Joseph Salmon, Professeur
IMAG, Université de Montpellier
CC51, place Eugène Bataillon
34095 Montpellier cedex 5
http://josephsalmon.eu/
joseph.salmon@umontpellier.fr

Montpellier, June 12 2023

Report on Tomasz SKALSKI doctoral dissertation.

This dissertation addresses several important aspects of modern statistics. The references are up to date and the writing style is very clear and concise. Overall the quality of the dissertation is excellent.

Chapters 1 and 2 of the thesis introduce the context, background and useful notation for convex analysis, statistics and probability. In particular, the central estimator of the dissertation is introduced, the Slope estimator, which is a convex optimization problem addressing sparse regression, and that can be written :

$$\hat{\beta}^{SLOPE} \in \underset{b \in \mathbb{R}^p}{\arg\min} \left[\frac{1}{2} \|Y - Xb\|^2 + \sum_{j=1}^{p} \lambda_j |\beta|_{(j)} \right] \quad ,$$

where the tuning parameters satisfy $\lambda_1 \geq \cdots \geq \lambda_p \geq 0$, and the parenthesis index notation refers to a reordering of the coefficients in decreasing magnitudes.

In Chapter 3, the author investigates the pattern recovery properties of the Slope estimator in the context of high-dimensional in the orthogonal design case (when $X^\top X = nI_p$) For the Lasso (ℓ_1) penalty, the pattern is classical and well-known (it is simply the support of the estimated coefficient), while for the Slope penalty, the pattern is more complex : it combines information on the sign, support and grouping patterns of the coefficients. The author provides a detailed analysis of the pattern recovery properties of the Slope estimator, and shows that it can recover the true pattern asymptotically (in the standard additive white Gaussian noise model), under some conditions on the tuning parameters : the largest parameter λ_1/n must converge to 0, while the difference between consecutive parameters $\lambda_j - \lambda_{j+1}$ must be lower bounded by $C \cdot \sqrt{n} \log(n)^{1/2}$. This result is very interesting and is a first step toward understanding the pattern recovery properties of the Slope estimator. An extension could be provided to account for non-asymptotic results, illustrating the impact of the noise level and of the true pattern structure itself for identification. In particular, a possible choice of the parameters is $\lambda_j = c \cdot (p + 1 - j)n^{2/3}$. Moreover, when $\lim_{n\to\infty} \lambda_1/n = \lambda_0 > 0$, the Slope estimator does not converge to the true parameter anymore.

Chapter 4 is to me the most interesting part of the dissertation, and presents results on support recovery, this time without orthogonal constraints on the design matrix. An extension of the Lasso irrepresentable condition is introduced to provide a necessary and sufficient condition for pattern recovery. This condition is expressed in terms of the Slope dual norm and the fact that the slope tuning pattern belongs to the span of the design matrix restricted on the columns associated with the true pattern. Conditions for asymptotic pattern recovery are also provided when p is fixed and n goes to infinity. Geometrical interpretation and visualization are also given to ease the understanding of the results. At the end of the chapter, a simulation study is provided illustrating the recovery results, choosing this time the tuning parameters as functions of the re-ordered statistics of i.i.d standard Gaussian variables. In Chapter 5, pattern recovery is investigated for general penalties, as long as they are polyhedral gauges, included when a thresholding step is added. The latter is known to be beneficial for the Lasso, and similar results are extended for the case of gauges.

The last two chapters focus on different topics, and are less related to the previous ones, falling far from my expertise. Chapter 6 is concerned with exponential families over finite sets. Likelihood and log-likelihood are investigated in such a context, showing that the Maximum likelihood does not necessarily exist. This can be tested by solving a linear program, and the author provides a detailed analysis of the problem. Applications to random graph models and Rademacher functions are also provided.

The last chapter focuses on graphical models. More precisely, the author relates the zeros of the precision matrix of the features X, to the zeros of the Laplacian matrix of an underlying graph, and instantiates its results to several classes of popular graphs.

To my knowledge, the thesis makes original and significant contributions to the above research problems. The writing style is great, and the quality of the provided illustrations is remarkable. Moreover, the student has already published his results in two statistics and probability journals. In summary, I am confident to say that this thesis contains original and significant results which constitute a substantial contribution to the knowledge of high-dimensional statistics. Therefore, I have no doubts to say that the student should be awarded a Ph.D. degree.

Sincerely yours,
Joseph Salmon