Report on CSDA-D-17-00692
"Sparsity by Worst-Case Penalties"

The authors proposed a new optimization algorithm to solve the sparse linear regression with a norm penalty term, including group-lasso and elastic-net. The proposed method is both accurate in estimation and extremely computationally fast, as shown in extensive simulation studies. It is very good that the authors also provided the accompanying R-package (**quadrupen**) on CRAN.

The paper is interesting in computational aspects and fit the theme of the journal. However, I think a few concerns should be addressed before publication is possible.

Major comments:

1. As the authors also admitted in the paper, the proposed algorithm is highly similar to the active-set methods, except the procedures taken in Step 3. I understand that the authors argue that **quadrupen** is more flexible in handling a wider range of penalties and look at the problem from a different perspective. However, either methodological innovations or numerical improvement should be proved to make the contribution enough to consider it as an "alternative" algorithm.

2. In the simulation studies, under the same tuning parameter, a more conservative selection of the **quadrupen** was observed in the QTL example comparing to that of **glmnet**, but not in the examples showed in Figure 8. I was wondering if this is an universal phenomenon? If it is the authors should provide some theoretical insights on why does this happen and maybe run some more monte carlo simulations to verify it, since basically they are solving the same underlying optimization problem.

3. For solving the elastic-net problem, the authors compared the proposed method to other two optimization strategies. It seems that the proposed method only gains obvious efficiency in the regime where λ_1 and λ_2 are very small. In the case if we choose parameters probably (according the theoretical rate or simply by cross-validation), what is the advantage of using **quadrupen**?

Minor comments:

1. The presentation of the figures in the paper is not very straightforward to readers. For example, I found Figure 1,2,3,6 kind of hard to understand by reading the captions. More basic introductions should be made in the captions or in the context describing the figures. Moreover, please pay attention to the size of fonts in the labs, titles and legends across the figures. Currently they are quite differing.

2. In the section 5.2.3 and 5.2.4, the authors benchmarked the methods by Lasso, letting $\lambda_2 = 0$. Is it possible to consider the more general elastic-net problem (say for a fixed λ_2) for at least some of the packages, especially in measuring the model selection accuracy? I understand that the authors have pointed out that they have different rules in determining tuning parameters.