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 48-years-old heuristic performed by eye. Formally, this is a matrix denoising problem, in which one recovers an unknown matrix $X$ from a noisy observation $Y=X+Z$. We show that, for white noise and appropriate asymptotic frameworks, random matrix theory successfully describes the random behavior of the singular values and vectors of $Y$. It delivers simple, convincing answers to a range of fundamental questions, such as the location of the optimal singular value threshold $(2.309)$ and the shape of the optimal singular value shrinker (reflected Quarter Circle density). Our framework has been used to discover optimal eigenvalue shrinkers for high-dimensional covariance estimation, and seems to apply to various other estimators that rely on eigendecomposition of signal or data matrices. Moreover, several methods for low-rank matrix recovery from incomplete observations rely on iterative matrix denoising; we discuss evidence that improved matrix denoising can lead to improved matrix compressed sensing.