Tuning Parameter Selection Consistency in an Ultrahigh Dimensional Setup: A Comment on the Sure Independence Screening Rule

Chenlei Leng and Hansheng Wang *

July 8, 2008

We congratulate Professors Fan and Lv for a thought-provoking paper, which provides us deep understanding about variable selection in an ultrahigh dimensional setup. We would like to supply our comments as follows.

The important work of Breiman (1996) and Tibshirani (1996) demonstrated clearly that shrinkage estimation is a promising solution for variable selection. The first paper on the asymptotic results of the lasso appeared in Knight and Fu (2000). However, the important question regarding whether those shrinkage methods are consistent in model selection (Shao, 1997) was not clear. In a seminal paper, Fan and Li (2001) developed SCAD and, more importantly, introduced a general theoretical framework to understand the asymptotic behavior of various shrinkage methods. As a consequence, Fan and Li (2001) is also partially responsible for the recent development of the adaptive LASSO methods (Zou, 2006; Wang *et al.*, 2007a; Zhang and Lu, 2007).

Note that the oracle properties defined in Fan and Li (2001) depend on an appropriate selection of tuning parameters, for which prediction based criteria such as GCV have been commonly used in practice. Nevertheless, Leng *et al.* (2006) and Wang *et al.* (2007b) showed that this practice leads to seriously overfitted model. For model selection consistency, a BIC-type criterion is a justifiable alternative. Results were established for SCAD (Wang *et al.*, 2007b) and adaptive LASSO (Wang and Leng, 2007) with a fixed dimension, and also for these two methods with diverging model dimensions (Wang *et al.*, 2008).

*Leng is with Department of Statistics and Applied Probability, National University of Singapore. Leng's research is supported in part by NUS research grants (Email: stalc@nus.edu.sg). Wang is with Guanghua School of Management, Peking University, China (Email: hansheng@gsm.pku.edu.cn). Wang's research is supported in part by a NSFC grant (#10771006).

It is very natural to ask whether similar results can be established in an ultrahigh dimensional setup. In particular, we are very interested in knowing:

(1) How can the parameter γ in the first stage of SIS be automatically tuned? The authors' numerical studies suggest that $[\kappa n/\log n]$ might be a good choice, with a reasonably range of κ (e.g., $\kappa = 1$, 2, etc). However, we still believe a completely data driven choice can make SIS more attractive for real practitioners.

(2) How about the stochastic error involved in SIS's first stage screening? Is it ignorable in its second stage shrinkage estimation? Are the BIC criteria developed in the exisiting literature still applicable? We believe that research along those directions will further enhance the applicability of SIS in an ultrahigh dimensional setting.

Lastly, we want to conclude this comment by congratulating the authors again for such a wonderful piece of work!

References

Breiman, L. (1996) Heuristics of instability and stabilization in model selection. *The Annals of Statistics*, **24**, 2350–2383.

Fan, J. and Li, R. (2001) Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, **96**, 1348–1360.

Fan, J. and Peng, H. (2004) On non-concave penalized likelihood with diverging number of parameters. *The Annals of Statistics*, **32**, 928–961.

Knight, K. and Fu, W. (2000) Asymptotics for lasso-type estimators. *The Annals of Statistics*, **28**, 1356–1378.

Leng, C., Lin, Y. and Wahba, G. (2006) A note on lasso and related procedures in model selection. *Statistica Sinica*, **16**, 1273–1284.

Shao, J. (1997) An asymptotic theory for linear model selection. *Statistica Sinica*, **7**, 221–264.

Tibshirani, R. J. (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, **58**, 267–288.

Wang, H. and Leng, C. (2007) Unified lasso estimation via least squares approximation. *Journal of the American Statistical Association*, **101**, 1418–1429.

Wang, H., Li, B. and Leng, C. (2008) Shrinkage tuning parameter selection with a diverging number of parameters. *Technical Report*.

Wang, H., Li, G. and Tsai, C. L. (2007a) Regression coefficient and autoregressive order shrinkage and selection via lasso. *Journal of Royal Statistical Society, Series B*, **69**, 63–78.

Wang, H., Li, R. and Tsai, C. L. (2007b) On the consistency of scad tuning parameter selector. *Biometrika*, **94**, 553–558.

Zhang, H. H. and Lu, W. (2007) Adaptive lasso for cox's proportional hazard model. *Biometrika*, To appear.

Zou, H. (2006) The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, **101**, 1418–1429.