2023
ACTIVITY REPORT

Project-Team

OCKHAM

Optimization, pHysical Knowledge, Algorithms and Models

IN COLLABORATION WITH: Laboratoire de l'Informatique du Parallélisme (LIP)

DOMAIN

Applied Mathematics, Computation and Simulation

THEME

Optimization, machine learning and statistical methods

Ínría

Contents

Project-Team OCKHAM

Creation of the Project-Team: 2023 March 01

Keywords

Computer sciences and digital sciences

A3.4.1. – Supervised learning

A3.4.4. – Optimization and learning

A3.4.6. – Neural networks

A3.4.7. – Kernel methods

A3.4.8. – Deep learning

A3.5. – Social networks

A3.5.1. – Analysis of large graphs

A5.3.2. – Sparse modeling and image representation

A5.9. – Signal processing

A5.9.4. – Signal processing over graphs

A5.9.5. – Sparsity-aware processing

A5.9.6. – Optimization tools

A6.3.1. – Inverse problems

A8.2. – Optimization

A8.6. – Information theory

A8.12. – Optimal transport

Other research topics and application domains

B2.6. – Biological and medical imaging

B6.6. – Embedded systems

B7.2.1. – Smart vehicles

B9.5.1. – Computer science

B9.5.2. – Mathematics

B9.5.6. – Data science

B9.10. – Privacy

1 Team members, visitors, external collaborators

Research Scientists

- Remi Gribonval [Team leader, INRIA, Senior Researcher, HDR]

- Paulo Goncalves [INRIA, Senior Researcher, HDR]

- Mathurin Massias [INRIA, Researcher]

- Titouan Vayer [INRIA, Researcher]

Faculty Members

- Marion Foare [CPE LYON, Associate Professor]

- Elisa Riccietti [ENS DE LYON, Chaire Inria]

Post-Doctoral Fellows

- Ayoub Belhadji [ENS DE LYON, Post-Doctoral Fellow, until Nov 2023]

- Etienne Lasalle [ENS DE LYON, Post-Doctoral Fellow]

- Rémi Vaudaine [ENS DE LYON, Post-Doctoral Fellow]

PhD Students

- Anne Gagneux [UNIV LYON I, from Oct 2023, also M2 intern from April to September]

- Antoine Gonon [ENS DE LYON]

- Clément Lalanne [ENS DE LYON, until Sep 2023]

- Guillaume Lauga [INRIA]

- Quoc-Tung Le [ENS DE LYON]

- Sibylle Marcotte [ENS PARIS]

- Can Pouliquen [ENS DE LYON]

- Léon Zheng [Valeo AI, CIFRE]

Technical Staff

- Pascal Carrivain [INRIA, Permanent SED Engineer]

- Badr Moufad [INRIA, Engineer]

Administrative Assistants

- Solène Audoux [Inria, until Aug 2023]

- Leslie Dussollier [INRIA, from Dec 2023]

Visiting Scientist

- Filippo Marini [UNIV BOLOGNE, from Mar 2023 until Jun 2023]

External Collaborator

- Márton Karsai [UNIV CEU, from Mar 2023, HDR]

2 Overall objectives

Building on a culture at the interface of signal modeling, mathematical optimization and statistical machine learning, the global objective of OCKHAM is to develop **computationally efficient and mathematically founded methods and models to process high-dimensional data**. Our ambition is to develop **frugal signal processing and machine learning methods** able to **exploit structured models**, intrinsically associated to **resource-efficient implementations**, and endowed with **solid statistical guarantees**.

Challenge 1: Developing frugal methods with robust expressivity. The idea of frugal approaches means algorithms relying on a controlled use of computing resources, but also methods whose expressivity and flexibility provably relies on the versatile notion of sparsity. This is expected to avoid the current pitfalls of costly over-parameterizations and to robustify the approaches with respect to adversarial examples and overfitting. More specifically, it is essential to contribute to the understanding of methods based on neural networks, in order to improve their performance and most of all, their efficiency in resource-limited environments.

Challenge 2: Integrating models in learning algorithms. To make statistical machine learning both more frugal and more interpretable, it is important to develop techniques able to exploit not only high-dimensional data but also models in various forms when available. When some partial knowledge is available about some phenomena related to the processed data, e.g. under the form of a physical model such as a partial differential equation, or as a graph capturing local or non-local correlations, the goal is to use this knowledge as an inspiration to adapt machine learning algorithms. The main challenge is to flexibly articulate a priori knowledge and data-driven information, in order to achieve a controlled extrapolation of predicted phenomena much beyond the particular type of data on which they were observed, and even in applications where training data is scarce.

Challenge 3: Guarantees on interpretability, explainability, and privacy. The notion of sparsity and its structured avatars –notably via graphs– is known to play a fundamental role in ensuring the identifiability of decompositions in latent spaces, for example for high-dimensional inverse problems in signal processing. The team's ambition is to deploy these ideas to ensure not only frugality but also some level of explainability of decisions and an interpretability of learned parameters, which is an important societal stake for the acceptability of "algorithmic decisions". Learning in small-dimensional latent spaces is also a way to spare computing resources and, by limiting the public exposure of data, it is expected to enable tunable and quantifiable tradeoffs between the utility of the developed methods and their ability to preserve privacy.

3 Research program

This project is resolutely at the interface of signal modeling, mathematical optimization and statistical machine learning, and concentrates on scientific objectives that are both ambitious –as they are difficult and subject to a strong international competition– and realistic thanks to the richness and complementarity of skills they mobilize in the team.

Sparsity constitutes a backbone for this project, not only as a target to ensure resource-efficiency and privacy, but also as prior knowledge to be exploited to ensure the identifiability of parameters and the interpretability of results. **Graphs** are its necessary *alter ego*, to flexibly model and exploit relations between variables, signals, and phenomena, whether these relations are known a priori or to be inferred from data. Lastly, **advanced large-scale optimization** is a key tool to handle in a statistically controlled and algorithmically efficient way the dynamic and incremental aspects of learning in varying environments.

The scientific activity of the project is articulated around the three axes described below. A common endeavor to these three axes consists in designing structured low-dimensional models, algorithms of bounded complexity to adjust these models to data through learning mechanisms, and a control of the performance of these algorithms to exploit these models on tasks ranging from low-level signal processing to the extraction of high-level information.

3.1   Axis 1: Sparsity for high-dimensional learning.

As now widely documented, the fact that a signal admits a sparse representation in some signal dictionary [57] is an enabling factor not only to address a variety of inverse problems with high-dimensional signals and images, such as denoising, deconvolution, or declipping, but also to speedup or decrease the cost of the acquisition of analog signals in certain scenarios compatible with compressive sensing [58, 51]. The flexibility of the models, which can incorporate learned dictionaries [75], as well as structured and/or low-rank variants of the now-classical sparse modeling paradigm [64], has been a key factor of the success of these approaches. Another important factor is the existence of algorithms of bounded complexity with provable performance, often associated to convex regularization and proximal strategies [47, 54], allowing to identify latent sparse signal representations from low-dimensional indirect observations.

While being now well-mastered (and in the core field of expertise of the team), these tools are typically constrained to relatively rigid settings where the unknown is described either as a sparse vector or a low-rank matrix or tensor in high (but finite) dimension. Moreover, the algorithms hardly scale to the dimensions needed to handle inverse problems arising from the discretization of physical models (e.g., for 3D wavefield reconstruction). A major challenge is to establish a comprehensive algorithmic and theoretical toolset to handle continuous notions of sparsity [52], which have been identified as a way to potentially circumvent these bottlenecks. The other main challenge is to extend the sparse modeling paradigm to resource-efficient and interpretable statistical machine learning. The methodological and conceptual output of this axis provides tools for Axes 2 and 3, which in return fuel the questions investigated in this axis.

- **1.1 Versatile and efficient sparse modeling.** The goal is to propose flexible and resource-efficient sparse models, possibly leveraging classical notions of dictionaries and structured factorization, but also the notion of sparsity in continuous domains (e.g. for sketched clustering, mixture model estimation, or image super-resolution), low-rank tensor representations, and neural networks with sparse connection patterns.
  Besides the empirical validation of these models and of the related algorithms on a diversity of targeted applications, the challenge is to determine conditions under which their success can be mathematically controlled, and to determine the fundamental tradeoffs between the expressivity of these models and their complexity.

- **1.2 Sparse optimization.** The main objectives are: a) to define cost functions and regularization penalties that integrate not only the targeted learning tasks, but also a priori knowledge, for example under the form of conservation laws or as relation graphs, cf Axis 2; b) to design efficient and scalable algorithms [4, 20] to optimize these cost functions in a controlled manner in a large-scale setting. To ensure the resource-efficiency of these algorithms, while avoiding pitfalls related to the discretization of high-dimensional problems (aka curse of dimensionality), we investigate the notion of "continuous" sparsity (i.e., with sparse measures), of hierarchies (along the ideas of multilevel methods), and of reduced precision (cf also Axis 3). The nonconvexity and non-smoothness of the problems are key challenges, and the exploitation of proximal algorithms and/or convexifications in the space of Borelian measures are privileged approaches.

- **1.3 Identifiability of latent sparse representations.** To provide solid guarantees on the interpretability of sparse models obtained via learning, one needs to ensure the identifiability of the latent variables associated to their parameters. This is particularly important when these parameters bear some meaning due to the underlying physics. Vice-versa, physical knowledge can guide the choice of which latent parameters to estimate. By leveraging the team's know-how obtained in the field of inverse problems, compressive sensing and source separation in signal processing, we aim at establishing theoretical guarantees on the uniqueness (modulo some equivalence classes to be

characterized) of the solutions of the considered optimization problems, on their stability in the presence of random or adversarial noise, and on the convergence and stability of the algorithms.

3.2   Axis 2: Learning on graphs and learning of graphs.

Graphs provide synthetic and sparse representations of the interactions between potentially high-dimensional data, whether in terms of proximity, statistical correlation, functional similarity, or simple affinities. One central task in this domain is how to infer such discrete structures, from the observations, in a way that best accounts for the ties between data, without becoming too complex due to spurious relationships. The graphical lasso [59] is among the most popular and successful algorithm to build a sparse representation of the relations between time series (observed at each node) and that unveils relevant patterns of the data. Recent works (e.g. [65]) strived to emphasize the clustered structure of the data by imposing spectral constraints to the Laplacian of the sought graphs, with the aim to improve the performance of spectral approaches to unsupervised classification. In this direction, several challenges remain, such as for instance the transposition of the framework to graph-based semi-supervised learning [1], where natural models are stochastic block models rather than strictly multi-component graphs (e.g. Gaussian mixtures models). As it is done in [80], the standard l_1-norm penalization term of graphical lasso could be questioned in this case. On another level, when low-rank (precision) matrices and / or when preservation of privacy are important stakes, one could be inspired by the sketching techniques developed in [62] and [53] to work out a *sketched graphical lasso*. There exists other situations where the graph is known a priori and does not need to be inferred from the data. This is for instance the case when the data naturally lie on a graph (e.g. social networks or geographical graphs) and so, one has to combine this data structure with the attributes (or measures) carried by the nodes or the edges of these graphs. Graph signal processing (GSP) [72] [9], which underwent methodological developments at a very rapid pace in recent years, is precisely an approach to jointly exploit algebraically these structures and attributes, either by filtering them, by re-organizing them, or by reducing them to principal components. However, as it tends to be more and more the case, data collection processes yield very large data sets with high dimensional graphs. In contrast to standard digital signal processing that relies on regular graph structures (cycle graph or cartesian grid) treating complex structured data in a global form is not an easily scalable task [60]. Hence, the notion of distributed GSP [55, 56] has naturally emerged. Yet, very little has been done on graph signals supported on dynamical graphs that undergo vertices/edges editions.

- **2.1 Learning of graphs.** When the graphical structure of the data is not known a priori, one needs to explore how to build it or to infer it. In the case of partially known graphs, this raises several questions in terms of relevance with respect to sparse learning. For example, a challenge is to determine which edges should be kept, whether they should be oriented, and how attributes on the graph could be taken into account (in particular when considering time-series on graphs) to better infer the nature and structure of the un-observed interactions. We strive to adapt known approaches such as the *graphical lasso* to estimate the covariance under a sparsity constraint (integrating also temporal priors), and investigate diffusion approaches to study the identifiability of the graphs. In connection with Axis 1.2, a particular challenge is to incorporate a priori knowledge coming from physical models that offer concise and interpretable descriptions of the data and their interactions.

- **2.2 Distributed and adaptive learning on graphs.** The availability of a known graph structure underlying training data offers many opportunities to develop distributed approaches, open perspectives where graph signal processing and machine learning can mutually fertilize each other.

  Some classifiers can be formalized as solutions of a constrained optimization problem, and an important objective is then to reduce their global complexity by developing distributed versions of these algorithms. Compared to costly centralized solutions, distributing the operations by restricting them to local node neighborhoods will enable solutions that are both more frugal and more privacy-friendly. In the case of dynamic graphs, the idea is to get inspiration from adaptive processing techniques to make the algorithms able to track the temporal evolution of data, either in terms of structural evolution or of temporal variations of the attributes. This aspect finds a natural continuation in the objectives of Axis 3.

3.3 Axis 3: Dynamic and frugal learning.

With the resurgence of neural networks approaches in machine learning, training times of the order of days, weeks, or even months are common. Mainstream research in deep learning somehow applies it to an increasingly large class of problems and uses the general wisdom to improve the models prediction accuracy by "stacking more layers", making the approach ever more resource-hungry. Underpinning theory on which resources are needed for a network architecture to achieve a given accuracy is still in its infancy. Efficient scaling of such techniques to massive sample sizes or dimensions in a resource-restricted environment remains a challenge and is a particularly active field of academic and industrial R&D, with recent interest in techniques such as sketching, dimension reduction, and approximate optimization.

A central challenge is to develop novel approximate techniques with reduced computational and memory imprint. For certain unsupervised learning tasks such as PCA, unsupervised clustering, or parametric density estimation, random features (e.g. random Fourier features [70]) allow to compute aggregated *sketches* guaranteed to preserve the information needed to learn, and no more: this has led to the *compressive learning* framework, which is endowed with statistical learning guarantees [62] as well as privacy preservation guarantees [53]. A sketch can be seen as an embedding of the empirical probability distribution of the dataset with a particular form of kernel mean embedding [73]. Yet, designing random features given a learning task remains something of an art, and a major challenge is to design provably good end-to-end sketching pipelines with controlled complexity for supervised classification, structured matrix factorization, and deep learning.

Another crucial direction is the use of dynamical learning methods, capable of exploiting wisely multiple representations at different scales of the problem at hand. For instance, many low and mixed-precision variants of gradient-based methods have been recently proposed [78, 77], which are however based on a static reduced precision policy, while a dynamic approach can lead to much improved energy-efficiency. Also, despite their massive success, gradient-based training methods still possess many weaknesses (low convergence rate, dependence on the tuning of the learning parameters, vanishing and exploding gradients) and the use of dynamical information promises to allow for the development of alternative methods, such as second-order or multilevel methods, which are as scalable as first-order methods but with faster convergence guarantees [71, 79].

The overall objective in this axis is to adapt in a controlled manner the information that is extracted from datasets or data streams and to dynamically use such information in learning, in order to optimize the tradeoffs between statistical significance, resource-efficiency, privacy-preservation and integration of a priori knowledge.

- **3.1 Compressive and privacy-preserving learning.** The goal is to compress training datasets as soon as possible in the processing workflow, before even starting to learn. In the spirit of compressive sensing, this is desirable not only to ensure the frugal use of ressources (memory and computation), but also to preserve privacy by limiting the diffusion of raw datasets and controlling the information that could actually be extracted from the targeted compressed representations, called *sketches*, obtained by well-chosen nonlinear random projections. We aim to build on a *compressive learning* framework developed by the team with the viewpoint that sketches provide an embedding of the data distribution, which should preserve some metrics, either associated to the specific learning task or to more generic optimal transport formulations. Besides ensuring the identifiability of the task-specific information from a sketch (cf Axis 1.3), an objective is to efficiently extract this information from a sketch, for example via algorithms related to avatars of continuous sparsity as studied in Axis 1.2. A particular challenge, connected with Axis 2.1 when inferring dynamic graphs from correlation of non-stationary times series, and with Axis 3.2 below, is to dynamically adapt the sketching mechanism to the analyzed data stream.

- **3.2 Sequential sparse learning.** Whether aiming at dynamically learning on data streams (cf. Axes 2.1 and 2.2), at integrating *a priori* physical knowledge when learning, or at ensuring domain adaptation for transfer learning, the objective is to achieve a statistically near-optimal update of a model from a sequence of observations whose content can also dynamically vary. When considering time-series on graphs, to preserve resource-efficiency and increase robustness, the algorithms further need to update the current models by dynamically integrating the data stream.

- **3.3 Dynamic-precision learning.** The goal is to propose new optimization algorithms to overcome the cost of solving large scale problems in learning, by dynamically adapting the precision of the data. The main idea is to exploit multiple representations at different scales of the problem at hand. We explore in particular two different directions to build the scales of problems: a) exploiting ideas coming from multilevel optimization to propose dynamical hierarchical approaches exploiting representations of the problem of progressively reduced dimension; b) leveraging the recent advances in hardware and the possibility of representing data at multiple precision levels provided by them. We aim at improving over state-of-the-art training strategies by investigating the design of scalable multilevel and mixed-precision second-order optimization and quantization methods, possibly derivative-free.

4 Application domains

The primary objectives of this project, which is rooted in Signal Processing and Machine Learning methodology, are to develop flexible methods, endowed with solid mathematical foundations and efficient algorithmic implementations, that can be adapted to numerous application domains. We are nevertheless convinced that such methods are best developed in strong and regular connection with concrete applications, which are not only necessary to validate the approaches but also to fuel the methodological investigations with relevant and fruitful ideas. The following application domains are primarily investigated in partnership with research groups with the relevant expertise.

4.1 Frugal AI on embedded devices

There is a strong need to drastically compress signal processing and machine learning models (typically, but not only, deep neural networks) to fit them on embedded devices. For example, on autonomous vehicles, due to strong constraints (reliability, energy consumption, production costs), the memory and computing resources of dedicated high-end image-analysis hardware are two orders of magnitude more limited than what is typically required to run state-of-the-art deep network models in real-time. The research conducted in the OCKHAM project finds direct applications in these areas, including: compressing deep neural networks to obtain low-bandwidth video-codecs that can run on smartphones with limited memory resources; sketched learning and sparse networks for autonomous vehicles; or sketching algorithms tailored to exploit optical processing units for energy efficient large-scale learning.

4.2 Imaging in physics and medicine

Many problems in imaging involve the reconstruction of large scale data from limited and noise-corrupted measurements. In this context, the research conducted in OCKHAM pays a special attention to modeling domain knowledge such as physical constraints or prior medical knowledge. This finds applications from physics to medical imaging, including: multiphase flow image characterization; near infrared polarization imaging in circumstellar imaging; compressive sensing for joint segmentation and high-resolution 3D MRI imaging; or graph signal processing for radio astronomy imaging with the Square Kilometer Array (SKA).

4.3 Interactions with computational social sciences

Based on collaborations with the relevant experts the team also regularly investigates applications in computational social science. For example, modeling infection disease epidemics requires efficient methods to reduce the complexity of large networked datasets while preserving the ability to feed effective and realistic data-driven models of spreading phenomena. In another area, estimating the vote transfer matrices between two elections is an ill-posed problem that requires the design of adapted regularization schemes together with the associated optimization algorithms.

5   Social and environmental responsibility

Machine learning methods achieve remarkable performance across various domains. However, the training of underlying models typically relies on significant computational resources, and consequently, energy resources. For the most high-performing models, these resources are far from negligible. Therefore, it becomes crucial to move towards more "frugality" and be capable of constructing learning models under resource constraints. We organized a workshop day titled "Frugality and Machine Learning", in partnership with IXXI, with the aim of discussing the feasibility of this objective from both technical and societal perspectives: how can we build models with minimal resources while maintaining good performance? Is it possible to surpass certain limits, such as those imposed by the rebound effect? This day brought together around fifty people, and the exchanges were fruitful. The various presentations enabled rich discussions that will contribute to reflections on the role of AI in society.

6   Highlights of the year

Our paper on conservation laws in deep neural training [23] was accepted as an oral presentation at the NeurIPS 2023 conference (67 orals were accepted out of 3218 accepted papers and 12343 submissions).

6.1   Awards

The 2023 IEEE SPS Sustained Impact Paper Award was granted to the paper "Performance Measurement in Blind Audio Source Separation", co-authored by Emmanuel Vincent, Rémi Gribonval, and Cédric Févotte, and published in the *IEEE Transactions on Audio, Speech, and Language Processing*, VOL. 14, NO. 4, JULY 2006 [76].

7   New software, platforms, open data

In an effort towards reproducible research, the default policy of the team is to release open-source code (typically python or matlab) associated to research papers that report experiments [24, 20, 23, 27]. When applicable and possible, more engineered software is developed and maintained over several years to provide more robust and consistent implementations of selected results.

7.1   New software

7.1.1   FAuST

Keywords: Matrix calculation, Multilayer sparse factorisation

Scientific Description: FAuST allows to approximate a given dense matrix by a product of sparse matrices, with considerable potential gains in terms of storage and speedup for matrix-vector multiplications.

Functional Description: FAUST is a C++ toolbox designed to decompose a given dense matrix into a product of sparse matrices in order to reduce its computational complexity (both for storage and manipulation).

Faust includes Matlab and Python wrappers and scripts to reproduce the experimental results of the following papers: - Le Magoarou L. and Gribonval R,. "Flexible multi-layer sparse approximations of matrices and applications", Journal of Selected Topics in Signal Processing, 2016. - Le Magoarou L., Gribonval R., Tremblay N. "Approximate fast graph Fourier transforms via multi-layer sparse", IEEE Transactions on Signal and Information Processing over Networks, 2018 - Quoc-Tung Le, Rémi Gribonval. Structured Support Exploration For Multilayer Sparse Matrix Factorization. ICASSP 2021 – IEEE International Conference on Acoustics, Speech and Signal Processing, Jun 2021, Toronto, Ontario, Canada. pp.1-5. - Sibylle Marcotte, Amélie Barbe, Rémi Gribonval, Titouan Vayer, Marc Sebban, et al.. Fast Multiscale Diffusion on Graphs. 2021.

Release Contributions: Faust 1.x contains Matlab routines to reproduce experiments of the PANAMA team on learned fast transforms.

Faust 2.x contains a C++ implementation with preliminary Matlab / Python wrappers.

Faust 3.x includes Python and Matlab wrappers around a C++ core with GPU acceleration, new algorithms.

URL: https://faust.inria.fr/

Publications: hal-03212764, hal-01416110, hal-01627434, hal-01167948, hal-01254108, tel-01412558, hal-01156478, hal-01104696, hal-01158057, hal-03132013

Contact: Remi Gribonval

Participants: Luc Le Magoarou, Nicolas Tremblay, Remi Gribonval, Nicolas Bellot, Adrien Leman, Hakim Hadj-djilani

7.1.2 skglm

Keywords: Optimization, Machine learning, Sparsity

Functional Description: skglm is a Python package that offers fast estimators for Generalized Linear Models (GLMs) that are compatible with scikit-learn. It is highly flexible and supports a wide range of GLMs. Its main feature is flexibility: you can implement virtually any estimator as a combination of datafit and penalty.

Thanks to this flexible design, skglm supports many missing models in scikit-learn while ensuring high performance. There are several reasons to opt for skglm:

- SUpport for many fast solvers able to tackle large datasets, either dense or sparse, with millions of features up to 100 times faster than scikit-learn - User-friendly API than enables composing custom estimators with any combination of existing datafits and penalties - Flexible design that makes it simple and easy to implement new datafits and penalties, a matter of few lines of code - Estimators fully compatible with the scikit-learn API and drop-in replacements of its GLM estimators

skglm is integrated into scikit-learn via the scikit-learn-contrib organization.

URL: https://contrib.scikit-learn.org/skglm/

Publication: hal-03819082

Contact: Mathurin Massias

Participants: Mathurin Massias, Badr Moufad

7.1.3 Benchopt

Keywords: Mathematical Optimization, Benchmarking, Reproducibility

Functional Description: BenchOpt is a package to simplify, make more transparent and more reproducible the comparisons of optimization algorithms. It is written in Python but it is available with many programming languages. So far it has been tested with Python, R, Julia and compiled binaries written in C/C++ available via a terminal command. If it can be installed via conda, it should just work!

BenchOpt is used through a simple command line and ultimately running and replicating an optimization benchmark should be as easy a cloning a repo and launching the computation with a single command line. For now, BenchOpt features benchmarks for around 10 convex optimization problems and we are working on expanding this to feature more complex optimization problems. We are also developing a website to display the benchmark results easily.

Release Contributions: https://github.com/benchopt/benchopt/releases/tag/1.5.1

Publication: [hal-03830604](#)

Contact: Thomas Moreau

Participants: Thomas Moreau, Alexandre Gramfort, Mathurin Massias, Badr Moufad

7.1.4 Celer

Keywords: Mathematical Optimization, Machine learning, Sparsity

Functional Description: celer is a Python package that solves Lasso-like problems and provides estimators that under the popular scikit-learn API. Thanks to a tailored implementation, celer provides a fast solver that tackles large-scale datasets with millions of features up to 100 times faster than scikit-learn. It handles Lasso, ElasticNet, Group Lasso, Multitask Lasso and Sparse Logistic regression, and comes with - automated parallel cross-validation - support of sparse and dense data - optional feature centering and normalization - unpenalized intercept fitting

celer also provides easy-to-use estimators as it is designed under the scikit-learn API.

URL: http://mathurinm.github.io/celer

Publications: [hal-02263500](#), [hal-01833398](#)

Contact: Mathurin Massias

Participants: Badr Moufad, Alexandre Gramfort

8 New results

8.1 Integrating Structured Models in Machine Learning and Signal Processing

8.1.1 Optimal Transport and Machine Learning on Graphs

Participants: Titouan Vayer.

Collaborations with Hugues Van Assel (PhD student, ENS Lyon), Cédric Vincent-Cuaz (post-doctoral researcher, EPFL), Rémi Flamary (CMAP, Ecole Polytechnique) and Nicolas Courty (IRISA, Université Bretagne Sud).

The Gromov-Wasserstein (GW) distance is derived from optimal transport (OT) theory. The interest of OT lies both in its ability to provide relationships, connections, between sets of points and distances between probability distributions. By modeling graphs as probability distributions GW has become an important tool in many ML tasks involving structured data. An interesting application case is that of the dimension reduction framework. It can be viewed as projecting, in the GW sense, a graph illustrating the relationships among data points in a high-dimensional space into a lower-dimensional space. Preliminary work has formally demonstrated these relationships and generalized them to define a new framework for dimension reduction, known as "Distributional Dimension Reduction," based on graphs and optimal transport [31].

8.1.2 Physics informed neural networks

Collaboration with Serge Gratton, Valentin Mercier (IRIT, Toulouse), Philippe Toint (U. Namur, Belgium), Stefania Bellavia (UNIFI, Italy), Hugo Passe (internship at UNIFI, Italy)

Physics informed neural networks (PINNs) are specialized network architectures designed for the solution of partial differential equations (PDEs) that take into account the underlying physics of the problem. We investigated their use both for direct and inverse problems involving PDEs.

In the context of the internship of Hugo Passe, we investigated their ability to deal with ill-posed inverse problems, focusing especially on parameter identification problems. We investigated the regularizing properties of PINNs and the use of regularising training procedures to correctly fit noisy data in such a context.

In the context of the Ph.D. work of Valentin Mercier, we focused on the direct problem. We studied the integration of a multigrid approach in their training, a large scale optimization problem involving complex solutions with multiple frequency components. The proposed training scheme ensures not only to reduce the training time, but also to improve the quality of the approximated solutions.

8.1.3 Bilevel and unrolled approaches for the learning of sparse covariance matrices

Participants: Can Pouliquen, Paulo Goncalves, Mathurin Massias, Titouan Vayer.

The PhD of Can Pouliquen is devoted to the dynamic inference of brain connectivity graphs for epileptic patients. We have adopted the mathematical framework of the Graphical Lasso, and pursue two directions. First, we have developed a bilevel optimization framework, that eases the tuning of individual correlation strengths in the Graphical Lasso penalty [28]. Second, we have introduced a new deep neural network architecture for sparse covariance matrix estimation, which guarantees a simultaneously sparse and positive definite output. This highly desirable property was so far a missing feature of existing architectures, and has many potential applications beyond neurosciences.

8.1.4 Precision Matrix Estimation with with Riemannian Optimization

Participants: Titouan Vayer.

Collaborations with Alexandre Hippert-Ferrer (MCF, LaSTIG), Florent Bouchard CR, CNRS), Ammar Mian (MCF, LISTIC) and Arnaud Breloy (PR, CNAM).

The estimation of precision matrices is a crucial problem that enables obtaining a compact representation, in the form of a graph, of complex data with interactions. Numerous optimization approaches aim to solve the underlying problem of Graphical Lasso (and its variants). In our work [18], we proposed a general Riemannian optimization framework for precision matrix estimation. The benefits of this framework are numerous: it allows solving robust variants of Graphical Lasso, its flexibility enables incorporating low-rank priors on covariances, and, finally, Riemannian optimization algorithms are particularly effective in solving the underlying optimization problems.

8.1.5 New penalties and proximal operators

Participants: Anne Gagneux, Remi Gribonval, Mathurin Massias.

Collaboration with Emmanuel Soubies (CNRS, IRIT, Toulouse).

During the internship of Anne Gagneux, we have studied the properties of sorted non convex penalties. Convex sorted penalties such as SLOPE are known to automatically cluster coefficients associated to correlated variables; non convex penalties on the other hand mitigate the well-known amplitude bias

of the L1 norm. Combining non-convexity with automatic grouping is therefore a promising venue. However the technical difficulties raised by such new penalties are many (non convexity, non smoothness, non Lipschitzianity). The goal of the internship was to compute the proximal operator of such penalties. We derived an algorithm based on the Pool Adjacent Violators Algorithm (PAVA) that is computes the exact proximal operator of these penalties in some cases (sorted MCP, sorted Log-sum), and are currently finalizing the case of sorted ℓ_q $(0 < q > 1)$ in order to publish this work.

8.1.6   Inverse problems for medical imaging

Participants: Marion Foare.

Collaboration with Luis Enrique Amador Arya Hélène Ratiney (Creatis, Villeurbanne), Hélène Ratiney (Creatis, Villeurbanne), Éric Van Reeth (Creatis, Villeurbanne), and Siemens Healthcare, Saint Denis

It is of particular interest in the field of medical imaging to quickly acquire low-resolution volumes (compromise between acquisition time, SNR and spatial resolution), and enhance their resolution as a post-processing step. The PhD of Luis Enrique Amador Araya aims at developing new techniques to build multi-constrasts super-resolution images for 3D Magnetic Resonance Imaging (MRI).

We propose to explore specialized piecewise smooth variational methods combining data fitting terms with geometric priors (e.g. the Discrete Mumford-Shah model) to build faithful super-resolution images. Preliminary work has been submitted to ISBI 2024.

8.2   Sparse deep neural networks : theory and algorithms

8.2.1   Mathematics of deep learning: rescaling invariances, generalization bounds, and conservation laws

Participants: Rémi Gribonval, Antoine Gonon, Elisa Riccietti, Sibylle Marcotte.

Collaborations with Nicolas Brisebarre (ARIC team, ENS de Lyon), with Gabriel Peyré (DMA, ENS, Paris), and with Yann Traonmilin (IMB, Bordeaux) and Samuel Vaiter (Laboratoire J. Dieudonné, Université Côte d'Azur, Nice)

Rescaling invariance in ReLU networks. Neural networks with the ReLU activation function are described by weights and bias parameters, and implemented into a piecewise linear continuous function. Natural *scalings* and permutations operations on the parameters leave the realization unchanged, leading to equivalence classes of parameters that yield the same realization.
Path-embedding and path-norm based generalization bounds. The path-embedding of parameters that we introduced last year [74] was invariant to such scalings but limited to strictly layered ReLU architectures. In the context of the PhD of Antoine Gonon, we extended it [36] to fully encompass general DAG ReLU networks with biases, skip connections and any operation based on the extraction of order statistics: max pooling, GroupSort etc. The norm of the resulting embedding is called a path-norm, and we established a general toolkit to obtain statistical generalization bounds for such modern neural networks. The resulting bounds are not only the most widely applicable path-norm based ones, but also recover or beat the sharpest known bounds of this type. These extended path-norms further enjoy the usual benefits of path-norms: ease of computation, invariance under the symmetries of the network, and improved sharpness on feed-forward networks compared to the product of operators' norms, another complex- ity measure most commonly used. The versatility of the toolkit and its ease of implementation allowed us to challenge the concrete promises of path-norm-based generalization bounds, by numerically evaluating the sharpest known bounds for ResNets on ImageNet.

Conservation laws. In the thesis of Sibylle Marcotte, the above path-embedding also served as a key enabler for the analysis of conservation laws in gradient descent dynamics of ReLU networks [23]. Understanding the geometric properties of gradient descent dynamics is a key ingredient in deciphering the recent success of very large machine learning models. A striking observation is that trained over-parameterized models retain some properties of the optimization initialization. This "implicit bias" is believed to be responsible for some favorable properties of the trained models and could explain their good generalization properties. The purpose of this work was threefold. First, we rigorously exposed the definition and basic properties of "conservation laws", which are maximal sets of independent quantities conserved during gradient flows of a given model (e.g. of a ReLU network with a given architecture) with any training data and any loss. Then we explained how to find the exact number of these quantities by performing finite-dimensional algebraic manipulations on the Lie algebra generated by the Jacobian of the model. Finally, we provided algorithms (implemented in SageMath) to: a) compute a family of polynomial laws; b) compute the number of (not necessarily polynomial) conservation laws. We provided showcase examples that we fully work out theoretically. Besides, applying the two algorithms confirmed for a number of ReLU network architectures that all known laws are recovered by the algorithm, and that there are no other laws. Such computational tools paved the way to understanding desirable properties of optimization initialization in large machine learning models. Current work involves exploring heir extension to flows with momentum and to general DAG ReLU network architectures, using the associated extended path-embedding.

8.2.2 Quantized networks: theory and algorithms

Participants: Rémi Gribonval, Elisa Riccietti, Antoine Gonon.

Collaborations with Nicolas Brisebarre (ARIC team, ENS de Lyon), with Silviu Filip and Paul Estano (IRISA, Rennes), and with Theo Mary (LIP6, Paris)

Quantization of neural networks: theory Motivated by the importance of quantizing networks besides pruning them to achieve sparsity, we studied the expressivity of *quantized* deep networks from an approximation theoretic perspective [12]. Our objective was to define and compare the corresponding approximation classes [7] with the unquantized ones. We also characterized the error of nearest-neighbour uniform quantization of ReLU networks and we investigated when ReLU networks can be expected, or not, to have better approximation properties than other classical approximation families.

Quantization of neural networks: algorithms From a more computational perspective, and as a first step towards a better understanding of nonlinear quantized networks, we studied the simpler linear case. Particularly, we investigated the problem of optimally quantizing low rank matrices by exploiting scaling invariances inherent to the optimization problem. We proposed [38, 27] an optimal solution algorithm with polynomial complexity in the dimension of the problem and exponential complexity in the number of bits. We showed that it provides much more accurate quantizations than the simple round to nearest strategy. Particularly we used this algorithm in combination with the hierarchical procedure in [68], to design an heuristic startegy to efficiently quantize the family of butterfly matrices, which very often occur in machine learning applications, for instance to sparsify dense neural networks. Our work may help to improve the compression rate in this context by coupling sparsification and quantization.

In order to further exploit the benefits of quantization in neural networks and the modern computer architectures, we studied the introduction of mixed precision in the training. Within the framework of the Ph.D. of Paul Estano, we studied stochastic gradient methods capable of exploiting multiple quantization levels. The proposed methods are supported by an error analysis, which suggests a good rule to switch among the available quantization levels, yielding a procedure that provides the same accuracy of classical training strategies but with a lower energy consumption.

8.2.3 Sparse regularization, unfolding, and approximation theory

Participants: Mathurin Massias.

Collaborations with Laura Thesing (Ludwig-Maximilians-Universität, Munich), and with Nelly Pustelnik (Physics lab, ENS de Lyon)

Unfolded or unrolled approaches consist in creating neural architectures inspired by the iterations of an optimization algorithm, in order to combine the expressivity of neural networks with an adequate inductive bias. Allowing end-to-end learning, they have become very popular, especially in imaging tasks. The theoretical results are scarcer: although generic fully connected neural networks are known to be universal approximators, it is still unclear what is gained, or lost, when using specific unrolled architectures. We are currently studying unrolled approach to solve the Lasso, for which the target function is piecewise affine, and which was one of the first applications of unrolled methods [61]. Several notions of unrolling are studied, in order to quantify the approximation speed of the underlying network classes.

In the PhD work of Hoang Trieu Vy Le, we investigated several unfolding strategies of standard proximal algorithms and their associated accelerated version in the context of image denoising, deconvolution. The goal was to study the impact of accelerated schemes on learning performance and robustness. Some preliminary works were also conducted to tackle the joint task of image restoration and edge detection.

8.2.4   Deep sparsity: from hardness to deformable butterfly algorithms

Participants: Rémi Gribonval, Elisa Ricietti, Pascal Carrivain, Léon Zheng, Quoc-Tung Le.

Collaboration with Patrick Perez and Gilles Puy (Valeo AI, Paris)

Matrix factorization with sparsity constraints plays an important role in many machine learning and signal processing problems such as dictionary learning, data visualization, dimension reduction.

In the context of the PhD thesis of Quoc-Tung Le [33] and Leon Zheng, building on our series of work on the hardness, tractability, and uniqueness properties of sparse matrix factorizations under various sparsity constraints [81, 67, 68] we extended our investigation into several directions.

First, we extended the tractable algorithm for so-called butterfly sparsity patterns (which somehow factorizes a given matrix essentially at the cost of a single matrix-vector multiplication, with exact recovery guarantees) to so-called *deformable butterlies* and studied its performance guarantees beyond the case of matrices admitting an exact factorization. This is the object of a paper to be submitted. The corresponding algorithm is being incorporated in the FAμST software library (see Section 7) and is subject to software optimizations to further speed it up. An optimized GPU implementation of deformable butterfly factors is notably on its way.

Second, the pitfalls that we had identified for certain sparse matrix factorization problems [68] were shown to also hold for certain sparse ReLU neural network training problems [22]. In particular, there exist settings where the optimization is necessarily unstable, in the sense that minimizing the loss function can only be achieved by letting some coefficients diverge to infinity.

Finally, we developed heuristics to handle butterfly approximations for matrices under unknown permutations of rows and/or columns [29]

8.3   Statistical learning, dimension reduction, and privacy preservation

8.3.1   Theoretical foundations of compressive learning: sketches, kernels, and optimal transport

Participants: Rémi Gribonval, Titouan Vayer, Paulo Goncalves, Ayoub Belhadji, Etienne Lassalle.

The compressive learning framework proposes to deal with the large scale of datasets by compressing them into a single vector of generalized random moments, called a *sketch*, from which the learning task is then performed. In past works we established statistical guarantees on the generalization error of this procedure, first in a general abstract setting illustrated on PCA [5], then for the specific case of compressive k-means and compressive Gaussian Mixture Modeling [63]. The overall framework is described in a tutorial paper from last year [6].

Theoretical guarantees in compressive learning fundamentally rely on comparing certain metrics between probability distributions. In [16] we established some conditions under which the Wasserstein distance can be controlled by Maximum Mean Discrepancy (MMD) norms, which are defined using reproducing kernel Hilbert spaces. Based on the relations between the MMD and optimal transport, we provided new guarantees for compressive statistical learning by introducing and studying the concept of Wasserstein learnability.

Compressive learning also exploits the ability to approximate certain kernels by finite dimensional quadratures. We revisited in [48] existing proofs of the Restricted Isometry Property of sketching operators with respect to certain mixtures models. We proposed an alternative analysis that circumvents the need to assume importance sampling when drawing random Fourier features to build random sketching operators. Our analysis is based on new deterministic bounds on the restricted isometry constant that depend solely on the set of frequencies used to define the sketching operator. This analysis opens the door to theoretical guarantees for structured sketching with frequencies associated to fast random linear operators [49].

Finally, by establishing a connection between graph learning and sketching methods, new results in [42] demonstrated how sketching techniques can be employed to estimate the precision matrix used in the Graphical Lasso algorithm. The central advantage lies in providing a graph estimation method with a limited amount of data compared to standard methods. Specifically, we theoretically demonstrated the feasibility of estimating such matrices with limited memory by employing a sketch based on (structured) rank-one measurements. Additionally, we proposed a quite effective reconstruction algorithm for the inverse problem based on the Graphical Lasso.

8.3.2   Practical exploration of sketching and methods with limited resources

Participants: Rémi Gribonval, Titouan Vayer, Ayoub Belhadji, Léon Zheng, Elisa Riccietti, Rémi Vaudaine.

Collaborations with Valeo AI, with Hughes Van Assel (UMPA, ENS de Lyon); with Marton Karsai (CEU, Vienne, Austria) and Pierre Borgnat (Physics Lab, ENS deLyon)

From a more empirical perspective, we pursued our efforts to make sketching for compressive learning and sketching more versatile and efficient. This notably involved investigating improved algorithms to learn from a sketch. In the context of compressive clustering, the standard heuristic is CL-OMPR, a variant of sliding Frank-Wolfe. We showed how this algorithm can fail to recover clusters even in advantageous scenarios, and showed how its deficiencies can be attributed to optimization difficulties related to the structure of a correlation function appearing at core steps of the algorithm. To address these limitations, we propose an alternative decoder offering substantial improvements over CL-OMPR. Its design was notably inspired from the mean shift algorithm, a classic approach to detect the local maxima of kernel density estimators. The proposed algorithm can extract clustering information from a sketch of the MNIST dataset that is 10 times smaller than previously. This work was submitted for a journal publication [35].

Sketching was also explored for temporal network compression [41]. In the context of temporal networks, which can model spreading processes such as epidemics, the out-component of a source node

is the set of nodes reachable from this node, and the distribution of the size of out-components is an important characteristics which computation can be demanding for large networks. We proposed both an exact online matrix algorithm with controlled complexity footprint to compute this distribution, and a sketching-based framework to estimate it from a highly compressed representation of the temporal network.

More generally, properties of kernels methods were also exploited in a more applicative context to reduce time and memory complexity: self-supervised learning of image representations. We introduced a regularization loss based on kernel mean embeddings with rotation-invariant kernels on the hypersphere, promoting the embedding distribution to be close to the uniform distribution on the hypersphere, with respect to the maximum mean discrepancy pseudometric [25]. Besides being fully competitive with the state of the art, our method significantly reduces the resources needed for training, making it implementable for very large embedding dimensions on existing devices and more easily adjustable than previous methods to settings with limited resources.

8.3.3 Dimension reduction and optimal transport

Participants: Titouan Vayer.

Collaborations with Hugues Van Assel (PhD student, ENS Lyon), Cédric Vincent-Cuaz (post-doctoral researcher, EPFL), Rémi Flamary (CMAP, Ecole Polytechnique), Nicolas Courty (IRISA, Université Bretagne Sud), Antoine Collas (postdoctoral researcher, MIND) and Arnaud Breloy (PR, CNAM).

Exploring and analyzing high-dimensional data is a core problem of data science that requires building low-dimensional and interpretable representations of the data through dimensionality reduction (DR). In a series of work we provide new methods an analysis for DR, inspired from optimal transport (OT).

A key requirement for dimensionality reduction is to incorporate global dependencies among original and embedded samples while preserving clusters in the embedding space. To achieve this, we combine in [17] the principles of OT and principal component analysis and provide a new simple linear DR method which seeks the best linear subspace that minimizes the reconstruction entropic OT error, which naturally encodes the neighborhood information of the samples.

In another more comprehensive study [24], we introduced and explored an innovative nonlinear dimension reduction method by utilizing the optimal transport framework and entropic affinities. Our research extends well-known techniques such as t-distributed stochastic neighbor embedding (t-SNE) and brings about numerous empirical and theoretical advantages. Notably, affinities in methods like t-SNE are inherently asymmetric and row-wise stochastic. Still, in DR approaches, they are commonly used following heuristic symmetrization. We unveil a new interpretation of these affinities as optimal transport problems, enabling a natural symmetrization that can be efficiently computed. The novel affinity matrix gains benefits from symmetric doubly stochastic normalization, enhancing clustering performance and effectively controlling the entropy of each row. This makes it particularly robust against varying levels of noise. Subsequently, we introduce a new DR algorithm, SNEkhorn, which leverages this novel affinity matrix. We demonstrate its superiority over state-of-the-art approaches using various indicators on both synthetic and real-world datasets.

Finally, these works naturally give rise to "adaptive regularizations" of OT problems, a topic we started investigating in [30].

8.3.4 Formal differential privacy preservation

Participants: Rémi Gribonval, Clément Lalanne.

Collaborations with Aurélien Garivier (UMPA, ENS de Lyon) and SARUS, Paris

Producing statistics that respect the privacy of the samples while still maintaining their accuracy is an important topic of research that we addressed under the framework of *differential privacy* with two complementary perspectives, on selected statistical problems : the design of concrete mechanims with controlled statistical utility and provable differential privacy guarantees; and the exhibition of lower-bounds on the achievable statistical performance of *any* mechanism with constrained differential privacy guarantees.

In the context of the PhD thesis of Clément Lalanne [32] we addressed the problem of differentially private estimation of multiple quantiles (MQ) of a dataset, a key building block in modern data analysis. First, we showed [15] how to implement the non-smoothed Inverse Sensitivity (IS) mechanism for this specific problem and established that the resulting method is closely related to the recent JointExp algorithm, sharing in particular the same computational complexity and a similar efficiency. We also identified pitfalls of the two approaches on certain peaked distributions, and proposed a fix. Numerical experiments showed that the empirical efficiency of the resulting algorithms is similar to the non-smoothed methods for non-degenerate datasets, but orders of magnitude better on real datasets with repeated values. We then refined the analysis [19] and notably showed that when the number of quantiles to estimate is large, it is better to estimate the density rather than the quantile function at specific points.

We studied minimax lower bounds when the class of estimators is restricted to the differentially private ones [14]. In particular, we showed that characterizing the power of a distributional test under differential privacy can be done by solving a transport problem. With specific coupling constructions, this observation allowed us to derivate Le Cam-type and Fano-type inequalities for both regular definitions of differential privacy and for divergence-based ones (based on Renyi divergence). We illustrated our results on three simple, fully worked out examples. For some problems, we showed that privacy leads to a provable degradation only when the rate of the privacy parameters is small enough whereas for other problems, the degradation systematically occurs under much looser hypotheses on the privacy parameters. Finally, we showed the near minimax optimality of the known guarantees for DP-SGLD, a private convex solver for maximum likelihood estimation on log-concave models. Based on these approaches, we studied the fundamenta statistical-privacy tradeoffs of density estimation [13].

8.3.5   Privacy and sparsity

Participants:	Can Pouliquen, Clément Lalanne, Antoine Gonon, Anne Gagneux, Léon Zheng, Quoc-Tung Le.

Sparse neural networks are mainly motivated by ressource efficiency since they use fewer parameters than their dense counterparts but still reach comparable accuracies. [26] empirically investigates whether sparsity could also improve the privacy of the data used to train the networks. The experiments show positive correlations between the sparsity of the model, its privacy, and its classification error. Simply comparing the privacy of two models with different sparsity levels can yield misleading conclusions on the role of sparsity, because of the additional correlation with the classification error. From this perspective, some caveats are raised about previous works that investigate sparsity and privacy.

8.4   Large-scale convex and nonconvex optimization

8.4.1   Multilevel schemes for image restoration

Participants:	Elisa Riccietti, Paulo Gonçalves, Guillaume Lauga.

Collaboration with Nelly Pustelnik (CNRS, ENS de Lyon), Nils Laurent (ENS de Lyon)

In the context of the Ph.D. work of Guillaume Lauga, we pursued the work started last year on the study of the combination of proximal methods and multiresolution analysis in large-scale image denoising problems. In the spirit of multilevel gradient methods [3] we developed a family of multilevel inertial

proximal methods, tailored for problems arising in imaging, which exploit wavelets-based transfer operators. Their ability to accelerate proximal algorithms was shown in several large dimensional problems [66] [21], and particularly in real-word problems arising in radio-interferometry and involving hyperspectral images. We also studied the link between multilevel and block coordinate methods and their convergence analysis.

In the context of the postdoc of Nils Laurent, we also started investigating the use of multilevel schemes in conjunction with plug and play methods. As these methods involve neural networks, the strategy to integrate multilevel schemes is naturally different from the one used so far in classical image denoising problems.

8.4.2 Subsampling methods for problems involving large datasets

Participants: Elisa Riccietti.

Collaboration with Margehrita Porcelli and Filippo Marini (U. Bologna, Italy)

Training problems usually involve a large number of data. The choice of the batch size in stochastic methods affects their convergence and their efficiency. We started to study a variant of stochastic variance reduction methods inspired by multilevel schemes, which should be less dependant on the choice of the hyperparameters.

8.4.3 Reproducible benchmarking of optimization algorithms

Participants: Mathurin Massias, Badr Moufad.

Collaboration with Thomas Moreau (MIND, Inria Saclay).

The team continues working on reproducible optimisation benchmarks, with Benchopt [69], a collaborative framework to automate, reproduce and publish optimization benchmarks in machine learning across programming languages and hardware architectures.

We continued to publish open source implementations of state-of-the-art solvers on major ML problems, and a detailed comparison of the regimes in which they succeed and fail respectively.

8.4.4 Algorithms for large scale sparse linear models

Participants: Mathurin Massias, Badr Moufad.

Collaboration with Quentin Bertrand (MILA, Montréal).

Based on our seminal works in [8] and [2], we continued to develop and implement new state-of-the-art solvers for optimization problems with millions of variables in the context of sparse linear models [50], implemented in the `skglm` package (see Section 7.1.2), that was integrated into the ecosystem of the `scikit-learn` package.

9   Bilateral contracts and grants with industry

9.1   Bilateral grants with industry

- CIFRE contract with Valeo AI, Paris on Frugal learning with applications to autonomous vehicles

    Participants: Rémi Gribonval, Elisa Riccietti, Léon Zheng.

    Duration: 3 years (2021-2024)

    Partners: Valeo AI, Paris; ENS de Lyon

    Funding: Valeo AI, Paris; ANRT

    Context: Chaire IA AllegroAssai 10.1.2

    The overall objective of this thesis is to develop machine learning methods exploiting low-dimensional sketches and sparsity to address perception-based learning tasks in the context of autonomous vehicles.

- Funding from Facebook Artificial Intelligence Research, Paris

    Participants: Rémi Gribonval.

    Duration: 4 years (2021-2024)

    Partners: Facebook Artificial Intelligence Research, Paris; ENS de Lyon

    Funding: Facebook Artificial Intelligence Research, Paris

    Context: Chaire IA AllegroAssai 10.1.2

    This is supporting the research conducted in the framework of the Chaire IA AllegroAssai.

10   Partnerships and cooperations

10.1   National initiatives

10.1.1   PEPR IA project : SHARP

Participants: Rémi Gribonval *(correspondant)*, Paulo Gonçalves, Elisa Ricietti, Marion Foare, Mathurin Massias, Titouan Vayer.

Partnership with LAMSADE (PSL); LIGM (ENPC); GENESIS (Inria London & University College London); IRISA; CEA List; ISIR (Sorbonne Université)

Duration of the project: **2023 - 2027**.

The vision of the SHARP proposal is that the resources required to train ML models can be decreased by several orders of magnitude, with negligible performance loss compared to the state of the art. This means significantly reducing the dimensionality of predictors (to reduce inference costs) and of their gradients (to reduce training and bandwidth costs in distributed settings), the amount of data needed to learn (to address data scarce settings up to zero-shot learning, and incremental learning scenarios), and compressing datasets before learning (to reduce storage and compute requirements, and address privacy concerns).

10.1.2   ANR IA Chaire : AllegroAssai

Participants:	Rémi Gribonval *(correspondant)*, Paulo Gonçalves, Elisa Ricietti, Marion Foare, Mathurin Massias, Léon Zheng, Quoc-Tung Le, Antoine Gonon, Titouan Vayer, Ayoub Belhadji, Clement Lalanne, Can Pouliquen.

Past members: Luc Giffon.

Duration of the project: **2020 - 2025**.

AllegroAssai focuses on the design of machine learning techniques endowed both with statistical guarantees (to ensure their performance, fairness, privacy, etc.) and provable resource-efficiency (e.g. in terms of bytes and flops, which impact energy consumption and hardware costs), robustness in adversarial conditions for secure performance, and ability to leverage domain-specific models and expert knowledge. The vision of AllegroAssai is that the versatile notion of sparsity, together with sketching techniques using random features, are key in harnessing these fundamental tradeoffs. The first pillar of the project is to investigate sparsely connected deep networks, to understand the tradeoffs between the approximation capacity of a network architecture (ResNet, U-net, etc.) and its "trainability" with provably-good algorithms. A major endeavor is to design efficient regularizers promoting sparsely connected networks with provable robustness in adversarial settings. The second pillar revolves around the design and analysis of provably-good end-to-end sketching pipelines for versatile and resource-efficient large-scale learning, with controlled complexity driven by the structure of the data and that of the task rather than the dataset size.

10.1.3   ANR DataRedux

Participants:	Paulo Gonçalves *(correspondant)*, Rémi Gribonval, Marion Foare, Rémi Vaudaine.

Duration of the project: **February 2020 - January 2024**.

DataRedux puts forward an innovative framework to reduce networked data complexity while preserving its richness, by working at intermediate scales ("mesoscales"). Our objective is to reach a fundamental breakthrough in the theoretical understanding and representation of rich and complex networked datasets for use in predictive data-driven models. Our main novelty is to define network reduction techniques in relation with the dynamical processes occurring on the networks. To this aim, we will develop methods to go from data to information and knowledge at different scales in a human-accessible way by extracting structures from high-resolution, diverse and heterogeneous data. Our methodology will involve the identification of the most relevant subparts of time-resolved datasets while remapping the remaining parts of the system, the simultaneous structural-temporal representations of time-varying networks, the development of parsimonious data representations extracting meaningful structures at mesoscales ("mesostructures"), and the building of models of interactions that include mesostructures of various types. Our aim is to identify data aggregation methods at intermediate scales and new types of data representations in relation with dynamical processes, that carry the richness of information of the original data, while keeping their most relevant patterns for their manageable integration in data-driven numerical models for decision making and actionable insights.

10.1.4   ANR Darling

Participants:	Paulo Gonçalves *(correspondant)*, Rémi Gribonval, Marion Foare.

Duration of the project: **February 2020 - January 2024**.

This project meets the compelling demand of developing a unified framework for distributed knowledge extraction and learning from graph data streaming using in-network adaptive processing, and adjoining powerful recent mathematical tools to analyze and improve performances. The project draws on three major parallel directions of research: network diffusion, signal processing on graphs, and random matrix theory which DARLING aims at unifying into a holistic dynamic network processing framework. Signal processing on graphs has recently provided a comprehensive set of basic instruments allowing for signal on graph filtering or sampling, but it is limited to static signal models. Network diffusion on the opposite inherently assumes models of time varying graphs and signals, and has pursued the path of proposing and understanding the performance of distributed dynamic inference on graphs. Both areas are however limited by their assuming either deterministic graph or signal models, thereby entailing often inflexible and difficult-to-grasp theoretical results. Random matrix theory for random graph inference has taken a parallel road in explicitly studying the performance, thereby drawing limitations and providing directions of improvement, of graph-based algorithms (e.g., spectral clustering methods). The ambition of DARLING lies in the development of network diffusion-type algorithms anchored in the graph signal processing lore, rather than heuristics, which shall systematically be analyzed and improved through random matrix analysis on elementary graph models. We believe that this original communion of as yet remote areas has the potential to path the pave to the emergence of the critically needed future field of dynamical network signal processing.

10.1.5   ANR JCJC MASSILIA

Participants: Titouan Vayer.

Duration of the project: **December 2021 - December 2025**.

Collaboration with Arnaud Breloy (PI of the project, Univ. Paris Nanterre), Florent Bouchard (CentraleSupélec), Cédric Richard (Univ. Côte d'Azur), Rémi Flamary (Ecole Polytechnique) and Ammar Mian (Univ. Savoie Mont Blanc)

This project aims at tackling current problems related to graph learning and its applications in a unified way centered around the spectral decomposition of the graph Laplacian and/or adjacency matrices. The central objective of this project is to model graph structures (distributions on spectral parameters) and leverage this formalism in to two main directions 1) improve graph learning processes by directly learning structured spectral decompositions from the data 2) handle collections of graphs in order to compute structured graphs barycenters, compress graphs representations, and classify/cluster data using their graph as the main feature.

10.1.6   ANR JCJC Multisc-In

Participants: Marion Foare, Elisa Ricietti.

Collaboration with Nelly Pustelnik (PI of the project, ENS de Lyon), Laurent Condat (KAUST, Saudi Arabia), Luis Briceño-Arias (Univ. Téchnica Federico Santa Maria, Chili)

Interface detection is a challenging question in image processing, and more generally in graph processing, leading to a large panel of applications going from geophysics research to societal studies. The common point to these applications is the willingness to have an interface detection at a fine scale, in order to extract physical or societal parameters, from high resolution data. This project is devoted to

original image processing tools relying both on optimization and multiresolution techniques in order to provide a new paradigm for the interface detection on large scale data.

10.1.7 ANR JCJC EROSION

Participants: Mathurin Massias.

Collaboration with Emmanuel Soubies (PI of the project, CNRS, IRIT), Paul Escande (CR CNRS, I2M), Cédric Févotte (DR CNRS, IRIT), Henrique Goulart (MdC INP, IRIT) and Joseph Salmon (Prof. Université de Montpellier, IMAG)

The promise of EROSION is to push the frontiers of sparse and low-rank optimization by combining the strengths of exact relaxations and local optimization. More precisely, we propose to move away from the appealing convex relaxation requiring too strong assumptions to ensure the equivalence with the original problem. Instead, EROSION will address the following two research objectives. 1 : Deriving exact relaxations of ℓ_0 regression (= same global minimizers) which, although still non-convex, are more amenable to non-convex local optimization (e.g., less local minimizers, wider basins of attraction). 2 : Developing new local optimization strategies that exploit the nice properties of such exact relaxations so as to improve both the quality of reached local extrema and the convergence speed over existing solvers.

10.1.8 DI2A - Subvention Simone et Cino del Duca, Institut de France.

Participants: Elisa Riccietti, Marion Foare, Paulo Gonçalves.

Duration of the project: **December 2023 - December 2025**.
This project focuses on the physics-informed design of architectures and multiresolution deep learning techniques for large scale image restoration and data analysis for astronomy. With the term physics-informed design we refer to all the deep learning strategies in which the choice of the architecture, biases and activation functions of neural networks is guided by the underlying physics of data acquisition and/or from the optimization proximal schemes employed for the solution. From an application point of view, the project targets problems in astronomy and specifically the study of circumstellars environments through the instrument SPHERE/IRDIS. We aim to propose innovative reconstruction approaches partially supervised or even non supervised.

10.1.9 GDR ISIS project MOMIGS

Participants: Elisa Riccietti *(correspondant)*, Marion Foare, Paulo Gonçalves.

Duration of the project: **September 2021 - September 2023**.

This project focuses on large scale optimization problems in signal processing and imaging. A natural way to tackle them is to exploit their underlying structure, and to represent them at different resolution levels. The use of multiresolution schemes, such as wavelets transforms, is not new in imaging and is widely used to define regularization strategies. However, such techniques could be used to a wider extent, in order to accelerate the optimization algorithms used for their solution and to tackle large datasets. Techniques based on such ideas are usually called multilevel optimization methods and are well-known and widely used in the field of smooth optimization and especially in the solution of partial differential equations. Optimization problems arising in image reconstruction are however usually nonsmooth and thus solved by proximal methods. Such approaches are efficient for small-scale problems but still computationally demanding for problems with very high-dimensional data. The ambition of this project

is thus to combine proximal methods and multiresolution analysis not only as a regularization, but as a solution to accelerate proximal algorithms.

10.1.10   GDR ISIS project PROSSIMO

Participants: Mathurin Massias *(correspondant)*, Rémi Gribonval, Anne Gagneux, Emmanuel Soubies.

Duration of the project: **September 2023 - September 2025**.

Composite optimisation problems are ubiquitous in machine learning, signal, and image processing. With the proximal algorithms used to solve them, they have met with great success in applications and have been extensively studied. More recently, so-called 'plug-and-play' (PNP) methods, inspired by proximal algorithms, propose new iterative algorithms in which the application of the proximal operator of the regulariser is replaced by a pre-existing denoiser or a learned operator. Their flexibility, however, complicates their theoretical analysis, because in the general case the operator does not have the interesting properties of proximal operators. In the PROSSIMO project, we propose to implement and study PNP operators via neural networks, while guaranteeing that these operators have the same properties as proximal operators. We aim at combining the flexibility of PNP methods with the rigorous theoretical guarantees of model-based methods. In addition to implementing such networks, we propose to study their approximation capacity: what classes of function can they approximate, and at what speed?

10.2   Regional initiatives

10.2.1   Labex CominLabs LeanAI

Participants: Elisa Riccietti *(correspondant)*, Rémi Gribonval.

Duration of the project: October 2021-December 2024.

Collaboration with Silviu-Ioan Filip and Olivier Sentieys (IRISA, Rennes), Anastasia Volkova (LS2N Nantes)

The LeanAI project aims at developing a comprehensive and flexible framework for mixed-precision optimization. The project is motivated by the increasing demand for intelligent edge devices capable of on-site learning, driven by the recent developments in deep learning. The realization of such systems is a massive challenge due to the limited resources available in an embedded context and the massive training costs for state-of-the-art deep neural networks. In this project we attack these problems at the arithmetic and algorithmic levels by exploring the design of new mixed numerical precision algorithms, energy-efficient and capable of offering increased performance in a resource-restricted environment. The ambition of the project is to develop more flexible and faster techniques than existing reduced-precision gradient algorithms, by determining the best numeric formats to be used in combination with this kind of methods, rules to dynamically adjust the precision and extension of such techniques to second-order and multilevel strategies.

11   Dissemination

Participants: Rémi Gribonval, Paulo Gonçalves, Marion Foare, Mathurin Massias, Elisa Riccietti, Titouan Vayer.

11.1   Promoting scientific activities

11.1.1   Scientific events: organisation

Member of the organizing committees

- Mathurin Massias, Elisa Riccietti, Rémi Gribonval, Journées SMAI-MODE 2024, Lyon

- Mathurin Massias, Learning and Optimization in Luminy 2024, CIRM, Marseille

- Titouan Vayer, Graph Learning & Learning with Graphs, special session at GRETSI conference 2023.

- Elisa Riccietti, Marion Foare, Workshop Deep learning, image analysis, inverse problems, and optimization (DIPOpt), 2023, ENS de Lyon, Lyon

- Titouan Vayer, Mathurin Massias, GDR MIA Thematic day on dimensionality reduction, 2023, ENS de Lyon.

- Rémi Gribonval, Paulo Goncalves, Titouan Vayer, Mathurin Massias, IXXI thematic day on frugality in machine learning, 2023, ENS de Lyon.

11.1.2   Scientific events: selection

Member of the conference program committees

- Rémi Gribonval, GRETSI 2023

11.1.3   Journal

Member of the editorial boards

- Rémi Gribonval: Associate Editor for Constructive Approximation (Springer), Senior Area Editor for the IEEE Signal Processing Magazine

- Mathurin Massias: Associate editor for Computo (French Statistical Society)

11.1.4   Invited talks

- Titouan Vayer and Mathurin Massias: OLISSIPO Winter school (Lisbon), Dimensionality reduction, 6 h.

- Rémi Gribonval: Journées de Recherche en Apprentissage Frugal, Grenoble, Dec 13-14

- Rémi Gribonval: Workshop DIPOpt (Deep learning, image analysis, inverse problems, and optimization), Lyon, Nov 27-30

- Rémi Gribonval: Workshop A Multiscale tour of Harmonic Analysis and Machine Learning (Mallat's 60th), IHES, April 19-21 2023

11.1.5   Leadership within the scientific community

- Rémi Gribonval is a member of the Scientific Committee of RT MIA (formerly GDR MIA)

- Rémi Gribonval is a member of the Comité de Liaison SIGMA-SMAI

- Rémi Gribonval is a member of the Cellule ERC of INS2I, and joined the Celulle ERC of Inria, mentoring for ERC candidates in the STIC domain at the national level

11.1.6   Scientific expertise

- Rémi Gribonval is a member of the Scientific Advisory Board (vice-president) of the Acoustics Research Institute of the Austrian Academy of Sciences, and a member of the Commission Prospective of Institut de Mathématiques de Marseille

- Elisa Riccietti is a member of the "commission formation" of the labex MILyon

11.1.7   Research administration

- R. Gribonval and P. Gonçalves are both members of the executive committee of SCIDOLYSE research group.

- R. Gribonval and P. Gonçalves were both members of the drafting panel of AILYS proposal to the AI-Cluster call for projects.

- P. Gonçalves is member of the steering committee for the ShapeMed@Lyon consortium's Data for Health workshop

- Paulo Gonçalves is Deputy Scientific Director of Inria Lyon and member of the Inria Evaluation Committee.

11.2   Teaching - Supervision - Juries

11.2.1   Teaching

- Master :

  - Rémi Gribonval: Inverse problems and high dimension; Mathematical foundations of deep neural networks; Concentration of measure in probability and high-dimensional statistical learning; M2, ENS Lyon

  - Mathurin Massias: Large scale optimization for Machine Learning (M2, ENS Lyon); Python for Datascience (M1, Ecole Polytechnique/HEC); Optimisation (M1, ENS Lyon)

  - Titouan Vayer: Machine Learning for Graphs and on Graphs (M2, ENS Lyon)

  - Elisa Riccietti: Fundamentals of Machine Learning (M1, ENS Lyon). 19h of tutor responsibility at ENS Lyon

- Engineer cycle (Bac+3 to Bac+5):

  - Paulo Gonçalves: Traitement du Signal (déterministe, aléatoire, numérique), Estimation statistique. 80 heures Eq. TD. CPE Lyon, France

  - Marion Foare: Traitement du Signal (déterministe, numérique, aléatoire), Traitement et analyse d'images, Optimisation, Compression, Projets. 280 heures Eq. TD. CPE Lyon, France

- Other formations: "Fondements et pratique du machine learning et du deep learning", CNRS formation for Dassault Systèmes, 3 x 3 days (18h) with Mathurin Massias, Titouan Vayer and Aurélien Garivier.

11.2.2   Supervision

All PhD students of the team are co-supervised by at least one team member. In addition, some team members are involved in the co-supervision of students hosted in other labs.

- Marion Foare is involved in the co-supervision of the Ph.D. of Hoang Trieu Vy Le since 2021 (Laboratoire de Physique, Lyon, defended in December 2023).

- Elisa Riccietti is involved in the co-supervision of the Ph.D. of Valentin Mercier since 2021 (IRIT, Toulouse).

- Elisa Riccietti is involved in the co-supervision of the Ph.D. of Paul Estano since 2022 (IRISA, Rennes).

- Rémi Gribonval is involved in the co-supervision of the Ph.D. of Sibylle Marcotte since 2022 (Center for Data Science, ENS Paris).

- Marion Foare is involved in the co-supervision of the Ph.D. of Luis Enrique Amador Araya since 2023 (Siemens Healthcare, Saint Denis, and Creatis, Villeurbanne).

- Elisa Riccietti is involved in the co-supervision of the postdoc of Nils Laurent (ENS de Lyon).

PhD defenses in OCKHAM in 2023:

- Clément Lalanne

- Quoc-Tung Le

11.2.3 Juries

Members of the OCKHAM team participated to the following juries:

- PhD juries: Cédric Vincent-Cuaz (Université Côte d'Azur, member); Benoît Malezieux (Université Paris-Saclay, reviewer and president); Edouard Yvinec (Sorbonne Université, member); Joachim Bona-Pellissier (Université de Toulouse, reviewer)

- Habilitation juries: Xavier Luciani (Université de Toulon, president); Claire Boyer (Sorbonne Université, member)

12 Scientific production

12.1 Major publications

[1] E. Bautista, P. Abry and P. Gonçalves. 'L^γ -PageRank for Semi-Supervised Learning'. In: *Applied Network Science* 4.57 (2019), pp. 1–20. DOI: 10.1007/s41109-019-0172-x. URL: https://inria.hal.science/hal-02063780.

[2] Q. Bertrand, Q. Klopfenstein, M. Massias, M. Blondel, S. Vaiter, A. Gramfort and J. Salmon. 'Implicit differentiation for fast hyperparameter selection in non-smooth convex learning'. In: *Journal of Machine Learning Research* 23.149 (Apr. 2022), pp. 1–48. URL: https://hal.science/hal-03228663.

[3] H. Calandra, S. Gratton, E. Riccietti and X. Vasseur. 'On a multilevel Levenberg–Marquardt method for the training of artificial neural networks and its application to the solution of partial differential equations'. In: *Optimization Methods and Software* (2020), pp. 1–26. DOI: 10.1080/10556788.2020.1775828. URL: https://hal.science/hal-02956018.

[4] M. Foare, N. Pustelnik and L. Condat. 'Semi-Linearized Proximal Alternating Minimization for a Discrete Mumford-Shah Model'. In: *IEEE Transactions on Image Processing* (7th Oct. 2019), pp. 1–13. DOI: 10.1109/TIP.2019.2944561. URL: https://hal.science/hal-01782346.

[5] R. Gribonval, G. Blanchard, N. Keriven and Y. Traonmilin. 'Compressive Statistical Learning with Random Feature Moments'. In: *Mathematical Statistics and Learning* 3.2 (21st Aug. 2021), pp. 113–164. DOI: 10.4171/msl/20. URL: https://inria.hal.science/hal-01544609.

[6] R. Gribonval, A. Chatalic, N. Keriven, V. Schellekens, L. Jacques and P. Schniter. 'Sketching Data Sets for Large-Scale Learning: Keeping only what you need'. In: *IEEE Signal Processing Magazine* 38.5 (Sept. 2021), pp. 12–36. DOI: 10.1109/MSP.2021.3092574. URL: https://inria.hal.science/hal-03350599.

[7] R. Gribonval, G. Kutyniok, M. Nielsen and F. Voigtlaender. 'Approximation spaces of deep neural networks'. In: *Constructive Approximation*. special issue on "Deep Networks in Approximation Theory" (2021). DOI: 10.1007/s00365-021-09543-4. URL: https://inria.hal.science/hal-02117139.

[8] M. Massias, S. Vaiter, A. Gramfort and J. Salmon. 'Dual Extrapolation for Sparse Generalized Linear Models'. In: *Journal of Machine Learning Research* 21.234 (Oct. 2020), pp. 1–33. URL: `https://hal.science/hal-02263500`.

[9] B. Ricaud, P. Borgnat, N. Tremblay, P. Gonçalves and P. Vandergheynst. 'Fourier could be a Data Scientist: from Graph Fourier Transform to Signal Processing on Graphs'. In: *Comptes Rendus. Physique* (19th Sept. 2019), pp. 474–488. DOI: `10.1016/j.crhy.2019.08.003`. URL: `https://inria.hal.science/hal-02304584`.

[10] C. Vincent-Cuaz, R. Flamary, M. Corneli, T. Vayer and N. Courty. 'Semi-relaxed Gromov Wasserstein divergence with applications on graphs'. In: ICLR 2022 - 10th International Conference on Learning Representations. Virtual, France, 25th Apr. 2022, pp. 1–28. URL: `https://hal.science/hal-03832589`.

[11] C. Vincent-Cuaz, R. Flamary, M. Corneli, T. Vayer and N. Courty. 'Template based Graph Neural Network with Optimal Transport Distances'. In: NeurIPS 2022 – 36th Conference on Neural Information Processing Systems. New Orleans, United States, 2022. URL: `https://hal.science/hal-03839517`.

12.2 Publications of the year

International journals

[12] A. Gonon, N. Brisebarre, R. Gribonval and E. Riccietti. 'Approximation speed of quantized vs. unquantized ReLU neural networks and beyond'. In: *IEEE Transactions on Information Theory* 69.6 (1st June 2023), pp. 3960–3977. DOI: `10.1109/TIT.2023.3240360`. URL: `https://hal.science/hal-03672166`.

[13] C. Lalanne, A. Garivier and R. Gribonval. 'About the Cost of Central Privacy in Density Estimation'. In: *Transactions on Machine Learning Research Journal* (30th Aug. 2023). URL: `https://hal.science/hal-04133655`.

[14] C. Lalanne, A. Garivier and R. Gribonval. 'On the Statistical Complexity of Estimation and Testing under Privacy Constraints'. In: *Transactions on Machine Learning Research Journal* (Apr. 2023). URL: `https://hal.science/hal-03794374`.

[15] C. Lalanne, C. Gastaud, N. Grislain, A. Garivier and R. Gribonval. 'Private Quantiles Estimation in the Presence of Atoms'. In: *Information and Inference* (4th Aug. 2023). DOI: `10.1093/imaiai/iaad030`. URL: `https://hal.science/hal-03572701`.

[16] T. Vayer and R. Gribonval. 'Controlling Wasserstein Distances by Kernel Norms with Application to Compressive Statistical Learning'. In: *Journal of Machine Learning Research* 24.149 (1st Apr. 2023), pp. 1–51. URL: `https://hal.science/hal-03461492`.

International peer-reviewed conferences

[17] A. Collas, T. Vayer, R. Flamary and A. Breloy. 'Entropic Wasserstein component analysis'. In: IEEE International Workshop on Machine Learning for Signal Processing (MLSP). Rome, Italy, 17th Sept. 2023. URL: `https://hal.science/hal-04022713`.

[18] A. Hippert-Ferrer, F. Bouchard, A. Mian, T. Vayer and A. Breloy. 'Learning Graphical Factor Models with Riemannian Optimization'. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD 2023). Torino, Italy, 18th Sept. 2023. URL: `https://hal.science/hal-04249001`.

[19] C. Lalanne, A. Garivier and R. Gribonval. 'Private Statistical Estimation of Many Quantiles'. In: ICML 2023 - 40th International Conference on Machine Learning. Honolulu, United States, 24th July 2023. URL: `https://hal.science/hal-03986170`.

[20] J. Larsson, Q. Klopfenstein, M. Massias and J. Wallin. 'Coordinate Descent for SLOPE'. In: *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*. 26th International Conference on Artificial Intelligence and Statistics - AISTATS 2023. Valencia, Spain, Apr. 2023. URL: `https://hal.science/hal-04211867`.

[21]   G. Lauga, E. Riccietti, N. Pustelnik and P. Gonçalves. 'Multilevel FISTA for image restoration'. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*. ICASSP. Rhodes, Greece, 2023. DOI: 10.48550/arXiv.2210.15940. URL: https://hal.science/hal-03831180.

[22]   Q.-T. Le, E. Riccietti and R. Gribonval. 'Does a sparse ReLU network training problem always admit an optimum?' In: NeurIPS 2023 - Thirty-seventh Conference on Neural Information Processing Systems. Advances in Neural Information Processing Systems 36 (NeurIPS 2023). New Orleans (Lousiane), United States, Dec. 2023. URL: https://inria.hal.science/hal-04108849.

[23]   S. Marcotte, R. Gribonval and G. Peyré. 'Abide by the Law and Follow the Flow: Conservation Laws for Gradient Flows'. In: Thirty-seventh Conference on Neural Information Processing Systems. Advances in Neural Information Processing Systems 36 (NeurIPS 2023). New Orleans (Louisiane), United States, 30th June 2023. URL: https://inria.hal.science/hal-04150576.

[24]   H. Van Assel, T. Vayer, R. Flamary and N. Courty. 'SNEkhorn: Dimension Reduction with Symmetric Entropic Affinities'. In: Thirty-seventh Annual Conference on Neural Information Processing Systems (NeurIPS). New Orleans, United States, 2023. URL: https://hal.science/hal-04103326.

[25]   L. Zheng, G. Puy, E. Riccietti, P. Pérez and R. Gribonval. 'Self-supervised learning with rotation-invariant kernels'. In: The Eleventh International Conference on Learning Representations. Kigali, Rwanda, 1st May 2023. URL: https://hal.science/hal-03738466.

National peer-reviewed Conferences

[26]   A. Gonon, L. Zheng, C. Lalanne, Q.-T. Le, G. Lauga and C. Pouliquen. 'Can sparsity improve the privacy of neural networks?' In: GRETSI 2023 - XXIXème Colloque Francophone de Traitement du Signal et des Images. Grenoble, France, 18th Apr. 2023. URL: https://hal.science/hal-04062317.

[27]   R. Gribonval, T. Mary and E. Riccietti. 'Scaling is all you need: quantization of butterfly matrix products via optimal rank-one quantization'. In: 29ème Colloque sur le traitement du signal et des images (GRETSI). Actes du GRETSI 2023 2023-1193. Grenoble, France: GRETSI - Groupe de Recherche en Traitement du Signal et des Images, 28th Aug. 2023, pp. 497–500. URL: https://inria.hal.science/hal-04146706.

[28]   C. Pouliquen, P. Gonçalves, M. Massias and T. Vayer. 'Implicit Differentiation for Hyperparameter Tuning the Weighted Graphical Lasso'. In: GRETSI 2023 - XXIXème Colloque Francophone de Traitement du Signal et des Images. Grenoble (France), France, 28th Aug. 2023, pp. 1–4. URL: https://hal.science/hal-04151796.

[29]   L. Zheng, G. Puy, E. Riccietti, P. Pérez and R. Gribonval. 'Butterfly factorization by algorithmic identification of rank-one blocks'. In: XXIXème Colloque Francophone de Traitement du Signal et des Images. Grenoble, France, 28th Aug. 2023. URL: https://inria.hal.science/hal-04145743.

Conferences without proceedings

[30]   H. Van Assel, T. Vayer, R. Flamary and N. Courty. 'Optimal Transport with Adaptive Regularisation'. In: NeurIPS OTML Workshop. New Orleans, France, 4th Oct. 2023. URL: https://hal.science/hal-04229636.

[31]   H. Van Assel, C. Vincent-Cuaz, T. Vayer, R. Flamary and N. Courty. 'Interpolating between Clustering and Dimensionality Reduction with Gromov-Wasserstein'. In: NeurIPS OTML Workshop. New Orleans, United States, 5th Oct. 2023. URL: https://hal.science/hal-04356797.

Doctoral dissertations and habilitation theses

[32]   C. Lalanne. 'On the tradeoffs of statistical learning with privacy'. Ecole normale supérieure de lyon - ENS LYON, 4th Oct. 2023. URL: https://theses.hal.science/tel-04379624.

[33]   Q.-T. Le. 'Algorithmic and theoretical aspects of sparse deep neural networks'. Ecole Normale Supérieure de Lyon, 20th Dec. 2023. URL: https://inria.hal.science/tel-04329531.

Reports & preprints

[34] A. Belhadji, R. Bardenet and P. Chainais. *Signal reconstruction using determinantal sampling.* 14th Aug. 2023. URL: https://hal.science/hal-04181079.

[35] A. Belhadji and R. Gribonval. *Sketch and shift: a robust decoder for compressive clustering.* 15th Dec. 2023. URL: https://inria.hal.science/hal-04356977.

[36] A. Gonon, N. Brisebarre, E. Riccietti and R. Gribonval. *A path-norm toolkit for modern networks: consequences, promises and challenges.* Sept. 2023. URL: https://hal.science/hal-04225201.

[37] S. Gratton, V. Mercier, E. Riccietti and P. L. Toint. *A Block-Coordinate Approach of Multi-level Optimization with an Application to Physics-Informed Neural Networks.* 2023. URL: https://hal.science/hal-04104450.

[38] R. Gribonval, T. Mary and E. Riccietti. *Optimal quantization of rank-one matrices in floating-point arithmetic—with applications to butterfly factorizations.* 12th June 2023. URL: https://inria.hal.science/hal-04125381.

[39] G. Lauga, E. Riccietti, N. Pustelnik and P. Gonçalves. *IML FISTA: A Multilevel Framework for Inexact and Inertial Forward-Backward. Application to Image Restoration.* 27th June 2023. URL: https://inria.hal.science/hal-04075814.

[40] G. Lauga, E. Riccietti, N. Pustelnik and P. Gonçalves. *Multilevel methods for hyperspectral images restoration.* 28th June 2023. URL: https://inria.hal.science/hal-04067225.

[41] R. Vaudaine, P. Borgnat, P. Gonçalves, R. Gribonval and M. Karsai. *Temporal network compression via network hashing.* 2023. DOI: 10.48550/arXiv.2307.04890. URL: https://hal.science/hal-04228037.

[42] T. Vayer, E. Lasalle, R. Gribonval and P. Gonçalves. *Compressive Recovery of Sparse Precision Matrices.* 12th Dec. 2023. DOI: 10.48550/arXiv.2311.04673. URL: https://hal.science/hal-04275341.

12.3 Other

Softwares

[43] [SW] R. Gribonval, E. Riccietti and T. Mary, *Code for reproducible research for the article "OPTIMAL QUANTIZATION OF RANK-ONE MATRICES IN FLOATING-POINT ARITHMETIC—WITH APPLICATIONS TO BUTTERFLY FACTORIZATIONS",* 9th June 2023. LIC: BSD 2-Clause License. HAL: ⟨hal-04124171⟩, URL: https://inria.hal.science/hal-04124171, VCS: https://gitlab.inria.fr/ericciet/rank-1-quantization.git, SWHID: ⟨swh:1:dir:98eaa2697116f3815573f266465e4a7800d96af8;origin=https://hal.archives-ouvertes.fr/hal-04124171;visit=swh:1:snp:5c0e415b900b448623781225e10029b1e0e4201a;anchor=swh:1:rel:0c27febd089e0c3a3819886fddda683e400d4462;path=/⟩.

[44] [SW] Q.-T. Le, E. Riccietti and R. Gribonval, *Code for reproducible research: Does a sparse ReLU network training problemalways admit an optimum?,* 31st Oct. 2023. LIC: BSD 3-Clause License. HAL: ⟨hal-04233925⟩, URL: https://hal.science/hal-04233925, SWHID: ⟨swh:1:dir:bc925c94db38e22382161dead27e7f667d47f71a;origin=https://hal.archives-ouvertes.fr/hal-04233925;visit=swh:1:snp:03d26f344bc6823ea9f499da1ebbe9c37a7c2c8d;anchor=swh:1:rel:893f968c69665c2ca0d3844b48ca1b07d14bd845;path=/⟩.

[45] [SW] S. Marcotte, R. Gribonval and G. Peyré, *Code for reproducible research. Abide by the Law and Follow the Flow: Conservation Laws for Gradient Flows,* 8th Nov. 2023. LIC: BSD 3-Clause License. HAL: ⟨hal-04261339⟩, URL: https://hal.science/hal-04261339, VCS: https://github.com/sibyllema/Conservation_laws, SWHID: ⟨swh:1:dir:4ef0397467efbab3728558079522b7bf5b96bb90;origin=https://hal.archives-ouvertes.fr/hal-04261339;visit=swh:1:snp:78f6f8a174ff007bc6eb37ebcc0c3635aca44c1d;anchor=swh:1:rel:2c795a9be210d1bac58f9903a22b307b18299605;path=/⟩.

[46]   [SW] R. Vaudaine, R. Gribonval, P. Gonçalves, P. Borgnat and M. Karsai, *Code for the article Temporal Network Compression via Network Hashing*, 7th Dec. 2023. LIC: BSD 3-Clause Clear License. HAL: ⟨hal-04330187⟩, URL: https://inria.hal.science/hal-04330187, VCS: https://github .com/vaudaine/Temporal_network_compression_via_network_hashing.

12.4   Cited publications

[47]   H. H. Bauschke, P. L. Combettes et al. *Convex analysis and monotone operator theory in Hilbert spaces*. Vol. 408. Springer, 2011.

[48]   A. Belhadji. 'Des règles de quadrature dans les RKHSs à base de DPPs'. In: *GRETSI 2022 - XXVIIIème Colloque Francophone de Traitement du Signal et des Images*. Nancy, France, Sept. 2022, pp. 1–4. URL: https://hal.science/hal-03630086.

[49]   A. Belhadji and R. Gribonval. 'Revisiting RIP guarantees for sketching operators on mixture models'. working paper or preprint. Nov. 2022. URL: https://hal.science/hal-03872878.

[50]   Q. Bertrand, Q. Klopfenstein, P.-A. Bannier, G. Gidel and M. Massias. 'Beyond l1: Faster and better sparse models with skglm'. In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 38950–38965.

[51]   H. Boche, R. Calderbank, G. Kutyniok and J. Vybiral. *Compressed Sensing and its Applications*. English. Ed. by H. Boche, R. Calderbank, G. Kutyniok and J. Vybiral. Vol. Series: Applied and Numerical Harmonic Analysis. MATHEON Workshop 2013. ISSN: 2296-5009. Cham: Birkhäuser, Cham, 2015. DOI: 10.1007/978-3-319-16042-9. URL: http://books.google.cz/books?id =6KoYCgAAQBAJ%5C&pg=PA340%5C&dq=intitle:Compressed+Sensing+and+its+Applicati ons%5C&hl=%5C&cd=1%5C&source=gbs%5C_api.

[52]   Y. de Castro and F. Gamboa. 'Exact Reconstruction using Beurling Minimal Extrapolation'. In: *arXiv.org* (Mar. 2011). arXiv: 1103.4951v2. URL: http://arxiv.org/abs/1103.4951v2.

[53]   A. Chatalic, V. Schellekens, F. Houssiau, Y.-A. De Montjoye, L. Jacques and R. Gribonval. 'Compressive Learning with Privacy Guarantees'. In: *Information and Inference* (2021). URL: https://hal.i nria.fr/hal-02496896.

[54]   P. L. Combettes and J.-C. Pesquet. 'Proximal splitting methods in signal processing'. In: *Fixed-point algorithms for inverse problems in science and engineering*. Springer, 2011, pp. 185–212.

[55]   P. Di Lorenzo, P. Banelli, S. Barbarossa and S. Sardellitti. 'Distributed Adaptive Learning of Graph Signals'. In: *IEEE Transaction on Signal Processing* 65.16 (2017).

[56]   P. M. Djuric and R. C. *Cooperative and Graph Signal Processing: Principle and Applications*. Academic Press, 2018.

[57]   M. Elad. *Sparse and Redundant Representations*. English. From Theory to Applications in Signal and Image Processing. Springer, 2010. URL: http://books.google.fr/books?id=d5b6lJI9 BvAC%5C&printsec=frontcover%5C&dq=sparse+and+redundant+representations%5C&hl =%5C&cd=1%5C&source=gbs%5C_api.

[58]   S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. New York, NY: Springer, 2013. DOI: 10.1007/978-0-8176-4948-7. URL: http://link.springer.com/10.10 07/978-0-8176-4948-7.

[59]   J. Friedman, T. Hastie and R. Tibshirani. 'Sparse inverse covariance estimation with the graphical lasso'. In: *Biostatistics* 9.3 (2008), pp. 432–441.

[60]   B. Girault, P. Gonçalves and E. Fleury. 'Translation on Graphs: An Isometric Shift Operator'. In: *IEEE Signal Processing Letters* 22.12 (Dec. 2015), pp. 2416–2420. DOI: 10.1109/LSP.2015.2488279. URL: https://hal.inria.fr/hal-01221562.

[61]   K. Gregor and Y. LeCun. 'Learning fast approximations of sparse coding'. In: *Proceedings of the 27th international conference on international conference on machine learning*. 2010, pp. 399–406.

[62]   R. Gribonval, G. Blanchard, N. Keriven and Y. Traonmilin. 'Compressive Statistical Learning with Random Feature Moments'. en. In: *Mathematical Statistics and Learning* (2021). URL: https://ha l.inria.fr/hal-01544609 (visited on 15/04/2021).

[63]   R. Gribonval, G. Blanchard, N. Keriven and Y. Traonmilin. 'Statistical Learning Guarantees for Compressive Clustering and Compressive Mixture Modeling'. In: *Mathematical Statistics and Learning* 3.2 (Aug. 2021). This preprint results from a split and profound restructuring and improvements of of `https://hal.inria.fr/hal-01544609v2It` is a companion paper to `https://hal.inria.fr/hal-01544609v3`, pp. 165–257. DOI: `10.4171/msl/21`. URL: `https://hal.inria.fr/hal-02536818`.

[64]   R. Jenatton, J.-Y. Audibert and F. Bach. 'Structured Variable Selection with Sparsity-Inducing Norms'. English. In: *Journal of Machine Learning Research* 12 (2011). Publisher: Massachusetts Institute of Technology Press, pp. 2777–2824. URL: `http://hal.inria.fr/inria-00377732`.

[65]   S. Kumar, J. Ying, J. V. de M. Cardoso and D. Palomar. 'A unified Framework for Structured Graph Learning via Spectral Constraints'. In: *Journal of Machine Learning Research* 21 (2020), pp. 1–60.

[66]   G. Lauga, E. Riccietti, N. Pustelnik and P. Gonçalves. 'Méthodes proximales multi-niveaux pour la restauration d'images'. In: *GRETSI'22 - 28ème Colloque Francophone de Traitement du Signal et des Images*. Nancy, France, Sept. 2022. URL: `https://inria.hal.science/hal-03739496`.

[67]   Q.-T. Le, E. Riccietti and R. Gribonval. 'Spurious Valleys, NP-hardness, and Tractability of Sparse Matrix Factorization With Fixed Support'. In: *SIAM Journal on Matrix Analysis and Applications* (2022). URL: `https://hal.science/hal-03364668`.

[68]   Q.-T. Le, L. Zheng, E. Riccietti and R. Gribonval. 'Fast learning of fast transforms, with guarantees'. In: *ICASSP 2022 - IEEE International Conference on Acoustics, Speech and Signal Processing*. This paper is associated to code for reproducible research available at `https://hal.inria.fr/hal-03552956`. Singapore, Singapore, May 2022. DOI: `10.1109/ICASSP43922.2022.9747791`. URL: `https://inria.hal.science/hal-03438881`.

[69]   T. Moreau, M. Massias, A. Gramfort, P. Ablin, P.-A. Bannier, B. Charlier, M. Dagréou, T. Dupre la Tour, G. Durif, C. F. Dantas et al. 'Benchopt: Reproducible, efficient and collaborative optimization benchmarks'. In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 25404–25421.

[70]   A. Rahimi and B. Recht. 'Random features for large-scale kernel machines'. In: 2007.

[71]   F. Roosta-Khorasani and M. Mahoney. 'Sub-sampled Newton methods'. In: *Math. Program.* 174 (2019), pp. 293–326. DOI: `10.1007/s10107-018-1346-5`.

[72]   D. Shuman, S. Narang, P. Frossard, A. Ortega and P. Vandergheynst. 'The Emerging Field of Signal Processing on Graphs'. In: *IEEE Signal Processing Magazine* (May 2013), pp. 83–98.

[73]   B. K. Sriperumbudur, A. Gretton, K. Fukumizu, B. Schölkopf and G. R. G. Lanckriet. 'Hilbert Space Embeddings and Metrics on Probability Measures.' In: *JMLR* 11 (2010), pp. 1517–1561. URL: `http://dblp.org/rec/journals/jmlr/SriperumbudurGFSL10`.

[74]   P. Stock and R. Gribonval. 'An Embedding of ReLU Networks and an Analysis of their Identifiability'. In: *Constructive Approximation* 57 (2023), pages 853–899. DOI: `10.1007/s00365-022-09578-1`. URL: `https://hal.science/hal-03292203`.

[75]   I. Tosic and P. Frossard. 'Dictionary Learning'. In: *IEEE Signal Processing Magazine* 28.2 (), pp. 27–38. DOI: `10.1109/MSP.2010.939537`. URL: `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5714407`.

[76]   E. Vincent, R. Gribonval and C. Févotte. 'Performance measurement in blind audio source separation'. In: *IEEE Transactions on Audio, Speech and Language Processing* 14.4 (2006), pp. 1462–1469. URL: `https://inria.hal.science/inria-00544230`.

[77]   Y. Wang, Z. Jiang, X. Chen, P. Xu, Y. Zhao, Y. Lin and Z. Wang. 'E2-train: Training state-of-the-art cnns with over 80% energy savings'. In: *Advances in Neural Information Processing Systems*. 2019, pp. 5138–5150.

[78]   G. Yang, T. Zhang, P. Kirichenko, J. Bai, A. G. Wilson and C. De Sa. 'SWALP: Stochastic weight averaging in low precision training'. In: *International Conference on Machine Learning*. 2019, pp. 7015–7024.

[79]   Z. Yao, A. Gholami, S. Shen, K. Keutzer and M. W. Mahoney. 'ADAHESSIAN: An adaptive second order optimizer for machine learning'. In: *arXiv preprint arXiv:2006.00719* (2020).

[80] J. Ying, J. V. de M. Cardoso and D. Palomar. 'Nonconvex Sparse Graph Learning under Laplacian Constrained Graphical Model'. In: *34th Conference on Neural Information Processing Systems.* 2020.

[81] L. Zheng, E. Riccietti and R. Gribonval. 'Efficient Identification of Butterfly Sparse Matrix Factorizations'. In: *SIAM Journal on Mathematics of Data Science* (2022). URL: https://inria.hal.scien ce/hal-03362626.