On the robustness of the Generalized-Fused Lasso to prior specifications

Vivian Viallon[1], Sophie Lambert-Lacroix[2], Hölger Hoefling[3], and Franck Picard[4]

[1] Université de Lyon, F-69622, Lyon, France; Université Lyon 1, UMRESTTE, F-69373 Lyon; IFSTTAR, UMRESTTE, F-69675 Bron.
[2] UMR 5525 UJF-Grenoble 1/CNRS/UPMF TIMC-IMAG, Grenoble, F-38041
[3] Novartis Pharma, Basel, Switzerland.
[4] Laboratoire de Biométrie et Biologie Évolutive, UMR CNRS 5558 Univ. Lyon 1, F-69622 Villeurbanne.

Abstract. Using networks as *prior* knowledge to guide model selection is a way to reach structured sparsity. In particular, the fused lasso that was originally designed to penalize differences of coefficients corresponding to successive features has been generalized to handle features whose effects are structured according to a given network. As any *prior* information, the network provided in the penalty may contain misleading edges that connect coefficients whose difference is not zero, and the extent to which the performance of the method depend on the suitability of the graph has never been clearly assessed. In this work we investigate the theoretical and empirical properties of the adaptive generalized fused lasso in the context of generalized linear models. In the fixed p setting, we show that, asymptotically, adding misleading edges in the graph does not prevent the adaptive generalized fused lasso from enjoying asymptotic oracle properties, while forgetting suitable edges can be more problematic. These theoretical results are complemented by an extensive simulation study that assesses the robustness of the adaptive generalized fused lasso against misspecification of the network as well as its applicability when theoretical coefficients are not exactly equal. Our contribution is also to evaluate the applicability of the generalized fused lasso for the joint modeling of multiple sparse regression functions. Illustrations are provided on two real data examples.

Keywords: lasso, generalized linear models, joint modeling, model selection