

# Robust rank correlation based screening

Lixing Zhu

Department of Mathematics  
Hong Kong Baptist University

## Abstract

Independence screening is a variable selection method that uses a ranking criterion to select significant variables, particularly for statistical models with nonpolynomial dimensionality or “large  $p$ , small  $n$ ” paradigms when  $p$  can be as large as an exponential of the sample size  $n$ . In this paper, we propose a robust rank correlation screening (RRCS) method to deal with ultra-high dimensional data. The new procedure is based on the Kendall  $\tau$  correlation coefficient between response and predictor variables rather than the Pearson correlation of existing methods. The new method has four desirable features compared with existing independence screening methods. First, the sure independence screening property can hold only under the existence of a second order moment of predictor variables, rather than exponential tails or alikeness, even when the number of predictor variables grows as fast as exponentially of the sample size. Second, it can be used to deal with semiparametric models such as transformation regression models and single-index models under monotonic constraint to the link function without involving nonparametric estimation even when there are nonparametric functions in the models. Third, the procedure can be largely used against outliers and influence points in the observations. Last, the use of indicator functions in rank correlation screening greatly simplifies the theoretical derivation due to the boundedness of the resulting statistics, compared with previous studies on variable screening. Simulations are carried out for comparisons with existing methods and a real data example is analyzed.