



Sparse PCA for high-dimensional data with outliers

Mia Hubert*

KU Leuven, Department of Mathematics, Belgium – mia.hubert@wis.kuleuven.be

Tom Reynkens

KU Leuven, Department of Mathematics, Belgium – tom.reynkens@wis.kuleuven.be

Eric Schmitt

KU Leuven, Department of Mathematics, Belgium – eric.schmitt@wis.kuleuven.be

Tim Verdonck

KU Leuven, Department of Mathematics, Belgium – tim.verdonck@wis.kuleuven.be

A new sparse PCA algorithm is presented which is robust against outliers. The approach is based on the ROBPCA algorithm which generates robust but non-sparse loadings. The construction of the new ROSPCA method is detailed, as well as a selection criterion for the sparsity parameter. An extensive simulation study and a real data example are performed, showing that it is capable of accurately finding the sparse structure of datasets, even when challenging outliers are present. In comparison with a projection pursuit based algorithm, ROSPCA demonstrates superior robustness properties and comparable sparsity estimation capability, as well as significantly faster computation time.

Keywords: dimension reduction, outlier detection, robustness.