7. Extensions of our method

This section discusses extensions of our method on high-dimensional variable selection using Sure Independence Screening (SIS). It supplements the implementation of our procedure.

The key idea of the newly proposed SIS is using iteratively thresholded ridge regression to reduce dimensionality, a special case of which is applying a single componentwise regression. Three issues may come up with this approach. First, some unimportant predictors that are highly correlated with important predictors can have higher priorities selected by SIS than those that are relatively weakly related to the response variable but unrelated to those important predictors. Second, it might be possible that an important predictor is uncorrelated marginally, but correlated jointly with the response, and this variable can not be singled out by SIS and thus will not enter the estimated model. Third, the issue of collinearity adds difficulty to the problem of model selection and its implementation. These three issues are separately addressed further in the following two subsections. With these extensions, our procedure allows us to use more fully the joint information of the covariates, rather than the marginal information in variable selection.

7.1. Iterative application of SIS

Our method of high-dimensional model selection is a two-step procedure. When the model assumptions are satisfied, it has been demonstrated that SIS is applicable to reduce the dimensionality from untra high to a relatively large scale, e.g., below sample size. After dimension reduction, one can use SCAD, Dantzig selector, Lasso, or adaptive Lasso to select a suitable model. In some rare cases, the aforementioned first two issues arise. In these situations, it might be possible that SIS would miss some important predictors.

One way to solve this problem is introducing an iterative procedure. In the first step, we select a subset of $k_1$ variables $A_1 = \{X_{i_1}, \ldots, X_{i_{k_1}}\}$ using our method, i.e., SIS followed by a refined model selection procedure. Then, we have an $n$-vector of residuals from regressing the response $Y$ over $X_{i_1}, \ldots, X_{i_{k_1}}$. In the second step, we treat the residuals as new responses and apply our method of model selection to the remaining $p_n - k_1$ variables, resulting in a subset of $k_2$ variables $A_2 = \{X_{j_1}, \ldots, X_{j_{k_2}}\}$. Fitting the residuals using $\{X_1, \ldots, X_{p_n}\} \setminus A_1$ can weaken significantly the priorities of those that are uncorrelated with the response, but marginally highly correlated with the response variable through their associations with $X_{i_1}, \ldots, X_{i_{k_1}}$. This solves the first issue. It also makes those important predictors that are missed in the previous step possible to survive. This addresses the second issue above.

We can keep on doing this and finally get $\ell$ disjoint subsets $A_1, \ldots, A_\ell$. The collection of variables $A = \bigcup_{i=1}^{\ell} A_i$ would contain almost all the important variables. Finally, applying SCAD, Dantzig selector, Lasso, or adaptive Lasso to $A$ will produce a model that is very close to the true sparse model $M_{n,*}$.

Another possible way that is worthy of pursuing is using grouped variables. For instance, we can divide the pool of $p_n$ variables into disjoint groups each with 5 variables. The idea of
variable screening via SIS can be applied to select a small number of groups. In this way, there is less chance of missing important variables by taking advantage of the joint information among the predictor variables and thus a more reliable model can be constructed in the end.

7.2. **Transformation of input variables**

A notorious difficulty of variable selection lies in the collinearity among the variables. Effective ways of ruling out unimportant variables that are highly correlated with important predictors are being sought after. A good idea is transforming the predictor variables. Two possible ways stand out in this regard. One is subject related transformation and the other is statistical transformation.

Subject related transformation is a useful tool. In some cases, a simple linear transformation of the input variables can help weaken correlation among predictors. For example, in Somatotype studies, common sense tells us that the predictors such as the weights $w_1$, $w_2$ and $w_3$ at 2, 9 and 18 years are positively correlated. We could directly use $w_1$, $w_2$ and $w_3$ as input variables in a linear regression model, but a better way of model selection is using less correlated predictors $(w_1, w_2 - w_1, w_3 - w_2)^T$, which is a linear transformation of $(w_1, w_2, w_3)^T$ and specifies the changes of weights instead of the weights themselves. Another important example is the financial time series such as the prices of stocks or interest rates. Differencing weakens significantly the correlation among those variables.

Methods of statistical transformation include an application of a clustering algorithm such as the hierarchical clustering or $k$-mean algorithm using the correlation ‘metrics’ to first group variables into highly correlated groups, and then apply sparse principal components analysis (PCA) to construct weakly correlated predictors. Now, those weakly correlated predictors from each group can be regarded as new covariates and the SIS-refined model selection procedure can be employed to select them.

These statistical techniques can help identify important features and thus improve the effectiveness of the vanilla SIS-refined model selection strategy.