This paper demonstrates that sparse principal component analysis (Sparse PCA) can be computationally easier than standard PCA for large, real-world datasets. Sparse PCA finds a linear combination of a small number of features that maximizes variance across the data, providing better interpretability and statistical regularization compared to PCA.

The authors first formulate Sparse PCA as a semidefinite programming relaxation (DSPCA) of a cardinality-constrained PCA problem. They then present a rigorous safe feature elimination pre-processing step based on the formulation. This allows provably eliminating features with small variances from the problem before solving it, often dramatically reducing the problem size.

The key insight is that real-world datasets typically exhibit rapidly decaying feature variances. By eliminating features below a given threshold before solving DSPCA, the authors can work with a much smaller covariance matrix than the original data. For textdata with over 100,000 features, the reduced problem size was 150-200 times smaller than the original.

The authors develop a fast block coordinate ascent algorithm to solve the reduced DSPCA problem with computational complexity O(n^3), much better than the previous O(n^4√log n) first-order method. At each iteration, the new algorithm updates one row/column of the solution matrix by solving a box-constrained quadratic program and a 1-D problem.

Experimental results on two large text corpora - the NYTimes news articles (300K docs, 102K words) and PubMed abstracts (8.2M docs, 141K words) - demonstrate the efficacy of the approach. The authors set a target cardinality of 5 for the sparse principal components to aid interpretability.

For the NYTimes data, the top sparse PCs uncovered topics like business, sports, U.S. news, politics and education - perfectly corresponding to how the NYTimes classifies articles on its website, despite having no labeled data. For PubMed, coherent topics like medical terms, experimental treatments, human subjects and cancer were identified.

Most impressively, the safe feature elimination allowed reducing the full 102K x 102K and 141K x 141K covariance matrices to just 500x500 and 1000x1000 matrices respectively for the NYTimes and PubMed data when computing the low-cardinality sparse PCs. This reduction of 2-3 orders of magnitude in problem size made sparse PCA surprisingly easier than PCA itself for these datasets.

In conclusion, the safe feature elimination pre-processing coupled with the fast block coordinate ascent algorithm enabled, for the first time, sparse PCA on massive real-world datasets with hundreds of thousands of features. The interpretable low-dimensional representations discovered by sparse PCA provide a promising alternative to topic models for exploratory data analysis of large text corpora.