REJOINDER:
"INFLUENTIAL FEATURES PCA FOR HIGH DIMENSIONAL CLUSTERING"

By Jiashun Jin and Wanjie Wang

Carnegie Mellon University and National University of Singapore

"Screen first and estimate next" is a popular strategy for attacking many high dimensional problems. While the methods may vary from occurrence to occurrence, the high level ideas are all similar; below are some examples.

- *Screen and Clean*. Consider the variable selection problem where we have a large number of variables but most of them are 0. We may first screen out many variables and then focus on the small fraction of surviving variables. Success has been shown in, for example, Fan and Lv (2008), Jin, Zhang and Zhang (2014), Wasserman and Roeder (2009).
- *Screen and Classify*. Consider a classification problem where we have a large number of measured features but most of them are useless for the classification decision. We may first screen out many of them and use only the surviving ones for classification. Success has been shown in, for example, Donoho and Jin (2008), Efron (2009), Tibshirani et al. (2002).

Compared to popular penalization methods, the "screen first and estimate next" approach is usually computationally faster and sometimes easier to tune [e.g., Donoho and Jin (2008)]. It may also have the optimality that penalization methods do not have [e.g., Jin, Zhang and Zhang (2014), Ke, Jin and Fan (2014)]. Of course, we can always combine two approaches by applying some penalization methods to the post-screening data.

IF-PCA in our paper is one more example of the "screen first, estimate next" strategy: we first screen with the Kolmogorov–Smirnov (KS) statistic, and then cluster with the classical PCA. To set the threshold in the screening step, we use Efron's null correction and (Tukey's) Higher Criticism.

The focus of our paper is to find a balance between precise mathematical theory and practical feasibility, and to develop easy-to-use and yet effective methods that have minimum gaps between theory and real applications.

We would like to thank all the discussants for their very thoughtful and stimulating comments. Below are our responses.