris VII, Associate Professor, Delegation until Aug 2015]

PhD Students

Jean-Baptiste Alayrac [Ecole Polytechnique]
Dmitry Babichev [Inria, from Sep 2015]
Anaël Bonneton [ENS Paris]
Aymeric Dieuleveut [ENS Paris]
Christophe Dupuy [Technicolor]
Nicolas Flammarion [ENS Lyon]
Fajwel Fogel [CNRS]
Damien Garreau [Inria]
K. S. Sesh Kumar [Inria]
Remi Lajugie [Inria, until Oct 2015]
Loic Landrieu [Ecole des Ponts, until Sep 2015]
Remi Leblond [Inria, from Sep 2015]
Anastasia Podosinnikova [Inria]
Antoine Recanatì [CNRS, from Apr 2015]
Vincent Roulet [Ecole Polytechnique]
Thomas Schatz [Univ. Paris VI]
Damien Scieur [Inria, from Sep 2015]
Tatiana Shpakova [Inria, from Sep 2015]

Post-Doctoral Fellows

Amit Bermanis [CNRS]
Nicolas Boumal [Research in Paris]
Pascal Germain [Inria, from Nov 2015]
Anton Osokin [Inria]
Balamurugan Palaniappan [Inria]

Visiting Scientists

Mihai Cucuringu [UCLA, from Mar 2015 until Apr 2015]

Raman Sankaran [Indian Institute of Science, until Feb 2015]

Administrative Assistant

Lindsay Polienor [Inria]

Others

Remy Degenne [Inria, Masters intern, from May 2015 until Sep 2015]

Guillaume Obozinski [ENPC, associated member]

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 (this is the main topic of the ERC starting investigator grant awarded to F. Bach).

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

5. Highlights of the Year

5.1. Highlights of the Year

F. Bach has served as a program co-chair for the International Conference in Machine Learning (ICML) held in Lille, France, 2015.

6. New Software and Platforms

6.1. DICA: Moment Matching for Latent Dirichlet Allocation (LDA) and Discrete Independent Component Analysis (DICA)

The DICA package contains Matlab and C++ (via Matlab mex files) implementations of estimation in the LDA and closely related DICA models [21].

The implementation consists of two parts. One part contains the efficient implementation for construction of the moment/cumulant tensors, while the other part contains implementations of several so called joint diagonalization type algorithms used for matching the tensors. Any tensor type (see below) can be arbitrarily combined with one of the diagonalization algorithms (see below) leading, in total, to 6 algorithms.

Two types of tensors are considered: (a) the LDA moments and (b) the DICA cumulants. The diagonalization algorithms include: (a) the orthogonal joint diagonalization algorithm based on iterative Jacobi rotations, (b) the spectral algorithm based on two eigen decompositions, and (c) the tensor power method.

- Contact: Anastasia Podosinnikova
- URL: <https://github.com/anastasia-podosinnikova/dica>

6.2. LinearFW: Implementation of linearly convergent versions of Frank-Wolfe

This is the code to reproduce all the experiments in the NIPS 2015 paper: "On the Global Linear Convergence of Frank-Wolfe Optimization Variants" by Simon Lacoste-Julien and Martin Jaggi [17], which covers the global linear convergence rate of Frank-Wolfe optimization variants for problems described as in Eq. (1) in the paper. It contains the implementation of Frank-Wolfe, away-steps Frank-Wolfe and pairwise Frank-Wolfe on two applications.

- Contact: Simon Lacoste-Julien
- URL: <https://github.com/Simon-Lacoste-Julien/linearFW>

6.3. CNN-Head-Detection: Context-aware CNNs for person head detection

This is the code for the ICCV 2015 paper "Context-aware CNNs for person head detection" [23]. Person detection is a key problem for many computer vision tasks. While face detection has reached maturity, detecting people under a full variation of camera view-points, human poses, lighting conditions and occlusions is still a difficult challenge. In this work we focus on detecting human heads in natural scenes. Starting from the recent local R-CNN object detector, we extend it with two types of contextual cues. First, we leverage person-scene relations and propose a Global CNN model trained to predict positions and scales of heads directly from the full image. Second, we explicitly model pairwise relations among objects and train a Pairwise CNN model using a structured-output surrogate loss. The Local, Global and Pairwise models are combined into a joint CNN framework. To train and test our full model, we introduce a large dataset composed of 369,846 human heads annotated in 224,740 movie frames. We evaluate our method and demonstrate improvements of person head detection against several recent baselines in three datasets. We also show improvements of the detection speed provided by our model.

- Contact: Anton Osokin
- URL: https://github.com/aosokin/cnn_head_detection

7. New Results

7.1. On the Global Linear Convergence of Frank-Wolfe Optimization Variants

Participant: Simon Lacoste-Julien [correspondent].

Collaboration with Martin Jaggi (ETH Zurich).

The Frank-Wolfe (FW) optimization algorithm has lately re-gained popularity thanks in particular to its ability to nicely handle the structured constraints appearing in machine learning applications. However, its convergence rate is known to be slow (sublinear) when the solution lies at the boundary. A simple less-known fix is to add the possibility to take 'away steps' during optimization, an operation that importantly does not require a feasibility oracle. In this paper [17], we highlight and clarify several variants of the Frank-Wolfe optimization algorithm that have been successfully applied in practice: away-steps FW, pairwise FW, fully-corrective FW and Wolfe's minimum norm point algorithm, and prove for the first time that they all enjoy global linear convergence, under a weaker condition than strong convexity of the objective. The constant in the convergence rate has an elegant interpretation as the product of the (classical) condition number of the function with a novel geometric quantity that plays the role of a 'condition number' of the constraint set. We provide pointers to where these algorithms have made a difference in practice, in particular with the flow polytope, the marginal polytope and the base polytope for submodular optimization.

7.2. Barrier Frank-Wolfe for Marginal Inference

Participant: Simon Lacoste-Julien [correspondent].

Collaboration with Rahul G. Krishnan [correspondent] and David Sontag (NYU).

In [16], we introduce a globally-convergent algorithm for optimizing the tree-reweighted (TRW) variational objective over the marginal polytope. The algorithm is based on the conditional gradient method (Frank-Wolfe) and moves pseudomarginals within the marginal polytope through repeated maximum a posteriori (MAP) calls. This modular structure enables us to leverage black-box MAP solvers (both exact and approximate) for variational inference, and obtains more accurate results than tree-reweighted algorithms that optimize over the local consistency relaxation. Theoretically, we bound the sub-optimality for the proposed algorithm despite the TRW objective having unbounded gradients at the boundary of the marginal polytope. Empirically, we demonstrate the increased quality of results found by tightening the relaxation over the marginal polytope as well as the spanning tree polytope on synthetic and real-world instances.

7.3. Sequential Kernel Herding: Frank-Wolfe Optimization for Particle Filtering

Participants: Simon Lacoste-Julien [correspondent], Francis Bach.

Collaboration with Fredrik Lindsten (University of Cambridge).

Recently, the Frank-Wolfe optimization algorithm was suggested as a procedure to obtain adaptive quadrature rules for integrals of functions in a reproducing kernel Hilbert space (RKHS) with a potentially faster rate of convergence than Monte Carlo integration (and “kernel herding” was shown to be a special case of this procedure). In this paper [18], we propose to replace the random sampling step in a particle filter by Frank-Wolfe optimization. By optimizing the position of the particles, we can obtain better accuracy than random or quasi-Monte Carlo sampling. In applications where the evaluation of the emission probabilities is expensive (such as in robot localization), the additional computational cost to generate the particles through optimization can be justified. Experiments on standard synthetic examples as well as on a robot localization task indicate indeed an improvement of accuracy over random and quasi-Monte Carlo sampling.

7.4. Variance Reduced Stochastic Gradient Descent with Neighbors

Participant: Simon Lacoste-Julien [correspondent].

Collaboration with Thomas Hofmann [correspondent], Aurelien Lucchi and Brian McWilliams (ETH Zurich).

Stochastic Gradient Descent (SGD) is a workhorse in machine learning, yet its slow convergence can be a computational bottleneck. Variance reduction techniques such as SAG, SVRG and SAGA have been proposed to overcome this weakness, achieving linear convergence. However, these methods are either based on computations of full gradients at pivot points, or on keeping per data point corrections in memory. Therefore speed-ups relative to SGD may need a minimal number of epochs in order to materialize. This paper [15] investigates algorithms that can exploit neighborhood structure in the training data to share and re-use information about past stochastic gradients across data points, which offers advantages in the transient optimization phase. As a side-product we provide a unified convergence analysis for a family of variance reduction algorithms, which we call memorization algorithms. We provide experimental results supporting our theory.

7.5. Rethinking LDA: Moment Matching for Discrete ICA

Participants: Anastasia Podosinnikova [correspondent], Francis Bach, Simon Lacoste-Julien.

In [21], we consider moment matching techniques for estimation in latent Dirichlet allocation (LDA). By drawing explicit links between LDA and discrete versions of independent component analysis (ICA), we first derive a new set of cumulant-based tensors, with an improved sample complexity. Moreover, we reuse standard ICA techniques such as joint diagonalization of tensors to improve over existing methods based on the tensor power method. In an extensive set of experiments on both synthetic and real datasets, we show that our new combination of tensors and orthogonal joint diagonalization techniques outperforms existing moment matching methods.

7.6. Tensorizing Neural Networks

Participant: Anton Osokin [correspondent].

Collaboration with Alexander Novikov, Dmitry Podoprikin and Dmitry Vetrov.

Deep neural networks currently demonstrate state-of-the-art performance in several domains. At the same time, models of this class are very demanding in terms of computational resources. In particular, a large amount of memory is required by commonly used fully-connected layers, making it hard to use the models on low-end devices and stopping the further increase of the model size. In this paper [20], we convert the dense weight matrices of the fully-connected layers to the Tensor Train format such that the number of parameters is reduced by a huge factor and at the same time the expressive power of the layer is preserved. In particular, for the Very Deep VGG networks we report the compression factor of the dense weight matrix of a fully-connected layer up to 200000 times leading to the compression factor of the whole network up to 7 times.

7.7. Context-Aware CNNs for Person Head Detection

Participant: Anton Osokin [correspondent].

Collaboration with Tuan-Hung Vu [correspondent] and Ivan Laptev from the Willow project-team.

Person detection is a key problem for many computer vision tasks. While face detection has reached maturity, detecting people under a full variation of camera view-points, human poses, lighting conditions and occlusions is still a difficult challenge. In this work [23], we focus on detecting human heads in natural scenes. Starting from the recent local R-CNN object detector, we extend it with two types of contextual cues. First, we leverage person-scene relations and propose a Global CNN model trained to predict positions and scales of heads directly from the full image. Second, we explicitly model pairwise relations among objects and train a Pairwise CNN model using a structured-output surrogate loss. The Local, Global and Pairwise models are combined into a joint CNN framework. To train and test our full model, we introduce a large dataset composed of 369,846 human heads annotated in 224,740 movie frames. We evaluate our method and demonstrate improvements of person head detection against several recent baselines in three datasets. We also show improvements of the detection speed provided by our model.

7.8. Unsupervised Learning from Narrated Instruction Videos

Participants: Jean-Baptiste Alayrac [correspondent], Simon Lacoste-Julien.

Collaboration with Piotr Bojanowski, Josef Sivic and Ivan Laptev from the Willow project-team, and Nishant Agrawal.

In [29], we address the problem of automatically learning the main steps to complete a certain task, such as changing a car tire, from a set of narrated instruction videos. The contributions of this paper are three-fold. First, we develop a new unsupervised learning approach that takes advantage of the complementary nature of the input video and the associated narration. The method solves two clustering problems, one in text and one in video, applied one after each other and linked by joint constraints to obtain a single coherent sequence of steps in both modalities. Second, we collect and annotate a new challenging dataset of real-world instruction videos from the Internet. The dataset contains about 800,000 frames for five different tasks that include complex interactions between people and objects, and are captured in a variety of indoor and outdoor settings. Third, we experimentally demonstrate that the proposed method can automatically discover, in an unsupervised manner, the main steps to achieve the task and locate the steps in the input videos.

7.9. On Pairwise Cost for Multi-Object Network Flow Tracking

Participant: Simon Lacoste-Julien.

Collaboration with Visesh Chari, Ivan Laptev [correspondent] and Josef Sivic from the Willow project-team.

Multi-object tracking has been recently approached with the min-cost network flow optimization techniques. Such methods simultaneously resolve multiple object tracks in a video and enable modeling of dependencies among tracks. Min-cost network flow methods also fit well within the “tracking-by-detection” paradigm where object trajectories are obtained by connecting per-frame outputs of an object detector. Object detectors, however, often fail due to occlusions and clutter in the video. To cope with such situations, we propose in [13] an approach that regularizes the tracker by adding second order costs to the min-cost network flow framework. While solving such a problem with integer variables is NP-hard, we present a convex relaxation with an efficient rounding heuristic which empirically gives certificates of small suboptimality. Results are shown on real-world video sequences and demonstrate that the new constraints help selecting longer and more accurate tracks improving over the baseline tracking-by-detection method.

7.10. Multi-utility Learning: Structured-Output Learning with Multiple Annotation-Specific Loss Functions

Participant: Anton Osokin [correspondent].

Collaboration with Roman Shapovalov, Dmitry Vetrov and Pushmeet Kohli.

Structured-output learning is a challenging problem; particularly so because of the difficulty in obtaining large datasets of fully labelled instances for training. In this paper [22], we try to overcome this difficulty by presenting a multi-utility learning framework for structured prediction that can learn from training instances with different forms of supervision. We propose a unified technique for inferring the loss functions most suitable for quantifying the consistency of solutions with the given weak annotation. We demonstrate the effectiveness of our framework on the challenging semantic image segmentation problem for which a wide variety of annotations can be used. For instance, the popular training datasets for semantic segmentation are composed of images with hard-to-generate full pixel labellings, as well as images with easy-to-obtain weak annotations, such as bounding boxes around objects, or image-level labels that specify which object categories are present in an image. Experimental evaluation shows that the use of annotation-specific loss functions dramatically improves segmentation accuracy compared to the baseline system where only one type of weak annotation is used.

7.11. Convex Optimization for Parallel Energy Minimization

Participants: K. S. Sesh Kumar [correspondent], Francis Bach.

Collaboration with Alvaro Barbero, Stefanie Jegelka and Suvrit Sra.

Energy minimization has been an intensely studied core problem in computer vision. With growing image sizes (2D and 3D), it is now highly desirable to run energy minimization algorithms in parallel. But many existing algorithms, in particular, some efficient combinatorial algorithms, are difficult to parallelize. By exploiting results from convex and submodular theory, we reformulate in [47] the quadratic energy minimization problem as a total variation denoising problem, which, when viewed geometrically, enables the use of projection and reflection based convex methods. The resulting min-cut algorithm (and code) is conceptually very simple, and solves a sequence of TV denoising problems. We perform an extensive empirical evaluation comparing state-of-the-art combinatorial algorithms and convex optimization techniques. On small problems the iterative convex methods match the combinatorial max-flow algorithms, while on larger problems they offer other flexibility and important gains: (a) their memory footprint is small; (b) their straightforward parallelizability fits multi-core platforms; (c) they can easily be warm-started; and (d) they quickly reach approximately good solutions, thereby enabling faster “inexact” solutions. A key consequence of our approach based on submodularity and convexity is that it allows to combine *any arbitrary combinatorial or convex methods as subroutines*, which allows one to obtain hybrid combinatorial and convex optimization algorithms that benefit from the strengths of both.

7.12. Active-set Methods for Submodular Optimization

Participants: K. S. Sesh Kumar [correspondent], Francis Bach.

In [46], we consider submodular optimization problems such as submodular function minimization (SFM) and quadratic problems regularized by the Lovász extension; for cut functions, this corresponds respectively to graph cuts and total variation (TV) denoising. Given a submodular function with an SFM oracle, we propose a new active-set algorithm for total variation denoising, which is more flexible than existing ones; the algorithm may be seen as a local descent algorithm over ordered partitions with explicit convergence guarantees. For functions that decompose into the sum of two functions F_1 and F_2 with efficient SFM oracles, we propose a new active-set algorithm for total variation denoising (and hence for SFM by thresholding the solution at zero). This algorithm also optimizes over ordered partitions and improves over existing ones based on TV or SFM oracles for F_1 and F_2 .

7.13. Spectral Norm Regularization of Orthonormal Representations for Graph Transduction

Participant: Francis Bach [correspondent].

Collaboration with the Indian Institute of Science, Bangalore, India.

Recent literature suggests that embedding a graph on a unit sphere leads to better generalization for graph transduction. However, the choice of optimal embedding and an efficient algorithm to compute the same remains open. In this paper [25], we show that orthonormal representations, a class of unit-sphere graph embeddings are PAC learnable. Existing PAC-based analysis do not apply as the VC dimension of the function class is infinite. We propose an alternative PAC-based bound, which do not depend on the VC dimension of the underlying function class, but is related to the famous Lovasz function. The main contribution of the paper is SPORE, a SPECTRAL regularized ORTHONORMAL Embedding for graph transduction, derived from the PAC bound. SPORE is posed as a non-smooth convex function over an ellipsope. These problems are usually solved as semi-definite programs (SDPs) with time complexity $O(n^6)$. We present, Infeasible Inexact proximal (IIP): an Inexact proximal method which performs subgradient procedure on an approximate projection, not necessarily feasible. IIP is more scalable than SDP, has an $O(1/\sqrt{T})$ convergence, and is generally applicable whenever a suitable approximate projection is available. We use IIP to compute SPORE where the approximate projection step is computed by FISTA, an accelerated gradient descent procedure. We show that the method has a convergence rate of $O(1/\sqrt{T})$. The proposed algorithm easily scales to 1000's of vertices, while the standard SDP computation does not scale beyond few hundred vertices. Furthermore, the analysis presented here easily extends to the multiple graph setting.

7.14. On the Equivalence between Quadrature Rules and Random Features

Participant: Francis Bach [correspondent].

In [31], we show that kernel-based quadrature rules for computing integrals can be seen as a special case of random feature expansions for positive definite kernels, for a particular decomposition that always exists for such kernels. We provide a theoretical analysis of the number of required samples for a given approximation error, leading to both upper and lower bounds that are based solely on the eigenvalues of the associated integral operator and match up to logarithmic terms. In particular, we show that the upper bound may be obtained from independent and identically distributed samples from a specific non-uniform distribution, while the lower bound is valid for any set of points. Applying our results to kernel-based quadrature, while our results are fairly general, we recover known upper and lower bounds for the special cases of Sobolev spaces. Moreover, our results extend to the more general problem of full function approximations (beyond simply computing an integral), with results in L_2 - and L_∞ -norm that match known results for special cases. Applying our results to random features, we show an improvement of the number of random features needed to preserve the generalization guarantees for learning with Lipschitz-continuous losses.

7.15. Preconditioning of a Generalized Forward-Backward Splitting and Application to Optimization on Graphs

Participant: Loïc Landrieu [correspondent].

Collaboration with Hugo Raguet.

In [41], we present a preconditioning of a generalized forward-backward splitting algorithm for finding a zero of a sum of maximally monotone operators $\sum_{i=1}^n A_i + B$ with B cocoercive, involving only the computation of B and of the resolvent of each A_i separately. This allows in particular to minimize functionals of the form $\sum_{i=1}^n g_i + f$ with f smooth, using only the gradient of f and the proximity operator of each g_i separately. By adapting the underlying metric, such preconditioning can serve two practical purposes: first, it might accelerate the convergence, or second, it might simplify the computation of the resolvent of A_i for some i . In addition, in many cases of interest, our preconditioning strategy allows the economy of storage and computation concerning some auxiliary variables. In particular, we show how this approach can handle large-scale, non-smooth, convex optimization problems structured on graphs, which arises in many image processing or learning applications, and that it compares favourably to alternatives in the literature.

7.16. A Riemannian Low-Rank Method for Optimization over Semidefinite Matrices with Block-Diagonal Constraints

Participant: Nicolas Boumal [correspondent].

In [34], we propose a new algorithm to solve optimization problems of the form $\min f(X)$ for a smooth function f under the constraints that X is positive semidefinite and the diagonal blocks of X are small identity matrices. Such problems often arise as the result of relaxing a rank constraint (lifting). In particular, many estimation tasks involving phases, rotations, orthonormal bases or permutations fit in this framework, and so do certain relaxations of combinatorial problems such as Max-Cut. The proposed algorithm exploits the facts that (1) such formulations admit low-rank solutions, and (2) their rank-restricted versions are smooth optimization problems on a Riemannian manifold. Combining insights from both the Riemannian and the convex geometries of the problem, we characterize when second-order critical points of the smooth problem reveal KKT points of the semidefinite problem. We compare against state of the art, mature software and find that, on certain interesting problem instances, what we call the staircase method is orders of magnitude faster, is more accurate and scales better. Code is available.

7.17. Tightness of the Maximum Likelihood Semidefinite Relaxation for Angular Synchronization

Participant: Nicolas Boumal [correspondent].

Collaboration with Afonso S. Bandeira and Amit Singer.

Many maximum likelihood estimation problems are, in general, intractable optimization problems. As a result, it is common to approximate the maximum likelihood estimator (MLE) using convex relaxations. Semidefinite relaxations are among the most popular. Sometimes, the relaxations turn out to be tight. In this paper [33], we study such a phenomenon. The angular synchronization problem consists in estimating a collection of n phases, given noisy measurements of some of the pairwise relative phases. The MLE for the angular synchronization problem is the solution of a (hard) non-bipartite Grothendieck problem over the complex numbers. It is known that its semidefinite relaxation enjoys worst-case approximation guarantees. In this paper, we consider a stochastic model on the input of that semidefinite relaxation. We assume there is a planted signal (corresponding to a ground truth set of phases) and the measurements are corrupted with random noise. Even though the MLE does not coincide with the planted signal, we show that the relaxation is, with high probability, tight. This holds even for high levels of noise. This analysis explains, for the interesting case of angular synchronization, a phenomenon which has been observed without explanation in many other settings. Namely, the fact that even when exact recovery of the ground truth is impossible, semidefinite relaxations for the MLE tend to be tight (in favorable noise regimes).

7.18. Coherent Diffractive Imaging Using Randomly Coded Masks

Participant: Alexandre d'Aspremont [correspondent].

Collaboration with Matthew H. Seaberg and Joshua J. Turner.

Coherent diffractive imaging (CDI) provides new opportunities for high resolution X-ray imaging with simultaneous amplitude and phase contrast. Extensions to CDI broaden the scope of the technique for use in a wide variety of experimental geometries and physical systems. Here [44], we experimentally demonstrate a new extension to CDI that encodes additional information through the use of a series of randomly coded masks. The information gained from the few additional diffraction measurements removes the need for typical object-domain constraints; the algorithm uses prior information about the masks instead. The experiment is performed using a laser diode at 532.2 nm, enabling rapid prototyping for future X-ray synchrotron and even free electron laser experiments. Diffraction patterns are collected with up to 15 different masks placed between a CCD detector and a single sample. Phase retrieval is performed using a convex relaxation routine known as "PhaseCut" followed by a variation on Fienup's input-output algorithm. The reconstruction quality is judged via calculation of phase retrieval transfer functions as well as by an object-space comparison between reconstructions and a lens-based image of the sample. The results of this analysis indicate that with enough masks (in this case 3 or 4) the diffraction phases converge reliably, implying stability and uniqueness of the retrieved solution.

7.19. Renegar's Condition Number and Compressed Sensing Performance

Participants: Vincent Roulet, Nicolas Boumal, Alexandre d'Aspremont [correspondent].

Renegar's condition number is a data-driven computational complexity measure for convex programs, generalizing classical condition numbers in linear systems. In [42], we provide evidence that for a broad class of compressed sensing problems, the worst case value of this algorithmic complexity measure taken over all signals matches the restricted eigenvalue of the observation matrix, which controls compressed sensing performance. This means that, in these problems, a single parameter directly controls computational complexity and recovery performance.

7.20. Supervised Clustering in the Data Cube

Participants: Vincent Roulet [correspondent], Fajwel Fogel, Alexandre d'Aspremont, Francis Bach.

In [43], we study a supervised clustering problem seeking to cluster either features, tasks or sample points using losses extracted from supervised learning problems. We formulate a unified optimization problem handling these three settings and derive algorithms whose core iteration complexity is concentrated in a k-means clustering step, which can be approximated efficiently. We test our methods on both artificial and realistic data sets extracted from movie reviews and 20NewsGroup.

7.21. Convex Relaxations for Permutation Problems

Participants: Fajwel Fogel [correspondent], Francis Bach, Alexandre d'Aspremont.

Collaboration with Rodolphe Jenatton.

Seriation seeks to reconstruct a linear order between variables using unsorted similarity information. It has direct applications in archeology and shotgun gene sequencing for example. In [4], we prove the equivalence between the seriation and the combinatorial 2-sum problem (a quadratic minimization problem over permutations) over a class of similarity matrices. The seriation problem can be solved exactly by a spectral algorithm in the noiseless case and we produce a convex relaxation for the 2-sum problem to improve the robustness of solutions in a noisy setting. This relaxation also allows us to impose additional structural constraints on the solution, to solve semi-supervised seriation problems. We present numerical experiments on archeological data, Markov chains and gene sequences.

7.22. Phase Recovery, MaxCut and Complex Semidefinite Programming

Participant: Alexandre d’Aspremont [correspondent].

Collaboration with Irène Waldspurger and Stéphane Mallat.

Phase retrieval seeks to recover a signal x from the amplitude $|Ax|$ of linear measurements. We cast the phase retrieval problem as a non-convex quadratic program over a complex phase vector and formulate a tractable relaxation (called PhaseCut) similar to the classical MaxCut semidefinite program. In [10], we solve this problem using a provably convergent block coordinate descent algorithm whose structure is similar to that of the original greedy algorithm in Gerchberg-Saxton, where each iteration is a matrix vector product. Numerical results show the performance of this approach over three different phase retrieval problems, in comparison with greedy phase retrieval algorithms and matrix completion formulations.

7.23. Choice of V for V -Fold Cross-Validation in Least-Squares

Participant: Sylvain Arlot [correspondent].

Collaboration with Matthieu Lerasle.

The paper [30] studies V -fold cross-validation for model selection in least-squares density estimation. The goal is to provide theoretical grounds for choosing V in order to minimize the least-squares loss of the selected estimator. We first prove a non-asymptotic oracle inequality for V -fold cross-validation and its bias-corrected version (V -fold penalization). In particular, this result implies that V -fold penalization is asymptotically optimal in the nonparametric case. Then, we compute the variance of V -fold cross-validation and related criteria, as well as the variance of key quantities for model selection performance. We show that these variances depend on V like $1 + 4/(V - 1)$, at least in some particular cases, suggesting that the performance increases much from $V = 2$ to $V = 5$ or 10, and then is almost constant. Overall, this can explain the common advice to take $V = 5$ —at least in our setting and when the computational power is limited—, as supported by some simulation experiments. An oracle inequality and exact formulas for the variance are also proved for Monte-Carlo cross-validation, also known as repeated cross-validation, where the parameter V is replaced by the number B of random splits of the data.

7.24. Gains and Losses are Fundamentally Different in Regret Minimization: The Sparse Case

Participant: Vianney Perchet [correspondent].

Collaboration with Joon Kwon.

In [38], we demonstrate that, in the classical non-stochastic regret minimization problem with d decisions, gains and losses to be respectively maximized or minimized are fundamentally different. Indeed, by considering the additional sparsity assumption (at each stage, at most s decisions incur a nonzero outcome), we derive optimal regret bounds of different orders. Specifically, with gains, we obtain an optimal regret guarantee after T stages of order $\sqrt{T \log s}$, so the classical dependency in the dimension is replaced by the sparsity size. With losses, we provide matching upper and lower bounds of order $\sqrt{Ts \log(d)/d}$, which is decreasing in d . Eventually, we also study the bandit setting, and obtain an upper bound of order $\sqrt{Ts \log(d/s)}$ when outcomes are losses. This bound is proven to be optimal up to the logarithmic factor $\sqrt{\log(d/s)}$.

7.25. Batched Bandit Problems

Participant: Vianney Perchet [correspondent].

Collaboration with Philippe Rigollet, Sylvain Chassang and Erik Snowberg.

Motivated by practical applications, chiefly clinical trials, we study in [39] the regret achievable for stochastic bandits under the constraint that the employed policy must split trials into a small number of batches. Our results show that a very small number of batches gives close to minimax optimal regret bounds. As a byproduct, we derive optimal policies with low switching cost for stochastic bandits.

7.26. Online Learning in Repeated Auctions

Participant: Vianney Perchet [correspondent].

Collaboration with Jonathan Weed and Philippe Rigollet.

Motivated by online advertising auctions, in [40] we consider repeated Vickrey auctions where goods of unknown value are sold sequentially and bidders only learn (potentially noisy) information about a good's value once it is purchased. We adopt an online learning approach with bandit feedback to model this problem and derive bidding strategies for two models: stochastic and adversarial. In the stochastic model, the observed values of the goods are random variables centered around the true value of the good. In this case, logarithmic regret is achievable when competing against well behaved adversaries. In the adversarial model, the goods need not be identical and we simply compare our performance against that of the best fixed bid in hindsight. We show that sublinear regret is also achievable in this case and prove matching minimax lower bounds. To our knowledge, this is the first complete set of strategies for bidders participating in auctions of this type.

8. Bilateral Contracts and Grants with Industry

8.1. 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/>.

8.2. Bilateral Grants with Industry

- A. d’Aspremont, IdR EMMA, Euroclear – Institut Louis Bachelier.
- A. d’Aspremont, IBM Faculty Award.
- A. d’Aspremont and F. Bach, IdR AXA “Machine Learning”, chaire Havas-ILB, Economie des Nouvelles Données.
- A. d’Aspremont and F. Bach, Comité de pilotage, chaire Havas – Dauphine “Economie des nouvelles données”.
- S. Lacoste-Julien (with J. Sivic and I. Laptev in Willow project-team), Google Research Award “Structured Learning from Video and Natural Language”.

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. ANR: Calibration

Participant: Sylvain Arlot.

Titre: Statistical calibration

Coordinator: University Paris Dauphine

Leader: Vincent Rivoirard

Other members: 34 members, mostly among CEREMADE (Paris Dauphine), Laboratoire Jean-Alexandre Dieudonné (Université de Nice) and Laboratoire de Mathématiques de l’Université Paris Sud

Instrument: ANR Blanc

Duration: Jan 2012 - Dec 2015

Total funding: 240 000 euros

Webpage: <https://sites.google.com/site/anrcalibration/>

9.1.2. CNRS: BeFast

Participant: Sylvain Arlot.

Titre: BeFast

Coordinator: University Lille 1

Leader: Alain Celisse

Other members: Tristan Mary-Huard, Guillem Rigail, Guillemette Marot, and Julien Chiquet.

Instrument: PEPS

Duration: Mar 2015 – Dec 2015

Total funding: 9 000 euros

9.2. European Initiatives

9.2.1. FP7 & H2020 Projects

9.2.1.1. SIPA

Type: FP7

Defi: NC

Instrument: ERC Starting Grant

Duration: May 2011 - May 2016

Coordinator: A. d'Aspremont (CNRS)

Abstract: Interior point algorithms and a dramatic growth in computing power have revolutionized optimization in the last two decades. Highly nonlinear problems which were previously thought intractable are now routinely solved at reasonable scales. Semidefinite programs (i.e. linear programs on the cone of positive semidefinite matrices) are a perfect example of this trend: reasonably large, highly nonlinear but convex eigenvalue optimization problems are now solved efficiently by reliable numerical packages. This in turn means that a wide array of new applications for semidefinite programming have been discovered, mimicking the early development of linear programming. To cite only a few examples, semidefinite programs have been used to solve collaborative filtering problems (e.g. make personalized movie recommendations), approximate the solution of combinatorial programs, optimize the mixing rate of Markov chains over networks, infer dependence patterns from multivariate time series or produce optimal kernels in classification problems. These new applications also come with radically different algorithmic requirements. While interior point methods solve relatively small problems with a high precision, most recent applications of semidefinite programming in statistical learning for example form very large-scale problems with comparatively low precision targets, programs for which current algorithms cannot form even a single iteration. This proposal seeks to break this limit on problem size by deriving reliable first-order algorithms for solving large-scale semidefinite programs with a significantly lower cost per iteration, using for example subsampling techniques to considerably reduce the cost of forming gradients. Beyond these algorithmic challenges, the proposed research will focus heavily on applications of convex programming to statistical learning and signal processing theory where optimization and duality results quantify the statistical performance of coding or variable selection algorithms for example. Finally, another central goal of this work will be to produce efficient, customized algorithms for some key problems arising in machine learning and statistics.

9.2.1.2. SpaRTaN

Title: Sparse Representations and Compressed Sensing Training Network

Type: FP7

Defi: NC

Instrument: Initial Training Network

Duration: October 2014 to October 2018

Coordinator: Mark Plumbley (University of Surrey)

Inria contact: Francis Bach

Abstract: The SpaRTaN Initial Training Network will train a new generation of interdisciplinary researchers in sparse representations and compressed sensing, contributing to Europe's leading role in scientific innovation.

By bringing together leading academic and industry groups with expertise in sparse representations, compressed sensing, machine learning and optimisation, and with an interest in applications such as hyperspectral imaging, audio signal processing and video analytics, this project will create an interdisciplinary, trans-national and inter-sectorial training network to enhance mobility and training of researchers in this area.

SpaRTaN is funded under the FP7-PEOPLE-2013-ITN call and is part of the Marie Curie Actions — Initial Training Networks (ITN) funding scheme: Project number - 607290

9.2.1.3. *MacSeNet*

Title: Machine Sensing Training Network

Type: H2020

Instrument: Initial Training Network

Duration: January 2015 - January 2019

Coordinator: Mark Plumbley (University of Surrey)

Inria contact: Francis Bach

Abstract: The aim of this Innovative Training Network is to train a new generation of creative, entrepreneurial and innovative early stage researchers (ESRs) in the research area of measurement and estimation of signals using knowledge or data about the underlying structure.

We will develop new robust and efficient Machine Sensing theory and algorithms, together methods for a wide range of signals, including: advanced brain imaging; inverse imaging problems; audio and music signals; and non-traditional signals such as signals on graphs. We will apply these methods to real-world problems, through work with non-Academic partners, and disseminate the results of this research to a wide range of academic and non-academic audiences, including through publications, data, software and public engagement events.

MacSeNet is funded under the H2020-MSCA-ITN-2014 call and is part of the Marie Skłodowska-Curie Actions — Innovative Training Networks (ITN) funding scheme.

9.3. International Research Visitors

9.3.1. *Visits of International Scientists*

Visit from Chiranjib Bhattacharyya, Indian Institute of Science, Bangalore, May 2014.

9.3.1.1. *Internships*

Visit from Raman Sankaran, Indian Institute of Science, Bangalore, January 2014.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific events organisation

10.1.1.1. General chair, scientific chair

- F. Bach, program co-chair for International Conference on Machine Learning (ICML), 2015.

10.1.1.2. Member of the organizing committees

- A. d'Aspremont, workshop organizer, "Optimisation pour l'apprentissage statistique", École de Physique des Houches, January 2015.
- V. Perchet, local chair for the main conference in learning theory, the "Conference On Learning Theory – COLT'15", July 2015.
- V. Perchet, workshop organizer, "Challenges in Optimization for Data Science", July 2015.
- V. Perchet, organizer of the conference "French Symposium on Games - Theory and Applications", May 2015.

10.1.2. Scientific events selection

10.1.2.1. Area chairs

- S. Lacoste-Julien, International Conference on Machine Learning (ICML), 2015.

10.1.3. Journal

10.1.3.1. Member of the editorial boards

- A. d'Aspremont, SIAM Journal on Optimization, Associate Editor.
- F. Bach, Journal of Machine Learning Research, Action Editor.
- F. Bach, Information and Inference, Associate Editor.
- F. Bach, SIAM Journal on Imaging Sciences, Associate Editor.

10.1.4. Others

- A. d'Aspremont is on the scientific committee for Programme Gaspard Monge pour l'Optimisation, Fondation mathématique Jacques Hadamard.
- S. Lacoste-Julien received a NIPS 2015 Outstanding Reviewer Award.

10.1.5. Invited talks

- S. Arlot, "V-fold selection of kernel estimators", EMS 2015: European Meeting of Statisticians, session 'Recent advances in resampling methods' (Amsterdam, July 6, 2015).
- S. Arlot, "Cross-validation for estimator selection", Workshop "Information Based Complexity and Model Selection" (IHP, Paris, April 9, 2015).
- S. Arlot, "Comparaison de procédures de validation croisée (V-fold)", Conférence "Calibration statistique" (Université de Nice - Sophia Antipolis, February 20, 2015).
- S. Arlot, "Cross-validation for estimator selection", SMPGD 2015: Statistical Methods for Post Genomic Data (Ludwig-Maximilians University, February 13, 2015).
- S. Arlot, "Sélection d'estimateurs par validation croisée", Colloquium du MAP5 (Université Paris Descartes, November 13, 2015).
- S. Arlot, "Analyse du biais de forêts purement aléatoires", Séminaire de Probabilités et Statistiques (Lieu, Laboratoire JA Dieudonné, Université de Nice - Sophia Antipolis, January 8, 2015).
- A. d'Aspremont, "Ranking from Pairwise Comparisons using Seriation":

- Séminaire du Collège de France, June 2015.
- International Symposium on Mathematical Programming, July 2015.
- STATLEARN 2015.
- Fields Institute workshop on big data, Toronto, Feb. 2015.
- A. d’Aspremont, “Seriation, DNA Sequencing and Nanopores”, Datalead 2015.
- A. d’Aspremont, “Optimisation et apprentissage”, Conférence SCOR - Institut des actuaires, December 2015.
- F. Bach, Invited presentation at Workshop on Data-driven Algorithmics, Harvard, September 2015.
- F. Bach, Invited tutorial at Allerton Conference, Urbana-Champaign, 2015.
- F. Bach, Invited seminar at Ecole Polytechnique Federale de Lausanne, October 2015.
- D. Garreau, “Kernel change-point detection”, 45th Ecole d’été de Saint-Flour (Saint-Flour, July 7-17, 2015).
- S. Lacoste-Julien, “Sequential Kernel Herding: Frank-Wolfe Optimization for Particle Filtering”, invited talk at the Bayes in Paris Seminar at ENSAE, Paris, France, March 2015.
- S. Lacoste-Julien, “SAGA: a fast incremental gradient method”, invited talk in the Nonsmooth Optimization cluster at ISMP 2015 (22nd International Symposium on Mathematical Programming), Pittsburgh, PA, USA, July 2015.
- S. Lacoste-Julien, “Frank-Wolfe Optimization for Structured Machine Learning”:
 - McGill University, Montreal, Canada, December 2015.
 - Université de Montréal, Montreal, Canada, December 2015.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Master : S. Arlot and F. Bach, “Statistical learning”, 24h, M2, Université Paris-Sud.

Master : A. d’Aspremont is co-director of the Master MASH (Mathématiques, Apprentissage et Sciences Humaines) at Paris Sciences et Lettres.

Master : A. d’Aspremont, “Optimisation et apprentissage”, 21h, M2 (MVA), ENS Cachan.

Master : A. d’Aspremont, “Optimisation convexe”, 21h, M2, Ecole Normale Supérieure.

Master : S. Lacoste-Julien and F. Bach, “Apprentissage statistique”, 35h, M1, Ecole Normale Supérieure.

Master : S. Lacoste-Julien and F. Bach (together with G. Obozinski), 30h, M2 (MVA), ENS Cachan.

Master : S. Lacoste-Julien, “Projets informatiques”, 20h, M2 (MASH), Université de Paris-Dauphine.

10.2.2. Supervision

PhD : Fabian Pedregosa, “Feature extraction and supervised learning on fMRI: from practice to theory”, UPMC, February 20th, 2015, F. Bach, co-advised with A. Gramfort (Telecom).

PhD : Rémi Lajugie, “Structured prediction for sequential data”, UPMC, September 18th, 2015, S. Arlot and F. Bach.

PhD : Fajwel Fogel, “Convex and spectral relaxations for phase retrieval, seriation and ranking”, Ecole Polytechnique, November 18th, 2015, A. d’Aspremont and F. Bach.

PhD in progress: Thomas Schatz, September 2012, F. Bach, co-advised with E. Dupoux (ENS, cognitive sciences).

PhD in progress: Sesh Kumar, September 2013, F. Bach.

PhD in progress: Rafael Rezende, September 2013, F. Bach, co-advised with J. Ponce.

PhD in progress: Anastasia Podosinnikova, December 2013, F. Bach and S. Lacoste-Julien.
PhD in progress: Christophe Dupuy, January 2014, F. Bach, co-advised with C. Diot (Technicolor).
PhD in progress: Jean-Baptiste Alayrac, September 2014, S. Lacoste-Julien, co-advised with J. Sivic and I. Laptev.
PhD in progress: Aymeric Dieuleveut, September 2014, F. Bach.
PhD in progress: Nicolas Flammarion, September 2014, A. d'Aspremont and F. Bach.
PhD in progress: Damien Garreau, September 2014, S. Arlot (co-advised with G. Biau).
PhD in progress: Vincent Roulet, October 2014, A. d'Aspremont.
PhD in progress: Anaël Bonneton, December 2014, F. Bach, located in Agence nationale de la sécurité des systèmes d'information (ANSSI).
PhD in progress: Dmitry Babichev, September 2015, F. Bach, co-advised with A. Juditsky (Univ. Grenoble).
PhD in progress: Rémi Leblond, September 2015, S. Lacoste-Julien.
PhD in progress: Antoine Recanati, September 2015, A. d'Aspremont.
PhD in progress: Damien Scieur, September 2015, A. d'Aspremont and F. Bach.
PhD in progress: Tatiana Shpakova, September 2015, F. Bach.

10.2.3. *Juries*

S. Arlot, member of the PhD Committee for Maud Thomas at Université Paris Diderot on July 2, 2015.
F. Bach, member of the HDR committee of Jeremie Mary (Univ. Lille).

10.3. Popularization

S. Lacoste-Julien, general public talk "Apprentissage automatique et big data", at La Maison des Sciences, Châtenay-Malabry, France, September 2015.

11. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

- [1] F. FOGEL. *Convex and Spectral Relaxations for Phase Retrieval, Seriation and Ranking*, Ecole Doctorale de l'Ecole Polytechnique, November 2015, <https://hal.inria.fr/tel-01265606>
- [2] R. LAJUGIE. *Structured prediction for sequential data*, Ecole Normale Supérieure, September 2015, <https://tel.archives-ouvertes.fr/tel-01203438>

Articles in International Peer-Reviewed Journals

- [3] F. BACH. *Duality between subgradient and conditional gradient methods*, in "SIAM Journal of Optimization", January 2015, vol. 25, n^o 1, pp. 115-129 [DOI : 10.1137/130941961], <https://hal.archives-ouvertes.fr/hal-00757696>
- [4] F. FOGEL, R. JENATTON, F. BACH, A. D'ASPREMONT. *Convex Relaxations for Permutation Problems*, in "SIAM Journal on Matrix Analysis and Applications", December 2015, vol. 36, n^o 4 [DOI : 10.1137/130947362], <https://hal.archives-ouvertes.fr/hal-01239317>

- [5] R. GRIBONVAL, R. JENATTON, F. BACH. *Sparse and spurious: dictionary learning with noise and outliers*, in "IEEE Transactions on Information Theory", 2015, 22 p. , This is a substantially revised version of a first draft that appeared as a preprint titled "Local stability and robustness of sparse dictionary learning in the presence of noise", <http://hal.inria.fr/hal-00737152> [DOI : 10.1109/TIT.2015.2472522], <https://hal.archives-ouvertes.fr/hal-01025503>
- [6] R. GRIBONVAL, R. JENATTON, F. BACH, M. KLEINSTEUBER, M. SEIBERT. *Sample Complexity of Dictionary Learning and other Matrix Factorizations*, in "IEEE Transactions on Information Theory", 2015, 18 p. , forthcoming [DOI : 10.1109/TIT.2015.2424238], <https://hal.inria.fr/hal-00918142>
- [7] A. OSOKIN, D. P. VETROV. *Submodular Relaxation for Inference in Markov Random Fields*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", July 2015, vol. 37, n^o 7, 14 p. [DOI : 10.1109/TPAMI.2014.2369046], <https://hal.archives-ouvertes.fr/hal-01103488>
- [8] F. PEDREGOSA, M. EICKENBERG, P. CIUCIU, B. THIRION, A. GRAMFORT. *Data-driven HRF estimation for encoding and decoding models*, in "NeuroImage", January 2015, pp. 209–220, <https://hal.inria.fr/hal-00952554>
- [9] N. SHERVASHIDZE, F. BACH. *Learning the Structure for Structured Sparsity*, in "IEEE Transactions on Signal Processing", September 2015, vol. 63, n^o 18, pp. 4894 - 4902 [DOI : 10.1109/TSP.2015.2446432], <https://hal.inria.fr/hal-00986380>
- [10] I. WALDSPURGER, A. D'ASPREMONT, S. MALLAT. *Phase recovery, MaxCut and complex semidefinite programming*, in "Mathematical Programming", February 2015, vol. 149, n^o 1-2 [DOI : 10.1007/s10107-013-0738-9], <https://hal.archives-ouvertes.fr/hal-01239328>

International Conferences with Proceedings

- [11] A. BIETTI, F. BACH, A. CONT. *An online EM algorithm in hidden (semi-)Markov models for audio segmentation and clustering*, in "ICASSP 2015 - 40th IEEE International Conference on Acoustics, Speech and Signal Processing", Brisbane, Australia, April 2015, <https://hal.inria.fr/hal-01115826>
- [12] P. BOJANOWSKI, R. LAJUGIE, E. GRAVE, F. BACH, I. LAPTEV, J. PONCE, C. SCHMID. *Weakly-Supervised Alignment of Video With Text*, in "ICCV 2015 - IEEE International Conference on Computer Vision", Santiago, Chile, December 2015, <https://hal.inria.fr/hal-01154523>
- [13] V. CHARI, S. LACOSTE-JULIEN, I. LAPTEV, J. SIVIC. *On Pairwise Cost for Multi-Object Network Flow Tracking*, in "CVPR 2015 - 28th IEEE Conference on Computer Vision and Pattern Recognition", Boston, United States, June 2015, <https://hal.inria.fr/hal-01110678>
- [14] A. DÉFOSSEZ, F. BACH. *Constant Step Size Least-Mean-Square: Bias-Variance Trade-offs and Optimal Sampling Distributions*, in "International Conference on Artificial Intelligence and Statistics (AISTATS)", San Diego, United States, 2015, <https://hal.inria.fr/hal-01081578>
- [15] T. HOFMANN, A. LUCCHI, S. LACOSTE-JULIEN, B. MCWILLIAMS. *Variance Reduced Stochastic Gradient Descent with Neighbors*, in "NIPS 2015 - Advances in Neural Information Processing Systems 28", Montreal, Canada, December 2015, <https://hal.inria.fr/hal-01248672>

- [16] R. G. KRISHNAN, S. LACOSTE-JULIEN, D. SONTAG. *Barrier Frank-Wolfe for Marginal Inference*, in "NIPS 2015 - Advances in Neural Information Processing Systems 28", Montreal, Canada, December 2015, <https://hal.inria.fr/hal-01248674>
- [17] S. LACOSTE-JULIEN, M. JAGGI. *On the Global Linear Convergence of Frank-Wolfe Optimization Variants*, in "NIPS 2015 - Advances in Neural Information Processing Systems 28", Montreal, Canada, December 2015, <https://hal.inria.fr/hal-01248675>
- [18] S. LACOSTE-JULIEN, F. LINDSTEN, F. BACH. *Sequential Kernel Herding: Frank-Wolfe Optimization for Particle Filtering*, in "18th International Conference on Artificial Intelligence and Statistics (AISTATS)", San Diego, United States, JMLR Workshop and Conference Proceedings, May 2015, vol. 38, <https://hal.archives-ouvertes.fr/hal-01099197>
- [19] R. LAJUGIE, P. BOJANOWSKI, P. CUVILLIER, S. ARLOT, F. BACH. *A weakly-supervised discriminative model for audio-to-score alignment*, in "41st International Conference on Acoustics, Speech, and Signal Processing (ICASSP)", Shanghai, China, March 2016, <https://hal.archives-ouvertes.fr/hal-01251018>
- [20] A. NOVIKOV, D. PODOPRIKHIN, A. OSOKIN, D. VETROV. *Tensorizing Neural Networks*, in "The 29-th Conference on Natural Information Processing Systems (NIPS)", Montréal, Canada, December 2015, <https://hal.archives-ouvertes.fr/hal-01237600>
- [21] A. PODOSSINIKOVA, F. BACH, S. LACOSTE-JULIEN. *Rethinking LDA: Moment Matching for Discrete ICA*, in "NIPS 2015 - Advances in Neural Information Processing Systems 28", Montreal, Canada, December 2015, <https://hal.inria.fr/hal-01225271>
- [22] R. SHAPOVALOV, D. VETROV, A. OSOKIN, P. KOHLI. *Multi-utility Learning: Structured-Output Learning with Multiple Annotation-Specific Loss Functions*, in "Energy Minimization Methods in Computer Vision and Pattern Recognition", Hong Kong, China, Lecture Notes in Computer Science, January 2015, vol. 8932, pp. 406-420 [DOI : 10.1007/978-3-319-14612-6_30], <https://hal.archives-ouvertes.fr/hal-01237571>
- [23] T.-H. VU, A. OSOKIN, I. LAPTEV. *Context-aware CNNs for person head detection*, in "ICCV 2015 - IEEE International Conference on Computer Vision", Santiago, Chile, December 2015, <https://hal.archives-ouvertes.fr/hal-01237618>

Conferences without Proceedings

- [24] S. EL AOUAD, C. DUPUY, R. TEIXEIRA, C. DIOT, F. BACH. *Exploiting crowd sourced reviews to explain movie recommendation*, in "2nd Workshop on Recommendation Systems for Television and Online Video", Vienna, Austria, September 2015, <https://hal.inria.fr/hal-01193308>
- [25] R. SHIVANNA, B. CHATTERJEE, R. SANKARAN, C. BHATTACHARYYA, F. BACH. *Spectral Norm Regularization of Orthonormal Representations for Graph Transduction*, in "Neural Information Processing Systems", Montreal, Canada, December 2015, <https://hal.archives-ouvertes.fr/hal-01265524>
- [26] B. THIRION, F. PEDREGOSA, M. EICKENBERG, G. VAROQUAUX. *Correlations of correlations are not reliable statistics: implications for multivariate pattern analysis*, in "ICML Workshop on Statistics, Machine Learning and Neuroscience (Stamline 2015)", Lille, France, Bertrand Thirion, Lars Kai Hansen, Sanmi Koyejo, July 2015, <https://hal.inria.fr/hal-01187297>

Research Reports

- [27] A. DIEULEVEUT, F. BACH. *Non-parametric Stochastic Approximation with Large Step sizes*, DIENS Inria, July 2015, <https://hal.archives-ouvertes.fr/hal-01053831>

Scientific Popularization

- [28] G. VAROQUAUX, E. GOUILLART, O. VAHTRAS, V. HAENEL, N. P. ROUGIER, R. GOMMERS, F. PEDREGOSA, Z. JĘDRZEJEWSKI-SZMEK, P. VIRTANEN, C. COMBELLES, D. PINTE, R. CIMRMAN, A. ESPAZE, A. CHAUVE, C. BURNS. *Scipy Lecture Notes*, Zenodo, 2015 [DOI : 10.5281/ZENODO.31736], <https://hal.inria.fr/hal-01206546>

Other Publications

- [29] J.-B. ALAYRAC, P. BOJANOWSKI, N. AGRAWAL, J. SIVIC, I. LAPTEV, S. LACOSTE-JULIEN. *Unsupervised Learning from Narrated Instruction Videos*, June 2015, improved NLP method and bigger dataset, <https://hal.inria.fr/hal-01171193>
- [30] S. ARLOT, M. LERASLE. *Choice of V for V-Fold Cross-Validation in Least-Squares Density Estimation*, October 2015, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-00743931>
- [31] F. BACH. *On the Equivalence between Kernel Quadrature Rules and Random Feature Expansions*, November 2015, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01118276>
- [32] F. BACH. *Submodular Functions: from Discrete to Continuous Domains*, October 2015, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01222319>
- [33] A. S. BANDEIRA, N. BOUMAL, A. SINGER. *Tightness of the maximum likelihood semidefinite relaxation for angular synchronization*, October 2015, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01240565>
- [34] N. BOUMAL. *A Riemannian low-rank method for optimization over semidefinite matrices with block-diagonal constraints*, June 2015, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01213337>
- [35] N. FLAMMARION, F. BACH. *From Averaging to Acceleration, There is Only a Step-size*, March 2015, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01136945>
- [36] R. LAJUGIE, P. BOJANOWSKI, S. ARLOT, F. BACH. *Semidefinite and Spectral Relaxations for Multi-Label Classification*, June 2015, working paper or preprint, <https://hal.inria.fr/hal-01159321>
- [37] F. PEDREGOSA, F. BACH, A. GRAMFORT. *On the Consistency of Ordinal Regression Methods*, September 2015, working paper or preprint, <https://hal.inria.fr/hal-01054942>
- [38] V. PERCHET, J. KWON. *Gains and Losses are Fundamentally Different in Regret Minimization: The Sparse Case*, November 2015, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01265075>
- [39] V. PERCHET, P. RIGOLLET, S. CHASSANG, E. SNOWBERG. *Batched Bandit Problems*, July 2015, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01265077>

-
- [40] V. PERCHET, P. RIGOLLET, J. WEED. *Online learning in repeated auctions*, November 2015, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01265076>
- [41] H. R. RAGUET, L. LANDRIEU. *Preconditioning of a Generalized Forward-Backward Splitting and Application to Optimization on Graphs*, April 2015, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01144566>
- [42] V. ROULET, N. BOUMAL, A. D'ASPREMONT. *Renegar's Condition Number and Compressed Sensing Performance*, December 2015, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01239310>
- [43] V. ROULET, F. FOGEL, A. D'ASPREMONT, F. BACH. *Supervised Clustering in the Data Cube*, December 2015, Submitted to NIPS, <https://hal.archives-ouvertes.fr/hal-01239305>
- [44] M. H. SEABERG, A. D'ASPREMONT, J. J. TURNER. *Coherent Diffractive Imaging Using Randomly Coded Masks*, December 2015, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01239315>
- [45] G. SEGUIN, P. BOJANOWSKI, R. LAJUGIE, I. LAPTEV. *Instance-level video segmentation from object tracks*, January 2016, working paper or preprint, <https://hal.inria.fr/hal-01255765>
- [46] K. S. SESH KUMAR, F. BACH. *Active-set Methods for Submodular Optimization*, June 2015, working paper or preprint, <https://hal.inria.fr/hal-01161759>
- [47] K. S. SESH KUMAR, A. BARBERO, S. JEGELKA, S. SRA, F. BACH. *Convex Optimization for Parallel Energy Minimization*, March 2015, working paper or preprint, <https://hal.inria.fr/hal-01123492>