



IN PARTNERSHIP WITH:
CNRS

**Ecole normale supérieure de
Paris**

Activity Report 2011

Project-Team SIERRA

Statistical Machine Learning and Parsimony

IN COLLABORATION WITH: Laboratoire d'Informatique de l'Ecole Normale Supérieure (LIENS)

RESEARCH CENTER
Paris - Rocquencourt

THEME
**Optimization, Learning and Statistical
Methods**

Table of contents

1. Members	1
2. Overall Objectives	1
2.1. Statement	1
2.2. Highlights	2
3. Scientific Foundations	2
3.1. Supervised Learning	2
3.2. Unsupervised Learning	2
3.3. Parsimony	2
3.4. Optimization	2
4. Application Domains	2
5. Software	3
5.1. inlinedocs	3
5.2. directlabels	3
5.3. clusterpath	3
5.4. UGM	3
5.5. alphaBeta	4
5.6. Matlab Software from “Graphical Model Structure Learning with L1-Regularization”	4
5.7. Hybrid deterministic-stochastic methods for data fitting	4
5.8. Multi-task regression using minimal penalties	5
6. New Results	5
6.1. Local Component Analysis	5
6.2. Weakly Supervised Learning of Foreground-Background Segmentation using Masked RBMs	5
6.3. Clusterpath: an algorithm for clustering using convex fusion penalties	6
6.4. Convex and Network Flow Optimization for Structured Sparsity	6
6.5. Multi-scale Mining of fMRI data with Hierarchical Structured Sparsity	6
6.6. Convergence Rates of Inexact Proximal-Gradient Methods for Convex Optimization	7
6.7. Generalized Fast Approximate Energy Minimization via Graph Cuts: Alpha-Expansion Beta-Shrink Moves	7
6.8. Hybrid Deterministic-Stochastic Methods for Data Fitting	7
6.9. Multi-task regression using minimal penalties	7
6.10. Ask the locals: multi-way local pooling for image recognition	7
6.11. Trace Lasso: a trace norm regularization for correlated designs	8
6.12. Shaping level sets through submodular functions	8
6.13. Itakura-Saito Nonnegative Matrix Factorization with group sparsity	8
6.14. Online algorithms for nonnegative matrix factorization with the Itakura-Saito divergence	9
6.15. A Graph-matching Kernel for Object Categorization	9
6.16. Group Lasso with Overlaps: the Latent Group Lasso approach	9
6.17. Sparse Image Representation with Epitomes	9
6.18. Dictionary Learning for Deblurring and Digital Zoom	10
6.19. Robust linear least squares regression	10
6.20. Semantic hierarchies for image annotation	10
6.21. Deviations of Stochastic Bandit Regret	11
6.22. Minimax Policies for Combinatorial Prediction Games	11
6.23. Non-Asymptotic Analysis of Stochastic Approximation Algorithms for Machine Learning	11
7. Partnerships and Cooperations	12
7.1. National Initiatives	12
7.2. European Initiatives	12
7.3. International Initiatives	13
8. Dissemination	13

8.1. Animation of the scientific community	13
8.1.1. Conference and workshop organization	13
8.1.2. Editorial boards	13
8.1.3. Area chairs	14
8.1.4. Reviewing	14
8.1.5. PhD and HDR thesis committee	14
8.1.6. Other	14
8.2. Prizes and awards	14
8.3. Invited presentations	14
8.4. Teaching	16
9. Bibliography	17

Project-Team SIERRA

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

1. Members

Research Scientists

Francis Bach [Team leader, Senior researcher “Détaché” at INRIA from the Corps des Mines, HdR]

Sylvain Arlot [Junior Researcher CNRS]

Jean-Yves Audibert [until October; Chercheur at the Centre d’Enseignement et de Recherche en Technologies de l’Information et Systèmes (CERTIS) of the École Nationale des Ponts et Chaussées (ENPC), HdR]

Guillaume Obozinski [Ingénieur Expert de Recherche]

Technical Staff

Jean-Paul Chieze [from September; Ingénieur SED]

PhD Students

Louise Benoît

Florent Couzinié-Devy

Edouard Grave

Toby Hocking

Rodolphe Jenatton [graduated on November 28]

Armand Joulin

Augustin Lefèvre

Anil Nelakanti [Cifre Ph.D. with Xerox]

Thomas Schatz

Matthieu Solnon

Post-Doctoral Fellows

Simon Lacoste-Julien [from September; financed by Mairie de Paris]

Ronny Luss [from September; financed by Associated Team STATWEB]

Nicolas Le Roux [financed by ERC grant (INRIA)]

Mark Schmidt [financed by ERC grant (INRIA)]

Administrative Assistant

Cécile Espiègle [until May]

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.

2.2. Highlights

The SIERRA project-team was created on January, 1 , 2011.

3. Scientific Foundations

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: objet recognition, object detection, image segmentation, image/video processing, computational photography.
- Bioinformatics: cancer diagnosis, protein function prediction, virtual screening.
- Text processing: document collection modeling, language models.
- Audio processing: source separation, speech/music processing.
- Neuro-imaging: brain-computer interface (fMRI, EEG, MEG).

5. Software

5.1. inlinedocs

Participant: Toby Hocking [correspondant].

Generates Rd files from R source code with comments, providing for quick, sustainable package development. The syntax keeps code and documentation close together, and is inspired by the Don't Repeat Yourself principle.

See also the web page <http://inlinedocs.r-forge.r-project.org/>.

- Version: 1.8
- Contact: toby.hocking@inria.fr

5.2. directlabels

Participant: Toby Hocking [correspondant].

The directlabels package provides an extensible framework for automatically placing direct labels onto multicolor lattice or ggplot2 plots. It includes heuristics for examining "lattice" and "ggplot" objects and inferring an appropriate Positioning Method for placing the labels. Furthermore, the design of directlabels makes it simple to create Positioning Methods for specific plots or libraries of portable Positioning Methods that can be re-used.

See also the web page <http://directlabels.r-forge.r-project.org/>.

- Version: 2.2
- Contact: toby.hocking@inria.fr

5.3. clusterpath

Participant: Toby Hocking [correspondant].

The clusterpath package provides an R/C++ implementation of the algorithms described in [16].

See also the web page <http://clusterpath.r-forge.r-project.org/>.

- Version: 1.0
- Contact: toby.hocking@inria.fr

5.4. UGM

Participant: Mark Schmidt [correspondant].

UGM is a set of Matlab functions implementing various tasks in probabilistic undirected graphical models of discrete data with pairwise (and unary) potentials. Specifically, it implements a variety of methods for the following four tasks:

- Decoding: Computing the most likely configuration.
- Inference: Computing the partition function and marginal probabilities.
- Sampling: Generating samples from the distribution.
- Parameter Estimation: Given data, computing maximum likelihood (or MAP) estimates of the parameters.

The first three tasks are implemented for arbitrary discrete undirected graphical models with pairwise potentials. The last task focuses on Markov random fields and conditional random fields with log-linear potentials. The code is written entirely in Matlab, although more efficient mex versions of some parts of the code are also available.

See also the web page <http://www.di.ens.fr/~mschmidt/Software/UGM.html>.

- Version: 2011
- Contact: mark.schmidt@inria.fr

5.5. alphaBeta

Participant: Mark Schmidt [correspondant].

The code contains implementations of several available methods for the problem of computing an approximate minimizer of the sum of a set of unary and pairwise real-valued functions over discrete variables. This equivalent to the problem of MAP estimation, also known as decoding, in a pairwise undirected graphical model. The code focuses on scenarios where the pairwise energies encourage neighboring variables to take the same state. The particular methods contained in the package are iterated conditional mode, alpha-beta swaps, alpha-expansions, and alpha-expansion beta-shrink moves.

See also the web page <http://www.di.ens.fr/~mschmidt/Software/alphaBeta.html>.

- Version: 1
- Contact: mark.schmidt@inria.fr

5.6. Matlab Software from “Graphical Model Structure Learning with L1-Regularization”

Participant: Mark Schmidt [correspondant].

This package contains the code used to produce the results in Mark Schmidt’s thesis: Roughly, there are five components corresponding to five of the thesis chapters:

- Chapter 2: L-BFGS methods for optimizing differentiable functions plus an L1-regularization term.
- Chapter 3: L-BFGS methods for optimizing differentiable functions with simple constraints or regularizers.
- Chapter 4: An L1-regularization method for learning dependency networks, and methods for structure learning in directed acyclic graphical models.
- Chapter 5: L1-regularization and group L1-regularization for learning undirected graphical models, using either the L2, L_{inf}, or nuclear norm of the groups.
- Chapter 6: Overlapping group L1-regularization for learning hierarchical log-linear models, and an active set method for searching through the space of higher-order groups.

See also the web page <http://www.di.ens.fr/~mschmidt/Software/thesis.html>.

- Version: 1
- Contact: mark.schmidt@inria.fr

5.7. Hybrid deterministic-stochastic methods for data fitting

Participant: Mark Schmidt [correspondant].

Many structured data-fitting applications require the solution of an optimization problem involving a sum over a potentially large number of measurements. Incremental gradient algorithms (both deterministic and randomized) offer inexpensive iterations by sampling only subsets of the terms in the sum. These methods can make great progress initially, but often slow as they approach a solution. In contrast, full gradient methods achieve steady convergence at the expense of evaluating the full objective and gradient on each iteration. We explore hybrid methods that exhibit the benefits of both approaches. Rate of convergence analysis and numerical experiments illustrate the potential for the approach.

See also the web page <http://www.cs.ubc.ca/~mpf/2011-hybrid-for-data-fitting.html>.

- Version: 1
- Contact: mark.schmidt@inria.fr
- Participants outside of Sierra: Michael Friedlander (Scientific Computing Laboratory, Department of Computer Science, University of British Columbia)

5.8. Multi-task regression using minimal penalties

Participant: Matthieu Solnon [correspondant].

This toolbox implements statistical algorithms designed to perform multi-task kernel ridge regressions, as described in [33].

See also the web page http://www.di.ens.fr/~solnon/articles/multi-task_regression/multitask_minpen_en.html.

- Version: 1
- Contact: matthieu.solnon@ens.fr

6. New Results

6.1. Local Component Analysis

Participants: Francis Bach, Nicolas Le Roux.

Kernel density estimation, a.k.a. Parzen windows, is a popular density estimation method, which can be used for outlier detection or clustering. With multivariate data, its performance is heavily reliant on the metric used within the kernel. Most earlier work has focused on learning only the bandwidth of the kernel (i.e., a scalar multiplicative factor). In this paper, we propose to learn a full Euclidean metric through an expectation-minimization (EM) procedure, which can be seen as an unsupervised counterpart to neighbourhood component analysis (NCA). In order to avoid overfitting with a fully nonparametric density estimator in high dimensions, we also consider a semi-parametric Gaussian-Parzen density model, where some of the variables are modelled through a jointly Gaussian density, while others are modelled through Parzen windows. For these two models, EM leads to simple closed-form updates based on matrix inversions and eigenvalue decompositions. We show empirically that our method leads to density estimators with higher test-likelihoods than natural competing methods, and that the metrics may be used within most unsupervised learning techniques that rely on such metrics, such as spectral clustering or manifold learning methods. Finally, we present a stochastic approximation scheme which allows for the use of this method in a large-scale setting [30].

6.2. Weakly Supervised Learning of Foreground-Background Segmentation using Masked RBMs

Participant: Nicolas Le Roux.

Collaboration with: Nicolas Heess (School of Informatics, University of Edinburgh) and John Winn (Machine Learning and Perception, Microsoft research Cambridge).

We propose an extension of the Restricted Boltzmann Machine (RBM) that allows the joint shape and appearance of foreground objects in cluttered images to be modeled independently of the background. We present a learning scheme that learns this representation directly from cluttered images with only very weak supervision. The model generates plausible samples and performs foreground-background segmentation. We demonstrate that representing foreground objects independently of the background can be beneficial in recognition tasks [15].

6.3. Clusterpath: an algorithm for clustering using convex fusion penalties

Participants: Toby Hocking, Francis Bach, Armand Joulin.

Collaboration with: Jean-Philippe Vert (INSERM U900, Mines ParisTech, Institut Curie).

We present a new clustering algorithm by proposing a convex relaxation of hierarchical clustering, which results in a family of objective functions with a natural geometric interpretation. We give efficient algorithms for calculating the continuous regularization path of solutions, and discuss relative advantages of the parameters. Our method experimentally gives state-of-the-art results similar to spectral clustering for non-convex clusters, and has the added benefit of learning a tree structure from the data [16].

6.4. Convex and Network Flow Optimization for Structured Sparsity

Participants: Rodolphe Jenatton, Guillaume Obozinski, Francis Bach.

Collaboration with: Julien Mairal (Department of Statistics, University of California, Berkeley).

In [5], we consider a class of learning problems regularized by a structured sparsity-inducing norm defined as the sum of ℓ_2 - or ℓ_∞ -norms over groups of variables. Whereas much effort has been put in developing fast optimization techniques when the groups are disjoint or embedded in a hierarchy, we address here the case of general overlapping groups. To this end, we present two different strategies: On the one hand, we show that the proximal operator associated with a sum of ℓ_∞ -norms can be computed exactly in polynomial time by solving a quadratic min-cost flow problem, allowing the use of accelerated proximal gradient methods. On the other hand, we use proximal splitting techniques, and address an equivalent formulation with non-overlapping groups, but in higher dimension and with additional constraints. We propose efficient and scalable algorithms exploiting these two strategies, which are significantly faster than alternative approaches. We illustrate these methods with several problems such as CUR matrix factorization, multi-task learning of tree-structured dictionaries, background subtraction in video sequences, image denoising with wavelets, and topographic dictionary learning of natural image patches.

6.5. Multi-scale Mining of fMRI data with Hierarchical Structured Sparsity

Participants: Rodolphe Jenatton, Guillaume Obozinski, Francis Bach.

Collaboration with: Alexandre Gramfort, Vincent Michel, Evelyn Eger and Bertrand Thirion (Laboratoire de Neuroimagerie Assistée par Ordinateur (LNAO), CEA: DSV/I2BM/NEUROSPIN, PARIETAL (INRIA Saclay - Ile de France) and Neuroimagerie cognitive, INSERM: U992 – Université Paris Sud – CEA).

Inverse inference, or "brain reading", is a recent paradigm for analyzing functional magnetic resonance imaging (fMRI) data, based on pattern recognition and statistical learning. By predicting some cognitive variables related to brain activation maps, this approach aims at decoding brain activity. Inverse inference takes into account the multivariate information between voxels and is currently the only way to assess how precisely some cognitive information is encoded by the activity of neural populations within the whole brain. However, it relies on a prediction function that is plagued by the curse of dimensionality, since there are far more features than samples, i.e., more voxels than fMRI volumes. To address this problem, different methods have been proposed, such as, among others, univariate feature selection, feature agglomeration and regularization techniques. In this paper, we consider a sparse hierarchical structured regularization. Specifically, the penalization we use is constructed from a tree that is obtained by spatially-constrained agglomerative clustering. This approach encodes the spatial structure of the data at different scales into the regularization, which makes the overall prediction procedure more robust to inter-subject variability. The regularization used induces the selection of spatially coherent predictive brain regions simultaneously at different scales. We test our algorithm on real data acquired to study the mental representation of objects, and we show that the proposed algorithm not only delineates meaningful brain regions but yields as well better prediction accuracy than reference methods [24], [17].

6.6. Convergence Rates of Inexact Proximal-Gradient Methods for Convex Optimization

Participants: Mark Schmidt, Nicolas Le Roux, Francis Bach.

In [21], we consider the problem of optimizing the sum of a smooth convex function and a non-smooth convex function using proximal-gradient methods, where an error is present in the calculation of the gradient of the smooth term or in the proximity operator with respect to the non-smooth term. We show that both the basic proximal-gradient method and the accelerated proximal-gradient method achieve the same convergence rate as in the error-free case, provided that the errors decrease at appropriate rates. Using these rates, we perform as well as or better than a carefully chosen fixed error level on a set of structured sparsity problems.

6.7. Generalized Fast Approximate Energy Minimization via Graph Cuts: Alpha-Expansion Beta-Shrink Moves

Participant: Mark Schmidt.

Collaboration with: Karteek Alahari (Willow project-team, INRIA Paris-Rocquencourt).

In [20], we present alpha-expansion beta-shrink moves, a simple generalization of the widely-used alpha-beta swap and alpha-expansion algorithms for approximate energy minimization. We show that in a certain sense, these moves dominate both alpha-beta-swap and alpha-expansion moves, but unlike previous generalizations the new moves require no additional assumptions and are still solvable in polynomial-time. We show promising experimental results with the new moves, which we believe could be used in any context where alpha-expansions are currently employed.

6.8. Hybrid Deterministic-Stochastic Methods for Data Fitting

Participant: Mark Schmidt.

Collaboration with: Michael P. Friedlander (University of British Columbia).

Many structured data-fitting applications require the solution of an optimization problem involving a sum over a potentially large number of measurements. Incremental gradient algorithms offer inexpensive iterations by sampling only subsets of the terms in the sum. These methods can make great progress initially, but often slow as they approach a solution. In contrast, full gradient methods achieve steady convergence at the expense of evaluating the full objective and gradient on each iteration. We explore hybrid methods that exhibit the benefits of both approaches. Rate of convergence analysis shows that by controlling the size of the subsets in an incremental gradient algorithm, it is possible to maintain the steady convergence rates of full gradient methods. We detail a practical quasi-Newton implementation based on this approach, and numerical experiments illustrate its potential benefits [29].

6.9. Multi-task regression using minimal penalties

Participants: Matthieu Solnon, Sylvain Arlot, Francis Bach.

In [33] we study the kernel multiple ridge regression framework, which we refer to as multi-task regression, using penalization techniques. The theoretical analysis of this problem shows that the key element appearing for an optimal calibration is the covariance matrix of the noise between the different tasks. We present a new algorithm to estimate this covariance matrix, based on the concept of minimal penalty, which was previously used in the single-task regression framework to estimate the variance of the noise. We show, in a non-asymptotic setting and under mild assumptions on the target function, that this estimator converges towards the covariance matrix. Then plugging this estimator into the corresponding ideal penalty leads to an oracle inequality. We illustrate the behavior of our algorithm on synthetic examples.

6.10. Ask the locals: multi-way local pooling for image recognition

Participants: Nicolas Le Roux, Francis Bach.

Collaboration with: Y-Lan Boureau and Jean Ponce (Willow project-team, INRIA Paris-Rocquencourt) and Yann LeCun (Courant Institute of Mathematical Science (CIMS), New York University).

Invariant representations in object recognition systems are generally obtained by pooling feature vectors over spatially local neighborhoods. But pooling is not local in the feature vector space, so that widely dissimilar features may be pooled together if they are in nearby locations. Recent approaches rely on sophisticated encoding methods and more specialized codebooks (or dictionaries), e.g., learned on subsets of descriptors which are close in feature space, to circumvent this problem. In this work, we argue that a common trait found in much recent work in image recognition or retrieval is that it leverages locality in feature space on top of purely spatial locality. We propose to apply this idea in its simplest form to an object recognition system based on the spatial pyramid framework, to increase the performance of small dictionaries with very little added engineering. State-of-the-art results on several object recognition benchmarks show the promise of this approach [12].

6.11. Trace Lasso: a trace norm regularization for correlated designs

Participants: Edouard Grave, Guillaume Obozinski, Francis Bach.

Using the ℓ_1 -norm to regularize the estimation of the parameter vector of a linear model leads to an unstable estimator when covariates are highly correlated. In this paper, we introduce a new penalty function which takes into account the correlation of the design matrix to stabilize the estimation. This norm, called the trace Lasso, uses the trace norm of the selected covariates, which is a convex surrogate of their rank, as the criterion of model complexity. We analyze the properties of our norm, describe an optimization algorithm based on reweighted least-squares, and illustrate the behavior of this norm on synthetic data, showing that it is more adapted to strong correlations than competing methods such as the elastic net [14].

6.12. Shaping level sets through submodular functions

Participant: Francis Bach.

The concept of parsimony is central in many scientific domains. In the context of statistics, signal processing or machine learning, it may take several forms. Classically, in a variable or feature selection problem, a sparse solution with many zeros is sought so that the model is either more interpretable, cheaper to use, or simply matches available prior knowledge. In this work, we instead consider sparsity-inducing regularization terms that will lead to solutions with many equal values. A classical example is the total variation in one or two dimensions, which leads to piecewise constant solutions and can be applied to various image labelling problems, or change point detection tasks. In this work [9], we follow our earlier approach which consisted in designing sparsity-inducing norms based on non-decreasing submodular functions, as a convex approximation to imposing a specific prior on the supports of the predictors. Here, we show that a similar parallel holds for some other class of submodular functions, namely non-negative setfunctions which are equal to zero for the full and empty set. Our main instance of such functions are symmetric submodular functions and we show that the Lovász extension may be seen as the convex envelope of a function that depends on level sets (i.e., the set of indices whose corresponding components of the underlying predictor are greater than a given constant). By selecting specific submodular functions, we give a new interpretation to known norms, such as the total variation; we also define new norms, in particular ones that are based on order statistics with application to clustering and outlier detection, and on noisy cuts in graphs with application to change point detection in the presence of outliers.

6.13. Itakura-Saito Nonnegative Matrix Factorization with group sparsity

Participants: Augustin Lefèvre, Francis Bach.

Collaboration with: Cédric Févotte (Laboratoire traitement et communication de l'information (LTCI), CNRS: UMR5141 – Institut Télécom – Télécom ParisTech).

In [18], we propose an unsupervised inference procedure for audio source separation. Components in nonnegative matrix factorization (NMF) are grouped automatically in audio sources via a penalized maximum likelihood approach. The penalty term we introduce favors sparsity at the group level, and is motivated by the assumption that the local amplitude of the sources are independent. Our algorithm extends multiplicative updates for NMF; moreover we propose a test statistic to tune hyperparameters in our model, and illustrate its adequacy on synthetic data. Results on real audio tracks show that our sparsity prior allows to identify audio sources without knowledge on their spectral properties.

6.14. Online algorithms for nonnegative matrix factorization with the Itakura-Saito divergence

Participants: Augustin Lefèvre, Francis Bach.

Collaboration with: Cédric Févotte (Laboratoire traitement et communication de l'information (LTCI), CNRS: UMR5141 – Institut Télécom – Télécom ParisTech).

Nonnegative matrix factorization (NMF) is now a common tool for audio source separation. When learning NMF on large audio databases, one major drawback is that the complexity in time is $O(FKN)$ when updating the dictionary (where $(F;N)$ is the dimension of the input power spectrograms, and K the number of basis spectra), thus forbidding its application on signals longer than an hour. We provide an online algorithm with a complexity of $O(FK)$ in time and memory for updates in the dictionary. We show on audio simulations that the online approach is faster for short audio signals and allows to analyze audio signals of several hours [31].

6.15. A Graph-matching Kernel for Object Categorization

Participant: Armand Joulin.

Collaboration with: Olivier Duchenne and Jean Ponce (Willow project-team, INRIA Paris-Rocquencourt).

In [13], we address the problem of category-level image classification. The underlying image model is a graph whose nodes correspond to a dense set of regions, and edges reflect the underlying grid structure of the image and act as springs to guarantee the geometric consistency of nearby regions during matching. A fast approximate algorithm for matching the graphs associated with two images is presented. This algorithm is used to construct a kernel appropriate for SVM-based image classification, and experiments with the Caltech 101, Caltech 256, and Scenes datasets demonstrate performance that matches or exceeds the state of the art for methods using a single type of features.

6.16. Group Lasso with Overlaps: the Latent Group Lasso approach

Participant: Guillaume Obozinski.

Collaboration with: Laurent Jacob (Department of Statistics, University of California at Berkeley) and Jean-Philippe Vert (INSERM U900, Mines ParisTech, Institut Curie).

We study in [25] a norm for structured sparsity which leads to sparse linear predictors whose supports are unions of predefined overlapping groups of variables. We call the obtained formulation latent group Lasso, since it is based on applying the usual group Lasso penalty on a set of latent variables. A detailed analysis of the norm and its properties is presented and we characterize conditions under which the set of groups associated with latent variables are correctly identified. We motivate and discuss the delicate choice of weights associated to each group, and illustrate this approach on simulated data and on the problem of breast cancer prognosis from gene expression data.

6.17. Sparse Image Representation with Epitomes

Participants: Louise Benoît, Francis Bach.

Collaboration with: Julien Mairal (Department of Statistics, University of California, Berkeley) and Jean Ponce (Willow project-team, INRIA Paris-Rocquencourt).

Sparse coding, which is the decomposition of a vector using only a few basis elements, is widely used in machine learning and image processing. The basis set, also called dictionary, is learned to adapt to specific data. This approach has proven to be very effective in many image processing tasks. Traditionally, the dictionary is an unstructured "flat" set of atoms. In this work, we study structured dictionaries which are obtained from an epitome, or a set of epitomes. The epitome is itself a small image, and the atoms are all the patches of a chosen size inside this image. This considerably reduces the number of parameters to learn and provides sparse image decompositions with shift invariance properties. We propose a new formulation and an algorithm for learning the structured dictionaries associated with epitomes, and illustrate their use in image denoising tasks. This work has resulted in a CVPR'11 publication [11].

6.18. Dictionary Learning for Deblurring and Digital Zoom

Participant: Florent Couzinié-Devy.

Collaboration with: Julien Mairal (Department of Statistics, University of California, Berkeley) and Jean Ponce (Willow project-team, INRIA Paris-Rocquencourt).

The paper [23] proposes a novel approach to image deblurring and digital zooming using sparse local models of image appearance. These models, where small image patches are represented as linear combinations of a few elements drawn from some large set (dictionary) of candidates, have proven well adapted to several image restoration tasks. A key to their success has been to learn dictionaries adapted to the reconstruction of small image patches [35]. In contrast, recent works have proposed instead to learn dictionaries which are not only adapted to data reconstruction, but also tuned for a specific task [34]. We introduce here such an approach to deblurring and digital zoom, using pairs of blurry/sharp (or low-/high-resolution) images for training, as well as an effective stochastic gradient algorithm for solving the corresponding optimization task. Although this learning problem is not convex, once the dictionaries have been learned, the sharp/high-resolution image can be recovered via convex optimization at test time. Experiments with synthetic and real data demonstrate the effectiveness of the proposed approach, leading to state-of-the-art performance for non-blind image deblurring and digital zoom.

6.19. Robust linear least squares regression

Participant: Jean-Yves Audibert.

Collaboration with: Olivier Catoni (École Normale Supérieure, CNRS and INRIA Paris-Rocquencourt, Classic project-team)

In [26], we consider the problem of robustly predicting as well as the best linear combination of d given functions in least squares regression, and variants of this problem including constraints on the parameters of the linear combination. For the ridge estimator and the ordinary least squares estimator, and their variants, we provide new risk bounds of order d/n without logarithmic factor unlike some standard results, where n is the size of the training data. We also provide a new estimator with better deviations in presence of heavy-tailed noise. It is based on truncating differences of losses in a min-max framework and satisfies a d/n risk bound both in expectation and in deviations. The key common surprising factor of these results is the absence of exponential moment condition on the output distribution while achieving exponential deviations. All risk bounds are obtained through a PAC-Bayesian analysis on truncated differences of losses. Experimental results strongly back up our truncated min-max estimator. This work is to appear in the Annals of Statistics in 2012.

6.20. Semantic hierarchies for image annotation

Participant: Jean-Yves Audibert.

Collaboration with: Anne-Marie Tousch (École des Ponts and ONERA) and Stéphane Herbin (ONERA)

In the survey [6], we argue that using structured vocabularies is capital to the success of image annotation. We analyze literature on image annotation uses and user needs, and we stress the need for automatic annotation. We briefly expose the difficulties posed to machines for this task and how it relates to controlled vocabularies. We survey contributions in the field showing how structures are introduced. First we present studies that use unstructured vocabulary, focusing on those introducing links between categories or between features. Then we review work using structured vocabularies as an input and we analyze how the structure is exploited.

6.21. Deviations of Stochastic Bandit Regret

Participant: Jean-Yves Audibert.

Collaboration with: Antoine Salomon (École des Ponts)

The work [19] studies the deviations of the regret in a stochastic multi-armed bandit problem. When the total number of plays n is known beforehand by the agent, previous works exhibit a policy such that with probability at least $1 - 1/n$, the regret of the policy is of order $\log n$. They have also shown that such a property is not shared by the popular UCB1 policy. This work first answers an open question: it extends this negative result to any anytime policy. The second contribution of this paper is to design anytime robust policies for specific multi-armed bandit problems in which some restrictions are put on the set of possible distributions of the different arms.

6.22. Minimax Policies for Combinatorial Prediction Games

Participant: Jean-Yves Audibert.

Collaboration with: Sébastien Bubeck (Centre de Recerca Matemàtica of Barcelona) and Gabor Lugosi (ICREA and Pompeu Fabra University)

In [8], we address the online linear optimization problem when the actions of the forecaster are represented by binary vectors. Our goal is to understand the magnitude of the minimax regret for the worst possible set of actions. We study the problem under three different assumptions for the feedback: full information, and the partial information models of the so-called "semi-bandit", and "bandit" problems. We consider both L_∞ -, and L_2 -type of restrictions for the losses assigned by the adversary. We formulate a general strategy using Bregman projections on top of a potential-based gradient descent, which generalizes the ones studied in numerous recent works. We provide simple proofs that recover most of the previous results. We propose new upper bounds for the semi-bandit game. Moreover we derive lower bounds for all three feedback assumptions. With the only exception of the bandit game, the upper and lower bounds are tight, up to a constant factor. Finally, we answer an open question raised by W. M. Koolen, M. K. Warmuth, and J. Kivinen by showing that the exponentially weighted average forecaster is suboptimal against L_∞ adversaries.

6.23. Non-Asymptotic Analysis of Stochastic Approximation Algorithms for Machine Learning

Participant: Francis Bach.

Collaboration with: Eric Moulines (Telecom ParisTech)

In [10], we consider the minimization of a convex objective function defined on a Hilbert space, which is only available through unbiased estimates of its gradients. This problem includes standard machine learning algorithms such as kernel logistic regression and least-squares regression, and is commonly referred to as a stochastic approximation problem in the operations research community. We provide a non-asymptotic analysis of the convergence of two well-known algorithms, stochastic gradient descent (a.k.a. Robbins-Monro algorithm) as well as a simple modification where iterates are averaged (a.k.a. Polyak-Ruppert averaging). Our analysis suggests that a learning rate proportional to the inverse of the number of iterations, while leading to the optimal convergence rate in the strongly convex case, is not robust to the lack of strong convexity or the setting of the proportionality constant. This situation is remedied when using slower decays together with averaging, robustly leading to the optimal rate of convergence. We illustrate our theoretical results with simulations on synthetic and standard datasets.

7. Partnerships and Cooperations

7.1. National Initiatives

7.1.1. ANR Projects

7.1.1.1. DETECT

Participants: Sylvain Arlot, Francis Bach.

Title: New statistical approaches to computer vision and bioinformatics

Coordinator: École Normale Supérieure (Paris)

Leader of the project: Sylvain Arlot

Other members: J. Sivic (Willow project-team, ENS), A. Celisse (University Lille 1), T. Mary-Huard (AgroParisTech), E. Roquain and F. Villers (University Paris 6).

Instrument: ANR “Young researchers” Program

Duration: Sep 2009 - Aug 2012

Total funding: 70000 Euros

See also: <http://www.di.ens.fr/~arlot/ANR-DETECT.htm>

Abstract: The DETECT project aims at providing new statistical approaches for detection problems in computer vision (in particular, detecting and recognizing human actions in videos) and bioinformatics (e.g., simultaneously segmenting CGH profiles). These problems are mainly of two different statistical nature: multiple change-point detection (i.e., partitioning a sequence of observations into homogeneous contiguous segments) and multiple tests (i.e., controlling a priori the number of false positives among a large number of tests run simultaneously).

7.2. European Initiatives

7.2.1. FP7 Projects

7.2.1.1. SIERRA

Title: SIERRA – Sparse structured methods for machine learning

Type: IDEAS

Instrument: ERC Starting Grant

Duration: December 2009 - November 2014

Coordinator: INRIA (France)

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

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: the tenet of this proposal is that to achieve the expected breakthroughs, this two-stage paradigm should be replaced by an integrated process where the specific *structure* of a problem is taken into account explicitly in the learning process. Considering such structure appropriately allows the consideration of massive numbers of features or potentially the on-demand construction of relevant features, in both numerically efficient and theoretically

understood ways. Thus, one could get the benefits of very large numbers of features—e.g., better predictive performance—in a reasonable running time.

This problem will be attacked through the tools of regularization by *sparsity-inducing norms*, that have recently led to theoretical and algorithmic advances, as well as practical successes, in unstructured domains. The scientific objective is thus to *marry structure with sparsity*: this is particularly challenging because structure may occur in various ways (discrete, continuous or mixed) and the targeted applications in computer vision and audio processing lead to large-scale convex optimization problems.

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

INRIA principal investigator: Francis Bach

International Partner:

Institution: University of California Berkeley (United States)

Laboratory: EECS and IEOR Departments

Duration: 2011 - 2013

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.

8. Dissemination

8.1. Animation of the scientific community

8.1.1. Conference and workshop organization

- F. Bach: Banff workshop on Sparse Statistics, Optimization and Machine Learning, January 2011. Co-organizer.
- N. Le Roux: NIPS Workshop on Deep Learning and Unsupervised Feature Learning, December 2011. Co-organizer.
- F. Bach, G. Obozinski: ICML 2011 workshop "Structured Sparsity: Learning and Inference Workshop", July 2nd 2011, Bellevue, Washington, USA. Co-organizers.
- F. Bach, G. Obozinski: NIPS 2011 workshop Sparse Representation and Low-rank approximation, December 16th 2011, Sierra Nevada, Spain. Co-organizers.

8.1.2. Editorial boards

- F. Bach: Journal of Machine Learning Research, Action Editor
- F. Bach: IEEE Transactions on Pattern Analysis and Machine Intelligence, Associate Editor
- F. Bach: SIAM Journal on Imaging Science, Associate Editor

8.1.3. Area chairs

- J.-Y. Audibert: Conference on Learning Theory (COLT), 2011
- J.-Y. Audibert: International Joint Conference on Artificial Intelligence (IJCAI), 2011
- J.-Y. Audibert: Algorithmic Learning Theory (ALT), 2011
- F. Bach: International Conference on Machine Learning, 2011, Area chair, Workshop co-chair
- F. Bach: International Conference on Computer Vision, 2011, Area chair
- G. Obozinski: Fifteenth International Conference on Artificial Intelligence and Statistics

8.1.4. Reviewing

- International journals: Annals of Statistics, Artificial Intelligence , Computational Statistics and Data Analysis (CSDA), Electronic Journal of Statistics (EJS), IEEE Transactions on Information theory, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), IEEE Transaction on Signal Processing, Journal of Computational and Graphical Statistics (JCGS), Journal of Machine Learning Research (JMLR), Journal of the Royal Statistical Society (JRSS), Machine Learning, Signal Processing Elsevier, Statistics and Computing (STCO)
- International conferences: AISTATS, ALT, ICML, ICPRAM, NIPS

8.1.5. PhD and HDR thesis committee

- Pierre Connault, University Paris-Sud, 2011 (S. Arlot)
- Rodolphe Jenatton, ENS Cachan, 2011 (J.-Y. Audibert, F. Bach, G. Obozinski)
- Jean-Baptiste Monnier, Université Paris 7, 2011 (J.-Y. Audibert, F. Bach)
- Novi Quadrianto, Australian National University (F. Bach)
- Gilles Meyer, Université de liège (F. Bach)

8.1.6. Other

- S. Arlot is member of the board for the entrance exam in École Normale Supérieure (mathematics, voie B/L)
- J.-Y. Audibert, F. Bach, Co-organizers of the biweekly seminar “Statistical Machine Learning in Paris” (<http://sites.google.com/site/smileinparis/home>)

8.2. Prizes and awards

- CNRS Prime d’excellence scientifique (S. Arlot)
- S. Arlot: Ph.D. thesis prize from SFDS (Prix Marie-Jeanne Laurent-Duhamel 2011)
- T. Hocking: Best Student Poster at useR 2011 in Warwick, England for “Adding direct labels to plots”
- M. Schmidt: NSERC postdoctoral fellowship

8.3. Invited presentations

- S. Arlot, “Sélection d’estimateurs avec des pénalités aléatoires”, lecture for the Marie-Jeanne Laurent-Duhamel prize, 43e Journées de Statistique (Société Française de Statistique), Tunis, 2011.
- S. Arlot, “Data-driven calibration of linear estimators with minimal penalties, with an application to multi-task regression”, Seminar, Statistics Laboratory, University of Cambridge, 2011.
- S. Arlot, “Pénalités minimales et sélection d’estimateurs optimale”, Seminar, Colloquium du MAP 5, Paris 5, 2011.

- S. Arlot, “Calibration automatique d’estimateurs linéaires à l’aide de pénalités minimales, application à la régression multi-tâches”, Seminar, IRMAR, Rennes, 2011.
- J.-Y. Audibert, “Aggregation and robust estimation”, Seminar, Université de Lille, 2011.
- J.-Y. Audibert, “Bandits Adversarial: Équilibres de Nash Approchés”, Journée “Jeux Matriciels et Jeux Partiellement Observables”, Paris, 2011.
- J.-Y. Audibert, “Introduction to Bandits: Algorithms and Theory” (in collaboration with Rémi Munos, INRIA Sequel), ICML tutorial, Bellevue, United States, 2011.
- F. Bach, Small Workshop on Sparse Dictionary Learning (invited talk) - Queen Mary University, London - January 2011
- F. Bach, Ecole Polytechnique Fédérale de Lausanne (seminar) - February 2011
- F. Bach, Statlearn workshop on Challenging problems in Statistical Learning (invited talk) - Grenoble - March 2011
- F. Bach, Learning Theory/State of the Art (invited talk) - Institut Henri Poincaré, Paris - May 2011
- F. Bach, Congrès SMAI 2011 (invited talk in special session) - Lorient - May 2011
- F. Bach, Université de Lille (seminar) - June 2011
- F. Bach, Workshop on Signal Processing with Adaptive Sparse Structured Representations (invited talk) - Edimburg - June 2011
- F. Bach, Conference on Foundations of Computational Mathematics (two invited talks) - Budapest - July 2011
- F. Bach, Université de Liège (seminar) - September 2011
- F. Bach, IMA Workshop on High Dimensional Phenomena (invited tutorial) - Minneapolis - September 2011
- F. Bach, University of California, Berkeley (two seminars) - October 2011
- F. Bach, Machine Learning for NeuroImaging Workshop (invited talk) - November 2011
- F. Bach, Rencontres de Statistique Mathématique (invited talk) - Luminy - December 2011
- F. Bach, NIPS workshop on Discrete Optimization (invited talk) - Granada - December 2011
- F. Bach, NIPS workshop on Deep Learning (invited talk) - Granada - December 2011
- T. Hocking, “Collaborative R package development using R-Forge” and “Sustainable R package development using documentation generation”, Seminar on bioinformatics, Institut de Biologie de Lille, 9 June 2011.
- G. Obozinski, “Structured Sparse Coding: Efficient algorithms and applications”, invited talk, BIRS workshop "Sparse Statistics, Optimization and Machine Learning", Banff, January 2011
- G. Obozinski, “Convex relaxation for Combinatorial Penalties”, invited talk, Mini-workshop on "Mathematical and Computational Foundations of Learning Theory", Dagstuhl, July 2011
- G. Obozinski, “Convex relaxation for Combinatorial Penalties”, invited talk, Mini-Workshop: "Mathematics of Machine Learning", Oberwolfach, August 2011
- G. Obozinski, “Introduction to Statistical Learning”, tutorial talk, GDR Mascot-Num session "Statistical learning for computer experiments", Institut Henri Poincaré, May 2011
- M. Schmidt, “Structure Learning in Undirected Graphical Models”, Invited lecture at INRA workshop on network inference, Toulouse, January 2011.
- M. Schmidt, “Hybrid Deterministic-Stochastic Methods for Data Fitting”, Invited talk at Xerox Research Centre, Grenoble, July 2011.
- M. Schmidt, “Convex Optimization”, Practical session at MLSS 2011, Bordeaux, September 2011.

- M. Schmidt, “Inexact Gradient and Proximal-Gradient Methods”, Invited lecture at EPFL, November 2011.
- M. Solnon, “Multi-task ridge regression using minimal penalties”, Séminaire de l’unité MIA-Jouy, Jouy en Josas, France, November 2011.

8.4. Teaching

Licence:

- S. Arlot, J.-Y. Audibert, F. Bach and G. Obozinski, “Statistical learning”, 30h, L3, École Normale Supérieure (Paris), France.

Master:

- S. Arlot, “Leçons de Mathématiques: Classification”, 9h, M1, École Normale Supérieure (Paris), France
- S. Arlot and Francis Bach, “Statistical learning”, 24h, M2, Université Paris-Sud, France.
- J.-Y. Audibert, “Machine Learning and applications”, 30h, M2, École des Ponts ParisTech and Université Paris-Est Marne-la-Vallée, France
- J.-Y. Audibert, “Prédiction séquentielle”, 14h, M2, Université Paris 7, France
- F. Bach and G. Obozinski, “Probabilistic Graphical Models”, 30h, M2, Mastère MVA, ENS Cachan, France.
- N. Le Roux, “Neural Networks and Optimization Methods”, 3h, M2, Mastère MVA, ENS Cachan, France.
- G. Obozinski, Enseignement spécialisé “Apprentissage Artificiel”, 3h, M1 (Graduate 1st year level), Mines de Paris, April 29th 2011.

Doctorat:

- S. Arlot, “Sélection de modèles et sélection d’estimateurs pour l’Apprentissage statistique”, 8h, Collège de France, France
- S. Arlot, “Model selection and estimator selection for statistical learning”, 10h, Scuola Normale Superiore di Pisa, Italy
- G. Obozinski, Summer school on Sparsity and Model Selection, 10h, Centro de Matematica Montevideo Uruguay, Montevideo, Uruguay, Feb. 28 - March 4, 2011.

PhD & HdR:

PhD: Rodolphe Jenatton, “Normes Parcimonieuses Structurées: Propriétés Statistiques et Algorithmiques avec Applications à l’Imagerie Cérébrale”, ENS Cachan, November 2011, J.-Y. Audibert and F. Bach.

PhD in progress:

- Louise Benoît, 2009, F. Bach and J. Ponce
- Florent Couzinie-Devy, 2010, F. Bach and J. Ponce
- Edouard Grave, 2010, F. Bach and G. Obozinski
- Toby Hocking, 2009, F. Bach and Jean-Philippe Vert (Ecole des Mines de Paris)
- Armand Joulin, 2009, F. Bach and J. Ponce
- Augustin Lefèvre, 2009, F. Bach and Cédric Févotte (Telecom Paristech)
- Anil Nelakanti, 2010, Cédric Archambeau (Xerox) and F. Bach
- Fabian Pedregosa, 2011, F. Bach and Alexandre Gramfort (INRIA Saclay)
- Matthieu Solnon, Multi-task learning, September 2010, S. Arlot and F. Bach

9. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

- [1] R. JENATTON. *Normes Parcimonieuses Structurées: Propriétés Statistiques et Algorithmiques avec Applications à l'Imagerie Cérébrale*, Ecole Normale Supérieure de Cachan, 2011.

Articles in International Peer-Reviewed Journal

- [2] F. BACH, R. JENATTON, J. MAIRAL, G. OBOZINSKI. *Optimization with Sparsity-Inducing Penalties*, in "Foundations and Trends in Machine Learning", 2011, <http://hal.inria.fr/hal-00613125/en>.
- [3] R. JENATTON, J.-Y. AUDIBERT, F. BACH. *Structured Variable Selection with Sparsity-Inducing Norms*, in "Journal of Machine Learning Research", 2011, vol. 12, p. 2777-2824, <http://hal.inria.fr/inria-00377732/en>.
- [4] R. JENATTON, J. MAIRAL, G. OBOZINSKI, F. BACH. *Proximal Methods for Hierarchical Sparse Coding*, in "Journal of Machine Learning Research", July 2011, n^o 12, p. 2297-2334, <http://hal.inria.fr/inria-00516723/en>.
- [5] J. MAIRAL, R. JENATTON, G. OBOZINSKI, F. BACH. *Convex and Network Flow Optimization for Structured Sparsity*, in "Journal of Machine Learning Research", September 2011, vol. 12, p. 2681–2720, <http://hal.inria.fr/inria-00584817/en>.
- [6] A.-M. TOUSCH, S. HERBIN, J.-Y. AUDIBERT. *Semantic hierarchies for image annotation: A survey*, in "Pattern Recognition", 2011, vol. 45, n^o 1, p. Pages 333-345, 13 pages [DOI : 10.1016/J.PATCOG.2011.05.017], <http://hal.inria.fr/hal-00624460/en>.

Invited Conferences

- [7] J. MAIRAL, R. JENATTON, G. OBOZINSKI, F. BACH. *Learning Hierarchical and Topographic Dictionaries with Structured Sparsity*, in "SPIE Wavelets and Sparsity XIV", San Diego, United States, M. PAPADAKIS, D. VAN DE VILLE, V. K. GOYAL (editors), Proceedings of SPIE, SPIE, September 2011, vol. 8138 [DOI : 10.1117/12.893811], <http://hal.inria.fr/inria-00633983/en>.

International Conferences with Proceedings

- [8] J.-Y. AUDIBERT, S. BUBECK, G. LUGOSI. *Minimax Policies for Combinatorial Prediction Games*, in "COLT 2011 - Conference on Learning Theory", Budapest, Hungary, 2011, 23 pages, <http://hal.inria.fr/hal-00624463/en>.
- [9] F. BACH. *Shaping Level Sets with Submodular Functions*, in "Neural Information Processing Systems (NIPS)", Spain, 2011, <http://hal.inria.fr/hal-00542949/en>.
- [10] F. BACH, E. MOULINES. *Non-Asymptotic Analysis of Stochastic Approximation Algorithms for Machine Learning*, in "Neural Information Processing Systems (NIPS)", Spain, 2011, <http://hal.inria.fr/hal-00608041/en>.

- [11] L. BENOÎT, J. MAIRAL, F. BACH, J. PONCE. *Sparse Image Representation with Epitomes*, in "Computer Vision and Pattern Recognition", Colorado Springs, United States, June 2011, p. 2913 - 2920 [DOI : 10.1109/CVPR.2011.5995636], <http://hal.inria.fr/hal-00631652/en>.
- [12] Y.-L. BOUREAU, N. LE ROUX, F. BACH, J. PONCE, Y. LECUN. *Ask the locals: multi-way local pooling for image recognition*, in "13th International Conference on Computer Vision", Barcelone, Spain, 2011, <http://hal.inria.fr/hal-00646816/en>.
- [13] O. DUCHENNE, A. JOULIN, J. PONCE. *A Graph-matching Kernel for Object Categorization*, in "ICCV 2011 - 13th International Conference on Computer Vision", Barcelona, Spain, November 2011, 1056, <http://hal.inria.fr/hal-00650345/en>.
- [14] E. GRAVE, G. OBOZINSKI, F. BACH. *Trace Lasso: a trace norm regularization for correlated designs*, in "Neural Information Processing Systems (NIPS)", Spain, 2011, <http://hal.inria.fr/hal-00620197/en>.
- [15] N. HEES, N. LE ROUX, J. WINN. *Weakly Supervised Learning of Foreground-Background Segmentation using Masked RBMs*, in "ICANN 2011 - International Conference on Artificial Neural Networks", Espoo, Finland, June 2011, <http://hal.inria.fr/inria-00609681/en>.
- [16] T. D. HOCKING, A. JOULIN, F. BACH, J.-P. VERT. *Clusterpath An Algorithm for Clustering using Convex Fusion Penalties*, in "28th international conference on machine learning", United States, June 2011, 1, <http://hal.inria.fr/hal-00591630/en>.
- [17] R. JENATTON, A. GRAMFORT, V. MICHEL, G. OBOZINSKI, F. BACH, B. THIRION. *Multi-scale Mining of fMRI Data with Hierarchical Structured Sparsity*, in "PRNI 2011 - IEEE International Workshop on Pattern Recognition in NeuroImaging", Seoul, Korea, Republic Of, 2011, <http://hal.inria.fr/hal-00643901/en>.
- [18] A. LEFÈVRE, F. BACH, C. FÉVOTTE. *Itakura-Saito nonnegative matrix factorization with group sparsity*, in "36th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)", Prague, Czech Republic, 2011, <http://hal.inria.fr/hal-00567344/en>.
- [19] A. SALOMON, J.-Y. AUDIBERT. *Deviations of Stochastic Bandit Regret*, in "ALT'11 - Algorithmic Learning Theory", Espoo, Finland, 2011, <http://hal.inria.fr/hal-00624461/en>.
- [20] M. SCHMIDT, K. ALAHARI. *Generalized Fast Approximate Energy Minimization via Graph Cuts: Alpha-Expansion Beta-Shrink Moves*, in "UAI 2011 - 27th Conference on Uncertainty in Artificial Intelligence", Barcelona, Spain, July 2011, <http://hal.inria.fr/inria-00617524/en>.
- [21] M. SCHMIDT, N. LE ROUX, F. BACH. *Convergence Rates of Inexact Proximal-Gradient Methods for Convex Optimization*, in "NIPS'11 - 25 th Annual Conference on Neural Information Processing Systems", Grenada, Spain, December 2011, <http://hal.inria.fr/inria-00618152/en>.

Scientific Books (or Scientific Book chapters)

- [22] N. LE ROUX, Y. BENGIO, A. FITZGIBBON. *Improving First and Second-Order Methods by Modeling Uncertainty*, in "Optimization for Machine Learning", MIT PRESS (editor), Neural Information Processing, MIT Press, 2011, <http://hal.inria.fr/hal-00646214/en>.

Research Reports

- [23] F. COUZINIE-DEVY, J. MAIRAL, F. BACH, J. PONCE. *Dictionary Learning for Deblurring and Digital Zoom*, INRIA, September 2011, <http://hal.inria.fr/inria-00627402/en>.
- [24] R. JENATTON, A. GRAMFORT, V. MICHEL, G. OBOZINSKI, E. EGER, F. BACH, B. THIRION. *Multi-scale Mining of fMRI data with Hierarchical Structured Sparsity*, INRIA, May 2011, <http://hal.inria.fr/inria-00589785/en>.
- [25] G. OBOZINSKI, L. JACOB, J.-P. VERT. *Group Lasso with Overlaps: the Latent Group Lasso approach*, INRIA, October 2011, <http://hal.inria.fr/inria-00628498/en>.

Other Publications

- [26] J.-Y. AUDIBERT, O. CATONI. *Supplement to "Robust linear least squares regression"*, 2011, 19 pages, <http://hal.inria.fr/hal-00624459/en>.
- [27] F. BACH. *Learning with Submodular Functions: A Convex Optimization Perspective*, 2011, <http://hal.inria.fr/hal-00645271/en>.
- [28] F. BACH, R. JENATTON, J. MAIRAL, G. OBOZINSKI. *Structured sparsity through convex optimization*, 2011, <http://hal.inria.fr/hal-00621245/en>.
- [29] M. P. FRIEDLANDER, M. SCHMIDT. *Hybrid Deterministic-Stochastic Methods for Data Fitting*, 2011, 22 pages, <http://hal.inria.fr/inria-00626571/en>.
- [30] N. LE ROUX, F. BACH. *Local Component Analysis*, 2011, <http://hal.inria.fr/inria-00617965/en>.
- [31] A. LEFÈVRE, F. BACH, C. FÉVOTTE. *Online algorithms for Nonnegative Matrix Factorization with the Itakura-Saito divergence*, 2011, <http://hal.inria.fr/hal-00602050/en>.
- [32] A. SALOMON, J.-Y. AUDIBERT, I. EL ALAOU. *Regret lower bounds and extended Upper Confidence Bounds policies in stochastic multi-armed bandit problem*, 2011, <http://hal.inria.fr/hal-00652865/en>.
- [33] M. SOLNON, S. ARLOT, F. BACH. *Multi-task Regression using Minimal Penalties*, 2011, 33 pages, <http://hal.inria.fr/hal-00610534/en>.

References in notes

- [34] J. MAIRAL, F. BACH, J. PONCE. *Task-Driven Dictionary Learning*, INRIA, September 2010, n^o RR-7400, preprint version (final version is published in IEEE Transactions on Pattern Analysis of Machine Intelligence), <http://hal.inria.fr/inria-00521534/en/>.
- [35] J. MAIRAL. *Représentations parcimonieuses en apprentissage statistique, traitement d'image et vision par ordinateur*, Ecole Normale Supérieure de Cachan, 2010, <http://tel.archives-ouvertes.fr/tel-00595312/fr/>.