

RESEARCH CENTRE

Grenoble - Rhône-Alpes

IN PARTNERSHIP WITH:

Université Claude Bernard (Lyon 1),
Ecole normale supérieure de Lyon

2021

ACTIVITY REPORT

Project-Team

DANTE

**Dynamic Networks : Temporal and
Structural Capture Approach**

DOMAIN

**Applied Mathematics, Computation and
Simulation**

THEME

**Optimization, machine learning and
statistical methods**

Contents

Project-Team DANTE	1
1 Team members, visitors, external collaborators	2
2 Overall objectives	3
2.1 Evolution of the team and scope of this activity report	3
2.2 New objectives of the team	3
3 Research program	4
3.1 Axis 1: Sparsity for high-dimensional learning.	4
3.2 Axis 2: Learning on graphs and learning of graphs.	5
3.3 Axis 3: Dynamic and frugal learning.	6
4 Application domains	8
4.1 Frugal AI on embedded devices	8
4.2 Imaging in physics and medicine	8
4.3 Interactions with computational social sciences	8
5 Social and environmental responsibility	8
5.1 Contribution to the monitoring of the Covid-19 pandemic	8
6 Highlights of the year	8
7 New software and platforms	9
7.1 New software	9
7.1.1 FAuST	9
8 New results	10
8.1 Graph Signal Processing, Optimal Transport and Machine Learning on Graphs	10
8.1.1 Works on Gromov-Wasserstein: graph dictionary learning	10
8.1.2 Diffused Wasserstein Distance for Optimal transport between attributed graphs	10
8.2 Sparse deep neural networks : theory and algorithms	11
8.2.1 Mathematics of deep learning: approximation theory, scale-invariance, and regularization	11
8.2.2 Algorithms for quantized networks	12
8.2.3 Deep sparse factorizations: hardness, algorithms and identifiability	12
8.3 Statistical learning, dimension reduction, and privacy preservation	13
8.3.1 Theoretical and algorithmic foundations of compressive learning: sketches, kernels, and optimal transport	13
8.3.2 Privacy preservation	13
8.4 Large-scale convex and nonconvex optimization	14
9 Bilateral contracts and grants with industry	14
9.1 Bilateral grants with industry	14
10 Partnerships and cooperations	15
10.1 National initiatives	15
10.1.1 ANR IA Chaire : AllegroAssai	15
10.1.2 ANR DataRedux	16
10.1.3 ANR Darling	16
10.1.4 GDR ISIS project MOMIGS	17
10.2 Regional initiatives	17
10.2.1 Labex CominLabs LeanAI	17
10.2.2 Labex Emerging Topics	17

11 Dissemination	18
11.1 Promoting scientific activities	18
11.1.1 Scientific events: organisation	18
11.1.2 Scientific events: selection	18
11.1.3 Journal	18
11.1.4 Invited talks	19
11.1.5 Leadership within the scientific community	19
11.1.6 Scientific expertise	19
11.1.7 Research administration	19
11.2 Teaching - Supervision - Juries	19
11.2.1 Teaching	19
11.2.2 Supervision	19
11.2.3 Juries	20
12 Scientific production	20
12.1 Major publications	20
12.2 Publications of the year	21
12.3 Other	24
12.4 Cited publications	24

Project-Team DANTE

Creation of the Project-Team: 2015 January 01

Keywords

Computer sciences and digital sciences

- A1.2. – Networks
- A1.6. – Green Computing
- A3.4.1. – Supervised learning
- A3.4.4. – Optimization and learning
- A3.4.6. – Neural networks
- A3.4.8. – Deep learning
- A3.5. – Social networks
- A3.5.1. – Analysis of large graphs
- A5.9. – Signal processing
- A5.9.4. – Signal processing over graphs
- A5.9.5. – Sparsity-aware processing
- A5.9.6. – Optimization tools
- A8.8. – Network science
- A8.9. – Performance evaluation
- A9.2. – Machine learning
- A9.7. – AI algorithmics

Other research topics and application domains

- B2.6. – Biological and medical imaging
- B6.2. – Network technologies
- B6.4. – Internet of things
- B6.6. – Embedded systems
- B7.2.1. – Smart vehicles
- B9.5.1. – Computer science
- B9.5.2. – Mathematics
- B9.5.6. – Data science
- B9.10. – Privacy

1 Team members, visitors, external collaborators

Research Scientists

- Paulo Gonçalves [Team leader, Inria, Senior Researcher, HDR]
- Rémi Gribonval [Team leader, Inria, Senior Researcher, HDR]
- Mathurin Massias [Inria, Researcher, from Nov 2021]
- Philippe Nain [Inria, Senior Researcher, until Aug 2021, HDR]

Faculty Members

- Thomas Begin [Univ Claude Bernard, Associate Professor, HDR]
- Anthony Busson [Univ Claude Bernard, Associate Professor, HDR]
- Christophe Crespelle [Univ Claude Bernard, Associate Professor, until Sep 2021, HDR]
- Marion Foare [École supérieure de chimie physique électronique de Lyon, Associate Professor]
- Isabelle Guérin Lassous [Univ Claude Bernard, Professor, HDR]
- Elisa Riccietti [École Normale Supérieure de Lyon, Associate Professor, Chaire Inria]

Post-Doctoral Fellows

- Ayoub Belhadji [École Normale Supérieure de Lyon]
- Luc Giffon [École Normale Supérieure de Lyon, from Feb 2021]
- Vincent Schellekens [Inria, from Oct 2021 to Dec 2021]
- Marija Stojanova [Univ Claude Bernard, from Jul 2021 until Nov 2021]
- Titouan Vayer [École Normale Supérieure de Lyon]

PhD Students

- Lafdal Abdelwedoud [Gouvernement de Mauritanie, until Aug 2021]
- Dominique Barbe [École Normale Supérieure de Lyon]
- Anthony Bardou [École Normale Supérieure de Lyon, until Nov 2021]
- Nour El Houda Bouzouita [École Normale Supérieure de Lyon, until Nov 2021]
- Israel Campero-Jurado [École Normale Supérieure de Lyon, until Mar 2021]
- Sicheng Dai [École Normale Supérieure de Lyon]
- Antoine Gonon [École Normale Supérieure de Lyon, from Sep 2021]
- Clement Lalanne [École Normale Supérieure de Lyon]
- Guillaume Lauga [Inria, from Nov 2021]
- Quoc Tung Le [École Normale Supérieure de Lyon]
- Samir Si-Mohammed [Stakeo, until Nov 2021]
- Pierre Stock [Facebook, CIFRE, until Mar 2021]
- Leon Zheng [VALEO, CIFRE, from May 2021]

Technical Staff

- Hakim Hadj-Djilani [École Normale Supérieure de Lyon, until Jul 2021, Development Engineer]
- Leon Zheng [École Normale Supérieure de Lyon, Engineer, until May 2021]

Interns and Apprentices

- Manon Billet [École Normale Supérieure de Lyon, from May 2021 until Jul 2021]
- Amel Chadda [Inria, until Jul 2021]
- Antoine Gonon [École Normale Supérieure de Lyon, from Mar 2021 until Jul 2021]
- Hugo Gouttenegre [Univ de Lyon, from May 2021 until Aug 2021]
- Federico Grillini [École Normale Supérieure de Lyon, from Apr 2021 until Jun 2021]
- Esther Guerin [École Normale Supérieure de Lyon, until Jul 2021]
- Malasri Janumporn [Univ Claude Bernard, from Jul 2021 until Aug 2021]
- Giovanni Seraghiti [École Normale Supérieure de Lyon, from Sep 2021 until Nov 2021]

Administrative Assistant

- Solene Audoux [Inria]

External Collaborators

- Yohann De Castro [École centrale de Lyon, Professor, HDR]
- Eric Guichard [École nationale supérieure des sciences de l'information et des bibliothèques, until Oct 2021, HDR]
- Márton Karsai [Université d'Europe centrale Vienne-Autriche, HDR]

2 Overall objectives

2.1 Evolution of the team and scope of this activity report

After more than 9 years of existence of DANTE as a team focused on dynamic networks at large, and in the context of a strengthening of the research activities in statistical machine learning and signal processing in DANTE (and more largely on the academic site of Lyon Saint-Etienne), it was decided to split the DANTE team into two new teams. All activities related to network communications (with a focus on wireless networks and performance evaluation) are now part of the new team HowNet and are not covered in this scientific report. The new scientific contours of DANTE are described below and this activity report focuses on the machine learning and signal processing activities of DANTE in 2021.

2.2 New objectives of the team

Building on a culture at the interface of signal modeling, mathematical optimization and statistical machine learning, the global objective of DANTE 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 [62] 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 [63, 56].

The flexibility of the models, which can incorporate learned dictionaries [73], as well as structured and/or low-rank variants of the now-classical sparse modeling paradigm [66], 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 [55, 59], 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 [57], which have been identified as a way to potentially circumvent these bottlenecks. The other main challenge is to extend the sparse modeling paradigm to resource-efficient and interpretable statistical machine learning. The methodological and conceptual output of this axis provides tools for Axes 2 and 3, which in return fuel the questions investigated in this axis.

- **1.1 Versatile and efficient sparse modeling.** The goal is to propose flexible and resource-efficient sparse models, possibly leveraging classical notions of dictionaries and structured factorization, but also the notion of sparsity in continuous domains (e.g. for sketched clustering, mixture model estimation, or image super-resolution), low-rank tensor representations, and neural networks with sparse connection patterns.

Besides the empirical validation of these models and of the related algorithms on a diversity of targeted applications, the challenge is to determine conditions under which their success can be mathematically controlled, and to determine the fundamental tradeoffs between the expressivity of these models and their complexity.

- **1.2 Sparse optimization.** The main objectives are: a) to define cost functions and regularization penalties that integrate not only the targeted learning tasks, but also a priori knowledge, for example under the form of conservation laws or as relation graphs, cf Axis 2; b) to design efficient and scalable algorithms [4, 9] 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 [2], 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 [64] 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. [67]) strived to emphasize the clustered structure of the data by imposing spectral constraints to the Laplacian of the sought graphs, with the aim to improve the performance of spectral approaches to unsupervised classification. In this direction, several challenges remain, such as for instance the transposition of the framework to graph-based semi-supervised learning [1], where natural models are stochastic block models rather than strictly multi-component graphs (e.g. Gaussian mixtures models). As it is done in [77], 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 [65] and [58] 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) [70] [10], 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 [5]. Hence, the notion of distributed GSP [60, 61] 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 [68]) 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 [65] as well as privacy preservation guarantees [58]. A sketch can be seen as an embedding of the empirical probability distribution of the dataset with a particular form of kernel mean embedding [71]. 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 [75, 74], 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 [69, 76].

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 DANTE 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 DANTE pays a special attention to modeling domain knowledge such as physical constraints or prior medical knowledge. This finds applications from physics to medical imaging, including: multiphase flow image characterization; near infrared polarization imaging in circumstellar imaging; compressive sensing for joint segmentation and high-resolution 3D MRI imaging; or graph signal processing for radio astronomy imaging with the Square Kilometer Array (SKA).

4.3 Interactions with computational social sciences

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

5 Social and environmental responsibility

5.1 Contribution to the monitoring of the Covid-19 pandemic

Robust prediction of the spatio-temporal evolution of the reproduction number $R(t)$ of the Covid-19 pandemic from open data (Santé-Publique-France and the European Center for Disease Prevention).

Following our work of last year [54], where an algorithm exploiting sparsity and convex optimization was developed, and dynamic maps were proposed, we identified robustness to outliers as a critical issue.

This is addressed in a paper submitted for publication to a journal, using convex regularization [45].

6 Highlights of the year

P. Gonçalves was nominated Deputy Scientific Director of the new research center of Inria in Lyon.

R. Gribonval was a keynote speaker at the international conference EUSIPCO 2021 and an invited speaker at the national conference CAP21.

A survey paper on sketching for large-scale learning, summarizing in tutorial style a series of works of the team, was published in the September 2021 issue of the IEEE Signal Processing Magazine and made its front cover [7].

7 New software and platforms

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

7.1 New software

7.1.1 FAuST

Keywords: Learning, Sparsity, Fast transform, 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.

News of the Year: In 2021, new algorithms bringing improved precision and/or accelerations were incorporated into Faust, GPU support was completed together with a systematic optimization of the code (including the ability to run it in float instead of double precision), and PIP packages were made available to ease the installation of faust.

In 2020, major efforts were put into finalizing Python wrappers, producing tutorials using Jupyter notebooks and Matlab livescripts, as well as substantial refactoring of the code to optimize its efficiency and exploit GPUs.

In april 2018, a Software Development Initiative (ADT REVELATION) started in for the maturation of FAuST. A first step was to complete and robustify Matlab wrappers, to code Python wrappers with the same functionality, and to setup a continuous integration process. A second step was to simplify the parameterization of the main algorithms. The roadmap for next year includes showcasing examples and optimizing computational efficiency.

In 2017, new Matlab code for fast approximate Fourier Graph Transforms have been included, based on the approach described in the papers:

-Luc Le Magoarou, Rémi Gribonval, "Are There Approximate Fast Fourier Transforms On Graphs?", ICASSP 2016 .

-Luc Le Magoarou, Rémi Gribonval, Nicolas Tremblay, "Approximate fast graph Fourier transforms via multi-layer sparse approximations", IEEE Transactions on Signal and Information Processing over Networks,2017.

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

Publications: [hal-01416110](#), [hal-01627434](#), [hal-01167948](#), [hal-01254108](#), [tel-01412558](#), [hal-01156478](#), [hal-01104696](#), [hal-01158057](#), [hal-03132013](#)

Contact: Remi Gribonval

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

8 New results

8.1 Graph Signal Processing, Optimal Transport and Machine Learning on Graphs

8.1.1 Works on Gromov-Wasserstein: graph dictionary learning

Participants: Titouan Vayer.

Collaborations with Cédric Vincent-Cuaz (PhD student, MAASAI, Université Côte d'Azur), Rémi Flamary (CMAP, Ecole Polytechnique), Marco Corneli (MAASAI, Université Côte d'Azur) and Nicolas Courty (IRISA, Université Bretagne Sud).

The Gromov-Wasserstein (GW) distance is derived from optimal transport (OT) theory. The interest of OT lies both in its ability to provide relationships, connections, between sets of points and distances between probability distributions. By modeling graphs as probability distributions GW has become an important tool in many ML tasks involving structured data. Based on GW as a fidelity term, we proposed in [34] an efficient graph dictionary learning algorithm that allows to describe graphs as a simple composition of smaller graphs (atoms of the dictionary). We proposed a stochastic algorithm capable of learning a dictionary-like representation in the complex setting where the graphs in the dataset arrive progressively in time. We showed that these representations are particularly efficient for tasks such as change detection for structured data and clustering of graphs. We proposed an alternative approach in [48] which goal is to learn a single graph of large size whose subgraphs will best match (according to the GW criterion) the graphs of the dataset.

In another line of works, in collaboration with Clément Bonet (PhD student, IRISA, Université Bretagne-Sud), Nicolas Courty, François Septier (LMBA, Université de Bretagne Sud) and Lucas Drumetz (Lab-STICC OSE, IMT Atlantique), we proposed an extension of the GW framework for shape matching problems [11]. It consists in finding an optimal plan between the measures projected on a wisely chosen subspace and then completing it in a nearly optimal transport plan on the whole space. The advantage is to lower the computational complexity of the GW distance.

8.1.2 Diffused Wasserstein Distance for Optimal transport between attributed graphs

Participants: Paulo Gonçalves, Rémi Gribonval, Amélie Barbe, Titouan Vayer.

This work is a collaboration with Pierre Borgnat (CNRS) from the the Physics Lab of ENS de Lyon, Marc Sebban, Professor at the LabHC of University Jean Monet, and Sibylle Marcotte (student at ENS de Rennes).

In a series of recent articles, we proposed the Diffusion Wasserstein (DW) distance, a generalization of the standard Wasserstein to undirected and connected graphs where nodes are described by feature vectors. Using the heat diffusion equation constructed on the exponential kernel of the graph's Laplacian, we locally average the attributes of the nodes over a neighborhood that is controlled by the diffusion time. Like the fused Gromov-Wasserstein distance, this mixed distance allows to compute an optimal transport plan that captures both the structural and the feature information of the graphs. A big advantage of the Diffusion Distance though, is its computational cost that remains significantly inferior to that of the fused Gromov Wasserstein distance. Moreover, applied to different domain adaptation tasks, we experimentally showed that in many difficult situations, the DW distance was able to outperform the most recent concurrent methods.

To further reduce the computational cost of the diffusion Wasserstein distance, we proposed to use a Chebyshev approximation of the diffusion operator applied to the features vectors. In the course of this work, we were also able to tighten the theoretical approximation bounds, which in turn permits to significantly improve estimates of the polynomial order for a prescribed error [31].

Finally, to address the classical problem of tuning the diffusion time, the unique free parameter of DW distance, we devised a triplet loss based method that permits to find the best diffusion time in the context of domain adaptation tasks [27].

8.2 Sparse deep neural networks : theory and algorithms

8.2.1 Mathematics of deep learning: approximation theory, scale-invariance, and regularization

Participants: Rémi Gribonval, Pierre Stock, Antoine Gonon, Elisa Riccietti, Vincent Schellekens.

Collaborations with Facebook AI Research, Paris, with Nicolas Brisebarre (ARIC team, ENS de Lyon), and with Yann Traonmilin (IMB, Bordeaux) and Samuel Vaiter (JAD, Dijon)

Our paper studying the expressivity of sparse deep neural networks from an approximation theoretic perspective and highlighting the role of depth to enable efficient approximation of functions with very limited smoothness was published this year [8]. Motivated by the importance of quantizing networks besides pruning them to achieve sparsity, we started to investigate the approximation theoretic properties of *quantized* deep networks, with the objective of defining and comparing the corresponding approximation classes with the unquantized ones.

Neural networks with the ReLU activation function are described by weights and bias parameters, and realized 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. These considerations in turn lead to the notion of identifiability – the ability to recover (the equivalence class of) parameters from the sole knowledge of the realization of the corresponding network. We studied this problem in depth through the lens of a new embedding of ReLU neural network parameters of any depth. The proposed embedding is invariant to scalings and provides a locally linear parameterization of the realization of the network. Leveraging these two key properties, we derived some conditions under which a deep ReLU network is indeed locally identifiable from the knowledge of the realization on a finite set of samples. We studied the shallow case in more depth, establishing necessary and sufficient conditions for the network to be identifiable from a bounded subset [22].

An important challenge in deep learning is to promote sparsity during the learning phase using a regularizer. In the classical setting of linear inverse problems, it is well known that the ℓ^1 norm is a convex regularizer lending itself to efficient optimization and endowed with stable recovery guarantees.

A particular challenge is to understand to what extent using an ℓ^1 penalty in this context is also well-founded theoretically, and to possibly design alternate regularizers if possible. On the one hand, we

started investigating the properties of minimizers of the ℓ^1 norm in deep learning problems. On the other hand, we considered the abstract problem of recovering elements of a low-dimensional model set from under-determined linear measurements. Considering the minimization of a convex regularizer subject to a data fit constraint, we explored the notion of a "best" convex regularizer given a model set. This was formalized as a regularizer that maximizes a compliance measure with respect to the model. Several notions of compliance were studied and analytical expressions were obtained for compliance measures based on the best-known recovery guarantees with the restricted isometry property. This led to a formal proof of the optimality of the ℓ^1 -norm for sparse recovery and of the nuclear norm for low-rank matrix recovery for these compliance measures. We also investigated the construction of an optimal convex regularizer using the example of sparsity in levels [46].

8.2.2 Algorithms for quantized networks

Participants: Rémi Gribonval, Pierre Stock, Elisa Riccietti.

Collaboration with Facebook AI Research, Paris

From a more computational perspective, within the framework of the Ph.D. of Pierre Stock [40], we proposed last year a technique to drastically compress neural networks using product quantization [72], and this year an approach to learn networks that can be more efficiently quantized [33]. We also started to study efficient optimization algorithms to train quantized networks that leverage multiple quantization levels.

8.2.3 Deep sparse factorizations: hardness, algorithms and identifiability

Participants: Rémi Gribonval, Elisa Riccietti, Marion Foare, Léon Zheng, Quoc-Tung Le.

Collaboration with Valeo AI, Paris

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

Last year, from an algorithmic perspective, we analyzed and fixed a weakness of proximal algorithms in sparse matrix factorization. We also described a new tractable proximal operator called Generalized Hungarian Method, associated to so-called k -regular matrices, which are useful for the factorization of a class of matrices associated to fast linear transforms. We further illustrated the effectiveness of our proposals by numerical experiments on the Hadamard Transform and magnetoencephalography matrix factorization. This work was published this year in a conference [29], and the new proximal operator was implemented in the FA μ ST software library (see Section 7).

From a theoretical perspective, we considered the hardness and uniqueness properties of sparse matrix factorization. First, even with only two factors and a fixed, known support, we showed that optimizing the coefficients of the sparse factors can be an NP-hard problem. Besides, we studied the landscape of the corresponding optimization problem and exhibited "easy" instances where the problem can be solved to global optimality with an algorithm demonstrated to be orders of magnitude faster than classical gradient based methods [43]. In complement, we investigated the essential uniqueness of sparse matrix factorizations, both with two factors [50] and in a multi-layer setting [49]. We combined these results with a focus on so-called butterfly supports to achieve a multilayer sparse factorization algorithm able to learn fast transforms essentially at the cost of a single matrix-vector multiplication, with exact recovery guarantees [30]. A first version of the corresponding algorithm was incorporated in the FA μ ST software library (see Section 7) and is subject to software optimizations to further speed it up.

8.3 Statistical learning, dimension reduction, and privacy preservation

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

Participants: Rémi Gribonval, Titouan Vayer, Ayoub Belhadji, Vincent Schellekens, Luc Giffon, Léon Zheng.

Collaborations with Gilles Blanchard (Univ. Paris-Saclay), Yann Traonmilin (IMB, Bordeaux), Laurent Jacques and Vincent Schellekens (U. Louvain, Belgium), Nicolas Keriven (GIPSA-lab, Grenoble), Phil Schniter (Ohio State Univ.), and with Valeo AI

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. Our papers establishing statistical guarantees on the generalization error of this procedure, first in a general abstract setting illustrated on PCA [6], then for the specific case of compressive k -means and compressive Gaussian Mixture Modeling [16], were published this year. A tutorial paper on the principle and the main guarantees of compressive learning was also finalized and published this year [7].

Theoretical guarantees in compressive learning fundamentally rely on comparing certain metrics between probability distributions. This year we established some conditions under which the Wasserstein distance can be controlled by Maximum Mean Discrepancy (MMD) norms, which are defined using reproducing kernel Hilbert spaces. Based on the relations between the MMD and the Wasserstein distance, we provide new guarantees for compressive statistical learning by introducing and studying the concept of Wasserstein learnability of the learning task [47].

Dimension reduction in compressive learning exploits the ability to approximate certain kernels by finite dimensional quadratures. We studied a quadrature, proposed by Ermakov and Zolotukhin in the sixties, through the lens of kernel methods. The nodes of this quadrature rule follow the distribution of a determinantal point process, while the weights are defined through a linear system, similarly to the optimal kernel quadrature. We showed how these two classes of quadrature are related, and we proved a tractable formula of the expected value of the squared worst-case integration error on the unit ball of an RKHS of the former quadrature. In particular, this formula involves the eigenvalues of the corresponding kernel and leads to improving on the existing theoretical guarantees of the optimal kernel quadrature with determinantal point processes [28].

From a more empirical perspective, we pursued our efforts to make sketching for compressive learning more versatile and efficient. This notably involved exploring how to adapt the sketching pipeline to exploit optical processing units (OPUs) for energy-efficient fast random projection, and investigating the ability to exploit sketching in large-scale deep self-supervised learning scenarios.

Finally, making the connection between graph learning and sketching methods, we have recently started to study the practical possibility and theoretical limitations of using a sketching technique to estimate the precision matrix involved in the Graphical Lasso algorithm.

8.3.2 Privacy preservation

Participants: Rémi Gribonval, Clément Lalanne.

Collaborations with Aurélien Garivier (UMPA, ENS de Lyon) and SARUS, Paris; and with Laurent Jacques and Vincent Schellekens (U. Louvain, Belgium), Florimond Houssiau and Yves-Alexandre de Montjoye (Imperial College, London)

In the context of the Ph.D. thesis of Antoine Chatalic (in the PANAMA team in Rennes, defended last year) we showed [13] that a simple perturbation of the sketching mechanism with additive noise is

sufficient to satisfy differential privacy, a well-established formalism for defining and quantifying the privacy of a random mechanism. We combined this with a feature subsampling mechanism, which reduces the computational cost without damaging privacy. The framework was applied to the tasks of Gaussian modeling, k-means clustering and principal component analysis (PCA), for which sharp privacy bounds were derived. Empirically, the quality (for subsequent learning) of the compressed representation produced by this mechanism is strongly related with the induced noise level, for which we gave analytical expressions.

This year we also addressed problem of differentially private estimation of multiple quantiles (MQ) of a dataset, a key building block in modern data analysis. We showed how to implement the non-smoothed Inverse Sensitivity (IS) mechanism for this specific problem and established that the resulting method is closely related to the recent JointExp algorithm, sharing in particular the same computational complexity and a similar efficiency. We also identified pitfalls of the two approaches on certain peaked distributions, and proposed a fix. Numerical experiments showed that the empirical efficiency of the resulting algorithms is similar to the non-smoothed methods for non-degenerate datasets, but orders of magnitude better on real datasets with repeated values.

8.4 Large-scale convex and nonconvex optimization

Participants: Elisa Riccietti, Paulo Gonçalves, Federico Grillini, Giovanni Seraghiti, Guillaume Lauga.

Collaboration with Nelly Pustelnik (CNRS, ENS de Lyon)

In the context of the Ph.D. work of Guillaume Lauga and the previous internships of Federico Grillini and Giovanni Seraghiti, this year we started to study the combination of proximal methods and multiresolution analysis in large-scale image denoising problems. The use of multiresolution schemes, such as wavelets transforms, is not new in imaging and is widely used to define regularization strategies. We studied the use of such techniques to a wider extent, as a solution to accelerate proximal algorithms usually used for their solution and make them usable for problems of very large dimensions. In the fashion of multilevel gradient methods [3], popular techniques in smooth optimization, we designed multilevel versions of proximal algorithms employing wavelet transforms as transfer operators.

In the context of the internship of Hugo Gouttenegre, we also pursued the investigations in 3D MRI super-resolution using nonconvex optimization models. We provided a 3D extension of the Discrete Mumford-Shah, allowing to jointly perform a 3D super-resolution and a segmentation of the high-resolution volume. New phantom acquisitions were conducted, including a high-resolution groundtruth volume, to evaluate the quantitative performances of this approach. A numerical toolbox is under construction.

9 Bilateral contracts and grants with industry

9.1 Bilateral grants with industry

- CIFRE contract with Facebook Artificial Intelligence Research, Paris on Deep neural networks for large scale learning

Participants: Rémi Gribonval, Pierre Stock.

Duration: 3 years (2018-2021)

Partners: Facebook Artificial Intelligence Research, Paris; Inria-Grenoble

Funding: Facebook Artificial Intelligence Research, Paris; ANRT

The overall objective of this thesis [40] was to design, analyze and test large scale machine learning algorithms with applications to computer vision and natural language processing. A major challenge was to design compression techniques able to replace complex and deep neural networks with much more compact ones while preserving the capacity of the initial network to achieve the targeted task.

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

Participants: Rémi Gribonval, Elisa Riccietti, Léon Zheng.

Duration: 3 years (2021-2024)

Partners: Valeo AI, Paris; ENS de Lyon

Funding: Valeo AI, Paris; ANRT

Context: Chaire IA AllegroAssai 10.1.1

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

- Funding from Facebook Artificial Intelligence Research, Paris

Participants: Rémi Gribonval.

Duration: 4 years (2021-2024)

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

Funding: Facebook Artificial Intelligence Research, Paris

Context: Chaire IA AllegroAssai 10.1.1

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

10 Partnerships and cooperations

10.1 National initiatives

10.1.1 ANR IA Chaire : AllegroAssai

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

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

AllegroAssai focuses on the design of machine learning techniques endowed both with statistical guarantees (to ensure their performance, fairness, privacy, etc.) and provable resource-efficiency (e.g. in terms of bytes and flops, which impact energy consumption and hardware costs), robustness in adversarial conditions for secure performance, and ability to leverage domain-specific models and expert knowledge. The vision of AllegroAssai is that the versatile notion of sparsity, together with sketching techniques using random features, are key in harnessing these fundamental tradeoffs. The first pillar of the project is to investigate sparsely connected deep networks, to understand the tradeoffs between the

approximation capacity of a network architecture (ResNet, U-net, etc.) and its “trainability” with provably-good algorithms. A major endeavor is to design efficient regularizers promoting sparsely connected networks with provable robustness in adversarial settings. The second pillar revolves around the design and analysis of provably-good end-to-end sketching pipelines for versatile and resource-efficient large-scale learning, with controlled complexity driven by the structure of the data and that of the task rather than the dataset size.

10.1.2 ANR DataRedux

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

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

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

10.1.3 ANR Darling

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

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

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

remote areas has the potential to path the pave to the emergence of the critically needed future field of dynamical network signal processing.

10.1.4 GDR ISIS project MOMIGS

Participants: Elisa Riccietti (*correspondant*), Marion Foare, Trieu Vy Le Hoang, Paulo Gonçalves.

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

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

10.2 Regional initiatives

10.2.1 Labex CominLabs LeanAI

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

Duration of the project: **October 2021-December 2024.**

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

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

10.2.2 Labex Emerging Topics

Participants: Marion Foare (*correspondant*).

Duration of the project: **April 2019-December 2022.**

Collaboration with Eric Van Reeth (Creatis, Lyon)

Magnetic Resonance Imaging (MRI) is an extremely important anatomical and functional imaging technique, widely used by physicists to establish medical diagnosis. Acquiring high resolution volumes is desirable in many clinical and pre-clinical applications to accurately adapt the treatment to the measurements, or simply obtain highly resolved images of small anatomical structures. However, directly acquiring high-resolution volumes implies: i) long scanning times, which are often not tolerated by patients and children, and ii) images with low signal-to-noise ratio. Therefore, it is of particular interest to quickly acquire low-resolution volumes, and enhance their resolution as a post-processing step.

This project aims at developing new techniques to build super-resolution images for 3D MRI, that can take into account more physical constraints, such as prior medical knowledge, and to derive efficient machine learning algorithms suited for large scale data, with theoretical guarantees. In particular, we explore specialized piecewise smooth reconstruction variational methods, like the Mumford-Shah (MS) and the Total Variation (TV) variants, and to adapt their fitting terms as well as their optimization algorithms. The main originality of this project is to combine resolution enhancement and segmentation in MRI (usually performed as two distinct post-processing steps), starting from the MS model, a seminal tool originally designed for image denoising and segmentation tasks. This approach will improve the quality of the reconstruction both in terms of sharpness and smoothness, and help the doctors with reaching a diagnosis.

11 Dissemination

Participants: Rémi Gribonval, Paulo Gonçalves, Marion Foare, Elisa Riccietti.

11.1 Promoting scientific activities

11.1.1 Scientific events: organisation

Member of the organizing committees

- Rémi Gribonval, [Journées de Statistiques 2022](#), Lyon

11.1.2 Scientific events: selection

Member of the conference program committees

- Rémi Gribonval, [GRETSI 2022](#).
- Rémi Gribonval, 10th SMAI-SIGMA conference on [Curves and Surfaces](#)
- Rémi Gribonval, 2022 Spring School on Machine Learning ([EPIT22](#)), CIRM, Spring 2022
- Rémi Gribonval, MiLYON Spring School on Machine Learning, Saint-Etienne, Spring 2021 (postponed to 2022 then cancelled due to Covid-19)
- Rémi Gribonval, Conference on [Mathematics for Audio and Music Signal Processing](#), CIRM 2021 (cancelled due to Covid-19)

11.1.3 Journal

Member of the editorial boards

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

11.1.4 Invited talks

- R. Gribonval was a keynote speaker at the international conference EUSIPCO 2021 and an invited speaker at the national conference CAp21.
- E. Riccietti was an invited speaker at 13th JLESC Workshop.

11.1.5 Leadership within the scientific community

- Rémi Gribonval is a member of the Scientific Committee of **RT MIA** (formerly GDR MIA)
- Rémi Gribonval is a member of the Comité de Liaison SIGMA-SMAI
- Rémi Gribonval is a member of the Cellule ERC of INS2I, mentoring for ERC candidates in the STIC domain

11.1.6 Scientific expertise

- Rémi Gribonval is a member of the Scientific Advisory Board (vice-president) of the Acoustics Research Institute of the Austrian Academy of Sciences, and a member of the Commission Prospective of Institut de Mathématiques de Marseille
- Rémi Gribonval, member of the EURASIP Special Area Team (SAT) on Signal and Data Analytics for Machine Learning (SiG-DML) since 2015.

11.1.7 Research administration

- Paulo Gonçalves is Deputy Scientific Director of the new research center of Inria in Lyon.

11.2 Teaching - Supervision - Juries

11.2.1 Teaching

- Master :
 - Rémi Gribonval: Inverse problems and high dimension; Mathematical foundations of deep neural networks; Concentration of measure in probability and high-dimensional statistical learning; M2, ENS Lyon
- Engineer cycle (Bac+3 to Bac+5):
 - Paulo Gonçalves: Traitement du Signal (déterministe, aléatoire, numérique), Estimation statistique. 80 heures Eq. TD. CPE Lyon, France
 - Marion Foare: Traitement du Signal (déterministe, numérique, aléatoire), Traitement et analyse d'images, Optimisation, Compression, Projets. 280 heures Eq. TD. CPE Lyon, France
 - Elisa Riccietti: M1 course Optimization and Approximation (28h) and 19h of tutor responsibility at ENS Lyon

11.2.2 Supervision

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

- Marion Foare is involved in the co-supervision of the Ph.D. of Hoang Trieu Vy Le since 2021 (Laboratoire de Physique, Lyon).
- Elisa Riccietti is involved in the co-supervision of the Ph.D. of Valentin Mercier since 2021 (IRIT, Toulouse).

The following PhDs were defended in DANTE in 2021:

- Pierre Stock, Université de Lyon [40] (funded by ANRT and Facebook Artificial Intelligence Research; co-supervisors Rémi Gribonval and Hervé Jégou), *Efficiency and Redundancy in Deep Learning Models: Theoretical Considerations and Practical Applications*, April 2021
- Amélie Barbe, Université de Lyon (funded by ACADEMICS project, IdexLyon; co-supervisors Paulo Gonçalves, Pierre Borgnat and Marc Sebban), *Diffusion-Wasserstein distances for attributed graphs*, December 2021

11.2.3 Juries

Members of the DANTE team participated to the following juries

- PhD juries: Alexandre Araujo (Université Paris IX Dauphine, *member*); Marina Kremé (Aix-Marseille Université, *chair*); Pierre Humbert (University Paris-Saclay, *chair*); Raphaël Truffet (Université de Rennes I, *chair*); Vincent Schellekens (Université Catholique de Louvain, *reviewer*), PhD defence session at University of Florence (*member*)

12 Scientific production

12.1 Major publications

- [1] E. Bautista, P. Abry and P. Gonçalves. ‘ L^Y -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://hal.inria.fr/hal-02063780>.
- [2] Q. Bertrand, Q. Klopfenstein, M. Massias, M. Blondel, S. Vaiter, A. Gramfort and J. Salmon. *Implicit differentiation for fast hyperparameter selection in non-smooth convex learning*. 18th May 2021. URL: <https://hal.archives-ouvertes.fr/hal-03228663>.
- [3] H. Calandra, S. Gratton, E. Riccietti and X. Vasseur. ‘On a multilevel Levenberg–Marquardt method for the training of artificial neural networks and its application to the solution of partial differential equations’. In: *Optimization Methods and Software* (2020), pp. 1–26. DOI: [10.1080/10556788.2020.1775828](https://doi.org/10.1080/10556788.2020.1775828). URL: <https://hal.archives-ouvertes.fr/hal-02956018>.
- [4] M. Foare, N. Pustelnik and L. Condat. ‘Semi-Linearized Proximal Alternating Minimization for a Discrete Mumford-Shah Model’. In: *IEEE Transactions on Image Processing* (7th Oct. 2019), pp. 1–13. DOI: [10.1109/TIP.2019.2944561](https://doi.org/10.1109/TIP.2019.2944561). URL: <https://hal.archives-ouvertes.fr/hal-01782346>.
- [5] B. Girault, P. Gonçalves and É. 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>.
- [6] R. Gribonval, G. Blanchard, N. Keriven and Y. Traonmilin. ‘Compressive Statistical Learning with Random Feature Moments’. In: *Mathematical Statistics and Learning* (2021). Main novelties between version 1 and version 2: improved concentration bounds, improved sketch sizes for compressive k-means and compressive GMM that now scale linearly with the ambient dimension Main novelties of version 3: all content on compressive clustering and compressive GMM is now developed in the companion paper hal-02536818; improved statistical guarantees in a generic framework with illustration of the improvements on compressive PCA. URL: <https://hal.inria.fr/hal-01544609>.
- [7] 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://hal.inria.fr/hal-03350599>.
- [8] R. Gribonval, G. Kutyniok, M. Nielsen and F. Voigtlaender. ‘Approximation spaces of deep neural networks’. In: *Constructive Approximation*. special issue on “Deep Networks in Approximation Theory” (2020). URL: <https://hal.inria.fr/hal-02117139>.

- [9] 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.archives-ouvertes.fr/hal-02263500>.
- [10] B. Ricaud, P. Borgnat, N. Tremblay, P. Gonçalves and P. Vandergheynst. ‘Fourier could be a Data Scientist: from Graph Fourier Transform to Signal Processing on Graphs’. In: *Comptes Rendus. Physique* (19th Sept. 2019), pp. 474–488. DOI: [10.1016/j.crhy.2019.08.003](https://doi.org/10.1016/j.crhy.2019.08.003). URL: <https://hal.inria.fr/hal-02304584>.

12.2 Publications of the year

International journals

- [11] C. Bonet, T. Vayer, N. Courty, F. Septier and L. Drumetz. ‘Subspace Detours Meet Gromov-Wasserstein’. In: *Algorithms. Special Issue Optimal Transport: Algorithms and Applications* 14 (17th Dec. 2021), pp. 1–29. DOI: [10.3390/a14120366](https://doi.org/10.3390/a14120366). URL: <https://hal.archives-ouvertes.fr/hal-03500536>.
- [12] A. Chadda, M. Stojanova, T. Begin, A. Busson and I. Guérin-Lassous. ‘Assigning Channels in WLANs with Channel Bonding: A Fair and Robust Strategy’. In: *Computer Networks* (1st June 2021), pp. 1–17. URL: <https://hal.archives-ouvertes.fr/hal-03281096>.
- [13] 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). DOI: [10.1093/imaia/iaab005](https://doi.org/10.1093/imaia/iaab005). URL: <https://hal.inria.fr/hal-02496896>.
- [14] C. Gaultier, S. Kitić, R. Gribonval and N. Bertin. ‘Sparsity-based audio declipping methods: selected overview, new algorithms, and large-scale evaluation’. In: *IEEE/ACM Transactions on Audio, Speech and Language Processing* 29 (2021), pp. 1174–1187. DOI: [10.1109/TASLP.2021.3059264](https://doi.org/10.1109/TASLP.2021.3059264). URL: <https://hal.inria.fr/hal-02611226>.
- [15] 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://hal.inria.fr/hal-01544609>.
- [16] 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 (21st Aug. 2021), pp. 165–257. DOI: [10.4171/msl/21](https://doi.org/10.4171/msl/21). URL: <https://hal.inria.fr/hal-02536818>.
- [17] 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://hal.inria.fr/hal-03350599>.
- [18] R. Gribonval, G. Kutyniok, M. Nielsen and F. Voigtlaender. ‘Approximation spaces of deep neural networks’. In: *Constructive Approximation. special issue on "Deep Networks in Approximation Theory"* (2021). DOI: [10.1007/s00365-021-09543-4](https://doi.org/10.1007/s00365-021-09543-4). URL: <https://hal.inria.fr/hal-02117139>.
- [19] R. Grünblatt, I. Guérin-Lassous and O. Simonin. ‘A distributed antenna orientation solution for optimizing communications in a fleet of UAVs’. In: *Computer Communications* 181 (1st Jan. 2022), pp. 102–115. DOI: [10.1016/j.comcom.2021.09.020](https://doi.org/10.1016/j.comcom.2021.09.020). URL: <https://hal.archives-ouvertes.fr/hal-03400498>.
- [20] B. Jiang, P. Nain and D. Towsley. ‘Covert Cycle Stealing in a Single FIFO Server’. In: *ACM Transactions on Modeling and Performance Evaluation of Computing Systems* (2021), pp. 1–35. URL: <https://hal.inria.fr/hal-03216762>.
- [21] P. Nain, N. K. Panigrahy, P. Basu and D. Towsley. ‘One-dimensional Service Networks and Batch Service Queues’. In: *Queueing Systems* (2021). URL: <https://hal.inria.fr/hal-03066179>.

- [22] P. Stock and R. Gribonval. ‘An Embedding of ReLU Networks and an Analysis of their Identifiability’. In: *Constructive Approximation* (2022). URL: <https://hal.archives-ouvertes.fr/hal-03292203>.
- [23] M. Stojanova, T. Begin and A. Busson. ‘A Markov Model for Performance Evaluation of Channel Bonding in IEEE 802.11’. In: *Ad Hoc Networks* (2021). URL: <https://hal.archives-ouvertes.fr/hal-03158454>.
- [24] S. Unicomb, G. Iñiguez, J. Gleeson and M. Karsai. ‘Dynamics of cascades on burstiness-controlled temporal networks’. In: *Nature Communications* 12.1 (Dec. 2021), pp. 1–9. DOI: [10.1038/s41467-020-20398-4](https://doi.org/10.1038/s41467-020-20398-4). URL: <https://hal.inria.fr/hal-03117999>.
- [25] G. Vardoyan, S. Guha, P. Nain and D. Towsley. ‘On the Stochastic Analysis of a Quantum Entanglement Distribution Switch’. In: *IEEE Transactions on Quantum Engineering* (Feb. 2021). URL: <https://hal.inria.fr/hal-03175703>.

International peer-reviewed conferences

- [26] L. Abdelwedoud, A. Busson and I. Guérin-Lassous. ‘Use of a Weighted Conflict Graph in the Channel Selection Operation for Wi-Fi Networks’. In: WONS 2021 - 16th Wireless On-demand Network systems and Services Conference. Virtual Conference, France, 9th Mar. 2021, pp. 1–4. URL: <https://hal.inria.fr/hal-03266173>.
- [27] A. Barbe, P. Gonçalves, M. Sebban, P. Borgnat, R. Gribonval and T. Vayer. ‘Optimization of the Diffusion Time in Graph Diffused-Wasserstein Distances: Application to Domain Adaptation’. In: ICTAI 2021 - 33rd IEEE International Conference on Tools with Artificial Intelligence. Virtual conference, France: IEEE, 1st Nov. 2021, pp. 1–8. URL: <https://hal.inria.fr/hal-03353622>.
- [28] A. Belhadji. ‘An analysis of Ermakov-Zolotukhin quadrature using kernels’. In: NeurIPS 2021 - 35th Conference on Neural Information Processing Systems. Virtual-only Conference, Australia, 6th Dec. 2021, pp. 1–17. URL: <https://hal.archives-ouvertes.fr/hal-03405615>.
- [29] Q.-T. Le and R. Gribonval. ‘Structured Support Exploration For Multilayer Sparse Matrix Factorization’. In: ICASSP 2021 - IEEE International Conference on Acoustics, Speech and Signal Processing. Toronto, Ontario, Canada: IEEE, 6th June 2021, pp. 1–5. URL: <https://hal.inria.fr/hal-03132013>.
- [30] 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. Singapore, Singapore, 22nd May 2022. URL: <https://hal.inria.fr/hal-03438881>.
- [31] S. Marcotte, A. Barbe, R. Gribonval, T. Vayer, M. Sebban, P. Borgnat and P. Gonçalves. ‘Fast Multiscale Diffusion on Graphs’. In: ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing. Singapore, Singapore, 22nd May 2022. URL: <https://hal.archives-ouvertes.fr/hal-03212764>.
- [32] N. Santi, R. Grünblatt, B. Foubert, A. Hameed, J. Violos, A. Leivadreas and N. Mitton. ‘Automated and Reproducible Application Traces Generation for IoT Applications’. In: Q2SWinet 2021 - 17th ACM Symposium on QoS and Security for Wireless and Mobile Networks. Alicante, Spain: ACM, 22nd Nov. 2021, pp. 1–8. DOI: [10.1145/3479242.3487321](https://doi.org/10.1145/3479242.3487321). URL: <https://hal.archives-ouvertes.fr/hal-03390693>.
- [33] P. Stock, A. Fan, B. Graham, E. Grave, R. Gribonval, H. Jegou and A. Joulin. ‘Training with Quantization Noise for Extreme Model Compression’. In: International Conference on Learning Representations 2021. Vienna, Austria, 4th May 2021. URL: <https://hal.inria.fr/hal-03136442>.
- [34] C. Vincent-Cuaz, T. Vayer, R. Flamary, M. Corneli and N. Courty. ‘Online Graph Dictionary Learning’. In: ICML 2021 - 38th International Conference on Machine Learning. Virtual Conference, United States, 2021. URL: <https://hal.archives-ouvertes.fr/hal-03140349>.

Conferences without proceedings

- [35] A. Bardou, T. Begin and A. Busson. ‘Improving the Spatial Reuse in IEEE 802.11ax WLANs: A Multi-Armed Bandit Approach’. In: MSWiM’21 - 24th ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems. Alicante, Spain: ACM, 22nd Nov. 2021. URL: <https://hal.archives-ouvertes.fr/hal-03445712>.
- [36] A. Bonnefond, O. Simonin and I. Guérin-Lassous. ‘Extension des Modèles de Flocking aux Environnements avec Obstacles et Communications Dégradées’. In: JFSMA. Bordeaux, France, 28th June 2021. URL: <https://hal.archives-ouvertes.fr/hal-03369148>.
- [37] A. Bonnefond, O. Simonin and I. Guérin-Lassous. ‘Extension of Flocking Models to Environments with Obstacles and Degraded Communications’. In: IROS 2021 - IEEE/RSJ International Conference on Intelligent Robots and Systems. Prague / Virtual, Czech Republic: IEEE, 30th July 2021, pp. 1–7. URL: <https://hal.archives-ouvertes.fr/hal-03365129>.
- [38] E. Guérin, T. Begin, A. Busson and I. Guérin-Lassous. ‘Towards a Throughput and Energy Efficient Association Strategy for Wi-Fi/LiFi Heterogeneous Networks’. In: PE-WASUN 2021 - 18th ACM International Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, and Ubiquitous Networks. Alicante, Spain: ACM, 22nd Nov. 2021. DOI: [10.1145/1122445.1122456](https://doi.org/10.1145/1122445.1122456). URL: <https://hal.archives-ouvertes.fr/hal-03445734>.

Doctoral dissertations and habilitation theses

- [39] R. Grünblatt. ‘From WiFi Performance Evaluation to Controlled Mobility in Drone Networks’. Université Claude Bernard Lyon 1, 8th Jan. 2021. URL: <https://hal.inria.fr/tel-03126953>.
- [40] P. Stock. ‘Efficiency and Redundancy in Deep Learning Models : Theoretical Considerations and Practical Applications’. Université de Lyon, 9th Apr. 2021. URL: <https://tel.archives-ouvertes.fr/tel-03208517>.

Reports & preprints

- [41] R. Gribonval, A. Chatalic, N. Keriven, V. Schellekens, L. Jacques and P. Schniter. *Sketching Datasets for Large-Scale Learning (long version)*. 26th Jan. 2021. URL: <https://hal.inria.fr/hal-02909766>.
- [42] B. Jiang, P. Nain and D. Towsley. *Covert Cycle Stealing in a Single FIFO Server (extended version)*. 4th May 2021. URL: <https://hal.inria.fr/hal-02497557>.
- [43] Q.-T. Le, E. Riccietti and R. Gribonval. *Spurious Valleys, Spurious Minima and NP-hardness of Sparse Matrix Factorization With Fixed Support*. 31st May 2021. URL: <https://hal.inria.fr/hal-03364668>.
- [44] P. Nain, G. Vardoyan, S. Guha and D. Towsley. *Analysis of a Tripartite Entanglement Distribution Switch*. 16th Jan. 2022. URL: <https://hal.inria.fr/hal-03195985>.
- [45] B. Pascal, P. Abry, N. Pustelnik, S. g. Roux, R. Gribonval and P. Flandrin. *Nonsmooth convex optimization to estimate the Covid-19 reproduction number space-time evolution with robustness against low quality data*. 20th Sept. 2021. URL: <https://hal.archives-ouvertes.fr/hal-03348154>.
- [46] Y. Traonmilin, R. Gribonval and S. Vaïter. *A theory of optimal convex regularization for low-dimensional recovery*. 6th Dec. 2021. URL: <https://hal.archives-ouvertes.fr/hal-03467123>.
- [47] T. Vayer and R. Gribonval. *Controlling Wasserstein distances by Kernel norms with application to Compressive Statistical Learning*. 17th Dec. 2021. URL: <https://hal.archives-ouvertes.fr/hal-03461492>.
- [48] C. Vincent-Cuaz, R. Flamary, M. Corneli, T. Vayer and N. Courty. *Semi-relaxed Gromov Wasserstein divergence with applications on graphs*. 13th Oct. 2021. URL: <https://hal.archives-ouvertes.fr/hal-03376923>.

- [49] L. Zheng, E. Riccietti and R. Gribonval. *Efficient Identification of Butterfly Sparse Matrix Factorizations*. 15th Feb. 2022. URL: <https://hal.inria.fr/hal-03362626>.
- [50] L. Zheng, E. Riccietti and R. Gribonval. *Identifiability in Two-Layer Sparse Matrix Factorization*. 16th Nov. 2021. URL: <https://hal.inria.fr/hal-03362875>.

12.3 Other

Softwares

- [51] [SW] Q.-T. Le and R. Gribonval, *Code for the paper "Structured Support Exploration For Multilayer Sparse Matrix Factorization"*, 14th Feb. 2022. LIC: BSD-3 Clause License. HAL: [hal-03572265](https://hal.inria.fr/hal-03572265), URL: <https://hal.inria.fr/hal-03572265>, VCS: <https://github.com/sieunhanbom04/support-exploration-matrix-factorization/tree/master>, SWHID: [sw:1:dir:6794fd3f41d50a4e51e6e7f11ee721efafda9806;origin=https://hal.archives-ouvertes.fr/hal-03572265;visit=sw:1:snp:e5760cb4d925b5f3529486ffde997d8e3e9ea07e;anchor=sw:1:rel:a190fb7ab69d170eede1666e9e3edec33c56224b;path=/](https://hal.archives-ouvertes.fr/hal-03572265)).
- [52] [SW] S. Marcotte, A. Barbe, R. Gribonval, T. Vayer, M. Sebban, P. Borgnat and P. Gonçalves, *Code for reproducible research - Fast Multiscale Diffusion on Graphs*, 16th Feb. 2022. LIC: BSD 3-Clause License. HAL: [hal-03576498](https://hal.inria.fr/hal-03576498), URL: <https://hal.inria.fr/hal-03576498>, VCS: <https://github.com/sibyllema/Fast-Multiscale-Diffusion-on-Graphs>.
- [53] [SW] L. Zheng, Q.-T. Le, E. Riccietti and R. Gribonval, *Code for reproducible research - Fast learning of fast transforms, with guarantees*, 2nd Feb. 2022. LIC: BSD 3-Clause License. HAL: [hal-03552956](https://hal.inria.fr/hal-03552956), URL: <https://hal.inria.fr/hal-03552956>, VCS: <https://github.com/leonzheng2/butterfly>, SWHID: [sw:1:dir:42d42bf905c9f3bbcd6cf5e1e0cda6be2d0d63de;origin=https://hal.archives-ouvertes.fr/hal-03552956;visit=sw:1:snp:6a2256426a0f3a7491086fbf2da554d7589d1e18;anchor=sw:1:rel:f477382c82ada89647864605ea44b12383c04167;path=/](https://hal.archives-ouvertes.fr/hal-03552956)).

12.4 Cited publications

- [54] P. Abry, N. Pustelnik, S. G. Roux, P. Jensen, P. Flandrin, R. Gribonval, C.-G. Lucas, É. Guichard, P. Borgnat and N. B. Garnier. 'Spatial and temporal regularization to estimate COVID-19 reproduction number $R(t)$: Promoting piecewise smoothness via convex optimization'. In: *PLoS ONE* 15.8 (Aug. 2020), e0237901. DOI: [10.1371/journal.pone.0237901](https://doi.org/10.1371/journal.pone.0237901). URL: <https://hal.inria.fr/hal-02921836>.
- [55] H. H. Bauschke, P. L. Combettes et al. *Convex analysis and monotone operator theory in Hilbert spaces*. Vol. 408. Springer, 2011.
- [56] Boche, Holger, Calderbank, Robert, Kutyniok, Gitta and Vybiral, Jan. *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&pg=PA340&dq=intitle:Compressed+Sensing+and+its+Applications&hl=&cd=1&source=gbs_api.
- [57] 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>.
- [58] 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>.
- [59] 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.
- [60] 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).

- [61] P. M. Djuric and R. C. *Cooperative and Graph Signal Processing: Principle and Applications*. Academic Press, 2018.
- [62] 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=d5b6lJI9BvAC&printsec=frontcover&dq=sparse+and+redundant+representations&hl=&cd=1&source=gbs_api.
- [63] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. New York, NY: Springer, 2013. DOI: 10.1007/978-0-8176-4948-7. URL: <http://link.springer.com/10.1007/978-0-8176-4948-7>.
- [64] J. Friedman, T. Hastie and R. Tibshirani. ‘Sparse inverse covariance estimation with the graphical lasso’. In: *Biostatistics* 9.3 (2008), pp. 432–441.
- [65] 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).
- [66] 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>.
- [67] 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.
- [68] A. Rahimi and B. Recht. ‘Random features for large-scale kernel machines’. In: *NIPS*. 2007.
- [69] 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.
- [70] 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.
- [71] 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>.
- [72] P. Stock, A. Joulin, R. Gribonval, B. Graham and H. Jégou. ‘And the Bit Goes Down: Revisiting the Quantization of Neural Networks’. In: *ICLR 2020 - Eighth International Conference on Learning Representations*. Addis-Abeba, Ethiopia, Apr. 2020, pp. 1–11. URL: <https://hal.archives-ouvertes.fr/hal-02434572>.
- [73] I. Todic and P. Frossard. ‘Dictionary Learning’. In: *IEEE Signal Processing Magazine* 28.2 (), pp. 27–38. DOI: 10.1109/MSP.2010.939537. URL: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5714407>.
- [74] 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.
- [75] 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.
- [76] 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).
- [77] 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.