Sparse PCA from Sparse Linear Regression

Guy Bresler[1] Sung Min Park[1] Mădălina Persu[1,2]

[1]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology
[2]Two Sigma Investments, LP.

Summary

We show how to efficiently transform a black-box solver for **Sparse Linear Regression (SLR)** into an algorithm for **Sparse Principal Component Analysis (SPCA)**, two fundamental statistical problems for which analyses have been largely disjoint.

1. Background

- **Principal component analysis (PCA)** is a fundamental technique for dimension reduction used widely in data analysis. PCA projects data along a few directions that explain most of the variance of observed data.

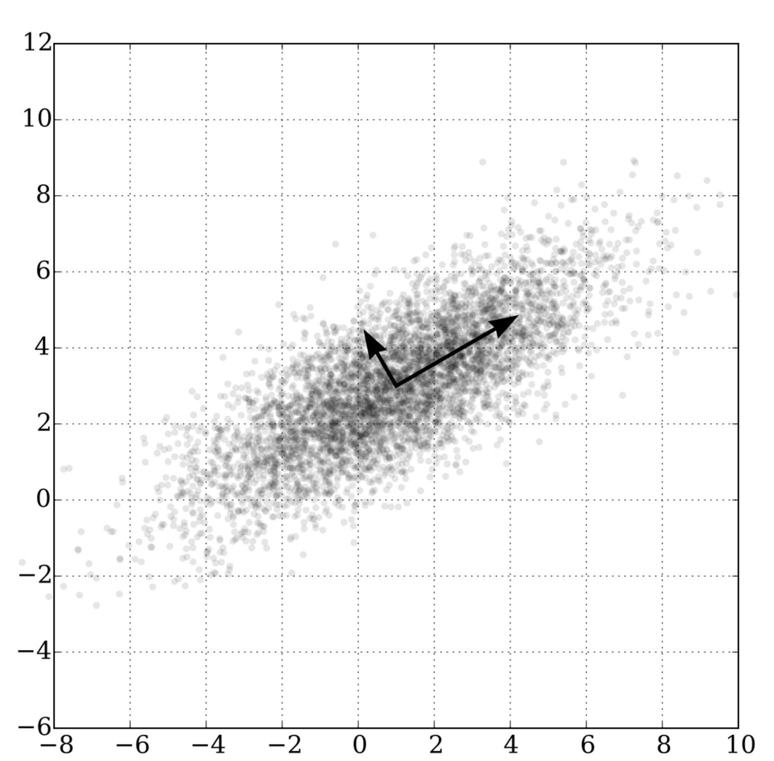

Figure 1: PCA for 2D data; arrows show principal components.

- Recent work in **high-dimensional statistics** has focused on **sparse principal component analysis (SPCA)**, as ordinary PCA estimates become inconsistent in this regime.
- Sparsity assumptions have played an important role in a variety of other problems in high-dimensional statistics, in particular **sparse linear regression**.
- Both problems exhibit similar statistical-computational trade-offs (i.e. gap in performance between information-theoretically optimal and computationally efficient procedures)

2. Question

What are statistical and algorithmic connections between sparse PCA and sparse linear regression?

3-i. SPCA model

We consider the **spiked covariance model**:

- d-dimensional Gaussiasn random variable
$$X \sim \mathcal{N}(0, I_{d\times d} + \theta \vec{u}\vec{u}^\top)$$
- spike $\vec{u}$ is sparse; at most $k \ll d$ non-zero entries
- n samples are observed; $\mathbf{X} \in \mathbb{R}^{d\times n}$
- signal-to-noise ratio θ

We study two different objectives:

- *hypothesis testing*: distinguish above distribution from isotropic distribution $N(0, I_{d\times d})$
- *support recovery*: recover the support of $\vec{u}$

3-ii. SLR model

Data is generated from the linear model
$$y = \mathbb{X}\beta^* + w$$

- Observed: $y \in \mathbb{R}^n$; design matrix $\mathbb{X} \in \mathbb{R}^{n\times d}$
- Goal is to recover β^* under some metric
- $w \in \mathbb{R}^n$ with i.i.d. $\mathcal{N}(0, \sigma^2)$ entries

We assume that our black-box oracle SLR satisfies the following *prediction error guarantee*: SLR$(y, \mathbb{X}, k)$ outputs $\hat{\beta}$ that is k-sparse and w.h.p:
$$\frac{1}{n}\|\mathbb{X}\hat{\beta} - \mathbb{X}\beta^*\|_2^2 \le \frac{c}{\gamma(\mathbb{X})^2}\frac{(\sigma^2 k \log d)}{n} \quad \forall \beta^* \in B_0(k)$$
where $\gamma(\mathbb{X})$ is the *restricted eigenvalue* constant.

4. Main Theorem

Given black-box access to an SLR oracle that satisfies the above prediction error guarantee, we can efficiently solve SPCA instances from the single spiked covariance model if the signal is strong enough ($\theta^2 \gtrsim \frac{k^2 \log d}{n}$).

5. Reduction

Algorithm 1 hypothesis testing

Input: $\mathbf{X} \in \mathbb{R}^{d\times n}, k$
Output: $\{0, 1\}$ ▷ 0 for null, 1 for spiked
for $i = 1, \dots, d$ **do**
  $\hat{\beta}_i = \mathrm{SLR}(\mathbf{X}_i, \mathbf{X}_{-i}, k)$ ▷ regress i^{th} coordinate on the rest
  $Q_i = \frac{1}{n}\|\mathbf{X}_i\|_2^2 - \frac{1}{n}\|\mathbf{X}_i - \mathbf{X}_{-i}\hat{\beta}_i\|_2^2$
  if $Q_i \gtrsim \frac{k \log \frac{d}{k}}{n}$ **then** return 1 **end if**
end for
return 0

Analysis boils down to analyzing the distribution of Q_i.

6. Experiments

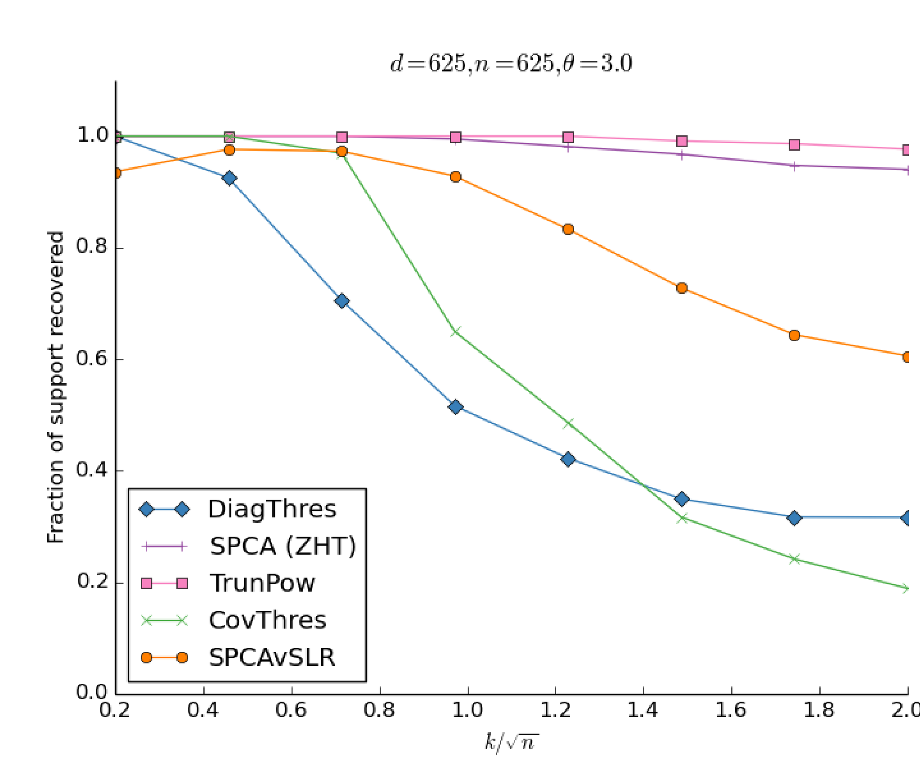

Figure 2: Performance of diagonal thresholding, SPCA (ZHT), truncated power method, covariance thresholding, and SPCAvSLR (ours) for support recovery at $n = d = 625$, $\theta = 3.0$

6. Experiments (cont.)

Many iterative approaches to SPCA are based on initialization by filtering for the highest variances. Such approaches are not robust to the rescaling.

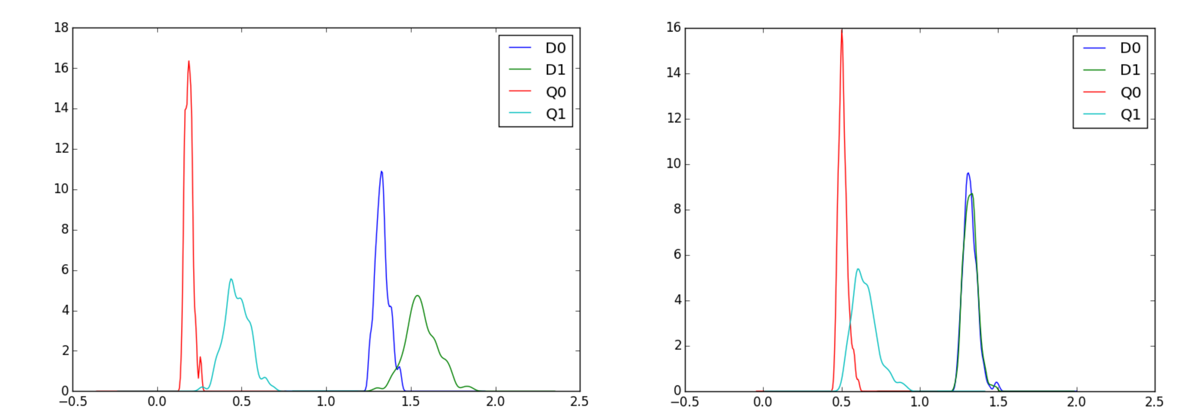

Figure 3: Performance of diagonal thresholding (D) vs. our statistic (Q) for hypothesis testing at $n = 200, d = 500, k = 30$; right is after rescaling all variances to one (note higher θ was used).

7. Related work

SPCA algorithm	approach	signal strength
Diagonal thresholding	variance filtering	nearly optimal
"SPCA" (Zou et al.)	sparse regression	??
(many papers)	SDP-relaxation	nearly optimal
Truncated Power	spectral/iterative	nearly optimal
Covariance Thresholding	spectral	optimal
SPCAvSLR (ours)	sparse regression	nearly optimal

Table 1: Various approaches to SPCA

8. Conclusion

- We demonstrated a mathematical connection between SPCA and SLR.
- New algorithmic framework performs comparably with increased robustness to scale of data.

Contact: {sp765, mpersu}@mit.edu