prof. Jan Mielniczuk
Institute of Computer Science, PAS
Faculty of Mathematics and Information Sciences, WUT

30.VI.2023

Review of Ph.D. thesis by Tomasz Skalski, M.Sc., entitled
Geometric and combinatorial aspects of statistical models

The following is a report on Ph.D. thesis by Tomasz Skalski, M.Sc., prepared at Wrocław University of Science and Technology and Université d'Angers and supervised jointly by Maciej Wilczyński and Piotr Graczyk, with Patrick Tardivel acting as a French co-advisor. The report has been prepared at the request of WUST.

Thematic content of the thesis. The main subjects of the dissertation are properties of Sorted ℓ_1 Penalised Estimator (SLOPE; Bogdan et al (2015), Zeng and Figuerido (2014)) considered in chapters 3 and 4 and its generalisations, when SLOPE penalty is replaced by polyhedral gauge penalty (chapter 5). Also, the results on existence of MLE estimators in exponential families are discussed (chapter 6) as well as interconnections between statistical graphical models with corresponding graph Laplacians and discretised Wiener processes (chapter 7).

The researched problems in chapters 3-5 revolve around characterisation of the solution of the SLOPE optimisation task for data following a linear regression model, pattern recovery of true parameter β by $\hat{\beta}_{SLOPE}$ (understood as recovery of the ranking of absolute values of coordinates of β, their ties and their signs) and asymptotic properties of $\hat{\beta}_{SLOPE}$ as well as analogous properties when SLOPE penalty is is replaced by a more general polyhedral gauge penalty. Finally, the thresholded versions of these estimators are investigated.

More specifically, in chapter 3 the dual problem for SLOPE optimisation is considered and the solution is found to be a saddle point of the corresponding objective function when the experimental matrix is of the full rank. Asymptotic strong consistency is investigated in terms of conditions placed on the largest weight penalty $\lambda_1^{(n)}$ and strong pattern consistency is studied in the orthogonal case. The main results of chapter 4 are algebraic conditions which are equivalent conditions for $\hat{\beta}_{SLOPE}$ and β having the same patterns: $patt(\hat{\beta}_{SLOPE}) = patt(\beta)$, its simplification for the noiseless case (irrepresentability condition) and refined results (in comparison with results of chapter 3) on asymptotic consistency and pattern discovery. Chapter 5 studies algebraic properties of penalized estimators using polyhedral gauge type penalty, in particular, accessibility property (meaning that the pattern of β is recovered for at least one value of response vector y) and conditions under which the solution of the optimisation problem is unique for all y. Also, sufficient conditions for pattern recovery by the thresholded estimator are proved.

Chapter 5 deals with existence of ML estimators in exponential families supported on finite sets, with the main results being the equivalent conditions for existence in the general case (Theorem 6.2.2) and for graph-valued iid samples (Theorem 6.4.1) with corollaries concerning in particular probability of existence of MLE. The last section concerns calculation of augmented Laplacian in some specific graph model, including trees.

The results of the dissertation are contained in 3 published papers (two papers in journals PMS and ALEA and one in the conference proceedings in LNCS series) and two preprints. All but one paper (in conference proceedings) are co-authored.

Evaluation of the thesis. General comments

The thesis contains a multitude of results concerning pattern recovery of SLOPE and its generalizations, as well as some results on uniqueness of MLEs for exponential families supported on finite sets and some properties of augmented Laplacians for certain graphs. I found the methodology introduced in Chapter 4 based on aggregated experimental matrices and weight vectors quite useful as it makes the treatment of the first problem in the case when the matrix $X'X$ is not invertible, not only feasible but effective. This is clearly shown by the proof of Proposition 4.2.1 which gives a clear description of subdifferential of SLOPE penalty $J_\Lambda(b)$. Especially insightful are Theorems 3.3.2 , 4.3.1, 4.6.1 , Corollary 4.3.1 and results on uniqueness of MLEs (Theorems 6.2.2 and 6.4.1). In particular Theorem 4.3.1 gives nice novel equivalent set of conditions for pattern discovery for possibly non-invertible $X'X$ matrix. *In general, I consider results of investigation of algebraic aspects of SLOPE solutions and its generalisations, the most interesting part of the thesis.*

The facts included in the dissertation are the results proven by several people: besides T. Skalski himself, his supervisors, B. Kołodziejek, K. Bogdan and M. Bosy. I obtained from T. Skalski detailed description of his input (official documents were not sufficiently specific in this respect), which I consider satisfactory. However, although it is understandable that some parts done by others have to be included in the dissertation to make it self-contained, I still fail to see the reason for including several parts of the thesis, such as e.g. section 4.3.2 on asymptotic properties, which are proved entirely by coauthors of T. Skalski, and can be removed from the thesis without any loss of coherence.

Moreover, there are several aspects of the dissertation, its layout as well as its contents, which are cause of additional concern to me and which I would like the author to address during his defence.

1. The thesis, disregarding the preliminaries, splits into three subject-wise disjoint parts (chapters 3-5), chapter 6 and chapter 7. It would be much better if one of the three subjects (presumably the first one) would have been selected and the author had concentrated solely on it. Putting all the presented subjects together results in cramming all author's results into Ph.D. thesis without much thought about its integrity. I agree the the title *Geometric and Combinatorial Aspects of Statistical Models* encompasses all thesis' themes, but it is also so broad that it borders on complete non-specificity;

2. The dissertation is based on five papers (3 published and 2 yet unpublished) put together and it shows. Despite the efforts of the author the exposition is uneven and sometimes repetitive. The main problem is lack of interconnections between the chapters, in particular between chapters 3,4 and 5. E.g. Theorem 4.3.1 has obvious connections with results of section 3.2.3, in particular equality (3.2.3), but those are neither compared or discussed. Proposition 4.4.2 reappears in a more general form as Proposition 5.4.2 with no comment;

3. Despite its considerable volume (146 pages) there are some aspects which are barely discussed. In particular, it at times not clear, which results are novel, and which are already known but proven differently. This concerns in particular Theorem 3.3.2 when compared with e.g. with Lemma 3 in Zeng and Figuerido (2014), results of section 4 compared with results on estimators using submodular penalties (cf e.g. Lemma 3 in Munami (2020)) and relation of Corollary 3.2.2 to results of Bellec et al (2017). It would have been beneficial if author had put his approach into some perspective as e.g. proximal projections on the dual ball of the norm $J_\Lambda(\cdot)$ seems well established method of studying SLOPE properties;

4. Pattern discovery (in particular ranking of variables based on absolute values of coefficients) is especially meaningful in the case of qualitative predictors and several popular methods have been investigated for this case such as e.g. DMLR (EJS (2015)) or method of Stokell et al (2021). Although they are referenced in the dissertation, no attempt to compare the performance of SLOPE with them is made.

5. Numerical experiments given in the chapters 3-5 should be regarded as a mere illustrations of the theoretical results and one can not draw solid conclusions how the pattern recovery by $\hat{\beta}_{SLOPE}$ works in practice. In particular, this concerns results of section 3.5, which although promising, are too limited for this purpose. Stating MSE results for *one* sample path (Table 3.1) could be misleading and should be strongly discouraged.

Specific comments

1. The end of the proof of Theorem 4.5.1 (i) requires more care as the parameter r satisfying the second to last displayed equation in the proof may depend on ω. (this has been satisfactorily justified by the author in an e-mail exchange);
2. Section 4.4.2: $patt(\beta) = M$ is needed for $\beta - \hat{\beta} \in col(U_M)$;
3. p. 77: accessibility is needed for sign recovery under tacit assumption that the vector of signs of β is estimated by the signs of $\hat{\beta}$;
4. Some proofs or justifications are omitted, e.g. p. 98, l. 4 and convexity of ι_β, remark 4.5.2 (b);
5. Section 3.5: Dimension of X matrix (300 rows, 100 predictors) do not match dimensions of the vectors which follow. Figure 4.11: 6. Comment is needed here that convergence in Theorem 4.5.3 is slow and probability of pattern recovery for samples sizes around 1000 does not exceed 0.2;
7. Remark 3.4.2 calls for precise description of those βs such that $\lambda_0||\beta||_\infty \leqslant \beta' C \beta$ for which $\hat{\beta}_{SLOPE} \rightarrow \beta$ a.s.;
8. Section 3.5: the reason why regularisation λ for Lasso differs significantly from λ_{CV} ($\lambda = 5\lambda_{CV}$) is left without any comment;
9. Theorem 4.5.1: the set-up of this result is rather uncommon and I can't think of any practical situation in which it may be applicable;
10. There are some typos (relatively few, but some important: p. 37: $||bS||_1 = k$ should be $||\mathbf{S}||_1 = k$, Corollary 3.2.2: $\cdot\hat{\beta}_{OLS}$ should read $n \times \hat{\beta}_{OLS}$ in the middle formula of displayed equation; p. 34: Theorem 3.2.1 does not exist, p. 65: Chapter ??);

some lapses: p. 37: 'true values of β' should be 'true values of coordinates of β'; p. 43, bottom: I guess that what is meant here is not a correlation between columns but correlation between coordinates of a row variable; p. 47: ε_A and ε_B do not have null covariance matrices, but their cross-covariance is null;
and some adventurous use of English: p.25:'descriptible', p. 69: 'which achieves the proof', p. 49: 'condition does not occur'.

Conclusion

It is my view that much more substantial and coherent thesis could have been prepared focusing solely on the novel results on SLOPE estimator proved by the author and on in-depth comparison with alternative proposals for variable clustering in the linear, and especially ANOVA, case.
Nevertheless, in my opinion the thesis 'Geometric and combinatorial aspects of statistical models' by Tomasz Skalski, M.Sc., fulfills the formal Polish requirements concerning Ph.D. theses in mathematics and I will vote in favour of him defending it.

Jan Mielniczuk