Discussion of "Sure Independence Screening for Ultra-High Dimensional Feature Space"

Hui Zou
Assistant Professor
School of Statistics, University of Minnesota
Minneapolis, MN 55455, USA
hzou@stat.umn.edu

I would like to congratulate Professors Fan and Lv for an excellent and stimulating paper which discusses several fundamental issues in high-dimensional data analysis. My comments will focus on the oracle properties after sure independence screening.

The size of reduced dimension. Fan and Peng (2004) extend the oracle properties of nonconcave penalized likelihood estimators in a finite dimension setting (Fan and Li, 2001) to the diverging dimension setting with $p = o(n^{1/3})$. Combined with SIS, this result allows us to reduce the dimension from $p \gg n$ to $d = o(n^{1/3})$ and we still have an oracle-like estimator. In real applications, we wish to use SIS to screen out noise features and also want to do so conservatively. Thus, it is of interest to know if Theorem 5 holds for larger d, i.e., $d = o(n^{\nu})$ and $\nu > \frac{1}{3}$. Some positive answers are reported in Zou and Zhang (2007) which show that, under reasonably week conditions, the adaptive elastic-net estimator enjoys the oracle properties for $p = O(n^{\nu})$ as long as $0 \le \nu < 1$. Hence, we believe SIS can be used in a more conservative way without sacrificing any theoretical optimality.

Which oracle estimator should be mimicked? In Fan and Li (2001) and Fan and Peng (2004) the likelihood estimators are considered, thus the oracle estimator should be MLE and the nonconcavely penalized likelihood estimator mimics the MLE oracle estimator. However, in linear regression models, we often do not wish to impose the error distribution assumption unless there is strong evidence for doing so. Thus the MLE estimator cannot be used to construct the oracle estimator. A popular oracle estimator is the least squares estimator, and the nonconcavely penalized least squares estimator mimics the least squares oracle estimator. When the error distribution is non-normal, then the least squares oracle can be inefficient. Zou and Yuan (2008) discuss the issues with the oracle in the oracle model selection theory and propose the *Composite Quantile Regression* (CQR) oracle estimator. Zou and Yuan (2008) show that relative efficiency of the CQR oracle compared to the least squares oracle is greater than 70% regardless the error distribution. Kai, Li and Zou (2008) further show that the efficiency lower bound can be as high as 86.4%. In the Gaussian model the relative efficiency is 95.5%. In a wide class of non-normal error models, the CQR oracle could be much more efficient and sometimes arbitrarily more efficient than the least squares oracle. Therefore, the CQR oracle is a safe and efficient alternative to the least square oracle.

Additional References

Zou, H. and Zhang, H.H. (2008) On the adaptive elastic-net with a diverging number of parameters. *The Annals of Statistics*, to appear.

Zou, H. and Yuan, M. (2008) Composite quantile regression and the oracle model selection theory. *The Annals of Statistics*, 36(3), 1108-1126.

Kai, B., Li, R. and Zou, H. (2008) Local polynomial composite quantile regression. *Technical Report 08-86*, The Methodology Center, The Pennsylvania State University, University Park.