



IN PARTNERSHIP WITH:  
**CNRS**

**Ecole normale supérieure de  
Paris**

Activity Report 2014

## **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

<b>1. Members</b>	<b>1</b>
<b>2. Overall Objectives</b>	<b>2</b>
<b>3. Research Program</b>	<b>2</b>
3.1. Supervised Learning	2
3.2. Unsupervised Learning	2
3.3. Parsimony	2
3.4. Optimization	2
<b>4. New Software and Platforms</b>	<b>3</b>
4.1. SAG	3
4.2. fMRI	3
<b>5. New Results</b>	<b>3</b>
5.1. An Optimal Affine Invariant Smooth Minimization Algorithm	3
5.2. SAGA: A Fast Incremental Gradient Method With Support for Non-Strongly Convex Composite Objectives	3
5.3. Non-parametric Stochastic Approximation with Large Step sizes	4
5.4. Adaptivity of averaged stochastic gradient descent to local strong convexity for logistic regression	4
5.5. Serialrank: Spectral Ranking using Seriation	4
5.6. Sequential Kernel Herding: Frank-Wolfe Optimization for Particle Filtering	4
5.7. Learning to Learn for Structured Sparsity	5
5.8. Analysis of purely random forests bias	5
5.9. Large-Margin Metric Learning for Constrained Partitioning Problems	5
5.10. Metric Learning for Aligning temporal sequences	5
5.11. Weakly Supervised Action Labeling in Videos Under Ordering Constraints	6
5.12. On Pairwise Cost for Multi-Object Network Flow Tracking	6
5.13. A Markovian approach to distributional semantics with application to semantic compositionality	6
5.14. A convex relaxation for weakly supervised relation extraction	6
5.15. Weakly supervised named entity classification	7
5.16. Fast imbalanced binary classification: a moment-based approach	7
<b>6. Bilateral Contracts and Grants with Industry</b>	<b>7</b>
6.1. Bilateral Contracts with Industry	7
6.2. Bilateral Grants with Industry	7
<b>7. Partnerships and Cooperations</b>	<b>8</b>
7.1. National Initiatives	8
7.1.1. ANR: Calibration	8
7.1.2. CNRS: Gargantua	8
7.2. European Initiatives	8
7.2.1.1. SIERRA	8
7.2.1.2. SIPA	9
7.2.1.3. SpaRTaN	9
7.3. International Initiatives	10
7.3.1. Inria Associate Teams	10
7.3.2. Inria International Partners	10
7.4. International Research Visitors	10
<b>8. Dissemination</b>	<b>10</b>
8.1. Scientific Animations	10
8.1.1. Editorial boards	10
8.1.2. Area chair	10

8.1.3.	Workshop and conference organization	10
8.1.4.	Awards	11
8.1.5.	Other	11
8.1.6.	Invited presentations	11
8.2.	Teaching - Supervision - Juries	12
8.2.1.	Teaching	12
8.2.2.	Supervision	13
8.2.3.	Juries	13
8.3.	Popularization	13
<b>9.</b>	<b>Bibliography</b> .....	<b>13</b>

## Project-Team SIERRA

**Keywords:** Machine Learning, Statistics, Convex Optimization, Data Mining

*Creation of the Project-Team:* 2011 January 01.

### 1. Members

#### Research Scientists

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

#### Faculty Member

Vianney Perchet [Univ. Paris VII, Associate Professor, Delegation from Sep 2014]

#### PhD Students

Thomas Schatz [Univ. Paris VI]  
Jean-Baptiste Alayrac [Ecole Polytechnique, from Oct 2014]  
Aymeric Dieuleveut [ENS Paris]  
Christophe Dupuy [Technicolor]  
Nicolas Flammarion [ENS Lyon]  
Fajwel Fogel [CNRS]  
Damien Garreau [Inria]  
Senanayak Karri [Inria]  
Rémi Lajugie [Inria]  
Loic Landrieu [Ecole des Ponts]  
Anil Nelakanti [Xerox, until Jun 2014]  
Fabian Pedregosa [Inria Saclay]  
Anastasia Podosinnikova [Inria]  
Vincent Roulet [Ecole Polytechnique, from May 2014]

#### Post-Doctoral Fellows

Anton Osokin [Inria, from Oct 2014]  
Amit Bermanis [Bourse Chateaubriand, from Sep 2014]  
Nicolas Boumal [Research in Paris, from Oct 2014]  
Edouard Grave [Inria]  
Nino Shervashidze [Inria, until Sep 2014]

#### Visiting Scientists

Aaron Defazio [NICTA, Australia, from Apr 2014 until Jun 2014]  
Maxime Sangnier [CEA, from Nov 2014]  
Raman Sankaran [Indian Institute of Science, from May 2014]

#### Administrative Assistants

Kadidiatou Barry [Inria, from Sep 2014]  
Lindsay Polienor [Inria]

#### Others

Alexandre Defossez [ENS Paris, Masters Intern, from Feb 2014 until Jul 2014]  
Isabella Lukasevitz [Inria, Masters intern, from Apr 2014 until Jul 2014]  
Felipe Andres Yanez Lang [Inria, Intern, until Sep 2014]  
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. New Software and Platforms

### 4.1. SAG

**Participant:** Mark Schmidt [correspondent].

SAG: Minimizing Finite Sums with the Stochastic Average Gradient.

The SAG code contains C implements (via Matlab mex files) of the stochastic average gradient (SAG) method detailed below, as well as several related methods, for the problem of L2-regularized logistic regression with a finite training set.

The specific methods available in the package are: SGD: The stochastic gradient method with (user-supplied) step-sizes, (optional) projection step, and (optional) (weighted-)averaging. ASGD: A variant of the above code that supports less features, but efficiently implements uniform averaging on sparse data sets. PCD: A basic primal coordinate descent method with step sizes set according the (user-supplied) Lipschitz constants. DCA: A dual coordinate ascent method with a numerical high-accuracy line-search. SAG: The stochastic average gradient method with a (user-supplied) constant step size. SAGlineSearch: The stochastic average gradient method with the line-search described in the paper. SAG-LipschitzLS: The stochastic average gradient method with the line-search and adaptive non-uniform sampling strategy described in the paper.

### 4.2. fMRI

**Participant:** Fabian Pedregosa [correspondent].

We showed that HRF estimation improves sensitivity of fMRI encoding and decoding models and propose a new approach for the estimation of Hemodynamic Response Functions from fMRI data. This is an implementation of the methods described in the paper.

## 5. New Results

### 5.1. An Optimal Affine Invariant Smooth Minimization Algorithm

**Participant:** Alexandre d'Aspremont.

We formulate an affine invariant implementation of the algorithm in Nesterov (1983). We show that the complexity bound is then proportional to an affine invariant regularity constant defined with respect to the Minkowski gauge of the feasible set. We also detail matching lower bounds when the feasible set is an  $\ell_p$  ball. In this setting, our bounds on iteration complexity for the algorithm in Nesterov (1983) are thus optimal in terms of target precision, smoothness and problem dimension. (in collaboration with Cristóbal Guzmán, Martin Jaggi)

### 5.2. SAGA: A Fast Incremental Gradient Method With Support for Non-Strongly Convex Composite Objectives

**Participants:** Simon Lacoste-Julien, Francis Bach.

In this work we introduce a new optimisation method called SAGA in the spirit of SAG, SDCA, MISO and SVRG, a set of recently proposed incremental gradient algorithms with fast linear convergence rates. SAGA improves on the theory behind SAG and SVRG, with better theoretical convergence rates, and has support for composite objectives where a proximal operator is used on the regulariser. Unlike SDCA, SAGA supports non-strongly convex problems directly, and is adaptive to any inherent strong convexity of the problem. Moreover, the proof of the convergence bounds is much simpler than the one of our earlier work SAG. (in collaboration with A. Defazio, ANU)

### 5.3. Non-parametric Stochastic Approximation with Large Step sizes

**Participants:** Aymeric Dieuleveut, Francis Bach.

We consider the random-design least-squares regression problem within the reproducing kernel Hilbert space (RKHS) framework. Given a stream of independent and identically distributed input/output data, we aim to learn a regression function within an RKHS  $\mathcal{H}$ , even if the optimal predictor (i.e., the conditional expectation) is not in  $\mathcal{H}$ . In a stochastic approximation framework where the estimator is updated after each observation, we show that the averaged unregularized least-mean-square algorithm (a form of stochastic gradient), given a sufficient large step-size, attains optimal rates of convergence for a variety of regimes for the smoothnesses of the optimal prediction function and the functions in  $\mathcal{H}$ .

### 5.4. Adaptivity of averaged stochastic gradient descent to local strong convexity for logistic regression

**Participant:** Francis Bach.

In this work, we consider supervised learning problems such as logistic regression and study the stochastic gradient method with averaging, in the usual stochastic approximation setting where observations are used only once. We show that after  $N$  iterations, with a constant step-size proportional to  $1/R^2\sqrt{N}$  where  $N$  is the number of observations and  $R$  is the maximum norm of the observations, the convergence rate is always of order  $O(1/\sqrt{N})$ , and improves to  $O(R^2/\mu N)$  where  $\mu$  is the lowest eigenvalue of the Hessian at the global optimum (when this eigenvalue is greater than  $R^2/\sqrt{N}$ ). Since  $\mu$  does not need to be known in advance, this shows that averaged stochastic gradient is adaptive to *unknown local* strong convexity of the objective function. Our proof relies on the generalized self-concordance properties of the logistic loss and thus extends to all generalized linear models with uniformly bounded features.

### 5.5. Serialrank: Spectral Ranking using Seriation

**Participants:** Fajwel Fogel, Alexandre d'Aspremont.

We describe a seriation algorithm for ranking a set of  $n$  items given pairwise comparisons between these items. Intuitively, the algorithm assigns similar rankings to items that compare similarly with all others. It does so by constructing a similarity matrix from pairwise comparisons, using seriation methods to reorder this matrix and construct a ranking. We first show that this spectral seriation algorithm recovers the true ranking when all pairwise comparisons are observed and consistent with a total order. We then show that ranking reconstruction is still exact even when some pairwise comparisons are corrupted or missing, and that seriation based spectral ranking is more robust to noise than other scoring methods. An additional benefit of the seriation formulation is that it allows us to solve semi-supervised ranking problems. Experiments on both synthetic and real datasets demonstrate that seriation based spectral ranking achieves competitive and in some cases superior performance compared to classical ranking methods. (in collaboration with Milan Vojnovic, Microsoft Research).

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

**Participants:** Simon Lacoste-Julien, Francis Bach.

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, 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. (in collaboration with Fredrik Lindsten, Cambridge University)



## 5.7. Learning to Learn for Structured Sparsity

**Participants:** Nino Shervashidze, Francis Bach.

Structured sparsity has recently emerged in statistics, machine learning and signal processing as a promising paradigm for learning in high-dimensional settings. All existing methods for learning under the assumption of structured sparsity rely on prior knowledge on how to weight (or how to penalize) individual subsets of variables during the subset selection process, which is not available in general. Inferring group weights from data is a key open research problem in structured sparsity.

In this work, we propose a Bayesian approach to the problem of group weight learning. We model the group weights as hyperparameters of heavy-tailed priors on groups of variables and derive an approximate inference scheme to infer these hyperparameters. We empirically show that we are able to recover the model hyperparameters when the data are generated from the model, and we demonstrate the utility of learning weights in synthetic and real denoising problems.

## 5.8. Analysis of purely random forests bias

**Participant:** Sylvain Arlot.

Random forests are a very effective and commonly used statistical method, but their full theoretical analysis is still an open problem. As a first step, simplified models such as purely random forests have been introduced, in order to shed light on the good performance of random forests. In this paper, we study the approximation error (the bias) of some purely random forest models in a regression framework, focusing in particular on the influence of the number of trees in the forest. Under some regularity assumptions on the regression function, we show that the bias of an infinite forest decreases at a faster rate (with respect to the size of each tree) than a single tree. As a consequence, infinite forests attain a strictly better risk rate (with respect to the sample size) than single trees. Furthermore, our results allow to derive a minimum number of trees sufficient to reach the same rate as an infinite forest. As a by-product of our analysis, we also show a link between the bias of purely random forests and the bias of some kernel estimators. (In collaboration with Robin Genauer, Université de Bordeaux)

## 5.9. Large-Margin Metric Learning for Constrained Partitioning Problems

**Participants:** Rémi Lajugie, Sylvain Arlot, Francis Bach.

We consider unsupervised partitioning problems based explicitly or implicitly on the minimization of Euclidean distortions, such as clustering, image or video segmentation, and other change-point detection problems. We emphasize on cases with specific structure, which include many practical situations ranging from mean-based change-point detection to image segmentation problems. We aim at learning a Mahalanobis metric for these unsupervised problems, leading to feature weighting and/or selection. This is done in a supervised way by assuming the availability of several (partially) labeled datasets that share the same metric. We cast the metric learning problem as a large-margin structured prediction problem, with proper definition of regularizers and losses, leading to a convex optimization problem which can be solved efficiently. Our experiments show how learning the metric can significantly improve performance on bioinformatics, video or image segmentation problems.

## 5.10. Metric Learning for Aligning temporal sequences

**Participants:** Damien Garreau, Rémi Lajugie, Sylvain Arlot, Francis Bach.

In this work, we propose to learn a Mahalanobis distance to perform alignment of multivariate time series. The learning examples for this task are time series for which the true alignment is known. We cast the alignment problem as a structured prediction task, and propose realistic losses between alignments for which the optimization is tractable. We provide experiments on real data in the audio to audio context, where we show that the learning of a similarity measure leads to improvements in the performance of the alignment task. We also propose to use this metric learning framework to perform feature selection and, from basic audio features, build a combination of these with better performance for the alignment.

## 5.11. Weakly Supervised Action Labeling in Videos Under Ordering Constraints

**Participants:** Rémi Lajugie, Francis Bach.

We are given a set of video clips, each one annotated with an ordered list of actions, such as “walk” then “sit” then “answer phone” extracted from, for example, the associated text script. We seek to temporally localize the individual actions in each clip as well as to learn a discriminative classifier for each action. We formulate the problem as a weakly supervised temporal assignment with ordering constraints. Each video clip is divided into small time intervals and each time interval of each video clip is assigned one action label, while respecting the order in which the action labels appear in the given annotations. We show that the action label assignment can be determined together with learning a classifier for each action in a discriminative manner. We evaluate the proposed model on a new and challenging dataset of 937 video clips with a total of 787720 frames containing sequences of 16 different actions from 69 Hollywood movies. (in collaboration with Piotr Bojanowski, Ivan Laptev, Jean Ponce, Cordelia Schmid and Josef Sivic)

## 5.12. On Pairwise Cost for Multi-Object Network Flow Tracking

**Participant:** Simon Lacoste-Julien.

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 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. (in collaboration with Visesh Chari, Ivan Laptev, Josef Sivic).

## 5.13. A Markovian approach to distributional semantics with application to semantic compositionality

**Participants:** Edouard Grave, Francis Bach, Guillaume Obozinski.

In this work, we describe a new approach to distributional semantics. This approach relies on a generative model of sentences with latent variables, which takes the syntax into account by using syntactic dependency trees. Words are then represented as posterior distributions over those latent classes, and the model allows to naturally obtain in-context and out-of-context word representations, which are comparable. We train our model on a large corpus and demonstrate the compositionality capabilities of our approach on different datasets.

## 5.14. A convex relaxation for weakly supervised relation extraction

**Participant:** Edouard Grave.

A promising approach to relation extraction, called weak or distant supervision, exploits an existing database of facts as training data, by aligning it to an unlabeled collection of text documents. Using this approach, the task of relation extraction can easily be scaled to hundreds of different relationships. However, distant supervision leads to a challenging multiple instance, multiple label learning problem. Most of the proposed solutions to this problem are based on non-convex formulations, and are thus prone to local minima. In this article, we propose a new approach to the problem of weakly supervised relation extraction, based on discriminative clustering and leading to a convex formulation. We demonstrate that our approach outperforms state-of-the-art methods on a challenging dataset introduced in 2010.

## 5.15. Weakly supervised named entity classification

**Participant:** Edouard Grave.

In this paper, we describe a new method for the problem of named entity classification for specialized or technical domains, using distant supervision. Our approach relies on a simple observation: in some specialized domains, named entities are almost unambiguous. Thus, given a seed list of names of entities, it is cheap and easy to obtain positive examples from unlabeled texts using a simple string match. Those positive examples can then be used to train a named entity classifier, by using the PU learning paradigm, which is learning from positive and unlabeled examples. We introduce a new convex formulation to solve this problem, and apply our technique in order to extract named entities from financial reports corresponding to healthcare companies.

## 5.16. Fast imbalanced binary classification: a moment-based approach

**Participant:** Edouard Grave.

In this paper, we consider the problem of imbalanced binary classification in which the number of negative examples is much larger than the number of positive examples. The two mainstream methods to deal with such problems are to assign different weights to negative and positive points or to subsample points from the negative class. In this paper, we propose a different approach: we represent the negative class by the two first moments of its probability distribution (the mean and the covariance), while still modeling the positive class by individual examples. Therefore, our formulation does not depend on the number of negative examples, making it suitable to highly imbalanced problems and scalable to large datasets. We demonstrate empirically, on a protein classification task and a text classification task, that our approach achieves similar statistical performance than the two mainstream approaches to imbalanced classification problems, while being more computationally efficient. (in collaboration with Laurent El Ghaoui, U.C. Berkeley)

# 6. Bilateral Contracts and Grants with Industry

## 6.1. 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-Rocquencourt and Grenoble) and four MSR sites (Cambridge, New England, Redmond, New York).

Technicolor, CIFRE PhD student: "User profiling from unstructured data".

## 6.2. Bilateral Grants with Industry

A. d'Aspremont, AXA, "mécénat scientifique, chaire Havas-Dauphine", machine learning.

A. d'Aspremont, Société Générale - fondation ENS, "mécénat scientifique".

A. d'Aspremont, Scientific committee, Thales Alenia Space. Evaluation program in control, signal processing, etc.

A. d'Aspremont, Projet EMMA at Institut Louis Bachelier. Collaboration with Euroclear on REPO markets.

## 7. Partnerships and Cooperations

### 7.1. National Initiatives

#### 7.1.1. 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/>

#### 7.1.2. CNRS: *Gargantua*

**Participants:** Sylvain Arlot, Francis Bach, Simon Lacoste-Julien, Alexandre d'Aspremont.

Titre: Big data; apprentissage automatique et optimisation mathématique pour les données gigantesques

Coordinator: Laboratoire Jean Kuntzmann (UMR 5224)

Leader: Zaid Harchaoui

Other members: 13 members: S. Arlot, F. Bach, S. Lacoste-Julien, A. d'Aspremont and researchers from Laboratoire Jean Kuntzmann, Laboratoire d'Informatique de Grenoble (Université Joseph Fourier) and Laboratoire Paul Painlevé (Université Lille 1).

Instrument: défi MASTODONS du CNRS

Duration: May 2013-Dec 2014

Total funding: 60 000 euros for the two years

Webpage: <http://lear.inrialpes.fr/people/harchaoui/projects/gargantua/index.html>

### 7.2. European Initiatives

#### 7.2.1. FP7 & H2020 Projects

##### 7.2.1.1. SIERRA

Type: FP7

Defi: NC

Instrument: ERC Starting Grant

Duration: December 2009 - November 2014

Coordinator: F. Bach

Abstract: Machine learning is now a core part of many research domains, where the abundance of data has forced researchers to rely on automated processing of information. The main current paradigm of application of machine learning techniques consists in two sequential stages: in the representation phase, practitioners first build a large set of features and potential responses for model building or prediction. Then, in the learning phase, off-the-shelf algorithms are used to solve the appropriate data processing tasks. While this has led to significant advances in many domains, the potential of machine learning techniques is far from being reached.

### 7.2.1.2. SIPA

Type: FP7

Defi: NC

Instrument: ERC Starting Grant

Objectif: NC

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.

### 7.2.1.3. SpaRTaN

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

## 7.3. International Initiatives

### 7.3.1. Inria Associate Teams

#### 7.3.1.1. STATWEB

Title: Fast Statistical Analysis of Web Data via Sparse Learning

International Partner (Institution - Laboratory - Researcher):

University of California Berkeley (ÉTATS-UNIS)

Duration: 2011 - 2014

See also: <http://www.di.ens.fr/~fbach/statweb.html>

The goal of the proposed research is to provide web-based tools for the analysis and visualization of large corpora of text documents, with a focus on databases of news articles. We intend to use advanced algorithms, drawing from recent progresses in machine learning and statistics, to allow a user to quickly produce a short summary and associated timeline showing how a certain topic is described in news media. We are also interested in unsupervised learning techniques that allow a user to understand the difference between several different news sources, topics or documents.

### 7.3.2. Inria International Partners

#### 7.3.2.1. International Partners

IFCAM: Collaboration with Indian Institute of Science, Bangalore (Chiranjib Battacharya). 10000 Euros for visits from/to India.

## 7.4. International Research Visitors

### 7.4.1. Internships

Visit from Raman Sankaran Indian Institute of Science, Bangalore, May-July 2014.

## 8. Dissemination

### 8.1. Scientific Animations

#### 8.1.1. Editorial boards

A. d'Aspremont, Associate Editor, Optimization Methods & Software (2010-2014)

A. d'Aspremont, Associate Editor, SIAM Journal on Optimization (2013-...)

F. Bach, Journal of Machine Learning Research, Action Editor.

F. Bach, IEEE Transactions on Pattern Analysis and Machine Intelligence, Associate Editor.

F. Bach, Information and Inference, Associate Editor.

F. Bach, SIAM Journal on Imaging Sciences, Associate Editor.

F. Bach, International Journal of Computer Vision, Associate Editor

#### 8.1.2. Area chair

F. Bach, International Conference on Machine Learning, 2013

#### 8.1.3. Workshop and conference organization

A. d'Aspremont, Workshop preparation for les Houches in Jan 2015: Optimization and Statistical Learning, with Zaid Harchaoui, LEAR, Inria and LJK, Anatoli Juditsky, LJK, Université Joseph Fourier, Jérôme Malick, CNRS and LJK et Philippe Rigollet, ORFE, Princeton University, USA.

A. d'Aspremont, session organized at SIOPT 2014 in San Diego.

F. Bach, Workshop co-chair for NIPS 2014.

V.Perchet, Organizer, summer school "Ecole d'été pluridisciplinaire de Théorie des Jeux", Aussois (7-13 September 2014).

#### **8.1.4. Awards**

F. Bach, 10-year best paper award, ICML 2014.

S. Lacoste-Julien, MCMCSKi IV Honorable Mention Poster Prize (January 2014).

#### **8.1.5. Other**

A. d'Aspremont, Scientific committee, programme Gaspard Monge pour l'Optimisation.

#### **8.1.6. Invited presentations**

S. Arlot, "Kernel change-point detection", Workshop "Kernel methods for big data" (Université Lille 1, March 31 - April 2, 2014)

S. Arlot, "Optimal model selection with V-fold cross-validation: how should V be chosen?", Seventh International Conference on High Dimensional Probability (Institut d'Études Scientifiques de Cargèse, May 26-31, 2014).

A. d'Aspremont, "Spectral Ranking using Seriation." Journée big data, SPOC seminar, LIP6, Université de Paris VI.

A. d'Aspremont, "Spectral Ranking using Seriation." Workshop on Semidefinite Optimization, Approximation and Applications, Simons Institute, Berkeley, Sept. 2014.

A. d'Aspremont, "Spectral Ranking using Seriation." Cambridge statistics seminar, November 2014.

A. d'Aspremont, "Convex Relaxations for Permutation Problems." Oxford robotics seminar, March 2014

A. d'Aspremont, "Convex Relaxations for Permutation Problems." Department of Mathematical Engineering seminar, UCL, Louvain-la-Neuve, February 2014.

A. d'Aspremont, "Convex Relaxations for Permutation Problems." Colloque CNRS MASTODONS, January 2014.

A. d'Aspremont, "Convex Relaxations for Permutation Problems." Lunteren Conference on the Mathematics of Operations Research, January 2014.

A. d'Aspremont, "An Optimal Affine Invariant Smooth Minimization Algorithm." Lunteren Conference on the Mathematics of Operations Research, January 2014.

A. d'Aspremont, "Phase Recovery, MaxCut and Complex Semidefinite Programming." SLAC Photon Science Seminar, Stanford, March 2014.

A. d'Aspremont, "Optimisation et apprentissage." Colloquium Jacques Morgenstern, Inria Sophia-Antipolis, Avril 2014.

F. Bach, Invited talk at workshop on stochastic gradient methods, IPAM, Los Angeles (February 2014).

F. Bach, Seminar at University of Logano (March 2014).

F. Bach, Invited tutorial at Eurandom Workshop, Eindhoven (March 2014).

F. Bach, Invited talk at the Centre de Recerca Matemàtica (CRM), Barcelona (April 2014).

F. Bach, Seminars at Oxford University (May 2014).

F. Bach, Invited talk at "Journées de la SFDS", Rennes (June 2014).

F. Bach, Invited talk at the conference CAP, Saint-Etienne (July 2014).

F. Bach, Invited tutorial at the IFCAM Summer school, Bangalore, India (July 2014).

F. Bach, Invited talk at the Duke/UCL workshop, London (September 2014).

- F. Bach, Keynote talk at the conference ECML, Nancy (September 2014).
- F. Bach, Seminar at the University of Vienna, Austria (November 2014).
- F. Bach, Invited talk at the Swiss Statistics Seminar, Berne (November 2014).
- F. Bach, Invited talk at the CIFAR meeting on Neural Computation & Adaptive Perception, Montreal, Canada (December 2014).
- N. Boumal, Journée conjointe des GDR, ISIS and MIA: optimisation géométrique sur les variétés, Nov. 21, Paris.
- N. Boumal, Bordeaux University GIO 2014 international workshop, on the Geometry of Information and Optimization, Dec. 4 & 5, Bordeaux.
- F. Fogel, "Convex Relaxations for Permutation Problems", Journées MAS de Toulouse, August.
- F. Fogel, workshop "Algorithm and Data Science" at Microsoft Research Cambridge, May 15.
- S. Lacoste-Julien, "Frank-Wolfe optimization insights in machine learning", Tsinghua University, Beijing, China, June.
- S. Lacoste-Julien, "Sequential Kernel Herding: Frank-Wolfe Optimization for Particle Filtering", invited talk at Journées MAS 2014, Institut de Mathématiques de Toulouse, Toulouse, France, August.
- S. Lacoste-Julien, "Recent Advances in Frank-Wolfe Optimization", invited talk at the 4th IMA Conference on Numerical Linear Algebra and Optimisation, Birmingham, UK, September.
- S. Lacoste-Julien, "Sequential Kernel Herding: Frank-Wolfe Optimization for Particle Filtering", McGill University, Montreal, Canada, December.
- R. Lajugie, invitation at the workshop "Kernel methods for big data" in Lille (March, 31st -April, 2nd).
- V. Perchet, Workshop "Optimal Cooperation, Communication, and Learning in Decentralized Systems", Banff, Canada (October 16, 2014). Blackwell Approachability in Absorbing Games.
- V. Perchet, Conference "PGMO-COPI", Paris (October 28, 2014). Online multiclass classification via Blackwell approachability.
- V. Perchet, Conference "PGMO-COPI", Paris (October 30, 2014). New results in bandits problems with applications.
- V. Perchet, Séminaire probas/stats, Orsay (December 4, 2014). From Bandits to Ethical Clinical Trials. Optimal Sample Size for Multi-Stage Problems.
- V. Perchet, Conference "New Procedures for New Data", CIRM, Marseille (December 15, 2014). From Bandits to Ethical Clinical Trials. Optimal Sample Size for Multi-Stage Problems.
- N. Shervashidze, "Representing graphs for machine learning", Séminaire de Statistiques, Institut Henri Poincaré, Paris, June.

## 8.2. Teaching - Supervision - Juries

### 8.2.1. Teaching

- A. d'Aspremont, Creation of a Master's program MASH (Mathématiques, Apprentissage et Sciences Humaines), with ENS - Paris Dauphine. Started September 2014.
- Licence : A. d'Aspremont, L3 course on Optimization: ENSAE, 24h
- Mastère: A. d'Aspremont, course on Optimization: MVA, ENS Cachan, 18h. item Mastère (M2) : F. Bach, G. Obozinski, Introduction aux modèles graphiques (30h), Master MVA (ENS Cachan).
- Mastère: S. Arlot and F. Bach, "Statistical learning", 24h, Mastère M2, Université Paris-Sud, France.
- Mastère (M1): S. Lacoste-Julien, F. Bach, R. Lajugie: "Apprentissage statistique", 35h, Ecole Normale Supérieure, Filière "Math-Info", deuxième année.



### 8.2.2. Supervision

PhD in progress: Vincent Roulet, October 2014, A. d'Aspremont.

PhD in progress: Nicolas Flammarion, September 2014, A. d'Aspremont and F. Bach.

PhD in progress: Fajwel Fogel, September 2012, A. d'Aspremont and F. Bach.

PhD in progress: Rémi Lajugie, September 2012, S. Arlot and F. Bach.

PhD in progress: Damien Garreau, September 2014, S. Arlot (co-advised with G. Biau).

PhD in progress: Anastasia Podosinnikova, December 2013, F. Bach and S. Lacoste-Julien.

PhD in progress: Jean-Baptiste Alayrac, September 2014, S. Lacoste-Julien, Josef Sivic and Ivan Laptev.

PhD in progress: Aymeric Dieuleveut, September 2014, F. Bach.

PhD in progress: Christophe Dupuy, January 2014, F. Bach, co-advised with Christophe Diot (Technicolor).

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

PhD in progress: Fabian Pedregosa, September 2012, F. Bach, co-advised with Alexandre Gramfort (Telecom).

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

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

### 8.2.3. Juries

A. d'Aspremont, member of the PhD Committee for Pierre-André Savalle at Ecole Central de Paris on Oct. 21 2014.

A. d'Aspremont, member of the PhD Committee for Nicolas Boumal at université Catholique de Louvain, Belgium on Feb. 14 2014.

F. Bach, member of the PhD committee of Samuel Vaiter (Dauphine), Rajen Shah (Cambridge), Anthony Bourrier (Rennes)

F. Bach, member of the HDR committee of Josef Sivic and Sylvain Arlot.

## 8.3. Popularization

S. Lacoste-Julien, Lecture "Apprentissage statistique et big data", Colloque Algorithmique et Programmation, CIRM, Luminy, France, April.

S. Lacoste-Julien, General public talk to high school students having participated in the mathematics olympiad of académie de Versailles "Apprentissage statistique et big data" at Inria-Rocquencourt, Rocquencourt, France, June.

S. Lacoste-Julien, Demi-heure de la science "Apprentissage automatique et big data", at Inria-Rocquencourt, Rocquencourt, France, November.

S. Lacoste-Julien, July-Dec: helped Sydo for making a popularization video for Inria on the theme of "Simulation and machine learning".

## 9. Bibliography

### Publications of the year

#### Doctoral Dissertations and Habilitation Theses

- [1] S. ARLOT. *Contributions to statistical learning theory: estimator selection and change-point detection*, Université Paris Diderot, December 2014, Rapporteurs: Stéphane Boucheron, Arnak Dalalyan, Gabor Lugosi, Habilitation à diriger des recherches, <https://tel.archives-ouvertes.fr/tel-01094989>

### Articles in International Peer-Reviewed Journals

- [2] F. BACH. *Adaptivity of averaged stochastic gradient descent to local strong convexity for logistic regression*, in "Journal of Machine Learning Research", 2014, vol. 15, pp. 595-627, <https://hal.archives-ouvertes.fr/hal-00804431>
- [3] F. BACH. *Duality between subgradient and conditional gradient methods*, in "SIAM Journal of Optimization", January 2015, vol. 25, n<sup>o</sup> 1, pp. 115-129 [DOI : 10.1137/130941961], <https://hal.archives-ouvertes.fr/hal-00757696>
- [4] T. D. HOCKING, V. BOEVA, G. RIGAILL, G. SCHLEIERMACHER, I. JANOUÉIX-LEROSEY, O. DELATRE, W. RICHER, F. BOURDEAUT, M. SUGURO, M. SETO, F. BACH, J.-P. VERT. *SegAnnDB: interactive web-based genomic segmentation*, in "Bioinformatics", 2014, vol. 30, n<sup>o</sup> 11, pp. 1539-1546 [DOI : 10.1093/BIOINFORMATICS/BTU072], <https://hal-mines-paristech.archives-ouvertes.fr/hal-01101868>
- [5] A. OSOKIN, D. VETROV. *Submodular relaxation for inference in Markov random fields*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", November 2014, 14 p. [DOI : 10.1109/TPAMI.2014.2369046], <https://hal.archives-ouvertes.fr/hal-01103488>
- [6] 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>

### International Conferences with Proceedings

- [7] P. BOJANOWSKI, R. LAJUGIE, F. BACH, I. LAPTEV, J. PONCE, C. SCHMID, J. SIVIC. *Weakly Supervised Action Labeling in Videos Under Ordering Constraints*, in "ECCV - European Conference on Computer Vision", Zurich, Switzerland, September 2014, pp. 628-643 [DOI : 10.1007/978-3-319-10602-1\_41], <https://hal.inria.fr/hal-01053967>
- [8] V. BORGHESANI, F. PEDREGOSA, E. EGER, M. BUIATTI, M. PIAZZA. *A perceptual-to-conceptual gradient of word coding along the ventral path*, in "Pattern Recognition in Neuroimaging", Tubingen, Germany, IEEE, June 2014, <https://hal.inria.fr/hal-00986606>
- [9] A. DEFAZIO, F. BACH, S. LACOSTE-JULIEN. *SAGA: A Fast Incremental Gradient Method With Support for Non-Strongly Convex Composite Objectives*, in "Advances In Neural Information Processing Systems", Montreal, Canada, November 2014, <https://hal.archives-ouvertes.fr/hal-01016843>
- [10] F. FOGEL, A. D'ASPREMONT, M. VOJNOVIC. *Spectral Ranking using Seriation*, in "Neural Information Processing Systems (NIPS)", Montréal, Canada, December 2014, <https://hal.archives-ouvertes.fr/hal-01092658>
- [11] D. GARREAU, R. LAJUGIE, S. ARLOT, F. BACH. *Metric Learning for Temporal Sequence Alignment*, in "Advances in Neural Information Processing Systems 27 (NIPS 2014)", Montréal, Canada, December 2014, <https://hal.inria.fr/hal-01062130>
- [12] E. GRAVE. *A convex relaxation for weakly supervised relation extraction*, in "Conference on Empirical Methods in Natural Language Processing (EMNLP)", Doha, Qatar, ACL SIGDAT, October 2014, <https://hal.inria.fr/hal-01080310>

- [13] E. GRAVE, G. OBOZINSKI, F. BACH. *A Markovian approach to distributional semantics with application to semantic compositionality*, in "International Conference on Computational Linguistics (Coling)", Dublin, Ireland, International Committee on Computational Linguistics (ICCL), August 2014, pp. 1447 - 1456, <https://hal.inria.fr/hal-01080309>
- [14] R. LAJUGIE, S. ARLOT, F. BACH. *Large-Margin Metric Learning for Constrained Partitioning Problems*, in "Proceedings of The 31st International Conference on Machine Learning", Beijing, China, June 2014, <https://hal.inria.fr/hal-00796921>
- [15] L. LANDRIEU, G. OBOZINSKI. *Continuously indexed Potts models on unoriented graphs*, in "UAI 2014 - 30th Conference on Uncertainty in Artificial Intelligence", Quebec, Canada, July 2014, pp. 459-468, <https://hal-npc.archives-ouvertes.fr/hal-01060957>
- [16] M. SEIBERT, M. KLEINSTEUBER, R. GRIBONVAL, R. JENATTON, F. BACH. *On The Sample Complexity Of Sparse Dictionary Learning*, in "SSP 2014 - IEEE Workshop on Statistical Signal Processing", Jupiters, Gold Coast, Australia, IEEE, June 2014, This work was partially supported Cluster of Excellence CoTeSys funded by the German DFG. This work was partially supported by the EU FP7, SMALL project, FET-Open grant number 225913, <https://hal.inria.fr/hal-00990684>
- [17] K. VERVIER, P. MAHÉ, A. D'ASPREMONT, J.-B. VEYRIERAS, J.-P. VERT. *On learning matrices with orthogonal columns or disjoint supports*, in "European Conference, ECML PKDD", Nancy, France, T. CALDER (editor), Machine Learning and Knowledge Discovery in Databases, September 2014, vol. LNCS 8726, pp. 274-289 [DOI : 10.1007/978-3-662-44845-8\_18], <https://hal-mines-paristech.archives-ouvertes.fr/hal-01101875>

### Conferences without Proceedings

- [18] E. GRAVE. *Weakly supervised named entity classification*, in "Workshop on Automated Knowledge Base Construction (AKBC)", Montréal, Canada, December 2014, <https://hal.inria.fr/hal-01095596>
- [19] R. GRIBONVAL, R. JENATTON, F. BACH, M. KLEINSTEUBER, M. SEIBERT. *Apprentissage de dictionnaire pour les représentations parcimonieuses*, in "46e Journées de Statistique", Rennes, France, Société Française de Statistique, June 2014, This is an abstract of the full preprint available at <http://hal.inria.fr/hal-00918142/>, <https://hal.inria.fr/hal-01054122>

### Scientific Books (or Scientific Book chapters)

- [20] J. MAIRAL, F. BACH, J. PONCE. *Sparse Modeling for Image and Vision Processing*, Foundations and Trends in Computer Graphics and Vision, now publishers, December 2014, vol. 8, n<sup>o</sup> 2-3, 216 p. [DOI : 10.1561/9781680830095], <https://hal.inria.fr/hal-01081139>

### Research Reports

- [21] S. ARLOT, R. GENUER. *Analysis of purely random forests bias*, July 2014, <https://hal.archives-ouvertes.fr/hal-01023596>
- [22] F. BACH. *Breaking the Curse of Dimensionality with Convex Neural Networks*, Inria Paris, December 2014, <https://hal.archives-ouvertes.fr/hal-01098505>

- [23] A. DIEULEVEUT, F. BACH. *Non-parametric Stochastic Approximation with Large Step sizes*, August 2014, <https://hal.archives-ouvertes.fr/hal-01053831>

### Other Publications

- [24] S. ARLOT, M. LERASLE. *Why  $V=5$  is enough in  $V$ -fold cross-validation*, July 2014, <https://hal.archives-ouvertes.fr/hal-00743931>
- [25] A. BIETTI. *Online learning for audio clustering and segmentation*, ENS Cachan, September 2014, <https://hal.inria.fr/hal-01064672>
- [26] V. CHARI, S. LACOSTE-JULIEN, I. LAPTEV, J. SIVIC. *On Pairwise Cost for Multi-Object Network Flow Tracking*, August 2014, <https://hal.inria.fr/hal-01110678>
- [27] A. DÉFOSSEZ, F. BACH. *Constant Step Size Least-Mean-Square: Bias-Variance Trade-offs and Optimal Sampling Distributions*, November 2014, <https://hal.inria.fr/hal-01081578>
- [28] E. GRAVE, L. E. GHAOUI. *Fast imbalanced binary classification: a moment-based approach*, November 2014, <https://hal.inria.fr/hal-01087452>
- [29] R. GRIBONVAL, R. JENATTON, F. BACH. *Sparse and spurious: dictionary learning with noise and outliers*, September 2014, 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>, <https://hal.archives-ouvertes.fr/hal-01025503>
- [30] R. GRIBONVAL, R. JENATTON, F. BACH, M. KLEINSTEUBER, M. SEIBERT. *Sample Complexity of Dictionary Learning and other Matrix Factorizations*, November 2014, forthcoming, <https://hal.inria.fr/hal-00918142>
- [31] S. LACOSTE-JULIEN, F. LINDSTEN, F. BACH. *Sequential Kernel Herding: Frank-Wolfe Optimization for Particle Filtering*, 2014, 23 pages initial draft, <https://hal.archives-ouvertes.fr/hal-01099197>
- [32] F. PEDREGOSA, F. BACH, A. GRAMFORT. *On the Consistency of Ordinal Regression Methods*, June 2014, <https://hal.inria.fr/hal-01054942>
- [33] N. SHERVASHIDZE, F. BACH. *Learning to Learn for Structured Sparsity*, October 2014, <https://hal.inria.fr/hal-00986380>
- [34] K. VERVIER, P. MAHÉ, A. D'ASPREMONT, J.-B. VEYRIERAS, J.-P. VERT. *On learning matrices with orthogonal columns or disjoint supports*, April 2014, 16 pages, <https://hal.archives-ouvertes.fr/hal-00985654>
- [35] F. YANEZ, F. BACH. *Primal-Dual Algorithms for Non-negative Matrix Factorization with the Kullback-Leibler Divergence*, October 2014, <https://hal.archives-ouvertes.fr/hal-01079229>