RESEARCH CENTRE

Inria Lyon Center

IN PARTNERSHIP WITH:

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

2022

ACTIVITY REPORT

Project-Team

DANTE

Dynamic Networks : Temporal and Structural Capture Approach

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

DOMAIN

Applied Mathematics, Computation and Simulation

THEME

Optimization, machine learning and statistical methods

Inria

Contents

Project-Team DANTE

Creation of the Project-Team: 2015 January 01

Keywords

Computer sciences and digital sciences

A3.4.1. – Supervised learning

A3.4.4. – Optimization and learning

A3.4.6. – Neural networks

A3.4.7. – Kernel methods

A3.4.8. – Deep learning

A3.5. – Social networks

A3.5.1. – Analysis of large graphs

A5.3.2. – Sparse modeling and image representation

A5.9. – Signal processing

A5.9.4. – Signal processing over graphs

A5.9.5. – Sparsity-aware processing

A5.9.6. – Optimization tools

A6.3.1. – Inverse problems

A8.2. – Optimization

A8.6. – Information theory

A8.12. – Optimal transport

Other research topics and application domains

B2.6. – Biological and medical imaging

B6.6. – Embedded systems

B7.2.1. – Smart vehicles

B9.5.1. – Computer science

B9.5.2. – Mathematics

B9.5.6. – Data science

B9.10. – Privacy

1 Team members, visitors, external collaborators

Research Scientists

- Remi Gribonval [Team leader, INRIA, Senior Researcher, HDR]
- Paulo Goncalves [INRIA, Senior Researcher, HDR]
- Mathurin Massias [INRIA, Researcher]
- Titouan Vayer [INRIA, Researcher, from Sep 2022]

Faculty Members

- Marion Foare [CPE LYON, Associate Professor]
- Elisa Riccietti [ENS DE LYON, Associate Professor, Chaire Inria]

Post-Doctoral Fellows

- Ayoub Belhadji [ENS DE LYON]
- Luc Giffon [ENS DE LYON, until Jun 2022]
- Rémi Vaudaine [ENS DE LYON]

PhD Students

- Antoine Gonon [ENS DE LYON]
- Clément Lalanne [ENS DE LYON]
- Guillaume Lauga [INRIA]
- Quoc-Tung Le [ENS DE LYON]
- Can Pouliquen [ENS DE LYON, from Nov 2022]
- Léon Zheng [ENS DE LYON]

Technical Staff

- Badr Moufad [ENS DE LYON, Engineer, from Apr 2022]

Administrative Assistant

- Solene Audoux [INRIA]

Visiting Scientist

- Sibylle Marcotte [ENS RENNES, from May 2022 until May 2022]

External Collaborator

- Márton Karsai [UNIV CEU, HDR]

2 Overall objectives

Building on a culture at the interface of signal modeling, mathematical optimization and statistical machine learning, the global objective of DANTE (and of its follow-up team Ockham which formal creation was process was formally started in 2022) 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 [51] 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 [52, 45]. The flexibility of the models, which can incorporate learned dictionaries [63], as well as structured and/or low-rank variants of the now-classical sparse modeling paradigm [57], 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 [43, 48], 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 [46], 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, 8] 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 [53] 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. [58]) 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 [69], 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 [55] and [47] 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) [61] [9], which underwent methodological developments at a very rapid pace in recent years, is precisely an approach to jointly exploit algebraically these structures and attributes, either by filtering them, by re-organizing them, or by reducing them to principal components. However, as it tends to be more and more the case, data collection processes yield very large data sets with high dimensional graphs. In contrast to standard digital signal processing that relies on regular graph structures (cycle graph or cartesian grid) treating complex structured data in a global form is not an easily scalable task [54]. Hence, the notion of distributed GSP [49, 50] 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 [59]) 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 [55] as well as privacy preservation guarantees [47]. A sketch can be seen as an embedding of the empirical probability distribution of the dataset with a particular form of kernel mean embedding [62]. 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 [67, 66], 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 [60, 68].

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

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

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

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

4 Application domains

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

4.1 Frugal AI on embedded devices

There is a strong need to drastically compress signal processing and machine learning models (typically, but not only, deep neural networks) to fit them on embedded devices. For example, on autonomous vehicles, due to strong constraints (reliability, energy consumption, production costs), the memory and computing resources of dedicated high-end image-analysis hardware are two orders of magnitude more limited than what is typically required to run state-of-the-art deep network models in real-time. The research conducted in the 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 past work [42], 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 using convex regularization in a journal paper published this year [14].

6 Highlights of the year

- Elisa Riccietti was a keynote speaker at the workshop Low-Rank Models and Applications (LRMA 2022) in Mons, Belgium.
- Three papers from the team were accepted at NeurIPS 2022 [11], [18] and [22]

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 [36, 37, 38, 39, 40, 41]. 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 2022, major efforts were put to optimize code efficiency (in particular for so-called butterly structures), and an anaconda package was made available. New Faust implementations of toeplitz, circulant, dct and dst matrices and more were made available.

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

Contact: Remi Gribonval

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

7.1.2 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.

News of the Year: In 2022 we added a fast solver based on coordinate descent for the Elastic Net problem.

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

Publications: hal-02263500, hal-01833398

Contact: Mathurin Massias

Participants: Badr Moufad, Alexandre Gramfort

7.1.3 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.

News of the Year: 2022: first release

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

Publication: hal-03819082

Contact: Mathurin Massias

Participants: Mathurin Massias, Badr Moufad

7.1.4 Benchopt

Keywords: Mathematical Optimization, Benchmarking, Reproducibility

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

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

Release Contributions: https://github.com/benchopt/benchopt/releases/tag/1.3.0

Publication: hal-03830604

Contact: Thomas Moreau

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

8 New results

8.1 Integrating Structured Models in Machine Learning and Signal Processing

8.1.1 Optimal Transport and Machine Learning on Graphs

Participants: Titouan Vayer.

Collaborations with 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. In a previous work [65] we proposed an efficient graph dictionary learning algorithm based on GW that allows to describe graphs as a simple composition of smaller graphs (atoms of the dictionary) and we showed that these representations are particularly efficient for tasks such as change detection for structured data and clustering of graphs. In [24] we proposed an alternative approach whose 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. This approach has the merit of being much more efficient to compute and more interpretable. We also validate this method for supervised learning tasks such as classification of multiple graphs [28].

In another line of works [25], we build upon the flexibility of the optimal transport framework and GW distance to define a novel graph neural network (GNN) architecture for graphs classification. More precisely, we propose a novel graph representation as GW distances to some learnable graph templates. We postulate that the vector of GW distances to a set of template graphs has a strong discriminative power, which is then fed to a non-linear classifier for final predictions. Distance embedding can be seen as a new layer, and can leverage on existing message passing techniques to promote sensible feature representations (and are learnt in an end-to-end fashion by differentiating through this layer). We empirically validate our claim on several synthetic and real life graph classification datasets, where our method is competitive or surpasses kernel and GNN state-of-the-art approaches.

8.1.2 Diffused Wasserstein Distance for Optimal Transport between Attributed Graphs

Participants: Paulo Gonçalves, Rémi Gribonval, 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).

Within the Ph.D work of A. Barbe (2018-2021), we introduced the Diffusion Wasserstein distance, a generalization of the standard Wasserstein to undirected and connected graphs where nodes are described by feature vectors. The last advance on this subject was to reduce the computational cost of the diffusion Wasserstein distance, by proposing 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 allowed to significantly improve estimates of the polynomial order for a prescribed error. This work led to a joint publication [21].

8.1.3 Structured Time Series Modeling

Participants: Titouan Vayer.

Collaborations with Romain Tavenard (IRISA, Université de Rennes 2), Laetitia Chapel (IRISA, Université Bretagne Sud), Rémi Flamary (CMAP, Ecole Polytechnique) and Nicolas Courty (IRISA, Université Bretagne Sud).

Multivariate time series are ubiquitous objects in signal processing, yet defining a distance or similarity between two such objects can be very difficult as soon as the temporal dynamics and the representation of the time series, i.e. the nature of the observed quantities, differ from one another. In the article [16], we propose a novel distance accounting for both feature space and temporal variabilities by learning a latent global transformation of the feature space together with a temporal alignment, cast as a joint optimization problem. The versatility of our framework allows for several variants depending on the structure of the time series at stake. Among other contributions, we define a differentiable loss for time series and present two algorithms for the computation of time series barycenters under this new geometry. We illustrate the interest of our approach on both simulated and real world data and show the robustness of our approach compared to state-of-the-art methods.

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, Antoine Gonon, Elisa Riccietti, Mathurin Massias.

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)

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 throught 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 the knowledge of its realization on some appropriate bounded domain. These results have been published this year [15].

Motivated by the importance of quantizing networks besides pruning them to achieve sparsity, we studied the expressivity of *quantized* deep networks from an approximation theoretic perspective [30]. Our objective is to define and compare the corresponding approximation classes [7] with the unquantized ones. We also characterize the error of nearest-neighbour uniform quantization of ReLU networks and we investigate when ReLU networks can be expected, or not, to have better approximation properties than other classical approximation families.

Another 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 lead 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 [34].

8.2.2 Algorithms for quantized networks

Participants: Rémi Gribonval, Elisa Riccietti.

Collaboration with Facebook AI Research Paris, Silviu Filip (IRISA, Rennes), Theo Mary (LIP6, Paris)

From a more computational perspective, we pursued the study of efficient optimization algorithms to solve problems involving quantized networks.

As a first step towards a better understanding of nonlinear quantized networks, we started from the linear case and investigated the problem of optimally quantizing low rank matrices. We showed that exploiting scaling invariances inherent to the optimization problem, much more accurate quantizations can be obtained than by a simple round to nearest strategy. We proposed an optimal solution algorithm with polynomial complexity in the dimension of the problem and exponential complexity in the number of bits.

Within the framework of the Ph.D. of Paul Estano, we studied the design of gradient-based training methods for neural networks, capable of exploiting multiple quantization levels. The proposed methods are supported by an error analysis, which suggests a good rule to switch among the available quantization levels, yielding a procedure that provides the same accuracy of classical training strategies but with a lower energy consumption.

8.2.3 Deep sparse factorizations: hardness, algorithms and identifiability

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

Collaboration with Valeo AI, Paris; Valérie Castin (M1 internship with DANTE)

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.

From a theoretical perspective, we pursued the study started last year on the hardness and uniqueness properties of sparse matrix factorization. Three papers have been published on this subject. First, in [13] we show that, even with only two factors and a fixed, known support, optimizing the coefficients of the sparse factors can be an NP-hard problem. Second, we study the landscape of the corresponding optimization problem and exhibite "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. Then, in [17] we investigate the essential uniqueness of sparse matrix factorizations in a multi-layer setting [17]. More details on the case with two factors can be found in the technical report [70]of last year. Third, in [20] we combine 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. A first version of the corresponding algorithm was incorporated in the FAμST software library (see Section 7) and is subject to software optimizations to further speed it up.

Finally, we investigated extensions of these results in several directions. To improve the flexibility of the algorithm of [20] for butterfly factorization, we adapted it to so-called *deformable butterlies* and studied its performance guarantees beyond the case of matrices admitting an exact factorization. To embrace deep ReLU neural networks with sparsity constraints, we showed that the identifiability results of [17, 20] [70]can be leveraged to identify (up to natural scaling ambiguities) the parameters of such networks with a prescribed butterfly structure. Finally, we investigated the closedness properties of the set of realizations of networks with prescribed support. These result are the objects of articles in preparation.

8.3 Statistical learning, dimension reduction, and privacy preservation

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

Participants: Rémi Gribonval, Titouan Vayer, Ayoub Belhadji.

Collaboration with Gilles Blanchard (Univ. Paris-Saclay)

The compressive learning framework proposes to deal with the large scale of datasets by compressing them into a single vector of generalized random moments, called a *sketch*, from which the learning task is then performed. In past works we established statistical guarantees on the generalization error of this procedure, first in a general abstract setting illustrated on PCA [5], then for the specific case of compressive k-means and compressive Gaussian Mixture Modeling [56]. A tutorial paper on the principle and the main guarantees of compressive learning was also finalized and published last year [6].

Theoretical guarantees in compressive learning fundamentally rely on comparing certain metrics between probability distributions. 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. The preprint submitted last year [64] is under revision.

Dimension reduction in compressive learning also exploits the ability to approximate certain kernels by finite dimensional quadratures. We revisited existing proofs of the Restricted Isometry Property of sketching operators with respect to certain mixtures models. We proposed an alternative analysis that circumvents the need to assume importance sampling when drawing random Fourier features to build random sketching operators. Our analysis is based on new deterministic bounds on the restricted isometry constant that depend solely on the set of frequencies used to define the sketching operator. Our analysis opens the door to theoretical guarantees for structured sketching with frequencies associated to fast random linear operators [29]. An other related approach that we investigated consists in exploiting Determinantal Point Processes (DPPs) to obtain quadrature rules for kernels in reproducing kernel Hilbert spaces [26].

8.3.2 Practical exploration of sketching and methods with limited resources

Participants: Rémi Gribonval, Titouan Vayer, Luc Giffon, Léon Zheng, Elisa Riccietti, Rémi Vaudaine.

Collaborations with Valeo AI; LightOn SAS; Hughes Van Assel (UMPA, ENS de Lyon); Marton Karsai (CEU, Vienne, Austria)

From a more empirical perspective, we pursued our efforts to make sketching for compressive learning and sketching 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 [27], and investigating the ability to exploit sketching in large-scale deep self-supervised learning scenarios [35].

Sketching was explored for temporal network compression. In the context of temporal networks, which can model spreading processes such as epidemics, the out-component of a source node is the set of nodes reachable from this node, and the distribution of the size of out-components is an important characteristics which computation can be demanding for large networks. We proposed both an exact online matrix algorithm with controlled complexity footprint to compute this distriution, and a sketching-based framework to estimate it from a highly compressed representation of the temporal network.

Moreover, making the connection between graph learning and sketching methods, we 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. In particular, we showed that it was possible to estimate such matrices with limited memory from a sketch based on Gaussian quadratic measurements. We pursed the practical applications of such result with structured rank-one measurements.

More generally, properties of kernels methods were also exploited in a more applicative context to reduce time and memory complexity: self-supervised learning of image representations. We introduced a regularization loss based on kernel mean embeddings with rotation-invariant kernels on the hypersphere, promoting the embedding distribution to be close to the uniform distribution on the hypersphere, with respect to the maximum mean discrepancy pseudometric [35]. Besides being fully competitive with the state of the art, our method significantly reduces the resources needed for training, making it implementable for very large embedding dimensions on existing devices and more easily adjustable than previous methods to settings with limited resources.

Finally, in collaboration with Hugues Van Assel (PhD student, UMPA), we proposed and investigated a novel dimension reduction method by leveraging the optimal transport framework and entropic affinities. Our work generalizes popular approaches such as t-distributed stochastic neighbor embedding (t-SNE) and has empirical benefits.

8.3.3 Privacy preservation

Participants: Rémi Gribonval, Clément Lalanne.

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

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

We addressed the problem of differentially private estimation of multiple quantiles (MQ) of a dataset [32], 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.

We studied minimax lower bounds when the class of estimators is restricted to the differentially private ones [31]. In particular, we showed that characterizing the power of a distributional test under differential privacy can be done by solving a transport problem. With specific coupling constructions, this observation allowed us to derivate Le Cam-type and Fano-type inequalities for both regular definitions of differential privacy and for divergence-based ones (based on Renyi divergence). We illustrated our results on three simple, fully worked out examples. For some problems, we showed that privacy leads to a provable degradation only when the rate of the privacy parameters is small enough whereas for other problems, the degradation systematically occurs under much looser hypotheses on the privacy parameters. Finally, we showed the near minimax optimality of the known guarantees for DP-SGLD, a private convex solver for maximum likelihood estimation on log-concave models.

8.4 Large-scale convex and nonconvex optimization

8.4.1 Multilevel schemes for image restoration

Participants: Elisa Riccietti, Paulo Gonçalves, Guillaume Lauga.

Collaboration with Nelly Pustelnik (CNRS, ENS de Lyon)

In the context of the Ph.D. work of Guillaume Lauga, we pursued the work started last year on the study of the combination of proximal methods and multiresolution analysis in large-scale image denoising problems. In the spirit of multilevel gradient methods [3] we developed a family of multilevel inertial proximal methods, tailored for problems arising in imaging, which exploit wavelets-based transfer operators. Our methods are capable of handling also problems in which the proximal operators cannot be computed explicitly. Their ability to accelerate proximal algorithms was shown in several large dimensional problems [19, 33].

8.4.2 Training of physics informed neural networks

Participants: Elisa Riccietti.

Collaboration with Serge Gratton, Valentin Mercier (IRIT, Toulouse), Stefania Bellavia (UNIFI, Italy), Mattéo Clémot (PLR internship with DANTE)

Physics informed neural networks (PINNs) are special network architectures designed for the solution of partial differential equations. We studied two aspects related to the training of these networks. On the one hand, in the context of the Ph.D. work of Valentin Mercier, we studied the integration of a multigrid approach in the training to improve the approximation of solutions with multiple frequency components. On the other hand, in the context of the internship of Mattéo Clémot, we investigated the ability of PINNs to solve ill-posed parameter identification inverse problems and the use of regularising training procedures to correctly fit noisy data in such a context.

8.4.3 Reproducible benchmarking of optimization algorithms

Participants: Mathurin Massias, Badr Moufad.

Collaborations with Thomas Moreau (MIND, Inria Saclay), Alexandre Gramfort (MIND, Inria Saclay).

To improve numerical reproducibility of optimisation benchmarks, we proposed Benchopt [22], a collaborative framework to automate, reproduce and publish optimization benchmarks in machine learning across programming languages and hardware architectures. This alleviates the burden of having many methods to reimplement, non-published code, and diverging stances on best practices. Benchopt (see also Section 7.1.4) simplifies benchmarking for the community by providing an off-the-shelf tool for running, sharing and extending experiments.

To demonstrate the benefits of using Benchopt, we showcased benchmarks on close to twenty standard machine learning tasks, such as ResNet18 training for image classification. We published open source implementations of state-of-the-art solvers on those problems, and a detailed comparison of the regimes in which they succeed and fail respectively.

8.4.4 Algorithms for large scale sparse linear models

Participants: Mathurin Massias, Badr Moufad.

Collaboration with Quentin Bertrand (MILA, Montréal), Quentin Klopfenstein (UMB, Dijon), Pierre-Antoine Bannier and Gauthier Gidel (MILA, Montréal), Samuel Vaiter (UMB, Dijon), Alexandre Gramfort (MIND, Inria Saclay), Joseph Salmon (IMAG, Montpellier).

In a series of works, we developed new fast algorithms that allow solving optimization problems with millions of variables in the context of sparse linear models. In [18] we proposed a new working set algorithm tailored to non convex sparse penalties such as the ℓ_q quasinorms ($0 < q < 1$). It exploits the support identification properties of coordinate descent to obtain faster convergence rates using Anderson acceleration [44] This algorithm was implemented in a high level python package, `skglm` (see Section 7.1.3), that was integrated into the ecosystem of the `scikit-learn` package. In [12], we proposed an efficient meta algorithm building on the previous work to automatically tune the regularization strength of sparse convex models such as the Lasso as sparse Logistic regression.

Finally, we provided dimension independent bound for stochastic gradient descent in a non convex setting, using Langevin dynamics in infinite dimension in [23].

9 Bilateral contracts and grants with industry

9.1 Bilateral grants with industry

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

  Participants: Rémi Gribonval, Elisa Riccietti, Léon Zheng.

  Duration: 3 years (2021-2024)

  Partners: Valeo AI, Paris; ENS de Lyon

  Funding: Valeo AI, Paris; ANRT

  Context: Chaire IA AllegroAssai 10.1.1

  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 Ricietti, Marion Foare, Mathurin Massias, Léon Zheng, Quoc-Tung Le, Antoine Gonon, Titouan Vayer, Ayoub Belhadji, Luc Giffon, Clement Lalanne, Can Pouliquen.

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

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

10.1.2 ANR DataRedux

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

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

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

10.1.3 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 ANR JCJC MASSILIA

Participants: Titouan Vayer.

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

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

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

10.1.5 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, Mathurin Massias, Elisa Riccietti, Titouan Vayer.

11.1 Promoting scientific activities

11.1.1 Scientific events: organisation

Member of the organizing committees

- Rémi Gribonval, Journées de Statistiques 2022, Lyon
- Mathurin Massias, Elisa Riccietti, Rémi Gribonval, Journées SMAI-MODE 2024, Lyon
- Mathurin Massias, Learning and Optimization in Luminy 2024, CIRM, Marseille
- Titouan Vayer, Graph Learning & Learning with Graphs, special session at GRETSI conference 2023.

11.1.2 Scientific events: selection

Member of the conference program committees

- Rémi Gribonval, GRETSI 2022, GRETSI 2023
- 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)

11.1.3 Journal

Member of the editorial boards

- Rémi Gribonval: Associate Editor for Constructive Approximation (Springer), Senior Area Editor for the IEEE Signal Processing Magazine
- Mathurin Massias: Associate editor for Computo (French Statistical Society)

11.1.4 Invited talks

- Titouan Vayer was an invited speaker at the 2nd Inria-DFKI European Summer School on AI (IDESSAI 2022) and at the conference Machine Learning and Signal Processing on Graphs (CIRM 2022).
- Mathurin Massias was an invited speaker at the Learning and Optimization in Luminy 2022 conference (CIRM) and an invited lecturer at the Computation and Modelling (Wrocław University of Science and Technology, Poland)
- Elisa Riccietti was an invited speaker at the workshop Low-Rank Models and Applications (LRMA 2022) in Mons, Belgium.

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

11.1.7 Research administration

- Paulo Gonçalves is Deputy Scientific Director of the new research center of Inria in Lyon and member of the Inria Evaluation Committee since sept. 2022.

11.2 Teaching - Supervision - Juries

11.2.1 Teaching

- Master :
  - Rémi Gribonval: Inverse problems and high dimension; Mathematical foundations of deep neural networks; Concentration of measure in probability and high-dimensional statistical learning; M2, ENS Lyon
  - Mathurin Massias: Large scale optimization for Machine Learning; M2, ENS Lyon
  - Mathurin Massias: Python for Datascience; M1, Ecole Polytechnique/HEC
- 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 courses on Optimization and Approximation and Fundamentals of Machine Learning. 19h of tutor responsibility at ENS Lyon
- Other formations: "Fondements et pratique du machine learning et du deep learning", CNRS formation, 5 days (20h) with Rémi Gribonval, Mathurin Massias and Titouan Vayer.

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

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

No PhD was defended in DANTE in 2022

11.2.3 Juries

Members of the DANTE team participated to the following juries:

- PhD juries: Antoine Mazarguil (Université Paris Cité, member); El-Mehdi Achour (Université de Toulouse, member); Nina Göttschling (University of Cambridge, external reviewer); Paul Viallard (Université de Lyon, chair); Pierre-Hugo Vial (Université de Toulouse, chair); Ruben Ohana (Sorbonne Université, reviewer); Thomas Debarre (EPFL, external examiner); Gilles Bareilles (Université Grenoble-Alpes, member); Florent Bascou (Université de Montpellier, member); Florian Mouret (Université de Toulouse, member); Mikhail Kamalov (Université Côte d'Azur, member)
- Habilitation juries: Antoine Liutkus (Université de Montpellier, reviewer); Vincent Duval (Université Paris-Dauphine PSL, 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. 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'. In: *Journal of Machine Learning Research* 23.149 (Apr. 2022), pp. 1–43. 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. 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. URL: https://hal.archives-ouvertes.fr/hal-01782346.

[5] 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 dimensionMain 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.

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

[7] R. Gribonval, G. Kutyniok, M. Nielsen and F. Voigtlaender. 'Approximation spaces of deep neural networks'. In: *Constructive Approximation*. special issue on "Deep Networks in Approximation Theory" (2020). URL: https://hal.inria.fr/hal-02117139.

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

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

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

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

12.2 Publications of the year

International journals

[12] Q. Bertrand, Q. Klopfenstein, M. Massias, M. Blondel, S. Vaiter, A. Gramfort and J. Salmon. 'Implicit differentiation for fast hyperparameter selection in non-smooth convex learning'. In: *Journal of Machine Learning Research* 23.149 (Apr. 2022), pp. 1–43. URL: https://hal.archives-ouvertes.fr/hal-03228663.

[13] 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.archives-ouvertes.fr/hal-03364668.

[14] 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'. In: *IEEE Transactions on Signal Processing* 70 (19th June 2022), pp. 2859–2868. URL: https://hal.archives-ouvertes.fr/hal-03348154.

[15] P. Stock and R. Gribonval. 'An Embedding of ReLU Networks and an Analysis of their Identifiability'. In: *Constructive Approximation* (2022). DOI: 10.1007/s00365-022-09578-1. URL: https://hal.archives-ouvertes.fr/hal-03292203.

[16] T. Vayer, R. Tavenard, L. Chapel, N. Courty, R. Flamary and Y. Soullard. 'Time Series Alignment with Global Invariances'. In: *Transactions on Machine Learning Research* (Oct. 2022). URL: https://hal.archives-ouvertes.fr/hal-02473959.

[17] 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://hal.inria.fr/hal-03362626.

International peer-reviewed conferences

[18] Q. Bertrand, Q. Klopfenstein, P.-A. Bannier, G. Gidel and M. Massias. 'Beyond L1: Faster and Better Sparse Models with skglm'. In: NeurIPS. New Orleans, United States, 27th Nov. 2022. URL: https://hal.archives-ouvertes.fr/hal-03819082.

[19] 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, 6th Sept. 2022. URL: https://hal.inria.fr/hal-03739496.

[20] 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. DOI: 10.1109/ICASSP43922.2022.9747791. URL: https://hal.inria.fr/hal-03438881.

[21] S. Marcotte, A. Barbe, R. Gribonval, T. Vayer, M. Sebban, P. Borgnat and P. Gonçalves. 'Fast Multiscale Diffusion on Graphs'. In: ICASSP 2022 - IEEE International Conference on Acoustics, Speech and Signal Processing. Singapore, Singapore, 22nd May 2022. DOI: 10.1109/ICASSP43922.2022.9746802. URL: https://hal.archives-ouvertes.fr/hal-03212764.

[22] T. Moreau, M. Massias, A. Gramfort, P. Ablin, P.-A. Bannier, B. Charlier, M. Dagréou, T. Dupré La Tour, G. Durif, C. F. Dantas, Q. Klopfenstein, J. Larsson, E. Lai, T. Lefort, B. Malézieux, 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.archives-ouvertes.fr/hal-03830604.

[23] B. Muzellec, K. Sato, M. Massias and T. Suzuki. 'Dimension-free convergence rates for gradient Langevin dynamics in RKHS'. In: COLT 2022 - 35th Annual Conference on Learning Theory. London, United Kingdom, 2nd July 2022. URL: https://hal.archives-ouvertes.fr/hal-03920387.

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

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

National peer-reviewed Conferences

[26] A. Belhadji. 'Des règles de quadrature dans les RKHSs à base de DPPs'. In: GRETSI 2022 - XXVIIIème Colloque Francophone de Traitement du Signal et des Images. Nancy, France, 5th Sept. 2022, pp. 1–4. URL: https://hal.science/hal-03630086.

[27] L. Giffon and R. Gribonval. 'Compressive Clustering with an Optical Processing Unit'. In: GRETSI 2022 - XXVIIIème Colloque Francophone de Traitement du Signal et des Images. Nancy, France, 6th Sept. 2022. URL: https://hal.inria.fr/hal-03690865.

[28] C. Vincent-Cuaz, R. Flamary, M. Corneli, T. Vayer and N. Courty. 'Semi-relaxed Gromov-Wasserstein divergence for graphs classification'. In: Colloque GRETSI 2022 - XXVIIIème Colloque Francophone de Traitement du Signal et des Images. Nancy, France, 6th Sept. 2022. URL: https://hal.science/hal-03839524.

Reports & preprints

[29] A. Belhadji and R. Gribonval. *Revisiting RIP guarantees for sketching operators on mixture models*. 29th Nov. 2022. URL: https://hal.science/hal-03872878.

[30] A. Gonon, N. Brisebarre, R. Gribonval and E. Riccietti. *Approximation speed of quantized vs. unquantized ReLU neural networks and beyond*. 19th May 2022. URL: https://hal.archives-ouvertes.fr/hal-03672166.

[31] C. Lalanne, A. Garivier and R. Gribonval. *On the Statistical Complexity of Estimation and Testing under Privacy Constraints*. 3rd Oct. 2022. URL: https://hal.archives-ouvertes.fr/hal-03794374.

[32] C. S. Lalanne, C. Gastaud, N. Grislain, A. Garivier and R. Gribonval. *Private Quantiles Estimation in the Presence of Atoms*. 14th Feb. 2022. URL: https://hal.archives-ouvertes.fr/hal-03572701.

[33] G. Lauga, E. Riccietti, N. Pustelnik and P. Gonçalves. *Multilevel FISTA for image restoration*. 27th Oct. 2022. URL: https://hal.archives-ouvertes.fr/hal-03831180.

[34] Y. Traonmilin, R. Gribonval and S. Vaiter. *A theory of optimal convex regularization for low-dimensional recovery*. 12th Dec. 2022. URL: https://hal.archives-ouvertes.fr/hal-03467123.

[35] L. Zheng, G. Puy, E. Riccietti, P. Pérez and R. Gribonval. *Self-supervised learning with rotation-invariant kernels*. 11th Oct. 2022. URL: https://hal.archives-ouvertes.fr/hal-03738466.

12.3 Other

Softwares

[36] [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⟩, URL: https://hal.inria.fr/hal-03572265, VCS: https://github.com/sieunhanbom04/support-exploration-matrix-factorization/tree/master, SWHID: ⟨swh:1:dir:6794fd3f41d50a4e51e6e7f11ee721efafda9806;origin=https://hal.archives-ouvertes.fr/hal-03572265;visit=swh:1:snp:e5760cb4d925b5f3529486ffde997d8e3e9ea07e;anchor=swh:1:rel:a190fb7ab69d170eede1666e9e3edec33c56224b;path=/⟩.

[37] [SW] Q.-T. Le, E. Riccietti and R. Gribonval, *Code for reproducible research - "Spurious Valleys, NP-hardness, and Tractability of Sparse Matrix Factorization With Fixed Support"*, 12th May 2022. LIC: BSD 2-Clause FreeBSD License. HAL: ⟨hal-03667186⟩, URL: https://hal.inria.fr/hal-03667186, SWHID: ⟨swh:1:dir:96dde633bde9eb800bc2e298ae98dba0579cb052;origin=https://hal.archives-ouvertes.fr/hal-03667186;visit=swh:1:snp:5d6cba12bff4367e05a63c41938941f627ed2946;anchor=swh:1:rel:76d45ab397d6879990a7aa51621191589440f04b;path=/⟩.

[38] [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⟩, URL: https://hal.inria.fr/hal-03576498, VCS: https://github.com/sibyllema/Fast-Multiscale-Diffusion-on-Graphs.

[39] [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⟩, URL: https://hal.inria.fr/hal-03552956, VCS: https://github.com/leonzheng2/butterfly, SWHID: ⟨swh:1:dir:42d42bf905c9f3bbcd6cf5e1e0cda6be2d0d63de;origin=https://hal.archives-ouvertes.fr/hal-03552956;visit=swh:1:snp:6a2256426a0f3a7491086fbf2da554d7589d1e18;anchor=swh:1:rel:f477382c82ada89647864605ea44b12383c04167;path=/⟩.

[40] [SW] L. Zheng, G. Puy, E. Riccietti, P. Pérez and R. Gribonval, *Code for reproducible research - Self-supervised learning with rotation-invariant kernels*, 25th July 2022. LIC: Apache License 2.0. HAL: ⟨hal-03737572⟩, URL: https://hal.inria.fr/hal-03737572, VCS: https://github.com/valeoai/sfrik, SWHID: ⟨swh:1:dir:0324646066b76298702273fed6baa6a64cfb5b01;origin=https://hal.archives-ouvertes.fr/hal-03737572;visit=swh:1:snp:c7c35f9273f8f23fd018d4dd6b0aaef7e600b0dc;anchor=swh:1:rel:7b216caea6e30f84fae2f12af6235dcbfbd1a85d;path=/⟩.

[41] [SW] L. Zheng, E. Riccietti and R. Gribonval, *Code for reproducible research - Efficient Identification of Butterfly Sparse Matrix Factorizations*, 30th Mar. 2022. LIC: BSD 3-Clause License. HAL: ⟨hal-03620052⟩, URL: https://hal.inria.fr/hal-03620052, VCS: https://github.com/leonzheng2/efficient-butterfly, SWHID: ⟨swh:1:dir:ed569ed027c484a53ca2d3d32aa6a05c79b50dbd;origin=https://hal.archives-ouvertes.fr/hal-03620052;visit=swh:1:snp:0706d7527a325d7ac8f46744c182056af46b4756;anchor=swh:1:rel:3884bd8154b9879c5743dfd2279d55102b60358c;path=/⟩.

12.4 Cited publications

[42] 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. URL: https://hal.inria.fr/hal-02921836.

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

[44] Q. Bertrand and M. Massias. 'Anderson acceleration of coordinate descent'. In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2021, pp. 1288–1296.

[45] 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. URL: http://books.google.cz/books?id=6KoYCgAAQBAJ&pg=PA340&dq=intitle:Compressed+Sensing+and+its+Applications&hl=&cd=1&source=gbs_api.

[46] 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.

[47] 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.

[48] 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.

[49] 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).

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

[51] 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.

[52] 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.

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

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

[55] 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).

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

[57] 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.

[58] 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.

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

[60] 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.

[61] 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.

[62] 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.

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

[64] T. Vayer and R. Gribonval. 'Controlling Wasserstein distances by Kernel norms with application to Compressive Statistical Learning'. working paper or preprint. Dec. 2021. URL: https://hal.archives-ouvertes.fr/hal-03461492.

[65] 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, July 2021. URL: https://hal.archives-ouvertes.fr/hal-03140349.

[66] 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.

[67] 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.

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

[69] 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.

[70] L. Zheng, E. Riccietti and R. Gribonval. 'Identifiability in Two-Layer Sparse Matrix Factorization'. working paper or preprint. Nov. 2021. URL: https://hal.inria.fr/hal-03362875.