# Discussion of "Sure Independence Screening for Ultrahigh Dimensional Feature Space" by Fan and Lv

Gareth M. James and Peter Radchenko*

We congratulate the authors on introducing a powerful new methodology for addressing an increasingly important problem. While the theoretical aspects of this work are impressive we have concentrated our discussion on the practical behavior of the authors' methodology. The basic moral of this paper is that when dealing with extremely large numbers of predictors one should use an iterative two step approach. At each iteration, one first uses a simple bivariate criterion to rank the predictors and hence obtain a "moderate" number of variables. Then a multivariate variable selection method is used to obtain the final set of predictors. The authors present convincing evidence that this approach can produce considerable improvements, both in terms of computational cost as well as statistical accuracy, over directly working with the full data set. The idea that a large number of variables can be discarded with little risk of eliminating important variables seems reasonable.

The authors work primarily with SCAD when implementing the ISIS approach. We were interested in the robustness of ISIS to different plug-in methods. Hence we reran the simulation results from Section 4.2.1 using two alternatives to the SCAD plug-in. The first replaced SCAD with the Lasso. The second replaced SCAD with a version of Forward Selection which selected the $K$ variables with largest correlations to the response. We utilized three values, $K = 1$, $K = n/4$ and $K = n/2$. In all other respects the setup was the same as for Section 4.2.1. Our results are provided in the following table.

For the $n = 50$ scenarios all methods gave almost perfect predictions. For the $p = 100$, $n = 20$ scenario we found that the iterative Forward method improved as $K$ grew with $K = n/2$ giving slightly superior results to the iterative Lasso approach. For the $p = 1,000$, $n = 20$ scenario the iterative Lasso outperformed the iterative Forward methods. However, interestingly, the iterative Lasso either gave the same performance as the standard Lasso or performed worse. In addition the iterative Lasso and Forward selection methods both substantially underperformed the iterative SCAD results reported by Fan and Lv. We drew the following conclusions from these results.

*University of Southern California, Los Angeles.

| $p$ | $n$ | ρ | ISIS | Lasso | Iterative Lasso | Forward$_1$ | Forward$_{n/4}$ | Forward$_{n/2}$ |
|---|---|---|---|---|---|---|---|---|
| 100 | 20 | 0 | 1.000 | 0.970 | 0.885 | 0.730 | 0.850 | 0.895 |
| | | 0.5 | 1.000 | 0.985 | 0.820 | 0.515 | 0.790 | 0.865 |
| | 50 | 0 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | 0.5 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 1000 | 20 | 0 | 1.000 | 0.340 | 0.305 | 0.250 | 0.275 | 0.235 |
| | | 0.5 | 1.000 | 0.556 | 0.180 | 0.025 | 0.130 | 0.165 |
| | 50 | 0 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | 0.5 | 1.000 | 1.000 | 0.985 | 0.940 | 0.990 | 1.000 |

Table 1: *Simulation comparison. ISIS and Lasso results are taken from Table 4 of Fan and Lv (2008). Iterative Lasso, Forward*$_1$*, Forward*$_{n/4}$ *and Forward*$_{n/2}$ *respectively replace SCAD in the ISIS method with Lasso, and Forward Selection using* $K = 1$, $K = n/4$ *and* $K = n/2$.

First, applying the iterative approach does not always cause an improvement, as demonstrated by the inferior performance of the iterative Lasso over standard Lasso. Second, at least in certain scenarios, the iterative approach seems to be sensitive to the plug-in method with SCAD providing significantly superior results to the Lasso and Forward Selection methods.