

Suggested papers for 2016-2017

I work in statistical problems that usually appear in the analysis of proteomics and genomics data. For example, in these applications, some variables may be measured with error, the number of features is much larger than the number of observations, the observations are not always independent, outliers are present in the data, among others. Some extensions of classical estimation approaches have been proposed to address some of these characteristics and others remain open problems for future research.

I suggest the following papers but you are also welcome to discuss with me other options of your interest. As a general goal, I expect the student to understand the theoretical contribution of the papers, to know the related literature, to reproduce and/or complete some of the theoretical results, to explain its relevance, and to test the proposed methodology in simulated and/or real data. Once the student selects one of these papers, I will give more specific details in an individual meeting with the student.

List of papers:

- 1) Haipeng Shen, Jianhua Z Huang (2008) Sparse principal component analysis via regularized low rank matrix approximation. *Journal of Multivariate Analysis* **99**, 1015-1034.
- 2) Christophe Croux, Peter Filzmoser, and Heinrich Fritz (2013) Robust Sparse Principal Component Analysis. *Technometrics* **55**, 202-214.
- 3) Wei Lin , Rui Feng , Hongzhe Li (2015) Regularization Methods for High-Dimensional Instrumental Variables Regression With an Application to Genetical Genomics. *Journal of the American Statistical Association* **110**, 270-288.
- 4) Tarr G, Muller S, Weber NC (2016) Robust estimation of precision matrices under cellwise contamination, *Computational Statistics and Data Analysis* **93**, 404-420