

2025 Activity Report

RESEARCH CENTRE: Inria Lyon Centre

IN PARTNERSHIP WITH: Ecole normale supérieure de Lyon, Université Claude Bernard (Lyon 1)

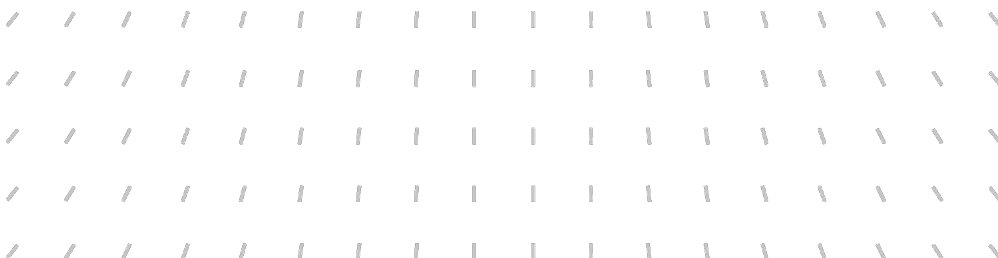

Project-Team

OCKHAM

Optimization, pHysical Knowledge, Algorithms and
Models



In collaboration with Laboratoire de l'Informatique du Parallélisme (LIP)



Project-Team OCKHAM

Creation of the Project-Team: 2023 March 01

Each year, Inria research teams publish an Activity Report presenting their work and results over the reporting period. These reports follow a common structure, with some optional sections depending on the specific team. They typically begin by outlining the overall objectives and research programme, including the main research themes, goals, and methodological approaches. They also describe the application domains targeted by the team, highlighting the scientific or societal contexts in which their work is situated. The reports then present the highlights of the year, covering major scientific achievements, software developments, or teaching contributions. When relevant, they include sections on software, platforms, and open data, detailing the tools developed and how they are shared. A substantial part is dedicated to new results, where scientific contributions are described in detail, often with subsections specifying participants and associated keywords. Finally, the Activity Report addresses funding, contracts, partnerships, and collaborations at various levels, from industrial agreements to international cooperations. It also covers dissemination and teaching activities, such as participation in scientific events, outreach, and supervision. The document concludes with a presentation of scientific production, including major publications and those produced during the year.

Keywords

Computer sciences and digital sciences

- A3.5. – Social networks
- A3.5.1. – Analysis of large graphs
- A5.3.2. – Sparse modeling and image representation
- A5.8. – Natural language processing
- 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
- A9.2.1. – Supervised learning
- A9.2.4. – Optimization and learning
- A9.2.6. – Neural networks
- A9.2.7. – Kernel methods
- A9.2.8. – Deep learning
- A9.11. – Generative AI

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

Contents

Project-Team OCKHAM	1
1 Team members, visitors, external collaborators	5
2 Overall objectives	6
3 Research program	6
3.1 Axis 1: Sparsity for high-dimensional learning.	7
3.2 Axis 2: Learning on graphs and learning of graphs.	8
3.3 Axis 3: Dynamic and frugal learning.	9
4 Application domains	10
4.1 Frugal AI on embedded devices	10
4.2 Imaging in physics and medicine	10
4.3 Interactions with computational social sciences	10
5 Highlights of the year	10
6 Latest software developments, platforms, open data	11
6.1 Latest software developments	11
6.1.1 skglm	11
6.1.2 Benchopt	11
6.1.3 lazylinop	12
6.1.4 Celer	12
6.1.5 TorchDR	12
6.1.6 FAuST	13
7 New results	13
7.1 Integrating Structured Models in Machine Learning and Signal Processing	13
7.1.1 Physics-informed neural networks	13
7.1.2 Differentiable and learning-based methods for structure representation: application to sparse precision matrices	14
7.1.3 New penalties and proximal operators	14
7.1.4 Inverse problems for medical imaging	14
7.1.5 Gromov hyperbolicity for tree representation of relational data	15
7.1.6 Contrastive pre-training of transformer encoders for SEEG-based seizure onset zone detection	15
7.2 Deep neural networks : theory and algorithms	15
7.2.1 Mathematics of deep learning: rescaling invariances, generalization bounds, and conservation laws	15
7.2.2 Quantized networks: theory and algorithms	17
7.2.3 Sparse regularization, unfolding, and approximation theory	18
7.2.4 Deep sparsity: from hardness to deformable butterfly algorithms	18
7.2.5 Plug and play methods	19
7.2.6 Generative models	19
7.3 Statistical learning, dimension reduction, and privacy preservation	20
7.3.1 Theoretical foundations of compressive learning: sketches, kernels, and optimal transport	20
7.3.2 Practical exploration of sketching and methods with limited resources	20
7.3.3 Dimensionality reduction and optimal transport	20
7.4 Large-scale convex and nonconvex optimization	21
7.4.1 Multilevel schemes for image restoration	21
7.4.2 Stochastic multilevel schemes	21
7.4.3 Reproducible benchmarking of optimization algorithms	22

7.4.4	Algorithms for large scale sparse linear models	22
8	Bilateral contracts and grants with industry	22
8.1	Bilateral grants with industry	22
9	Partnerships and cooperations	23
9.1	International research visitors	23
9.2	National initiatives	23
9.2.1	PEPR IA project : SHARP	23
9.2.2	ANR IA Chaire : AllegroAssai	24
9.2.3	ANR DataRedux	24
9.2.4	ANR JCJC EROSION	24
9.2.5	ANR JCJC MEPHISTO	25
9.2.6	Defi Hive Inria Cupseli	25
9.2.7	DI2A - Subvention Simone et Cino del Duca, Institut de France.	26
9.2.8	GDR ISIS project PROSSIMO	26
9.2.9	ANR TSIA BenchArk	26
9.2.10	ANR SEIZURE	27
10	Dissemination	27
10.1	Promoting scientific activities	27
10.1.1	Scientific events: organisation	27
10.1.2	Scientific events: selection	27
10.1.3	Journal	28
10.1.4	Invited talks	28
10.1.5	Leadership within the scientific community	28
10.1.6	Scientific expertise	28
10.1.7	Research administration	29
10.2	Teaching - Supervision - Juries - Educational and pedagogical outreach	29
10.2.1	Teaching	29
10.2.2	Supervision	29
10.2.3	Juries	29
11	Scientific production	30
11.1	Major publications	30
11.2	Publications of the year	31
11.3	Cited publications	34

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, Associate Professor, from Sep 2025]

Post-Doctoral Fellows

- Alice Brenon [ENS DE LYON]
- Etienne Lasalle [ENS DE LYON, Post-Doctoral Fellow, until Feb 2025]
- Guillaume Lauga [ENS DE LYON, Post-Doctoral Fellow, from Feb 2025 until Mar 2025]
- Hugo Lebeau [INRIA, Post-Doctoral Fellow, from Feb 2025]
- Manon Verbockhaven [ENS DE LYON, Post-Doctoral Fellow, from Dec 2025]

PhD Students

- Giuseppe Carrino [ENS DE LYON, from Nov 2025]
- Mael Chaumette [INRIA]
- Edgar Desainte-Mareville [ENS DE LYON]
- Anne Gagneux [UNIV LYON I]
- Arthur Lebourrier [ENS DE LYON]
- Sibylle Marcotte [ENS PARIS]
- Can Pouliquen [ENS DE LYON]

Technical Staff

- Pascal Carrivain [INRIA, Engineer]

Interns and Apprentices

- Ilias Bouhss [ENS DE LYON, Intern, from Jun 2025 until Aug 2025]
- Giuseppe Carrino [ENS DE LYON, Intern, from Mar 2025 until Jun 2025]
- Chady Essouabri [CNRS, Intern, from May 2025 until Aug 2025]
- Florian Kozikowski [CNRS, Intern, from Mar 2025 until Aug 2025]
- Damien Rouchouse [INRIA, Intern, from Apr 2025 until Sep 2025]

Administrative Assistant

- Emilie Gatignol [INRIA]

Visiting Scientist

- Laurent Jacques [Univ UCLouvain, from Sep 2025]

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 [66] 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 [68, 60]. The flexibility of the models, which can incorporate learned dictionaries [100], as well as structured and/or low-rank variants of the now-classical sparse modeling paradigm [78], 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 [56, 63], 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 [61], 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 [67, 80] 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 [69] 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. [79]) 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 [57], where natural models are stochastic block models rather than strictly multi-component graphs (e.g. Gaussian mixtures models). As it is done in [105], 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 [74] and [62] 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) [97] [8], 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 [71]. Hence, the notion of distributed GSP [64, 65] 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 [95]) 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 [74] as well as privacy preservation guarantees [62]. A sketch can be seen as an embedding of the empirical probability distribution of the dataset with a particular form of kernel mean embedding [98]. 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 [103, 102], 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 [96, 104].

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 resources (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 Highlights of the year

The paper “On the closed form of flow matching: generalization does not arise from stochasticity” [1] was accepted as an oral presentation at NeurIPS 2025 (top 0.3% of more than 22000 submissions).

The paper “Transformative or conservative? Conservation laws for ResNets and Transformers” [26] was accepted as an oral presentation at ICML 2025 (top 1% of about 12000 submissions)

The paper “Rapture of the deep: highs and lows of sparsity in a world of depths” [4] has been accepted in the Signal Processing Magazine.

Antoine Gonon, former Ph.D. student of the Ockham team, was awarded a *Honorable mention (2nd ex-aequo) of the 2025 Ph.D. award of the Société Savante Francophone en Apprentissage Machine*.

6 Latest software developments, platforms, open data

6.1 Latest software developments

6.1.1 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](https://hal.archives-ouvertes.fr/hal-03819082)

Contact: Mathurin Massias

Participant: 2 anonymous participants

6.1.2 Benchopt

Keywords: Benchmarking, Machine learning, Optimization

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 as 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](https://hal.archives-ouvertes.fr/hal-03830604)

Contact: Thomas Moreau

Participant: 4 anonymous participants

6.1.3 lazylinop

Name: lazylinop

Keywords: Signal processing, Numerical algorithm, Scientific computing

Scientific Description: lazylinop is an easy way to combine existing operators into more complex operators with direct access to its adjoint.

Functional Description: Lazy evaluation of linear operators applied to vectors or matrices. lazylinop aims at providing an easy way to combine existing operators into more complex operators with direct access to its adjoint. Thanks to the lazy computation paradigm, lazylinop offers potential performances gains and memory sparing.

Release Contributions: - Basic linear operators: Kronecker product, addition, diagonal, block-diagonal, concatenation ... - Polynomial of linear operators. - Usual signal processing linear operators. - Usual image processing linear operators. - Butterfly linear operators. - Near optimal Butterfly (real values) quantification. - Lazylinop operators take as input NumPy/CuPy arrays or torch tensors (via array-api)
Work-In-Progress: - Near optimal Butterfly (complex values) quantification.

URL: <https://lazylinop.inria.fr>

Contact: Pascal Carrivain

Participant: 4 anonymous participants

6.1.4 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

Participant: 2 anonymous participants

6.1.5 TorchDR

Keywords: Optimal transportation, Machine learning, Dimensionality reduction, High Dimensional Data

Scientific Description: TorchDR is an open-source dimensionality reduction (DR) library using PyTorch. Its goal is to accelerate the development of new DR methods by providing a common simplified framework.

Functional Description: TorchDR is an open-source dimensionality reduction (DR) library using PyTorch. Its goal is to accelerate the development of new DR methods by providing a common simplified framework.

URL: <https://torchdr.github.io/>

Contact: Titouan Vayer

Participant: 5 anonymous participants

6.1.6 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

Participant: 6 anonymous participants

7 New results

7.1 Integrating Structured Models in Machine Learning and Signal Processing

7.1.1 Physics-informed neural networks

Participants: Elisa Riccietti.

Collaboration with Alena Kopanicakova (IRIT, Toulouse), Stefania Bellavia and Mahsa Yousefi (UNIFI, Florence, 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 postdoc of Mahsa Yousefi, we pursued the work started last year on the investigation of their ability to deal with ill-posed inverse problems, focusing especially on parameter identification problems. We have proposed a two-step training strategy that first fits the available noisy observations and later adds the physics information. The strategy is shown to improve the solution of such ill-posed problems.

In collaboration with Alena Kopanicakova, we have proposed a book chapter on scientific machine learning with a focus on the training of physics-informed neural networks, guided by the neural tangent kernel theory to correct the spectral bias.

7.1.2 Differentiable and learning-based methods for structure representation: application to sparse precision matrices

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

The PhD of Can Pouliquen, defended successfully in December 2025, is devoted to the estimation of structures from signals, such as sparse precision matrices. For the latter problem we have adopted the mathematical framework of the Graphical Lasso, and pursued several directions. We have introduced SpodNet, a new deep neural network architecture for positive definite matrix estimation. In particular, it is the first architecture which can guarantee a simultaneously sparse and symmetric positive definite output. This highly desirable property was so far a missing feature of existing architectures, and has many potential applications in graph learning beyond neurosciences [28]. This work was accepted to ICLR 2025. We have also developed a bilevel optimization framework, that eases the tuning of individual correlation strengths in the Graphical Lasso penalty [94]. Finally, we have proposed a fast and modular benchmark for the Graphical Lasso, together with high quality open source implementations of fast solvers [35].

7.1.3 New penalties and proximal operators

Participants: Anne Gagneux, Remi Gribonval, Mathurin Massias.

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

Finishing the internship work 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). We have derived an algorithm based on the Pool Adjacent Violators Algorithm (PAVA) that computes the exact proximal operator of a first kind of sorted penalties (sorted MCP, sorted Log-sum). We have also extended it to compute the proximal operators of the sorted ℓ_q ($0 < q > 1$) penalties, which presented more difficulties due to non Lipschitzianity. This work has been submitted to IEEE TSP [44].

7.1.4 Inverse problems for medical imaging

Participants: Marion Foare.

Collaboration with Luis Enrique Amador Arya (Creatis, Villeurbanne), H el ene Ratiney (Creatis, Villeurbanne),  Eric 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. In particular, isotropic super-resolution (ISR) techniques consist in reconstructing an isotropic volume from the combination of several anisotropic volumes acquired with different orientations.

In the context of the PhD work of Luis Enrique Amador Araya, we pursued the development of 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 in 3D Magnetic Resonance Imaging (MRI).

In particular, we explored new regularization terms to extend this approach to multi-contrasts ISR, that is, to reconstruct isotropic and multi-contrasts high resolution images from multi-contrasts anisotropic acquisitions. Preliminary results were accepted for publication at the conference ISBI 2026.

7.1.5 Gromov hyperbolicity for tree representation of relational data

Participants: Titouan Vayer.

Collaboration with Pierre Houedry, Nicolas Courty, Florestan Martin-Baillon, Laetitia Chapel from Université Bretagne Sud.

Trees and the associated shortest-path tree metrics provide a powerful framework for representing hierarchical and combinatorial structures in data. Designing algorithms that can produce a tree from pairwise relationship between data points is a vivid subject of interest. However, most common approaches are either heuristical and lack guarantees, or perform moderately well. In [24] we develop a geometrical framework for learning such trees, based on the notion of Gromov hyperbolicity, that encodes to which extent a metric space deviate from a tree structure. We introduce a novel differentiable optimization framework, coined DeltaZero, that solves this problem. Experiments on synthetic and real-world datasets demonstrate that our method consistently achieves state-of-the-art distortion. This work was accepted in NeurIPS 2025.

7.1.6 Contrastive pre-training of transformer encoders for SEEG-based seizure onset zone detection

Participants: Paulo Goncalves.

Collaboration with Pierre Borgnat (ENS de Lyon), Julien Jung (Hôpital Neurologique, HCL, CRNL).

Within the context of his Master 2 internship, Zacharie Rodière pursued the work of Gaetan Frusque, a former PhD student in our group [70] on the clinical study of epilepsy. Zacharie developed a transformer encoder for the detection of Seizure Onset Zone (SOZ) from stereo-EEG. It integrates clinically grounded time-frequency features with spatial contrastive pre-training. While prior spatial transformer approaches analyse learned representations, the proposed method uniquely combines: (1) engineered time-frequency representations (TFRs) encoding epileptic spikes and oscillations, and (2) a contrastive objective leveraging anatomical relationships between the electrode contacts that are either inside the SOZ or outside. Attention heads provide interpretable connectivity patterns, bridging data-driven learning with the study of functional connectivity networks. Zacharie presented his preliminary results at the Graph Signal Processing Workshop 2025 [37].

7.2 Deep neural networks : theory and algorithms

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

Participants: Rémi Gribonval, Elisa Riccietti, Sibylle Marcotte, Arthur Lebeurier, Titouan Vayer.

Collaborations with Nicolas Brisebarre (ARIC team, ENS de Lyon), and with Gabriel Peyré (DMA, ENS, Paris)

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 in [99] was invariant to such scalings but limited to strictly layered ReLU architectures. In the context of the PhD of Antoine Gonon [72] (defended on 12/11/2024), we extended it [73] 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 complexity 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. Building on this toolkit, we more recently investigated a rescaling-invariant Lipschitz bound on the mapping from parameter space to function space and illustrated its potential for neural network pruning and quantization [22] in a paper published at ICML 2025.

Conservation laws. In the thesis of Sibylle Marcotte (defended on 21/11/2025), the above path-embedding also served as a key enabler for the analysis of conservation laws in gradient descent dynamics of ReLU networks [91]. Understanding the geometric properties of gradient descent dynamics is indeed 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.

Our initial work on this topic [91] was conducted with a motivation that 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.

We then studied [92] the notion of conservation law and the corresponding algorithms for optimization flows associated to non-Euclidean geometries and momentum-based dynamics. We characterized "all" conservation laws in this general setting. In stark contrast to the case of gradient flows, we proved that the conservation laws for momentum-based dynamics exhibit *temporal dependence*. Additionally, we often observed a "conservation loss" when transitioning from gradient flow to momentum dynamics. Specifically, for linear networks, our framework allowed us to identify all momentum conservation laws, which are less numerous than in the gradient flow case except in sufficiently over-parameterized regimes. With ReLU networks, no conservation law remains. This phenomenon also manifests in non-Euclidean metrics, used e.g. for Nonnegative Matrix Factorization (NMF): all conservation laws can be determined in the gradient flow context, yet none persists in the momentum case.

This year, we extended the analysis [26] to extensively cover ResNets and attention layers. For this, we first showed that basic building blocks such as ReLU (or linear) shallow networks, with or without convolution, have easily expressed conservation laws, and no more than the known ones. In the case of a single attention layer, we also completely described all conservation laws, and we showed that residual blocks

have the same conservation laws as the same block without skip connection. We then introduce the notion of conservation laws that depend only on a subset of parameters (corresponding e.g. to a pair of consecutive layers, to a residual block, or to an attention layer). We demonstrate that the characterization of such laws can be reduced to the analysis of the corresponding building block in isolation. Finally, we examined how these newly discovered conservation principles, initially established in the continuous gradient flow regime, persist under discrete optimization dynamics, particularly in the context of Stochastic Gradient Descent (SGD).

This year we investigated the consequences of conservation laws to characterize whether a (path)lifted representation has an intrinsic training dynamics [45], as a stepping stone to so-called implicit bias analysis. We expressed a so-called intrinsic dynamic property and showed how it is related to the study of conservation laws associated with the lifting function. This led to a simple criterion based on the inclusion of kernels of linear maps which yields a necessary condition for this property to hold. Applying our theory to general ReLU networks of arbitrary depth, with the path lifting, we showed that the dynamic is intrinsic for *any* initialization. In the case of linear networks with a natural lifting defined as the product of weight matrices, so-called *balanced* initializations were also known to enable such an intrinsic dynamic; we generalized this result to a broader class of *relaxed balanced* initializations, showing that, in certain configurations, these are the *only* initializations that ensure the intrinsic dynamic property. Finally, for the linear neural ODE associated with the limit of infinitely deep linear networks, with relaxed balanced initialization, we explicitly expressed the corresponding intrinsic dynamics.

Path-conditioning for faster training. Finally in the context of the PhD thesis of Arthur Lebeurrier, we are investigating how to leverage the path-lifting framework to better understand the dynamic of neural networks and to eventually accelerate the training of the parameters. We plan to submit this work for ICML 2026.

7.2.2 Quantized networks: theory and algorithms

Participants: Rémi Gribonval, Elisa Riccietti, Giuseppe Carrino, Mael Chaumette

Collaboration with Nicolas Brisebarre (ARIC team, ENS de Lyon), with Silviu Filip and El-Mehdi El arar (IRISA, Rennes), and with Theo Mary (LIP6, Paris)

Motivated by the importance of quantizing networks besides pruning them to achieve sparsity, we studied different aspects related to this topic.

Quantization of neural networks: the multi-linear case As a first step towards a better understanding of nonlinear quantized networks, we studied the simpler multi-linear case. Particularly, we investigated the problem of optimally quantizing low rank matrices by exploiting scaling invariances inherent to the optimization problem. We proposed [76, 77] 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 [90], to design a heuristic strategy to efficiently quantize the family of butterfly matrices, which very often occur in fast transforms and 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. The corresponding algorithms have been incorporated in the quantization module of the lazylinop library 6.1.3.

In the context of the thesis of Mael Chaumette we extended this approach to complex valued matrices [30]. This extension is important since most of the fast transforms that involve butterfly matrices, such as the Fourier transform, are complex valued and cannot be quantized by the previously proposed strategy. Building this extension has not been straightforward from the real case: this raised new questions and required to propose new algorithms. A journal version is in preparation as well as an implementation in the lazylinop library 6.1.3.

Quantization of neural networks: mixed-precision inference In order to further exploit the benefits of quantization in neural networks and the multiple reduced numerical formats made available by modern computer architectures, we studied the introduction of mixed precision in the inference of neural networks

[42]. We proposed an analysis on the propagation of the error in the forward pass of neural networks, which suggests a good rule to choose the numerical format of each line of the weight matrices, yielding a mixed-precision procedure that provides the same accuracy of classical inference but with a lower energy consumption.

Quantization of neural networks: mixed-precision training As a first step towards a mixed precision training of neural networks, in the context of the master internship and of the PhD thesis of Giuseppe Carrino, we have studied the convergence theory of the Newton’s method in finite precision [38]. This analysis allows for understanding the impact of the different errors on the convergence and thus to guide the choice of the precision in each step of the method, leading to a mixed-precision algorithm. Further research will deal with an extension to the stochastic case, which would be adapted to the training of neural networks.

7.2.3 Sparse regularization, unfolding, and approximation theory

Participants: Marion Foare.

Collaborations with Nelly Pustelnik (Physics lab, ENS de Lyon) and Audrey Repetti (Heriot-Watt University, Edinburgh).

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. Currently, we are studying various unrolling approaches to tackle the joint task of image restoration and edge detection. First, we proposed a two-step procedure mimicking the Blake-Zisserman minimization strategy, and relying on a smoothing Proximal Neural Network, followed by an edge detection layer ([85]).

On the other hand, we are working on the unrolling procedure of the (non-convex) Mumford-Shah model, which allows to jointly perform image restoration and edge detection using a single model-based proximal neural network. The proposed architecture is significantly lighter than recent learning models designed only for edge detection, both in terms of number of learnable parameters and inference time. This work was published in Eusipco 2025 [86].

7.2.4 Deep sparsity: from hardness to deformable butterfly algorithms

Participants: Rémi Gribonval, Elisa Riccietti, Pascal Carrivain.

Collaboration with Leon Zheng (Huawei), Quoc-Tung Le (TSE, Toulouse)

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.

We have deeply investigated this subject in the last years in the context of the thesis of Quoc-Tung Le [87] and Léon Zheng [106].

Building on this series of work on the hardness, tractability, and uniqueness properties of sparse matrix factorizations under various sparsity constraints [108, 89, 90], we prepared this year a tutorial paper [4] for the signal processing magazine (SPM) Special Issue ”Mathematics of Deep Learning”, in which we propose an overview on the role of sparsity in a deep learning context.

This work includes our previous results on the subject.

First of all, it includes the extension of 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 butterflies*. We have studied its performance guarantees beyond the case of matrices admitting an exact factorization [17]. The corresponding algorithm has been incorporated in the lazylinop software library 6.1.3.

Second, it includes also our study on the understanding on how to fully exploit the specific structure of butterfly factors and translate it into practical time gains, published at ICML 2025 [23]. Specifically, we have studied how to optimize memory access to the matrix elements and we implemented a CUDA kernel to multiply on GPU a dense matrix with a deformable butterfly factor. This is also available in lazylinop 6.1.3. In the paper we benchmark our implementation against existing matrix-vector multiplication algorithms to select the optimal one.

Going beyond the linear case, the paper also includes our results on neural networks. We have indeed shown that the pitfalls that we had identified for certain sparse matrix factorization problems [90] also hold for certain sparse ReLU neural network training problems [88]. In particular, there exist settings where the optimization is necessarily instable, in the sense that minimizing the loss function can only be achieved by letting some coefficients diverge to infinity.

Finally, the paper includes also our developed heuristics to handle butterfly approximations for matrices under unknown permutations of rows and/or columns [107].

7.2.5 Plug and play methods

Participants: Elisa Riccietti, Rémi Gribonval, Mathurin Massias, Anne Gagneux.

Collaboration with Emmanuel Soubies (CNRS, IRIT), Nelly Pustelnik and Julian Tachella (CNRS, ENS Lyon), Nils Laurent (LASPI Roanne)

In imaging tasks, Plug and Play (PnP) methods leverage the strength of pre-trained denoisers, often deep neural networks, by integrating them in optimization schemes, ensuring better reconstructions than classical variational methods.

In the early PhD work of Anne Gagneux, we have investigated the use of neural networks to implement convex functions. Learning convex functions has many applications in imaging (notably in Plug and Play methods) and in optimal transport. In [13] we have studied the expressive power of Input Convex Neural Networks (ICNNs), a special architectural constraint. In particular, we have shown that ICNNs are restrictive, and may require more neurons than unconstrained networks to implement a given convex function.

One of the main pitfalls of PnP methods is their slow rate of convergence and high computational cost. To overcome this, in the context of the postdoc of Nils Laurent, we have studied the use of multilevel schemes in conjunction with plug and play (PnP) methods. Since 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. We have proposed [18] a multilevel PnP method that leverages images of smaller sizes and lighter denoisers at coarse levels.

7.2.6 Generative models

Participants: Anne Gagneux, Rémi Gribonval, Mathurin Massias.

Collaboration with Quentin Bertrand, Rémi Emonet (INRIA Malice, Université Jean Monnet), Ségolène Martin, Paul Hagemann, Gabriele Steidl (TU Berlin).

Since mid 2024, the team has started to study generative modelling, with an initial focus on diffusion and flow matching methods for image generation. In [6] (work done a summer internship at TU Berlin), OCKHAM PhD student Anne Gagneux has proposed to use generative models, namely flow matching, in the PnP framework. This is achieved by defining a time-dependent denoiser using a pre-trained FM model. The algorithm alternates between gradient descent steps on the data-fidelity term, reprojections onto the learned FM path, and denoising. On tasks such as denoising, super-resolution, deblurring, and inpainting, the algorithm demonstrates superior results compared to existing PnP algorithms and Flow Matching based state-of-the-art methods. The algorithm has been released publicly on [GitHub](#).

In a collaboration with members of the Inria MALICE team, we have written an introductory blog post on flow matching, with is now considered as one of the reference materials on the topic [51].

In [1], we have shown that perfectly trained flow matching (and diffusion) models admit a closed-form solution, which can only generate points from their training data. We have shown that these models produce new data when they fail to perfectly learn their target, and that failure at small generation times was particularly important. This work was accepted as an oral presentation at NeurIPS 2025 (top 0.3% of submitted papers). We have pursued this research direction in [43], adopting a denoising perspective on the task of generating images: complementary to [6], we show how to build a generative model from a denoiser, and leverage this framework to produce new insights on the generation dynamics of flow matching.

We are currently pursuing several directions: conditional generation (text-to-image models), links with optimal transport through Schrodinger bridges, discrete flow matching for text generation, and application to molecule discovery.

7.3 Statistical learning, dimension reduction, and privacy preservation

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

Participants: Hugo Lebeau, Rémi Gribonval, Titouan Vayer.

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 [2], then for the specific case of compressive k -means and compressive Gaussian Mixture Modeling [75]. The overall framework is described in a tutorial paper [3].

Theoretical guarantees in compressive learning fundamentally rely on comparing certain metrics between probability distributions as explored in a previous paper [10]. Preliminary works on the relations between sketching and random matrix theory were conducted this year. We began to investigate the sharpness of the existing theoretical guarantees by looking at different metrics between probability distributions, which naturally arise when ones try to bound the excess risk of sketching methods.

7.3.2 Practical exploration of sketching and methods with limited resources

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

Collaborations with Rémi Vaudaine (previously postdoctoral researcher), Marton Karsai (CEU, Vienne, Austria) and Pierre Borgnat (Physics Lab, ENS deLyon)

We explored the sketching approach in the context of graph clustering, a key task in graph analysis. Many methods, like spectral clustering, are impractical for large graphs due to computational constraints. To address this, we introduced PASCO in [16], a sketching-based overlay that accelerates clustering algorithms. PASCO involves: 1- generating small, structure-preserving coarse graphs from the input graph, 2- running clustering algorithms in parallel on these graphs to produce partitions, and 3- aligning and merging these partitions using optimal transport. The PASCO framework is based on two key contributions: a novel global algorithm structure designed to enable parallelization and a fast, empirically validated graph coarsening algorithm that preserves structural properties. This work was published in the journal Machine Learning, 2025 and presented at ECML-PKDD 2025.

7.3.3 Dimensionality reduction and optimal transport

Participants: Titouan Vayer, Etienne Lasalle.

Collaborations with Franck Picard (DR CNRS, ENS Lyon), Chady Essouabri (intern, ENS Lyon), 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), Pascal Frossard (EPFL).

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. In a previous work [101], we introduced and explored an innovative nonlinear dimensionality reduction method by utilizing the optimal transport framework and entropic affinities.

Building on these results, we extended our work to generalize dimension reduction, as detailed in [9], accepted at TMLR 2025. Our approach leverages OT, specifically the Gromov-Wasserstein distance (GW), to propose a framework that simultaneously reduces both the dimensionality and the number of points in a dataset, enabling significant data compression. Notably, when the number of points is preserved, we demonstrated strong connections between our method and traditional dimensionality reduction techniques, such as spectral methods and t-SNE. We refer to our framework as "Distributional Dimension Reduction" which can be interpreted as projecting a distribution, and a geometry encoding the relationships among data points in high-dimensional space, into a lower-dimensional space using the GW perspective. Based on these principles, we developed a library for dimensionality reduction in Pytorch 6.1.5. Finally, we investigated the relations between OT and mixture models, and write a small tutorial on the subject in [48]. These works are at the core of further research on OT and self-supervised learning methods, as explored during the intership of Chady Essouabri in collaboration with Franck Picard.

7.4 Large-scale convex and nonconvex optimization

7.4.1 Multilevel schemes for image restoration

Participants: Elisa Riccietti, Paulo Gonçalves, Edgar Desainte-Mareville

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

In the context of the Ph.D. work of Guillaume Lauga (defended on the 18/12/2024), we studied the combination of multilevel schemes and proximal methods [5, 83, 84, 81, 82]. Pushing further in this direction, we studied the link between multilevel and block coordinate methods and their convergence analysis [33]. This line of research is also the object of the PhD thesis of Edgar Desainte-Mareville. Its aim is to investigate how to unroll such multilevel strategies in order to learn important ingredients such as the transfer operators. In order to do that, an improved understanding of the link between multilevel and block methods is essential.

7.4.2 Stochastic multilevel schemes

Participants: Elisa Riccietti.

Collaboration with Margherita Porcelli (UNIFI, Firenze, Italy) and Filippo Marini (UNIBO, Bologna, Italy)

Classical deterministic multilevel schemes are limited by the need of regularly handling the high level expensive objective function and are usited to solve stochastic problems such as expected risk minimization. We proposed a stochastic extension of the multilevel framework [46] that does not require the finest approximation to coincide with the original objective function along all the optimization process. This allows

for significantly decreasing the cost of the multilevel paradigm, for instance in data-fitting problems, where considering all the data at each iteration can be avoided.

7.4.3 Reproducible benchmarking of optimization algorithms

Participants: Mathurin Massias, Florian Kozikowski.

Collaboration with Thomas Moreau (MIND, Inria Saclay), Badr Moufad (Ecole Polytechnique), Nelly Pustelnik (CNRS, ENS de Lyon).

The team continues working on reproducible optimisation benchmarks, with Benchopt [7], a collaborative framework to automate, reproduce and publish 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. In 2025, thanks to the internship of Florian Kozikowski, we implemented new benchmarks (Poisson regression). We are currently planning to develop benchmarks related to generative models.

7.4.4 Algorithms for large scale sparse linear models

Participants: Mathurin Massias.

Collaboration with Quentin Bertrand (INRIA MALICE), Badr Moufad (Ecole Polytechnique)

Based on our seminal works in [93] and [59], 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 [58], implemented in the `skglm` package (see Section 6.1.1), that was integrated into the ecosystem of the `scikit-learn` package. In 2025, the internship work of Florian Kozikowski allowed implementing new solvers (Poisson, Group Poisson and Gamma regression) as well as a complete rewriting of the [documentation](#).

8 Bilateral contracts and grants with industry

8.1 Bilateral grants with industry

- CIFRE contract with CNES, Paris on "Optimized on-board decision with fast energy-efficient neural networks". This PhD thesis is in collaboration with Stéphane May, engineer at CNES.

Participants: Rémi Gribonval, Titouan Vayer, Arthur Lebeurrer.

Duration: 3 years (2024-2027)

Partners: CNES, Paris; ENS de Lyon

Funding: CNES, Paris; PEPR IA SHARP

Context: ANR Chaire IA AllegroAssai [9.2.2](#)

This thesis aims to develop compact, high-performance neural networks tailored to on-board constraints, enabling optimized decision-making on low-energy platforms. It includes an exploration of parsimony structures suited for deep networks and a comprehensive study of quantization and optimization techniques for neural networks.

- Funding from Facebook Artificial Intelligence Research, Paris

Participants: Rémi Gribonval.

Duration: 5 years (2021-2025)

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

Funding: Facebook Artificial Intelligence Research, Paris

Context: Chaire IA AllegroAssai [9.2.2](#)

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

9 Partnerships and cooperations

9.1 International research visitors

Laurent JACQUES

Status: researcher

Institution of origin: Université de Louvain

Country: Belgium

Dates: Sept. 1, 2025 till June 30, 2026

Context of the visit: Inria chair from the Collegium of Lyon

Mobility program/type of mobility: sabbatical

9.2 National initiatives

9.2.1 PEPR IA project : SHARP

Participants: Rémi Gribonval (*correspondant*), Paulo Goncalves, Elisa Ricietti, Marion Foare, Mathurin Massias, Titouan Vayer, Arthur Lebourrier, Mael Chaumette.

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

Duration of the project: **2023 - 2029**.

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).

9.2.2 ANR IA Chaire : AllegroAssai

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

9.2.3 ANR DataRedux

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

Collaboration with Marton Karsai (former PI, ECU Austria), Pierre Borgnat (ENS de Lyon)

Duration of the project: **February 2020 - January 2024 prolonged to March 31, 2026.**

DataRedux puts forward an innovative framework to reduce networked data complexity while preserving its richness, by working at intermediate scales (“mesoscales”). Our objective was to contribute to the theoretical understanding and representation of rich and complex networked datasets for use in predictive data-driven models. Our main novelty has been to define network reduction techniques in two particular usecases: one in relation with the dynamical processes occurring on the networks, and the second related to the clustering of large size graphs. Both approaches relied on the extracting information and knowledge at different scales in a human-accessible way by extracting structures from high-resolution, diverse and heterogeneous data.

Our guideline in the DataRedux project was to identify methods for aggregating data at intermediate scales and new types of data representations related to dynamic processes, which preserve the richness of information contained in the original data, while retaining their most relevant models for easy integration into data-based digital models to facilitate decision-making and obtain actionable information.

9.2.4 ANR JCJC EROSION

Participants: Mathurin Massias.

Duration of the project: **December 2023 - December 2026.**

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.

In OCKHAM, this collaboration has led to the internship of Anne Gagneux (co-supervised with Emmanuel Soubies), on the design of new sorted non-convex penalties and the computation of their proximal operators.

9.2.5 ANR JCJC MEPHISTO

Participants: Elisa Riccietti (*correspondant*).

Duration of the project: **November 2024 - November 2028.**

This project focuses on large scale optimization problems in signal processing and imaging. We consider a special class of such problem: those that admit a hierarchical structure. The aim of the project is to develop parsimonious methods for their solution by exploiting such underlying structure. We will focus on four different kinds of hierarchical structures: those arising from the geometry or physics of the problem (such as multiple resolutions in images or discretization of infinite dimensional problems); those that can be built by exploiting the analytical structure of some problems (training of neural networks, data-fitting problems); those that can be built exploiting the intrinsic structure of the algebraic tools involved (matrix, tensors, such as in matrix factorization problems); those that can be built exploiting multiple numerical formats (floating point numbers with reduced number of bits) .

The ambition of this project is thus to develop a large family of parsimonious multiresolution, multilevel and multiprecision algorithms that are not only efficient but that can also rely on solid mathematical foundations.

9.2.6 Defi Hive Inria Cupseli

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

Duration of the project: **September 2025-September 2028.**

The Cupseli challenge aims to demonstrate that it is possible to run complex applications on heterogeneous, distributed, and volatile resources, while achieving good parallel efficiency and preserving both accuracy and confidentiality. It explores algorithmic and system-level solutions to optimize computation, memory, and communication, while ensuring security and fault tolerance. The work is organized around three main axes: Frugality (adapting training and inference to limited and dynamic resources), Security and confidentiality (protecting data and models through encryption, secure enclaves, and defenses against attacks), and Volatility (ensuring robustness and performance despite the unpredictable arrival and departure of resources).

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

Participants: Elisa Riccietti, Marion Foare, Paulo Goncalves.

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 circumstellar environments through the instrument SPHERE/IRDIS. We aim to propose innovative reconstruction approaches partially supervised or even non supervised.

9.2.8 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?

9.2.9 ANR TSIA BenchArk

Participants: Mathurin Massias (*correspondant*).

Duration of the project: **October 2024 - October 2028.**

Collaboration with Thomas Moreau, Gaël Varoquaux (INRIA Saclay) and Joseph Salmon (INRIA Montpellier).

Numerical evaluation of novel methods, a.k.a. benchmarking, is a pillar of the scientific method in machine learning. However, due to practical and statistical obstacles, the reproducibility of published results is currently insufficient: many details can invalidate numerical comparisons, from insufficient uncertainty quantification to improper methodology. In 2022, the **Benchopt** initiative provided an open source Python package together with a framework to seamlessly run, reuse, share and publish benchmarks in numerical optimization. The BenchArk project aims at bringing Benchopt to the whole machine learning community, making it a new standard in benchmarking by empowering researchers and practitioners with efficient and valid benchmarking methods. Our goal is to ensure reproducibility and consistency in model evaluation. We will federate the machine learning community to develop informative and statistically valid benchmarks, while providing methods to reduce identified hurdles in implementing such practices.

9.2.10 ANR SEIZURE

Participants: Paulo Goncalves (*correspondant*), Can Pouliquen.

Duration of the project: **September 2024 - August 2028**

Collaboration with Carole Lartizien (PI of the project, CNRS, Insa de Lyon, CREATIS), Julien Jung (MD-PhD, Hospices Civils de Lyon, CRNL), Pierre Borgnat (CNRS, ENS de Lyon, Physics Lab).

“Seeing the Epileptogenic Zone through machine Learning on structural, functional and clinical neurological data”

This project deals with the multimodal detection and the characterisation of epileptic zones in neuroimaging and intracranial EEG (iEEG). Ockham is mainly involved in WP3 (P. Borgnat leader) that aims at analysing the propagation of biomarkers within the brain as an indicator of the dynamic interictal epileptogenic network. A detailed understanding of the brain network and its key hubs provides invaluable insights into surgical outcomes. In a previous PhD work (G. Frusque, 2017-2020) we derived graphical lasso techniques on iEEG data to infer graphs times series, as relevant connectivity networks. In Seizure, we envision to enrich our previous approaches with deep learning based models and more specifically with graph recurrent neural networks and neural implicit representations.

10 Dissemination

10.1 Promoting scientific activities

10.1.1 Scientific events: organisation

- Organization of the **COLT 2025 conference**, (30/06/25 – 04/07/25), Remi Gribonval
- Organization of the **GDR IASIS Thematic Day on Flow matching, diffusion and their applications** (24/10/25) Mathurin Massias
- Organization of the **GDR IASIS Thematic Day on optimal transport and machine learning** (17/02/25) Titouan Vayer
- **SMAI minisymposium** on generative modelling, optimal transport and image restoration, (02/06/25 – 06/06/25) Mathurin Massias
- **One-day workshop Sharp and Foundry: On Frugal and Robust Foundations for Machine Learning**, ENS Lyon (30/06/25) Remi Gribonval

10.1.2 Scientific events: selection

Member of the conference program committees

- Mathurin Massias – Area Chair for NeurIPS, ICML.
- Titouan Vayer – Member of the GRETSI program committee, area chair for ICML.
- Rémi Gribonval
 - **MIA’25 Program Committee**;
 - Organizer of a Minisymposium on "Mathematical aspects of deep learning", **Curves & Surfaces 2026**, St-Malo, June 8-12 2026;
 - Scientific board of **JRAF** (Journées de recherche en apprentissage frugal), Grenoble, Nov 26-27 2025;
 - Scientific board of a **workshop on the Mathematics of AI**, Institut de mathématiques de Bordeaux, Nov 4-6 2026

Organization of the weekly "**Machine Learning and Signal Processing (MLSP)**" seminar (about twenty presentations in 2025) Marion Foare; Paulo Goncalves; Remi Gribonval; Mathurin Massias; Elisa Riccietti; Titouan Vayer

10.1.3 Journal

Member of the editorial boards

- Mathurin Massias – Associate Editor for TMLR
- Remi Gribonval – Associate Editor for Constructive Approximation (Springer); founding member of the Editorial Board of Mathematical Foundations of Machine Learning (Springer), Senior Area Editor for the IEEE Signal Processing Magazine

10.1.4 Invited talks

- Elisa Riccietti – Journée AILYS, ENS Lyon, 14/02/2025
- Elisa Riccietti – MIA25 conference, Paris, 13/01/2025-15/01/2025.
- Anne Gagneux – Séminaire Imaging In Paris, 06/05/25
- Mathurin Massias – Séminaire Palaisien, 04/11/25
- Mathurin Massias – Séminaire IMAGINE, 05/11/25
- Remi Gribonval
 - **Rencontre nationale du RT Optimisation**, INSA Lyon, Nov 26-28 2025
 - **Workshop “(Blind) inverse problems in imaging: from foundations to applications”**, CIRM, Luminy, Sep 29-Oct 3 2025
 - **Festum Pi Mathematics Conference**, Chania, Crete, July 21-25 2025;
 - Workshop on Mathematics of Data Science, Cente Lagrange, Paris, May 13-15 2025
 - **PEPR IA Days**, CentraleSupélec, Mar 18th 2025
 - as well as invited seminars: DATASHAPE Team, Inria Saclay, Nov 6 2025; **Talk @MALGA Seminar - Genova, April 28th 2025**, Apr 28th 2025; **Journée AILYS** at ENS Lyon, Feb 14th 2025; Séminaire "mathématiques de l'IA", IMB, Bordeaux, Jan 30th 2025; **Séminaire MMCS de l'ICJ**, Lyon 1 , Jan 7th 2025

10.1.5 Leadership within the scientific community

- Remi Gribonval
 - Scientific Committee of **RT MAIAGES** (formerly RT/GDR MIA);
 - Comité de Liaison **SIGMA-SMAI**;
 - Board of the **GRETSI** association;
 - **Cellule ERC** of Inria, mentoring for ERC candidates in computer science and applied mathematics at the national Inria level
- Mathurin Massias – Secretary of the **MODE** group of **SMAI**

10.1.6 Scientific expertise

- Remi Gribonval – Scientific Advisory Board of the Acoustics Research Institute of the Austrian Academy of sciences
- Elisa Riccietti – Scientific Board of the Federation Informatique de Lyon (*Conseil Scientifique de la FIL*)

10.1.7 Research administration

- Paulo Goncalves
 - member of the steering committee for the ShapeMed@Lyon consortiums Data for Health workshop
 - Scientific Director of the Inria Centre of Lyon and member of the Inria Evaluation Committee.

10.2 Teaching - Supervision - Juries - Educational and pedagogical outreach

10.2.1 Teaching

- Master:
 - Elisa Riccietti – Optimisation (ENS Lyon) and Harnessing inexactness in scientific computing (ENS Lyon)
 - Mathurin Massias – Python for datascience (Ecole Polytechnique), Statistics (Ecole Polytechnique), Optimal Transport for Machine and Deep Learning (ENS Lyon), Fundamentals of Machine Learning (ENS Lyon), Generative Models (ENS Lyon)
 - Titouan Vayer – Optimal Transport for Machine and Deep Learning (ENS Lyon), Fundamentals of Machine Learning (ENS Lyon)
 - Marion Foare – Image and Signal Processing, Inverse problems and optimization (CPE Lyon)
 - Paulo Goncalves – Image and Signal Processing (CPE Lyon)
 - Remi 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

10.2.2 Supervision

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

- Elisa Riccietti – co-supervision of the PhD of Filippo Marini with Margherita Porcelli (Università di Bologna) – defence on 16/06/2025
- Remi Gribonval – co-supervision of the PhD of Sibylle Marcotte with Gabriel Peyré since 2022 (Center for Data Science, ENS Paris) – defense on 21/11/2025
- Marion Foare – co-supervision of the PhD of Luis Enrique Amador Arya with H el ene Ratiney and  Eric Van Reeth (Creatis, Villeurbanne) and Siemens Healthcare (Saint Denis) since 2023

PhD defenses in Ockham in 2025:

- Can Pouliquen

10.2.3 Juries

Members of the Ockham team participated in the following juries :

- Elisa Riccietti – PhD defence of Iskander Legheraba (Dauphine Universit e, Paris), CSI of Xavier Pillet (PhD student, Lyon 1 University)
- Mathurin Massias – CSI of Yu-Han Wu (PhD Student, Sorbonne Universit e)
- Paulo Goncalves – PhD defense of Valerian Mange (U Toulouse), CSI of Andr ea Ducos (PhD Student, Lyon 1 University)
- Titouan Vayer – Junjie Yang (07/04/2025, examiner), member of the CSI for Antonin Joly (PhD Student, IRISA), Antoine Monier (PhD Student, IRISA).

- Remi Gribonval – PhD defenses of: Armand Foucault (26/05/25, Université de Toulouse, reviewer); Blaise Delattre (16/2/25, Dauphine PSL, reviewer); Manon Verbockhaven (28/03/25, Université Paris-Saclay, reviewer); Maud Biquard (5/11/25, Université de Toulouse, president); Mimoun Mohamed (31/03/25, Aix-Marseille Université, examiner); Volodimir Mitarchuk (17/01/25, Université Jean Monnet Saint-Étienne, president); Pierre Warion (19/11/25, Aix-Marseille Université, examiner); Romain Verdière (8/12/25, Université Grenoble Alpes, president).

11 Scientific production

11.1 Major publications

- [1] Q. Bertrand, A. Gagneux, M. Massias and R. Emonet. ‘On the Closed-Form of Flow Matching: Generalization Does Not Arise from Target Stochasticity’. In: *NeurIPS 2025*. NeurIPS 2025 - 39th Annual Conference on Neural Information Processing Systems. San Diego (CA), United States, 1st Dec. 2025. URL: <https://hal.science/hal-05119547> (cit. on pp. 10, 20).
- [2] 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](https://doi.org/10.4171/msl/20). URL: <https://inria.hal.science/hal-01544609> (cit. on p. 20).
- [3] 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](https://doi.org/10.1109/MSP.2021.3092574). URL: <https://inria.hal.science/hal-03350599> (cit. on p. 20).
- [4] R. Gribonval, E. Riccietti, Q.-T. Le and L. Zheng. ‘Rapture of the deep: highs and lows of sparsity in a world of depths’. In: *IEEE Signal Processing Magazine* (20th June 2025), 22 p. URL: <https://inria.hal.science/hal-04954574> (cit. on pp. 11, 18).
- [5] G. Lauga, E. Riccietti, N. Pustelnik and P. Gonçalves. ‘IML FISTA: A Multilevel Framework for Inexact and Inertial Forward-Backward. Application to Image Restoration’. In: *SIAM Journal on Imaging Sciences* (1st June 2024). DOI: [10.1137/23M1582345](https://doi.org/10.1137/23M1582345). URL: <https://inria.hal.science/hal-04075814> (cit. on p. 21).
- [6] *PNP-FLOW: Plug-And-Play Image Restoration with Flow Matching*. International Conference on Learning Representations. Singapore, Singapore, 24th Apr. 2025. URL: <https://hal.science/hal-04881033> (cit. on pp. 19, 20).
- [7] T. Moreau, M. Massias, A. Gramfort, P. Ablin, P.-A. Bannier, B. Charlier, M. Dagr eou, T. Dupr e La Tour, G. Durif, C. F. Dantas, Q. Kloppenstein, J. Larsson, E. Lai, T. Lefort, B. Mal ezieux, B. Moufad, B. T. Nguyen, A. Rakotomamonjy, Z. Ramzi, J. Salmon and S. Vaiter. ‘Benchopt: Reproducible, efficient and collaborative optimization benchmarks’. In: *NeurIPS 2022 - 36th Conference on Neural Information Processing Systems*. New Orleans, United States, 28th Nov. 2022. URL: <https://hal.science/hal-03830604> (cit. on p. 22).
- [8] B. Ricaud, P. Borgnat, N. Tremblay, P. Gonalves 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](https://doi.org/10.1016/j.crhy.2019.08.003). URL: <https://inria.hal.science/hal-02304584> (cit. on p. 8).
- [9] H. Van Assel, C. Vincent-Cuaz, N. Courty, R. Flamary, P. Frossard and T. Vayer. ‘Distributional Reduction: Unifying Dimensionality Reduction and Clustering with Gromov-Wasserstein’. In: *Transactions on Machine Learning Research Journal* (27th June 2025). URL: <https://hal.science/hal-04726195> (cit. on p. 21).
- [10] 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> (cit. on p. 20).

11.2 Publications of the year

International journals

- [11] A. Belhadji, R. Bardenet and P. Chainais. ‘Signal reconstruction using determinantal sampling’. In: *Applied and Computational Harmonic Analysis* (2025). URL: <https://hal.science/hal-04181079>.
- [12] A. Brenon. ‘Les effets de la brièveté dans les entrées encyclopédiques’. In: *Syntaxe et sémantique* (2025). URL: <https://hal.science/hal-05381809>. In press.
- [13] A. Gagneux, M. Massias, E. Soubies and R. Gribonval. ‘Convexity in ReLU Neural Networks: beyond ICNNs?’ In: *Journal of Mathematical Imaging and Vision* 67.4 (19th June 2025), p. 40. DOI: [10.1007/s10851-025-01253-x](https://doi.org/10.1007/s10851-025-01253-x). URL: <https://hal.science/hal-04877619> (cit. on p. 19).
- [14] R. Gribonval, E. Riccietti, Q.-T. Le and L. Zheng. ‘Rapture of the deep: highs and lows of sparsity in a world of depths’. In: *IEEE Signal Processing Magazine* (20th June 2025), 22 p. URL: <https://inria.hal.science/hal-04954574>.
- [15] C. Lage, N. Pustelnik, J.-M. Arbona, B. Audit and R. Gribonval. ‘Identifying a piecewise affine signal from its nonlinear observation -application to DNA replication analysis’. In: *IEEE Transactions on Signal Processing* 73 (2025), pp. 1278–1292. DOI: [10.1109/TSP.2025.3544463](https://doi.org/10.1109/TSP.2025.3544463). URL: <https://hal.science/hal-04991122>.
- [16] E. Lasalle, R. Vaudaine, T. Vayer, P. Borgnat, R. Gribonval, P. Gonçalves and M. Karsai. ‘PASCO (PARallel Structured COarsening): an overlay to speed up graph clustering algorithms’. In: *Machine Learning* 114 (22nd Aug. 2025), p. 212. DOI: [10.1007/s10994-025-06837-7](https://doi.org/10.1007/s10994-025-06837-7). URL: <https://hal.science/hal-04837207> (cit. on p. 20).
- [17] Q.-T. Le, L. Zheng, E. Riccietti and R. Gribonval. ‘Butterfly factorization with error guarantees’. In: *SIAM Journal on Matrix Analysis and Applications* 46.4 (23rd July 2025). DOI: [10.1137/24M1708796](https://doi.org/10.1137/24M1708796). URL: <https://hal.science/hal-04763712> (cit. on p. 18).
- [18] J. Tachella, M. Terris, S. Hurault, A. Wang, L. Davy, J. Scanvic, V. Sechaud, R. Vo, T. Moreau, T. Davies, D. Chen, N. Laurent, B. Monroy, J. Dong, Z. Hu, M.-H. Nguyen, F. Sarron, P. Weiss, P. Escande, M. Massias, T. Modrzyk, B. Levac, T. I. Liaudat, M. Song, J. Hertrich, S. Neumayer and G. Schramm. ‘DeepInverse: A Python package for solving imaging inverse problems with deep learning’. In: *Journal of Open Source Software* 10.115 (19th Nov. 2025), p. 8923. DOI: [10.21105/joss.08923](https://doi.org/10.21105/joss.08923). URL: <https://hal.science/hal-05377313> (cit. on p. 19).
- [19] H. Van Assel, C. Vincent-Cuaz, N. Courty, R. Flamary, P. Frossard and T. Vayer. ‘Distributional Reduction: Unifying Dimensionality Reduction and Clustering with Gromov-Wasserstein’. In: *Transactions on Machine Learning Research Journal* (27th June 2025). URL: <https://hal.science/hal-04726195>.

Invited conferences

- [20] R. Gribonval and E. Riccietti. ‘Une brève histoire de la parcimonie : du traitement de signal à l’apprentissage profond’. In: GRETSI 2025 – XXXème Colloque Francophone de Traitement du Signal et des Images. Strasbourg, France, 25th Aug. 2025. URL: <https://inria.hal.science/hal-05248111>.

International peer-reviewed conferences

- [21] Q. Bertrand, A. Gagneux, M. Massias and R. Emonet. ‘On the Closed-Form of Flow Matching: Generalization Does Not Arise from Target Stochasticity’. In: *NeurIPS 2025*. NeurIPS 2025 - 39th Annual Conference on Neural Information Processing Systems. San Diego (CA), United States, 1st Dec. 2025. URL: <https://hal.science/hal-05119547>.

- [22] A. Gonon, N. Brisebarre, E. Riccietti and R. Gribonval. ‘A Rescaling-Invariant Lipschitz Bound Based on Path-Metrics for Modern ReLU Network Parameterizations’. In: *Proceedings of the 42nd International Conference on Machine Learning (ICML 2025)*. International Conference on Machine Learning. Vancouver (BC), Canada, 13th July 2025. URL: <https://hal.science/hal-04584311> (cit. on p. 16).
- [23] A. Gonon, L. Zheng, P. Carrivain and Q.-T. Le. ‘Fast Inference with Kronecker-Sparse Matrices’. In: *Proceedings of the 42nd International Conference on Machine Learning (ICML 2025)*. International Conference on Machine Learning. Vancouver (BC), Canada, 13th July 2025. URL: <https://hal.science/hal-04584450> (cit. on p. 19).
- [24] P. Houedry, N. Courty, F. Martin-Baillon, L. Chapel and T. Vayer. ‘Bridging Arbitrary and Tree Metrics via Differentiable Gromov Hyperbolicity’. In: *NeurIPS 2025 - 39th Annual Conference on Neural Information Processing Systems*. San diego (Californie), United States, 2nd Dec. 2025. URL: <https://hal.science/hal-05088161> (cit. on p. 15).
- [25] H. T. V. Le, M. Foare, A. Repetti and N. Pustelnik. ‘Unfolded discrete Mumford-Shah functional for joint image denoising and edge detection’. In: *EUSIPCO 2025 - 32nd European Signal Processing Conference*. Palerme, Italy, 8th Sept. 2025. URL: <https://hal.science/hal-05404060>.
- [26] S. Marcotte, R. Gribonval and G. Peyré. ‘Transformative or Conservative? Conservation laws for ResNets and Transformers’. In: *42nd International Conference on Machine Learning (ICML 2025)*. Vancouver (Canada), Canada, 13th July 2025. URL: <https://hal.science/hal-05108418> (cit. on pp. 10, 16).
- [27] S. Martin, A. Gagneux, P. Hagemann and G. Steidl. ‘PNP-FLOW: Plug-And-Play Image Restoration with Flow Matching’. In: *International Conference on Learning Representations 2025*. International Conference on Learning Representations. Singapore, Singapore, 24th Apr. 2025. URL: <https://hal.science/hal-04881033>.
- [28] C. Pouliquen, M. Massias and T. Vayer. ‘Schur’s Positive-Definite Network: Deep Learning in the SPD cone with structure’. In: *International Conference on Learning Representations*. International Conference on Learning Representations ICLR 2025. Singapore, Singapore, 22nd Jan. 2025. URL: <https://hal.science/hal-04726325> (cit. on p. 14).

National peer-reviewed Conferences

- [29] V. Castin and R. Gribonval. ‘Opening the Black Box: Reverse-Engineering of Sparse Neural Networks’. In: *30° colloque sur le traitement du signal et des images*. GRETSI 2025 – XXXème Colloque Francophone de Traitement du Signal et des Images. Strasbourg, France, 2025. URL: <https://hal.science/hal-05136482>.
- [30] M. Chaumette, R. Gribonval and E. Riccietti. ‘CROQuant: Complex Rank-One Quantization Algorithm’. In: *GRETSI 2025 – XXXème Colloque Francophone de Traitement du Signal et des Images*. Strasbourg, France, 25th Aug. 2025. URL: <https://hal.science/hal-05140371> (cit. on p. 17).
- [31] A. Gagneux, M. Massias, E. Soubies and R. Gribonval. ‘How to improve expressivity of convex ReLU neural networks?’ In: *GRETSI 2025*. GRETSI 2025 - XXXème Colloque Francophone de Traitement du Signal et des Images. Strasbourg, France, 2025. URL: <https://hal.science/hal-05216923>.
- [32] G. Lauga, M. Chaumette, E. Desainte-Maréville, É. Lasalle and A. Lebeurrer. ‘A multilevel approach to accelerate the training of Transformers’. In: *GRETSI’25, XXème Colloque Francophone de Traitement du Signal et des Images*. Strasbourg, France, 25th Aug. 2025. URL: <https://hal.science/hal-05044666>.
- [33] G. Lauga, E. Riccietti, L. Briceño-Arias, N. Pustelnik and P. Gonçalves. ‘Une équivalence entre algorithmes multiniveaux et algorithmes de descente par blocs’. In: *GRETSI. Colloque GRETSI’25, XXXe Colloque Francophone de Traitement du Signal et des Images*. Strasbourg, France, 25th Aug. 2025. URL: <https://hal.science/hal-05044669> (cit. on p. 21).

- [34] N. Laurent, E. Riccietti, J. Tachella and N. Pustelnik. ‘Algorithme multiniveau hybride pour la restauration d’images’. In: GRETSI 2025 - XXXème Colloque Francophone de Traitement du Signal et des Images. Strasbourg, France, 29th Aug. 2025. URL: <https://hal.science/hal-05219319>.
- [35] C. Pouliquen, P. Gonçalves, T. Vayer and M. Massias. ‘En quête de précision : Un benchmark open-source et un solveur polyvalent pour le Graphical Lasso’. In: GRETSI 2025 - XXXème Colloque Francophone de Traitement du Signal et des Images. Strasbourg, France, 2025, pp. 1–3. URL: <https://hal.science/hal-05128230> (cit. on p. 14).

Conferences without proceedings

- [36] A. Brenon and D. Vigier. ‘Propositions pour une approche quantitative-qualitative des discours traitant des métiers dans l’Encyclopédie de Diderot et d’Alembert: Proposals for a quantitative-qualitative approach to discourse on trades in Diderot’s and d’Alembert’s Encyclopedie’. In: Journées internationales de Linguistique de Corpus (JLC 2025). Lyon (ENS LSH), France, 21st Oct. 2025. URL: <https://hal.science/hal-05399212>.
- [37] Z. Rodière, P. Borgnat, P. Gonçalves and J. Jung. ‘Spatial contrastive pre-training of transformer encoders for seeg-based seizure onset zone detection’. In: GSP 2025 - Workshop on Graph Signal Processing. Montréal (Québec), Canada, May 2025, pp. 1–3. URL: <https://hal.science/hal-05364300> (cit. on p. 15).

Doctoral dissertations and habilitation theses

- [38] G. Carrino. ‘Frugality in second-order optimization: floating-point approximations for Newton’s method’. Bologna University, 7th Oct. 2025. URL: <https://hal.science/tel-05375552> (cit. on p. 18).
- [39] C. Pouliquen. ‘Differentiable and learning-based methods for structure representation’. Ecole Normale Supérieure de Lyon, 5th Dec. 2025. URL: <https://hal.science/tel-05430511>.

Reports & preprints

- [40] L. Briceño-Arias, P. Gonçalves, G. Lauga, N. Pustelnik and E. Riccietti. *A flexible block-coordinate forward-backward algorithm for non-smooth and non-convex optimization*. 29th Oct. 2025. URL: <https://inria.hal.science/hal-05337116>.
- [41] Y. de Castro, R. Gribonval and N. Jouvin. *Effective regions and kernels in continuous sparse regularisation, with application to sketched mixtures*. 14th July 2025. URL: <https://hal.science/hal-05162134>.
- [42] E.-M. El Arar, S.-I. Filip, T. Mary and E. Riccietti. *Mixed precision accumulation for neural network inference guided by componentwise forward error analysis*. 2025. URL: <https://hal.science/hal-04995708> (cit. on p. 18).
- [43] A. Gagneux, S. Martin, R. Gribonval and M. Massias. *The Generation Phases of Flow Matching: a Denoising Perspective*. 12th Jan. 2026. URL: <https://inria.hal.science/hal-05338361> (cit. on p. 20).
- [44] A. Gagneux, M. Massias and E. Soubies. *Proximal Operators of Sorted Nonconvex Penalties*. 13th June 2025. URL: <https://inria.hal.science/hal-05338351> (cit. on p. 14).
- [45] S. Marcotte, G. Peyré and R. Gribonval. *Intrinsic training dynamics of deep neural networks*. 10th Aug. 2025. URL: <https://inria.hal.science/hal-05458265> (cit. on p. 17).
- [46] F. Marini, M. Porcelli and E. Riccietti. *A multilevel stochastic regularized first-order method with application to finite sum minimization*. 13th Jan. 2025. URL: <https://inria.hal.science/hal-04884049> (cit. on p. 21).
- [47] D. Rouchouse, A. Gonon, R. Gribonval and B. Guedj. *Non-Vacuous Generalization Bounds: Can Rescaling Invariances Help?* 30th Sept. 2025. URL: <https://hal.science/hal-05290479>.

- [48] T. Vayer and E. Lasalle. *A note on the relations between mixture models, maximum-likelihood and entropic optimal transport*. 20th Jan. 2025. URL: <https://hal.science/hal-04901607> (cit. on p. 21).

Other scientific publications

- [49] L. Amador, M. Foare, O. Beuf, H. Ratiney and E. V. Reeth. ‘MULTI-CONTRAST SUPER-RESOLUTION FOR MRI WITH A VARIATIONAL APPROACH: DISENTANGLING THE CHALLENGES’. In: European Society for Magnetic Resonance in Medicine and Biology (ESMRMB). Marseille, France, 17th Oct. 2025. URL: <https://hal.science/hal-05319118>.
- [50] L. Amador, M. Foare, O. Beuf, H. Ratiney and E. V. Reeth. ‘SUPER-RESOLUTION ISOTROPE VARIATIONNELLE POUR L’IRM’. In: Septième congrès de la Société Française de Résonance Magnétique en Biologie et Médecine (SFRMBM). Saint Malo, France, 24th Mar. 2025. URL: <https://hal.science/hal-05012699>.
- [51] A. Gagneux, S. Martin, R. Emonet, Q. Bertrand and M. Massias. *A Visual Dive into Conditional Flow Matching*. Apr. 2025. URL: <https://hal.science/hal-05088592> (cit. on p. 19).

Software

- [52] [SW] V. Castin and R. Gribonval, *Code for reproducible research: "Opening the Black Box: Reverse-Engineering of Sparse Neural Networks"* 1st July 2025. LIC: BSD 3-Clause "New" or "Revised" License. HAL: [hal-05136299](https://hal.science/hal-05136299), URL: <https://hal.science/hal-05136299>, vcs: <https://github.com/vcastin/2025-reverse-engineering>, SWHID: [swh:1:dir:9b57025ded7a453e68284d5f650d7920ee693108](https://swh.io/dir/9b57025ded7a453e68284d5f650d7920ee693108); origin=<https://hal.archives-ouvertes.fr/hal-05136299>; visit=[swh:1:snp:9de311e78eb80c1059f72db355d32e887ac6021a](https://swh.io/snp/9de311e78eb80c1059f72db355d32e887ac6021a); anchor=[swh:1:rel:95fea2bbe20a3ebd7e0268781375e8ad055bc431](https://swh.io/rel/95fea2bbe20a3ebd7e0268781375e8ad055bc431); path=/).
- [53] [SW] M. Chaumette, R. Gribonval and E. Riccietti, *Code for reproducible research - CROQuant: Complex Rank-One Quantization Algorithm* 18th July 2025. LIC: BSD 3-Clause License. HAL: [hal-05133647](https://hal.science/hal-05133647), URL: <https://hal.science/hal-05133647>, SWHID: [swh:1:dir:31b2b28bf4b578180b0daaf03902808de35fb109](https://swh.io/dir/31b2b28bf4b578180b0daaf03902808de35fb109); origin=<https://hal.archives-ouvertes.fr/hal-05133647>; visit=[swh:1:snp:e895217205bbf60c1b4b749ed0fc30100d24e964](https://swh.io/snp/e895217205bbf60c1b4b749ed0fc30100d24e964); anchor=[swh:1:rel:879c672e56027dde5f1dcfa20f6797577524e662](https://swh.io/rel/879c672e56027dde5f1dcfa20f6797577524e662); path=/).
- [54] [SW] A. Gonon, N. Brisebarre, E. Riccietti and R. Gribonval, *Code for reproducible research - A rescaling-invariant Lipschitz bound using path-metrics* version 0.0.1, 13th June 2025. LIC: BSD 3-Clause "New" or "Revised" License. HAL: [hal-05088847](https://hal.science/hal-05088847), URL: <https://hal.science/hal-05088847>, vcs: https://github.com/agonon/pathnorm_toolkit, SWHID: [swh:1:dir:3ef1659778414497b9fc2469b8afad02c6eb9e68](https://swh.io/dir/3ef1659778414497b9fc2469b8afad02c6eb9e68); origin=<https://hal.archives-ouvertes.fr/hal-05088847>; visit=[swh:1:snp:4315baa0c8e513fa42b84634266d757d5e44fcbd](https://swh.io/snp/4315baa0c8e513fa42b84634266d757d5e44fcbd); anchor=[swh:1:rel:147f1cfa0d54dace5a3c59dad43d04b055217f5b](https://swh.io/rel/147f1cfa0d54dace5a3c59dad43d04b055217f5b); path=/).
- [55] [SW] E. Lasalle, R. Vaudaine and T. Vayer, *PASCO* 27th May 2025. LIC: BSD 3-Clause "New" or "Revised" License. HAL: [hal-05086352](https://hal.science/hal-05086352), URL: <https://hal.science/hal-05086352>, vcs: <https://github.com/elasalle/PASCO>, SWHID: [swh:1:dir:18c931d98f790cddb1dcfba6e2e03ff4427a8c24](https://swh.io/dir/18c931d98f790cddb1dcfba6e2e03ff4427a8c24); origin=<https://hal.archives-ouvertes.fr/hal-05086352>; visit=[swh:1:snp:369c0877aee64505a50b80345742b519c9dadb7b](https://swh.io/snp/369c0877aee64505a50b80345742b519c9dadb7b); anchor=[swh:1:rel:45b059a3bd793f910f722c5822c3bf5f69462f89](https://swh.io/rel/45b059a3bd793f910f722c5822c3bf5f69462f89); path=/).

11.3 Cited publications

- [56] H. H. Bauschke, P. L. Combettes et al. *Convex analysis and monotone operator theory in Hilbert spaces*. Vol. 408. Springer, 2011 (cit. on p. 7).
- [57] 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](https://doi.org/10.1007/s41109-019-0172-x). URL: <https://inria.hal.science/hal-02063780> (cit. on p. 8).

- [58] 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 (cit. on p. 22).
- [59] 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* 23.1 (Apr. 2022), pp. 6680–6722. URL: <https://hal.science/hal-03228663> (cit. on p. 22).
- [60] 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](https://doi.org/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+Applications%5C&hl=%5C&cd=1%5C&source=gbs%5C_api (cit. on p. 7).
- [61] 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> (cit. on p. 7).
- [62] 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.inria.fr/hal-02496896> (cit. on pp. 8, 9).
- [63] 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 (cit. on p. 7).
- [64] 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) (cit. on p. 8).
- [65] P. M. Djuric and R. C. *Cooperative and Graph Signal Processing: Principle and Applications*. Academic Press, 2018 (cit. on p. 8).
- [66] 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=d5b61J19BvAC%5C&printsec=frontcover%5C&dq=sparse+and+redundant+representations%5C&hl=%5C&cd=1%5C&source=gbs%5C_api (cit. on p. 7).
- [67] M. Foare, N. Pustelnik and L. Condat. ‘Semi-Linearized Proximal Alternating Minimization for a Discrete Mumford-Shah Model’. In: *IEEE Transactions on Image Processing* 29 (Oct. 2019), pp. 2176–2189. DOI: [10.1109/TIP.2019.2944561](https://doi.org/10.1109/TIP.2019.2944561). URL: <https://hal.science/hal-01782346> (cit. on p. 7).
- [68] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. New York, NY: Springer, 2013. DOI: [10.1007/978-0-8176-4948-7](https://doi.org/10.1007/978-0-8176-4948-7). URL: <http://link.springer.com/10.1007/978-0-8176-4948-7> (cit. on p. 7).
- [69] J. Friedman, T. Hastie and R. Tibshirani. ‘Sparse inverse covariance estimation with the graphical lasso’. In: *Biostatistics* 9.3 (2008), pp. 432–441 (cit. on p. 8).
- [70] G. Frusque. ‘Inférence et décomposition modale de réseaux dynamiques en neurosciences’. 2020LYSEN080. PhD thesis. 2020. URL: <http://www.theses.fr/2020LYSEN080/document> (cit. on p. 15).
- [71] 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](https://doi.org/10.1109/LSP.2015.2488279). URL: <https://hal.inria.fr/hal-01221562> (cit. on p. 8).
- [72] A. Gonon. ‘Harnessing symmetries for modern deep learning challenges : a path-lifting perspective’. Theses. Ecole normale supérieure de Lyon - ENS LYON, Nov. 2024. URL: <https://theses.hal.science/tel-04784426> (cit. on p. 16).

- [73] A. Gonon, N. Brisebarre, E. Riccietti and R. Gribonval. ‘A path-norm toolkit for modern networks: consequences, promises and challenges’. In: *International Conference on Learning Representations*. Erratum: in the published version there was a typo in the definition of the activation matrix in Definition A.3. This is fixed with this new version. Wien, Austria, May 2024. URL: <https://hal.science/hal-04225201> (cit. on p. 16).
- [74] 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://hal.inria.fr/hal-01544609> (visited on 15/04/2021) (cit. on pp. 8, 9).
- [75] 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-01544609v2> It 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> (cit. on p. 20).
- [76] R. Gribonval, T. Mary and E. Riccietti. ‘Optimal quantization of rank-one matrices in floating-point arithmetic—with applications to butterfly factorizations’. working paper or preprint. June 2023. URL: <https://inria.hal.science/hal-04125381> (cit. on p. 17).
- [77] R. Gribonval, T. Mary and E. Riccietti. ‘Scaling is all you need: quantization of butterfly matrix products via optimal rank-one quantization’. In: *Actes du GRETSI 2023*. Actes du GRETSI 2023 2023-1193. Grenoble, France: GRETSI - Groupe de Recherche en Traitement du Signal et des Images, Aug. 2023, pp. 497–500. URL: <https://inria.hal.science/hal-04146706> (cit. on p. 17).
- [78] 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> (cit. on p. 7).
- [79] 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 (cit. on p. 8).
- [80] 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*. Valencia, Spain, Apr. 2023. URL: <https://hal.science/hal-04211867> (cit. on p. 7).
- [81] G. Lauga, A. Repetti, E. Riccietti, N. Pustelnik, P. Gonçalves and Y. Wiaux. ‘A multilevel framework for accelerating uSARA in radio-interferometric imaging’. In: *European Signal Processing Conference (EUSIPCO)*. Lyon, France, Aug. 2024. DOI: 10.23919/EUSIPCO63174.2024.10715263. URL: <https://inria.hal.science/hal-04509285> (cit. on p. 21).
- [82] G. Lauga, E. Riccietti, N. Pustelnik and P. Gonçalves. ‘Méthodes multi-niveaux pour la restauration d’images hyperspectrales’. In: *Colloque GRETSI, September 2023* (2023) (cit. on p. 21).
- [83] 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> (cit. on p. 21).
- [84] G. Lauga, E. Riccietti, N. Pustelnik and P. Gonçalves. ‘Multilevel FISTA for image restoration’. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*. IEEE. Rhodes, Greece, June 2023. DOI: 10.48550/arXiv.2210.15940. URL: <https://hal.science/hal-03831180> (cit. on p. 21).
- [85] H. T. V. Le, M. Foare, A. Repetti and N. Pustelnik. ‘Embedding Blake-Zisserman Regularization in Unfolded Proximal Neural Networks for Enhanced Edge Detection’. 2024. URL: <https://hal.archives-ouvertes.fr/hal-04771534v1> (cit. on p. 18).
- [86] H. T. V. Le, M. Foare, A. Repetti and N. Pustelnik. ‘Unfolded discrete Mumford-Shah functional for joint image denoising and edge detection’. 2025. URL: <https://hal.science/hal-05404060/> (cit. on p. 18).

- [87] Q.-T. Le. ‘Algorithmic and theoretical aspects of sparse deep neural networks’. Theses. Ecole normale supérieure de Lyon - ENS LYON, Dec. 2023. URL: <https://inria.hal.science/tel-04329531> (cit. on p. 18).
- [88] Q.-T. Le, E. Riccietti and R. Gribonval. ‘Does a sparse ReLU network training problem always admit an optimum?’ In: *Advances in Neural Information Processing Systems 36 (NeurIPS 2023)*. Advances in Neural Information Processing Systems 36 (NeurIPS 2023). New Orleans (Louisiane), United States, Dec. 2023. URL: <https://inria.hal.science/hal-04108849> (cit. on p. 19).
- [89] 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> (cit. on p. 18).
- [90] 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](https://doi.org/10.1109/ICASSP43922.2022.9747791). URL: <https://inria.hal.science/hal-03438881> (cit. on pp. 17–19).
- [91] S. Marcotte, R. Gribonval and G. Peyré. ‘Abide by the Law and Follow the Flow: Conservation Laws for Gradient Flows’. In: *Advances in Neural Information Processing Systems 36 (NeurIPS 2023)*. Advances in Neural Information Processing Systems 36 (NeurIPS 2023). New Orleans (Louisiane), United States, Dec. 2023. URL: <https://inria.hal.science/hal-04150576> (cit. on p. 16).
- [92] S. Marcotte, R. Gribonval and G. Peyré. ‘Keep the Momentum: Conservation Laws beyond Euclidean Gradient Flows’. In: *Forty-first International Conference on Machine Learning*. Accepted to ICML 2024. Vienna, Austria, July 2024. URL: <https://hal.science/hal-04590932> (cit. on p. 16).
- [93] 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> (cit. on p. 22).
- [94] 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, Aug. 2023, pp. 1–4. URL: <https://hal.science/hal-04151796> (cit. on p. 14).
- [95] A. Rahimi and B. Recht. ‘Random features for large-scale kernel machines’. In: 2007 (cit. on p. 9).
- [96] 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](https://doi.org/10.1007/s10107-018-1346-5) (cit. on p. 9).
- [97] 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 (cit. on p. 8).
- [98] 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> (cit. on p. 9).
- [99] 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](https://doi.org/10.1007/s00365-022-09578-1). URL: <https://hal.science/hal-03292203> (cit. on p. 16).
- [100] I. Tosić and P. Frossard. ‘Dictionary Learning’. In: *IEEE Signal Processing Magazine* 28.2 (), pp. 27–38. DOI: [10.1109/MSP.2010.939537](https://doi.org/10.1109/MSP.2010.939537). URL: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5714407> (cit. on p. 7).
- [101] 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)*. NeurIPS 2023 conference paper. New Orleans, United States, Dec. 2023. URL: <https://hal.science/hal-04103326> (cit. on p. 21).
- [102] 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 (cit. on p. 9).

- [103] 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 (cit. on p. 9).
- [104] Z. Yao, A. Gholami, S. Shen, K. Keutzer and M. W. Mahoney. ‘ADAHESIAN: An adaptive second order optimizer for machine learning’. In: *arXiv preprint arXiv:2006.00719* (2020) (cit. on p. 9).
- [105] 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 (cit. on p. 8).
- [106] L. Zheng. ‘Data frugality and computational efficiency in deep learning’. Theses. Ecole normale supérieure de lyon - ENS LYON, May 2024. URL: <https://theses.hal.science/tel-04680306> (cit. on p. 18).
- [107] L. Zheng, G. Puy, E. Riccietti, P. Pérez and R. Gribonval. ‘Factorisation butterfly par identification algorithmique de blocs de rang un’. In: *XXIXème Colloque Francophone de Traitement du Signal et des Images*. Grenoble, France, Aug. 2023. URL: <https://inria.hal.science/hal-04145743> (cit. on p. 19).
- [108] 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.science/hal-03362626> (cit. on p. 18).