LETTER TO THE EDITOR

By Yuao Hu, Ye Tian and Heng Lian

Nanyang Technological University

The paper by Alfons, Croux and Gelper (2013), *Sparse least trimmed squares regression for analyzing high-dimensional large data sets*, considered a combination of least trimmed squares (LTS) and lasso penalty for robust and sparse high-dimensional regression. In a recent paper [She and Owen (2011)], a method for outlier detection based on a sparsity penalty on the mean shift parameter was proposed (designated by "SO" in the following). This work is mentioned in Alfons et al. as being an "entirely different approach." Certainly the problem studied by Alfons et al. is novel and interesting. However, there is actually a connection between the LTS approach and that of She and Owen (2011). This connection can be roughly seen from Theorem 4.1 and Proposition 4.1 of She and Owen (2011), where iterative thresholding was related to penalized regression and also to the M-estimator. In particular, although not explicitly mentioned in She and Owen (2011), from this one can derive the close relationship between hard thresholding, L_0 penalty and LTS [the relationship between hard thresholding and L_0 penalty was mentioned on page 630 of She and Owen (2011)]. Given that LTS regression is not directly posed as an M-estimator, the following proposition can be directly shown via elementary arguments.

PROPOSITION 1. *Using the notation of Alfons et al., if $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$ is a minimizer of $\sum_{i=1}^n (y_i - \mathbf{x}_i'\boldsymbol{\beta} - \gamma_i)^2 + \lambda_1\|\boldsymbol{\beta}\|_1 + \lambda_2\|\boldsymbol{\gamma}\|_0$ and $\|\boldsymbol{\gamma}\|_0 = n - h$, then $\hat{\boldsymbol{\beta}}$ is the minimizer of $\sum_{i=1}^h (\mathbf{r}^2(\boldsymbol{\beta}))_{i:n} + \lambda_1\|\boldsymbol{\beta}\|_1$.*

PROOF. Obviously we have $\hat{\gamma}_i = y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}$ if $(y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}})^2 > \lambda_2$ and $\hat{\gamma}_i = 0$ if $(y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}})^2 < \lambda_2$. Thus, we can profile out $\boldsymbol{\gamma}$ and get exactly the LTS problem. □

The result above says that a solution of SO is a solution of some LTS problem and, thus, the set of solutions that can be obtained by SO (by varying λ_1 and λ_2) is a subset that can be obtained by LTS (by varying λ_1 and h). Obviously, if for any fixed λ_1 and $h \geq n/2$, we can make $\|\boldsymbol{\gamma}\|_0 = n - h$ by choosing an appropriate value for λ_2, then the two will be the same. Numerically, we do find occasionally some values of $n - h$ cannot be obtained by $\|\boldsymbol{\gamma}\|_0$. In the numerical example below with sample size $n = 59$, $h = 45$ (25% trimmed) can be achieved in both cases.
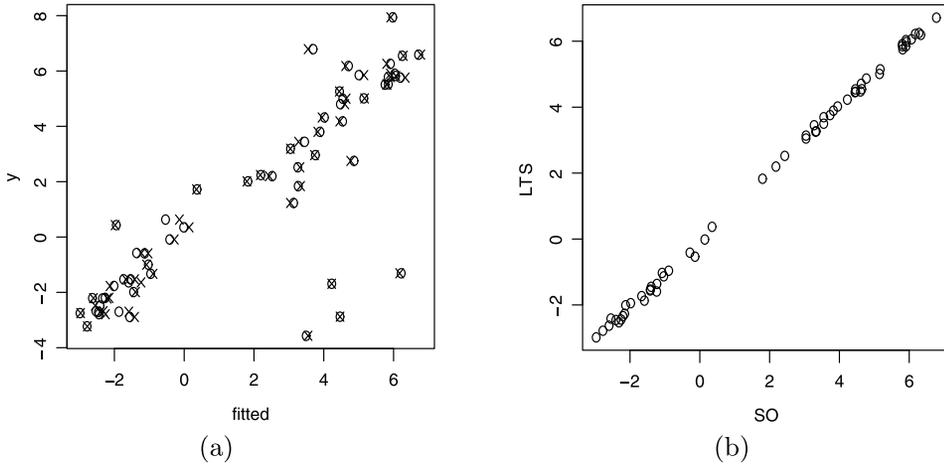
FIG. 1. *Results for the NCI data set with $n = 59$ and $p = 10$ of the 22,283 genes with the highest Spearman rank correlation coefficients.* (a) *Observed responses versus fitted responses, where results from LTS are denoted with the character "x" and results from "SO" points are denoted with the character "o."* (b) *Fitted responses from LTS versus those from SO.*

We use the same NCI-60 data to illustrate the similarities between the two approaches. Working with the whole data ($n = 59$, $p = 22{,}283$) using the R package *robustHD* on our desktop PC causes memory problems. Even with $p = 1000$ the program is quite slow (for both approaches). So we use only a small number of genes just to illustrate the similarities of the two approaches. We select p genes with the largest Spearman correlations with the response. We first use $p = 10$ and $\lambda_1 = 0$ to avoid the complications brought about by the lasso penalty. SO is implemented by initializing with $\gamma = 0$ and iteratively estimates β (by OLS) and γ (by hard thresholding). We use the default setting with $h = 45$ (25% trimmed). For SO, we set $\lambda_2 = 1.34$ which results in $\|\gamma\|_0 = 14$. The fitted response values of the two approaches are shown in Figure 1, demonstrating their similarity.

As a second illustration, we use $p = 500$ genes. We find that BIC values for the LTS approach decrease as $\lambda_1 \to 0$, possibly because we picked genes with the largest correlations with the response. So we just manually set the parameter for the lasso penalty to be 0.1 in the *sparseLTS*() function of the *robustHD* package. Based on equation (1.4) in Alfons, Croux and Gelper (2013), this actually should correspond to $\lambda_1 = h \times 0.1 = 4.5$. However, this value of λ_1 was too large for the SO implementation and resulted in $\beta = 0$. Thus, we perform a two-dimensional search to find the values of (λ_1, λ_2) that produce a similar solution (in particular, with the same number of outliers), and finally find $\lambda_1 = 0.26$, $\lambda_2 = 1.44$. The fitted response values for the two approaches are shown in Figure 2. There is a larger difference between the two approaches compared to Figure 1. The difference might be due to different initialization methods, numerical errors or convergence issues. We also note that the initialization method used for penalized LTS is random and
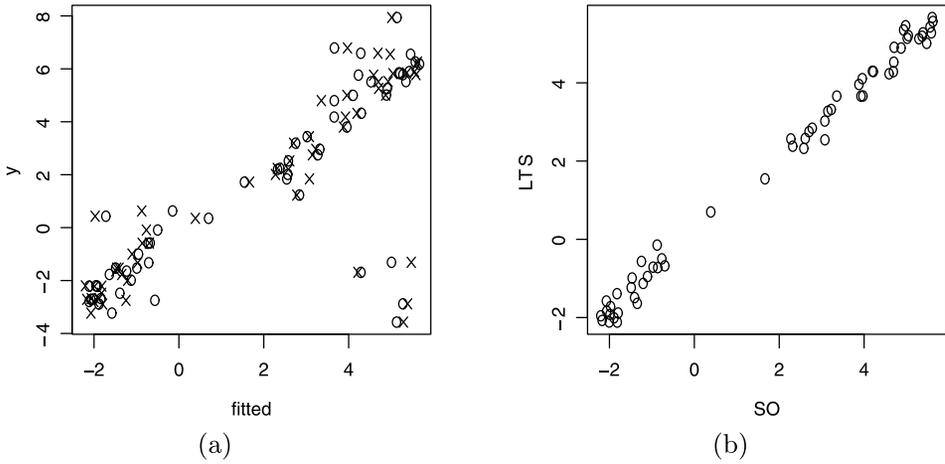
FIG. 2.   *Results for the NCI data set with n = 59 and p = 500 of the 22,283 genes with the highest Spearman rank correlation coefficients.* (a) *Observed responses versus fitted responses, where results from LTS are denoted with the character "*x*" and results from "SO" points are denoted with the character "*o.*" (b) *Fitted responses from LTS versus those from SO.*

multiple executions of the same function in *robustHD* will produce slightly different results.

REFERENCES

ALFONS, A., CROUX, C. and GELPER, S. (2013). Sparse least trimmed squares regression for analyzing high-dimensional large data sets. *Ann. Appl. Stat.* **7** 226–248.

SHE, Y. and OWEN, A. B. (2011). Outlier detection using nonconvex penalized regression. *J. Amer. Statist. Assoc.* **106** 626–639. MR2847975

DIVISION OF MATHEMATICAL SCIENCES
SCHOOL OF PHYSICAL AND MATHEMATICAL SCIENCES
NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE 637371
SINGAPORE
E-MAIL: henglian@ntu.edu.sg