Karim Lounici
Professor of Statistics
CMAP, Ecole Polytechnique, France
karim.lounici@polytechnique.edu

November 3rd, 2020

Review, in the perspective of the PhD thesis defence of Arshak Minasyan, of the thesis entitled: "Robust Estimation of Gaussian Mean within the Domain of Computational Tractability"

This thesis addresses the central issue of robustness in statistics. It tackles new important and difficult problems that appeared with the development of high-dimensional statistics where even "small" (random or malicious) imperfections in the data (called outliers) can lead to completely erroneous analysis if not taken into account. Indeed, high-dimensionality tends to make imperfections more frequent and also to amplify their impact on the behavior of classical estimators. There exists an important literature on robustness in the standard low-dimensional setting but these estimators are most of the time not useful in the high-dimensional setting as they are not computationally tractable.

The main results in this thesis concern the robust estimation of the mean of high-dimensional vectors in the presence of outliers and develops a robust procedure that is both computationally tractable and statistically optimal. These are highly important contributions in a field of statistics that will continue to take more an more importance in the future.

Structure of the thesis.

The thesis is extremely well-written and well-organized.

Chapter 1 gives a clear and complete introduction to the field of robust statistics. It describes the form of the desired results starting with the 1-dimensional case and then highlights the additional challenges of the high-dimensional case. An extensive review of the literature is also provided.

Chapter 2 provides the mathematical formalism required to tackle robustness problems in statistics. This includes a presentation of the main possible contamination scenarios and the relationship between them, as well as the definition of the minimax criterion in the context of robustness that will be used to assess the statistical optimality of the new procedures.

These first two chapters provide an excellent introduction for anyone interested in learning about robustness in statistics.

Chapter 3 is the core of this thesis. Arshak proposed an iterative reweighting (IR) scheme for the robust estimation of the mean and proved several strong properties of this procedure. Notably, among the main results, the following are particularly significant to me: (1) this procedure is computationally tractable via semi-definite programming; (2) they determined the breakdown point and the nearly-minimax optimal rate breakdown point of this procedure; (3) He also proved that this procedure is minimax optimal up to a logarithmic factor. The proofs of these results were rigorous and require the mastery of several fields of mathematics: concentration of random matrices, minimax theory in statistics and optimization.

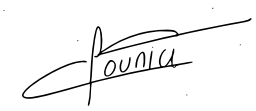
Note that the results were stated assuming that the outlier proportion ϵ and the noise covariance are known. But Arshak suggested that a version of the Lepski procedure could be used to handle unknown ϵ. Regarding the covariance operator, I believe that recent result on robust covariance matrix estimation could be exploited in relation to the IR scheme of this thesis. For these reasons, I believe that this procedure can have a real impact in practice.

Chapter 4 explores several algorithms to tackle important robust statistical learning problems including sparse estimation and Principal Component Analysis. Several of these methods are based on iterative reweighting schemes somehow related to the one studied in Chapter 3. The performances of these methods were investigated through numerical experiments. The theoretical analysis of these methods could be challenging but definitely interesting.

Overall assessment.

Arshak Minasyan already published 3 papers and have an additional preprint related to chapter 3 that I believe could be published in a top journal in statistics. This is an

excellent record for a finishing PhD student. In addition, this thesis suggests several futures directions of research that are relevant and exciting. I hope that Arshak will continue to work on these problems as any future progress on robust learning can potentially impact the way we use statistics in virtually everything fields of sciences that are plagued with corrupted data. For all these reasons, I find the thesis to be of **very high scientific quality and definitely worthy of defence**.



Karim Lounici