

# Deux points de vue sur l'ACP robuste

Projet M2 MIGS – 2016-17

Ce projet porte sur la comparaison, à l'aide d'exemples, de deux méthodes visant à rendre robuste l'analyse en composantes principales. L'analyse en composantes principales, qui est une méthode statistique incontournable en réduction de la dimension, repose sur la diagonalisation de la matrice de variance covariance des observations (ou encore la décomposition en valeurs singulières de la matrice des données). Elle est donc naturellement peu robuste, c'est-à-dire très sensible aux valeurs atypiques.

Plusieurs pistes ont été envisagées pour "robustifier" cette technique statistique et l'objectif de ce projet sera de comprendre et de comparer (en termes de robustesse, de précision, de rapidité, ...) deux de ces méthodes.

- Candès, E. J., Li, X., Ma, Y., & Wright, J. (2011). Robust principal component analysis? *Journal of the ACM*, 58(3), 11
- Cardot, H. & Godichon-Baggioni, A. (2015). Fast Estimation of the Median Covariation Matrix with Application to Online Robust Principal Components Analysis. [arXiv:1504.02852](https://arxiv.org/abs/1504.02852)

Des exemples basés sur données réelles (des images par exemple) et des données simulées seront utilisés pour comparer les deux techniques. Les étudiants pourront utiliser les packages `rpca` et `Gmedian` qui sont disponibles sous le langage  $\mathbb{R}$ .

Projet encadré par H. Cardot, idéalement pour un binôme (efficace), éventuellement pour un étudiant seul mais très motivé.