Sparse Principal Component Analysis, also known as Sparse PCA, is a problem which is a variant of the classical PCA problem, which performs a trade-off between the explained variance along a normalized vector and the number of non-zero components of that vector.

Sparse PCA is helpful because it not only brings better interpretation but also provides statistical regularization when the number of samples is less than the number of features. Numerous different formulations and algorithms have been proposed for this problem, ranging from factor rotation techniques and simple thresholding to greedy algorithms, regularized SVD method, and the generalized power method. These algorithms are based on non-convex formulations and may only converge to a local optimum. The l_1-norm based semidefinite relaxation DSPCA, however, guarantees global convergence, which makes it an appealing alternative to local methods. In fact, simple ad-hoc methods, as well as greedy, SCoTLASS, and SPCA algorithms often underperform DSPCA.

However, the downside is that the computational complexity of DSPCA is often too high for many large-scale data sets. Given this complexity estimate, this seems to indicate at first glance that solving sparse PCA would be much more expensive than PCA, since one principal component can be computed with a complexity of $O(n^2)$.

The paper's overall aim is to show that solving DSPCA is actually computationally easier than PCA, which would consequently mean it can be applied to very large-scale data sets. This is done by first viewing DSPCA as an approximation to a harder, cardinality-constrained optimization problem, and based on this formulation, describing a safe feature elimination method for that problem, which often leads to an important reduction in problem size.