



## Discussion on "Minimal penalties and the slope heuristic: a survey" by Sylvain Arlot

Emilie Devijver

### ► To cite this version:

Emilie Devijver. Discussion on "Minimal penalties and the slope heuristic: a survey" by Sylvain Arlot. Journal de la Societe Française de Statistique, 2019, 160 (3), pp.119-120. hal-03094408

HAL Id: hal-03094408

<https://hal.science/hal-03094408>

Submitted on 13 Apr 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Discussion on "Minimal penalties and the slope heuristic: a survey" by Sylvain Arlot

**Titre:** Discussion sur "Pénalités minimales et heuristique de pente" par Sylvain Arlot

Émilie Devijver<sup>1</sup>

In this elegant and complete paper, S. Arlot describes the minimal penalties and the slope heuristics from a theoretical and a practical viewpoint. This paper is very pedagogical, and gives a deep insight in the model selection paradigm through minimal penalties.

From a practical viewpoint, several ingredients are described, in order to illustrate what is done.

First of all, several definitions of the constant that should be put in front of the penalty term are introduced in the literature, but those may give different results. The survey provides an intense comparison between the several criteria. One should notice that, as highlighted, if the same model is selected by the different criteria, it leads to some trust in the model selection procedure; whereas different models would lead to consider another method to select a model. The slope heuristic is known to adapt itself to real dataset, in the sense that if the true model does not belong to the collection, the method will select a competitive model (whereas BIC, for example, will select a very complex model to overfit the data), and this adaptation may be measured by the difference between the several criteria.

However, a special attention must be paid when selecting a model from the union of subcollections having different approximations properties. The vizualisation of two criteria (the slope or the dimension jump) should be used to ensure the quality of model selection and to check the different heuristics. For example, if a model is defined among the collection by two attributes as, e.g., number of clusters and sparsity of the mean (for some examples, we refer to [Devijver \(2017\)](#) for mixture of Gaussian regressions for unsupervised learning and low rank estimation for the mean, or to [Devijver and Gallopin \(2018\)](#) for mixture of Gaussian regressions with sparse covariance matrix), the optimization problem is difficult to solve and the graphical criteria are not satisfied. Several solutions may be proposed: to consider the dimension with respect to each attribute, whereas considering a global dimension of the model which summarizes the several dimensions (and then consider the plane instead of the slope for complex models); or to express one attribute as a function of the others, and to select iteratively the value of the attributes. The second option, called nested minimal-penalty algorithm, has been preferred when there is an order of importance between the two attributes (the cluster structure has to be defined beforehand

<sup>1</sup> Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG

in the two examples). However, this method is empirical, and a theoretical optimization analysis would reinforce this method.

Finally, when the theoretical derivation of the optimal penalty involves several terms, and then several unknown constants, some solutions are also proposed. Numerical comparison between simplified penalty shape and calibration of two (or more) constants may advice for the need (or not) of all the terms. However, it should be noticed that it depends on the context, and a mathematical study may answer to this question. For example, we consider the case developed in [Devijver and Gallopin \(2018\)](#). The penalty has two terms, one proportional to the dimension and the other to the dimension up to a logarithmic term. As an oracle inequality and a minimax lower bound have been derived, this penalty is known to have the good shape. The numerical comparison has been done, reflecting similar performances for the simplified shape and the calibration of two constants. However, the simulation setting is not in high-dimension (we consider 200 variables through a sample of size 70) so we know that the penalty is equivalent, in that case, to consider only the dimension (the logarithm term is negligible). If one would use the shock method for really high dimension, with thousands of vertices, the logarithm term would not be negligible and the selected model would be different. Our conclusion on this analysis would be to always check with respect to the theoretical penalty if a simplified version is coherent.

Thus, to summarize, model selection is still an active and promising research area, as illustrated by the explosion of machine learning methods. This survey remarkably exposes theoretical and experimental tools to consider minimal penalties, which we think will stimulate further research in a broad domain.

## References

- Devijver, E. (2017). Joint rank and variable selection for parsimonious estimation in a high-dimensional finite mixture regression model. *J. Multivar. Anal.*, 157(C):1–13.  
 Devijver, E. and Gallopin, M. (2018). Block-diagonal covariance selection for high-dimensional gaussian graphical models. *Journal of the American Statistical Association*, 113(521):306–314.