Referee's Report on
Sparsity by worst-case penalties

Submitted to Computational Statistics and Data Analysis

In this paper, the authors propose a reformulation of sparsity-inducing penalties used in variable selection problems, based on the dual form of the original optimization. This method allows to define a "generic penalty" as a series of linear or quadratic penalties, which are easy to handle. A general algorithm is provided, and its efficiency and accuracy are compared to state-of-art benchmarks in numerical experiments with simulated and real-world data.

In general, I consider the paper well organized. The methodological part is well written, and the results are clearly presented and illustrated. The geometrical interpretation of the sparse problems is particularly interesting. The comparison of different existing algorithms is complete and made in a fair way. Below there are some comments that may help the authors to improve the contribution of the paper.

- The algorithm starts from a sparse initial guess, i.e., the active set $\mathbf{A} = \emptyset$. I think it may be useful to discuss briefly how the computational time changes for an experimental setup, in which the choice of the initial guess is randomized.
- In section 5, I would suggest to switch the results on simulated settings and on real-world data, such that the latter can be commented in light of the properties and conclusions obtained with synthetic data.
- In the simulated data section, I would suggest to add the F-measure between the support of the true coefficient vector $\beta^\star$ and the estimated one $\hat{\beta}$, as a performance measure to evaluate the selection properties of the different algorithms (see Section IV in Gasso et al., 2009).
- In line 244, the reference should not be in parenthesis.
- Figure 6 could be improved, as the axes labels are not easily readable and colors are not distinguishable.
- In line 367, I would suggest to explain the reason why *quadrupen, SPAMS-LARS* and *lars* are not sensitive to the level of correlation between features.
- In the caption of Figure 8, specify the x-axis.

Gasso G., Rakotomamonjy A., Canu S. (2009). Recovering sparse signals with a certain family of non convex penalties and DC programming. *IEEE Transactions on Signal Processing*, 57(12), 4686-4698.