

# **Journée Statistique et Informatique pour la Science des Données à Paris Saclay**

## **Rapport sur les contributions**

ID de Contribution: 1

Type: **Non spécifié**

# Computational Reproducibility in the Life Sciences and Research in Computer Science: Round Trip

*lundi 27 janvier 2020 10:00 (50 minutes)*

With the development of new experimental technologies, biologists are faced with an avalanche of data to be computationally analyzed for scientific advancements and discoveries to emerge. Faced with the complexity of analysis pipelines, the large number of computational tools, and the enormous amount of data to manage, there is compelling evidence that many (if not most) scientific discoveries will not stand the test of time. Increasing the reproducibility of computed results is of paramount importance.

While several elements of solutions are currently available, ensuring reproducible analyses relies on progress made in several areas of research in computer science including fundamental aspects.

After an introduction to the problem of computational reproducibility, we go on to discuss the challenges posed by this domain and describe the remaining opportunities of research in computer science.

**Orateur:** Prof. COHEN-BOULAKIA, Sarah (LRI, Paris-Sud)

ID de Contribution: 2

Type: **Non spécifié**

## Determinantal Point Processes in Machine Learning

*lundi 27 janvier 2020 11:20 (50 minutes)*

Determinantal point processes are a very powerful tool in probability theory, especially for integrable systems, because they allow to get very concise closed form formulas and simplify a lot of computations. This is one reason why they have become very attractive in machine learning. Another reason is that, when parametrized by a symmetric matrix, they allow to model repulsive interactions between finitely many items; They were even introduced as fermionic point processes by Odile Macchi in statistical physics in the 70's, in order to describe particles that tend to repel each other within same energy states. In this talk, I will define these point processes, give a few examples and properties, and list a few challenges that they pose in machine learning theory.

**Orateur:** Prof. BRUNEL, Victor-Emmanuel (ENSEA/CREST)

ID de Contribution: 3

Type: **Non spécifié**

## Orthogonal Greedy Algorithms for Sparse Reconstruction

*lundi 27 janvier 2020 14:20 (50 minutes)*

The past decade has witnessed a tremendous interest in the concept of sparse representations in signal and image processing. Inverse problems involving sparsity arise in many application fields such as nondestructive evaluation of materials, electroencephalography for brain activity analysis, biological imaging, or fluid mechanics, to name a few. In this lecture, I will introduce well-known greedy algorithms and show how they can be used to address ill-posed inverse problems regularized by sparsity. Orthogonal greedy algorithms are popular iterative schemes for sparse signal reconstruction. Their principle is to sequentially select atoms in a given dictionary and to update the sparse approximation coefficients by solving a least-square problem whenever a new atom is selected. Two classical greedy algorithms will be put forward, Orthogonal Matching Pursuit (OMP) and Orthogonal Least Squares (OLS). Their popularity relies on the fact that fast solvers are available, since least-square problems can be solved recursively. I will then introduce stepwise extension of greedy algorithms, where an early wrong atom selection can be counteracted by its further removal from the active set. I will also address non-negative extension of greedy algorithms for inverse problems regularized by both sparsity and non-negativity. In the latter algorithms, a series of non-negative least-squares subproblems are solved. I will then discuss how orthogonal greedy schemes can be adapted, and show that fast implementations are still possible, based on more involved strategies for recursively solving non-negative least-square problems. The last part of my talk will be dedicated to the theoretical analysis of greedy algorithms, aiming to characterize the performance of sparse algorithms in terms of exact recovery guarantees. I will give a flavor of the main concepts behind classical exact recovery analysis techniques.

**Orateur:** Prof. SOUSSEN, Charles (CentraleSupélec)

ID de Contribution: 4

Type: **Non spécifié**

## The Pre-image Problem from a Topological Perspective

*lundi 27 janvier 2020 12:10 (50 minutes)*

This talk will be a review of the efforts of the Topological Data Analysis (TDA) community to tackle the preimage problem. After a general introduction on TDA, the main focus will be on recent attempts to invert the TDA operator. While this line of work is still in its infancy, the hope on the long run is to use such inverses for feature interpretation. The mathematical tools involved in the analysis come mainly from metric geometry, spectral theory, and the theory of constructible functions—specific pointers will be given in the course of the exposition.

**Orateur:** Prof. OUDOT, Steve (INRIA)

ID de Contribution: 5

Type: **Non spécifié**

## Simultaneous Adaptation for Several Criteria Using an Extended Lepskii Principle

*lundi 27 janvier 2020 15:30 (50 minutes)*

In the setting of supervised learning using kernel methods, while the least-square (prediction) error is classically the performance measure of interest, if the true target function is assumed to be an element of a Hilbert space, one can also be interested in the norm of the error of an estimator in that space (reconstruction error); this is of particular relevance in inverse problems where the observed signal is the target after passing through a known linear operator. When the regularity (in a certain sense) of the target is known, a common regularization parameter can achieve optimal minimax error rates in both norms. When the regularity is unknown (which is usually the case), we address the question of data-dependent selection rule of a regularization parameter that is adaptive to the unknown regularity of the target function and is optimal both for the prediction error and for the reproducing kernel Hilbert space (reconstruction) norm error by proposing a modified Lepskii balancing principle using a varying family of norms. (Based on joint work with P. Mathé, N. Mücke).

**Orateur:** Prof. BLANCHARD, Gilles (IHES)

ID de Contribution: 6

Type: **Non spécifié**

## Quantitative Stability of Optimal Transport Maps and Linearization of the 2-Wasserstein Space

*lundi 27 janvier 2020 16:20 (50 minutes)*

This work studies an explicit embedding of the set of probability measures into a Hilbert space, defined using optimal transport maps from a reference probability density. This embedding linearizes to some extent the 2-Wasserstein space, and enables the direct use of generic supervised and unsupervised learning algorithms on measure data. Our main result is that the embedding is (bi-)Hölder continuous, when the reference density is uniform over a convex set, and can be equivalently phrased as a dimension-independent Hölder-stability results for optimal transport maps. Joint work with A. Delalande and F. Chazal.

**Orateur:** Prof. MERIGOT, Quentin (Paris-Sud)