

# Sparse local influence analysis

Jun Lu<sup>1</sup> and Lei Shi<sup>1</sup>

<sup>1</sup> *School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming, 650221 China*

## Abstract

Cook's (1986) local influence method is useful for identifying influential observations in statistical diagnostics and sensitivity analysis. However, it is often criticized for lack of a rigorous criterion to judge the influence magnitude from the elements of the main diagnostic. In this paper, we propose a new method, namely sparse local influence analysis, to detect the influential observations. We use the connection of local influence analysis with sparse principal component analysis and produce the modified local diagnostic with sparse elements, i.e. diagnostic with very few nonzero elements. With this method, influential observations can be efficiently detected by the remaining nonzero elements of the modified diagnostic. Two real data sets are used for illustration and a simulation example is conducted to confirm the efficiency of the proposed methodology.

## Keywords

Local influence analysis, Influential observations, Sparseness

## References

- [1] Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society, B* 48,133-169.
- [2] Qi,X., Luo, R., Zhao, H. (2013). Sparse principal component analysis by choice of norm. *Journal of Multivariate Analysis*, 114,127-160.
- [3] She,Y., Owen, A. (2011). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*, 85,633-639.
- [4] Shi, L., and Huang, J. (2008). Outlier detection using nonconvex penalized regression. *Journal of Multivariate Analysis*,106,626-639.
- [5] Shi, L. (1997). Local influence in principal component analysis. *Biometrika*, 84,175-186.
- [6] Tibshirani, R. (1996). Regression shrinkage and selection via Lasso. *Journal of the Royal Statistical Society, B*, 58,267-288.
- [7] Zou, H., Hastie, T., Tibshirani, R. (2006). Sparse principal component analysis. *Journal of Computational and Graphical Statistics*, 15,265-286.