ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Department of Mathematics
Seminar for Statistics

ETH Zurich
Prof. Dr. Sara van de Geer
HG G 15.1
Rämistrasse 101
8092 Zurich, Switzerland

Phone +41 44 632 22 52
geer@stat.math.ethz.ch

Zurich, May 15, 2023

Report on the Doctoral Dissertation by Tomasz Skalski

Geometric and Combinatorial Aspects of Statistical Models

This thesis clearly shows the candidate mathematical insights and broadness. It is novel and timely work, and the candidate's contributions are highly relevant in modern statistics and geometry.

After a nice introduction highlighting the main contributions of the thesis, and an exposition of the tools and notions needed, the work presents in Chapter 3 the results for the SLOPE estimator, primarily for the case of orthogonal design. The estimator is characterized as a saddle point, a so-called min-max solution. The candidate presents a geometric view of the dual unit ball of the SLOPE norm. The dual ball plays an important role in the theoretical analysis of SLOPE. For the case of i.i.d. Gaussian noise, the candidate shows strong consistency under certain conditions and also cases where consistency does not hold. The results are established by a careful comparison with the OLS estimator. Very importantly, the candidate also derives pattern recovery of SLOPE. Chapter 4 goes beyond orthogonal design and in fact shows necessary and sufficient conditions for pattern recovery, thus establishing superiority in this respect of the SLOPE as compared to the LASSO. For the noiseless case it gives a SLOPE-version of the irrepresentability condition, and then asympotics for the noisy case and various designs. The candidate gives original and very helpful geometric interpretations.

Chapter 5 goes beyond the special case of a SLOPE penalty by studying polyhedral norm penalties. It starts out with properties of normal cones, and the relation with sub-differentials and pattern recovery. The candidate derives the irrepresentability condition for polyhedral norm penalties in the noiseless case, which shows that pattern recovery for such penalized estimators can only happen under rather strong conditions. However, the candidate shows that the thresholded estimator can recover the pattern under the much weaker accessibility condition. Here, uniform uniqueness is assumed, and characterized.

Chapter 6 is about a different theme: existence of maximum likelihood estimators (MLE's) in discrete exponential families. This is an important problem: one may think for example of logistic regression where the MLE does not exist if the data are separable. It is shown that existence can be checked by solving a linear programming problem, thus relating the question to (convex) geometry. Moreover, in the i.i.d. case, the existence is shown to be guaranteed when having a large enough sample. As an important special case, the candidate considers random graphs. For example, it is shown that the MLE exists if the number of samples is at least $\log N$, where N is the number of nodes. The results are also applied to exponential families with Walsh functions, where the Ising model is a special case.

The last chapter of the thesis examines the Laplacian of graphs. It shows that the augmented graph Laplacian is the inverse of the corresponding covariance matrix of a random vector. Moreover. when there is only one root, the determinant is the number of spanning trees. This has important consequences for the Gaussian case.

The candidate carried out several interesting simulation studies to illustrate the theoretical results.

The candidate published his work in

T. Skalski, P. Graczyk, B. Kołodziejek and M. Wilczyński: Pattern recovery and signal denoising by slope when the design is orthogonal. *Probability and Mathematical Statistics* 42 (2022)

K. Bogdan, M. Bosy and T. Skalski: Maximum likelihood estimation for discrete exponential families and random graphs. *ALEA* (2022)

T. Skalski: Remarks on Laplacian of graphical models in various graphs. *Geometric Science of Information*, LNCS (2021)

There are moreover two preprints:

M. Bogdan, P. Graczyk, B. Kołodziejek, T. Skalski, P. Tardivel and M. Wilczyński: Pattern recovery by SLOPE, *ArXiv* 2203.12086 (2022)

P. Graczyk, U. Schneider, T. Skalski andP. Tardivel: Pattern recovery in penalized and thresholded estimation and its geometry. *HAL* hal–03262087 (2021)

I strongly support the acceptance of this thesis and as Doctoral Dissertation at Wrocław University of Science and Technology and Université D'Angers.

Yours sincerely,

Prof. S. A. van de Geer