Abstract:

It is common practice in multivariate and matrix-valued data analysis to reduce dimensionality by performing a Singular Value Decomposition or Principal Component Analysis, and keeping only $r$ singular values or principal components, the rest being presumably associated with noise. However, the literature does not propose a disciplined criterion to determine $r$; most practitioners still look for the "elbow in the Scree Plot", a 50-years-old heuristics performed by eye. I'll review a line of work which develops a systematic approach to eigenvalue and singular value thresholding. This approach assumes that the signal is low-rank and that the noise is rotationally invariant. Recent results derive optimal thresholds in the presence of quite general noise distributions.

Joint work with David Donoho, Iain Johnstone and Edgar Dobriban (Stanford).