Scalable Sparse PCA: A Tractable MIP under Statistical Assumptions

Kayhan Behdin and Rahul Mazumder

Massachusetts Institute of Technology

Problem Definition and Motivation

- In Sparse PCA (SPCA), we are given n independent samples from a mean zero p-dimensional Gaussian distribution with a spiked covariance model, i.e.,

$$X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(0, G^*) \quad \text{and} \quad G^* = I_p + \theta u^* u^{*T} \tag{1}$$

  where $u^* \in \mathbb{R}^p$ is a sparse vector with s-nonzeros and θ is the SNR. The goal is to estimate u^* from the data.

- The natural SPCA problem is given by:

$$\max_{u \in \mathbb{R}^p} \quad u^T X^T X u \quad \text{s.t.} \quad \|u\|_2 \leq 1; \|u\|_0 \leq s. \tag{2}$$

  A solution $\hat{u}$ to problem (2) is known to enjoy optimal statistical properties under model (1).

- Current MIP formulations of SPCA [1, 2, 3] provide small/moderate optimality gap $\sim 20\%$ when $p \sim 2000$.

Problem Formulation

- By utilizing Gaussian graphical models and properties of model (1), we reformulate SPCA as the MIQP:

$$\min_{\beta, z} \quad \sum_{j=1}^{p} \|x_j - \sum_{i \neq j} \beta_{i,j} x_i\|_2^2 \tag{3}$$
$$\text{s.t. } z \in \{0,1\}^p; \ |\beta_{i,j}| \leq \min(z_i, z_j), \ i, j \in [p]; \ \sum_i z_i \leq s$$

  where x_i is the i-th column of the data matrix $X \in \mathbb{R}^{n \times p}$.

- We also consider a perspective formulation of (3):

$$\min_{\beta, z, q} \quad \sum_{j=1}^{p} \|x_j - \sum_{i \neq j} \beta_{i,j} x_i\|_2^2 + \lambda \sum_{j \in [p]} \sum_{i \neq j} q_{i,j} \tag{4}$$
$$\text{s.t.} \quad z \in \{0,1\}^p; \quad q_{i,j} \geq 0; \quad |\beta_{i,j}| \leq \min(z_i, z_j) \ i, j \in [p]$$
$$\beta_{i,j}^2 \leq q_{i,j} z_j, \ i, j \in [p]; \quad \sum_i z_i \leq s.$$

- Big-M in our formulation under statistical model (1) is 1.

Optimization Algorithm

- We show problems (3) and (4) can be reformulated as

$$\min_{z} \quad F(z) \ \text{ s.t. } \ z \in \{0,1\}^p; \ \sum_{i=1}^{p} z_i \leq s, \tag{5}$$

  where $F : [0,1]^p \to \mathbb{R}$ is a convex subdifferentiable function.

- At each iteration, we minimize a piecewise linear lower bound of F under the constraints of problem (5):

$$\min_{z \in \{0,1\}^p} \ \max_{i=0,\cdots,t-1} \{F(z^i) + g_{z^i}^T (z - z^i)\} \text{ s.t. } \sum_{i=1}^{p} z_i \leq s, \tag{6}$$

  where g_z is a subgradient of F at z and z^i is the minimizer of (6) at iteration i. Note that problem (6) is an MILP.

- A subgradient of $F(z)$ can be efficiently computed by solving s QPs/SOCPs in parallel, each with s variables. For e.g., for (3) we need to solve (7) for all j such that $z_j = 1$:

$$\min_{\beta_j} \ \frac{1}{2}\|x_j - \sum_{i \neq j} \beta_{i,j} x_i\|_2^2 \quad \text{s.t.} \quad |\beta_{i,j}| \leq z_i, i \in [p]. \tag{7}$$

- We use first-order methods to solve problem (7). As s is small, first-order algorithms are efficient.

Statistical Theory

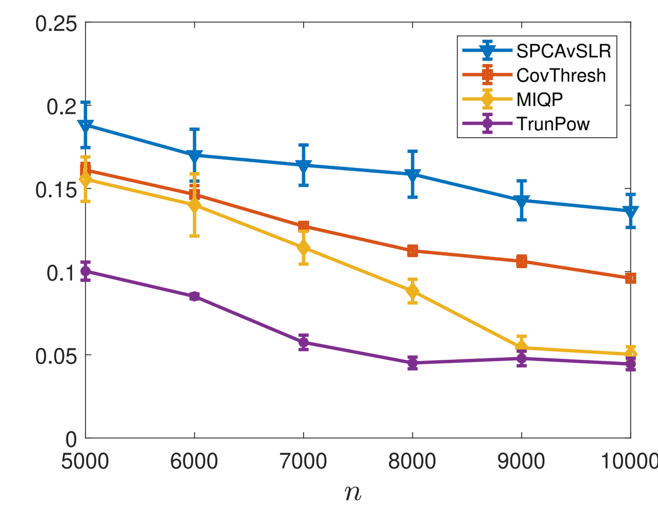
- **Theorem 1:** The optimal solution of problem (3) can be used to estimate $\hat{u}$ such that with high probability,

$$\sin^2 \angle(\hat{u}, u^*) \lesssim \frac{s^2 \log(p/s)}{n\theta^2}.$$

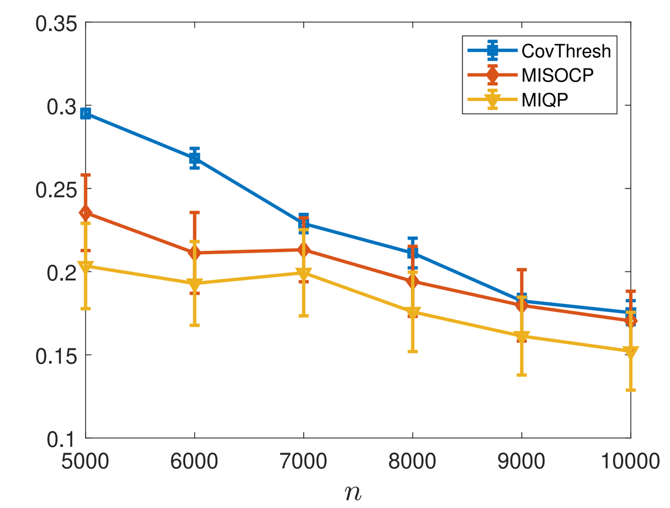
- **Theorem 2:** Suppose for $i \in [p]$ such $u_i^* \neq 0$, $|u_i^*| \gtrsim \frac{1}{\sqrt{s}}$. Then, if $n \gtrsim s^2 \log(p/s)/\theta^2$, problem (3) recovers the support of u^* correctly with high probability.
- Polynomial-time algorithms for SPCA achieve the same rate up to logarithmic factors [4].
- The minimax optimal error rate for (2) is $s \log(p/s)/n\theta^2$. We lose a factor s in lieu of a *simpler* MIP.

Numerical Experiments

- Experiments are done on a personal desktop with runtime limited to 20 minutes.
- Problem (4) (perspective formulation) leads to smaller MIP-gaps compared to (3).



Estimation Error ($s = 5$)



Estimation Error ($s = 10$)

Proposed vs polynomial-time and heuristic methods. ($p = 10^4$)

- Both our formulations provide optimality gap smaller than 10% and 20% for $s = 5$ and $s = 10$, respectively.

Conclusion

- We present simplified MIPs under statistical assumptions to solve SPCA problem with $p \sim 10^4$ in tens of minutes. Current MIP algorithms for SPCA can provide moderate optimality gap for $p \sim 2000$.
- Our framework enjoys statistical guarantees on par with polynomial-time algorithms, but with significantly improved statistical performance.
- Our algorithm provides numerical performance close to heuristic algorithms which, unlike our algorithm, have limited theoretical guarantees.

References:

[1] Yongchun Li and Weijun Xie. Exact and approximation algorithms for sparse pca. arXiv preprint arXiv:2008.12438, 2020

[2] Santanu S Dey, Rahul Mazumder, and Guanyi Wang.A convex integer programming approach for optimal sparse PCA. arXiv preprint arXiv:1810.09062, 2018

[3] Dimitris Bertsimas, Ryan Cory-Wright, and Jean Pauphilet.Solving large-scale sparse pca to certifiable (near) optimality.arXiv preprint arXiv:2005.05195, 2020.

[4] Yash Deshpande and Andrea Montanari. Sparse pca via covariance thresholding. The Journal of Machine Learning Research, 17(1):4913–4953, 2016.