

CRIStAL, UMR CNRS 9189

Centre de recherche en informatique, signal et automatique

École centrale de Lille

Cité scientifique, CS 20048

59651 Villeneuve d'Ascq Cedex, France

Pierre Chainais, MCF HDR École Centrale, Équipe SigMA

pierre.chainais@ec-lille.fr

Rémi Bardenet, chercheur CR2 CNRS, Équipe SigMA

remi.bardenet@ec-lille.fr

February 3, 2017,

Objet: Description of M2 internship / PhD project *Adaptive subspace discovery using determinantal point processes for signal processing*

Efficient signal processing often relies on efficient representations before processing. Various approaches dimension reduction, such as principal component analysis, matrix factorization, dictionary learning (Tosic and Frossard [2011]), etc. A host of methods have in common the search for a family of vectors (or functions) that are relevant to linearly approximate the data subspace. This dimension reduction step plays a crucial role when the original data are high-dimensional, helping fight the curse of dimensionality or giving access to massive datasets. This PhD project aims at **turning determinantal point processes into probabilistic approaches to dimensionality reduction in signal processing** applications such as face recognition and linear unmixing.

Determinantal point processes (DPPs, Lavancier et al. [2015]) on \mathbb{R}^d are stochastic point processes with repulsive interactions between points. In other words, a DPP describes a random set of points in \mathbb{R}^d that tend to appear far from each other in each realization. The interaction between these points is encoded by a positive definite kernel function, which also determines most interesting statistical quantities on the DPP. For this reason, DPPs can legitimately be considered the “kernel machine of point processes”. In particular, random sets of points drawn from a DPP correspond to almost orthogonal vectors in a higher-dimensional so-called *feature space*. This should naturally lead to efficient probabilistic strategies to build subspaces for dimension reduction. In this research project, we investigate two directions that follow up on this intuition.

First, we will study to what extent DPPs can provide a set of adaptive random projection subspaces. Training a regressor or a classifier on each subspace would then yield an ensemble of experts, where each expert sees data under a different angle. The machinery of DPPs provides interesting knobs that the user can tweak to design the random subspaces and combine the outputs of different experts, such as the choice of kernel and closed-form integration formulas.

Second, we will go one step further and use DPPs as the main ingredient of probabilistic generative models for specific classification or inverse signal processing problems, such as linear unmixing. Signals are often assumed to live in some low-dimensional subspace of a functional space. DPPs can model this very naturally, and the idea would be to incorporate adaptive dimensionality reduction by learning a generative model built around a DPP. The random number of points would correspond to the a priori unknown dimension of the subspace where the signals live, in the spirit of a Bayesian non parametric approach [Gershman and Blei, 2012].

Université de Lille is a very favourable setting for this thesis on cutting-edge probabilistic tools in machine learning and signal processing. The campus of Université de Lille has recently seen the merging

of several labs into CRIStAL [\[link\]](#), which, together with closely located Inria Lille and Laboratoire Painlevé, make Lille a stimulating place for computational statisticians and applied probabilists. Cross-disciplinary workgroups such as our “**determinantal point processes**” [workgroup](#) [\[link\]](#), also contribute to making Université de Lille the right place for this cross-disciplinary project, at the interface of data science, signal processing, statistics, and probability.

This Master internship project may be prolonged by a PhD.

References

- S.J. Gershman and D.M. Blei. A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, 56:1–12, 2012.
- F. Lavancier, J. Møller, and E. Rubak. Determinantal point process models and statistical inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(4):853–877, 2015. ISSN 1467-9868. doi: 10.1111/rssb.12096. URL <http://dx.doi.org/10.1111/rssb.12096>.
- I. Tosić and P. Frossard. Dictionary learning. *IEEE Signal Processing Magazine*, 28(2):27–38, march 2011. ISSN 1053-5888. doi: 10.1109/MSP.2010.939537.