Complexity Theoretic Lower Bounds for Sparse Principal Component Detection

Conference On Learning Theory - 2013, Princeton

Quentin Berthet, Philippe Rigollet

Princeton University
Sparse principal component detection

\(X_1, \ldots, X_n \in \mathbb{R}^d\) independent, centered Gaussian with unknown covariance.

Isotropy: \(\mathcal{N}(0, I_d)\)

Sparse PC: \(\mathcal{N}(0, I_d + \theta vv^\top)\)

Active topic: Amini and Wainwright (09), Vu and Lei (12), Cai and Ma (13)

Detection problem: Is there a sparse principal component?
Sparse principal component detection

Testing problem between two hypotheses.

\[
\begin{cases}
    H_0 : X \sim \mathcal{N}(0, I_d) \\
    H_1 : X \sim \mathcal{N}(0, I_d + \theta vv^\top), \quad v \in \mathcal{B}_0(k)
\end{cases}
\]

\(v\) is a \(k\)-sparse unit vector. \(\mathcal{B}_0(k) = \{v \in \mathbb{R}^p : |v|_2 = 1, |v|_0 \leq k\}\).

Isotropy: \(\mathcal{N}(0, I_d)\)  
Sparse PC: \(\mathcal{N}(0, I_d + \theta vv^\top)\)
Optimal testing - Gaussian setting

Minimax setting: $\theta^*$ is the \textbf{optimal rate of detection} when

- For $\theta > \bar{c} \theta^*$, testing is possible, with test $\psi$
  \[
P_0 \otimes n (\psi = 1) \lor \max_{v \in B_0(k)} P_v \otimes n (\psi = 0) \leq \delta.
  \]

- For $\theta < \underline{c} \theta^*$, testing is impossible, for all tests $\phi$
  \[
P_0 \otimes n (\phi = 1) \lor \max_{v \in B_0(k)} P_v \otimes n (\phi = 0) \geq \delta.
  \]
Optimal testing - Robust setting

Minimax setting: $\theta^*$ is the \textbf{optimal rate of detection} when

- For $\theta > \bar{c}\theta^*$, testing is possible, with test $\psi$
  \[
  \sup_{P_0, P_1} \left\{ P_0^\otimes n(\psi = 1) \lor P_1^\otimes n(\psi = 0) \right\} \leq \delta.
  \]

- For $\theta < c\theta^*$, testing is impossible, for all tests $\phi$
  \[
  \sup_{P_0, P_1} \left\{ P_0^\otimes n(\phi = 1) \lor P_1^\otimes n(\phi = 0) \right\} \geq \delta.
  \]
Optimal testing - Robust setting

Minimax setting: $\theta^*$ is the **optimal rate of detection over the class** $\mathcal{T}$ when

- For $\theta > \bar{c}\theta^*$, testing is possible, with test $\psi \in \mathcal{T}$

$$\sup_{P_0, P_1} \left\{ P_0^\otimes n(\psi = 1) \lor P_1^\otimes n(\psi = 0) \right\} \leq \delta.$$  

- For $\theta < c_\phi \theta^*$, testing is impossible, for all tests $\phi \in \mathcal{T}$

$$\sup_{P_0, P_1} \left\{ P_0^\otimes n(\phi = 1) \lor P_1^\otimes n(\phi = 0) \right\} \geq \delta.$$
Minimax testing

Classical theory, no restriction on testing function.

$$\theta^* = \sqrt{\frac{k \log(d)}{n}}$$

- For $\theta > \bar{c} \theta^*$, for an explicit test $\psi$

  $$\sup_{P_0, P_1} \left\{ P_0^\otimes n(\psi = 1) \lor P_1^\otimes n(\psi = 0) \right\} \leq \delta.$$  

- For $\theta < \underline{c} \theta^*$, testing is impossible, for all tests $\phi$

  $$\sup_{P_0, P_1} \left\{ P_0^\otimes n(\phi = 1) \lor P_1^\otimes n(\phi = 0) \right\} \geq \delta.$$  

Test $\psi$ based on a sparse eigenvalue statistic $\lambda_{\text{max}}^k(\hat{\Sigma})$, NP-hard to compute.
Polynomial-time testing

We only consider tests in $\mathcal{T} = \mathcal{P}$, running in polynomial time.

$$\tilde{\theta} = \sqrt{\frac{k^2 \log(d)}{n}}$$

- For $\theta > \bar{c} \tilde{\theta}$, for several explicit tests $\psi$

$$\sup_{P_0, P_1} \{P_0^\otimes n(\psi = 1) \lor P_1^\otimes n(\psi = 0)\} \leq \delta.$$

Tests: Diagonal method - Johnstone (01), SDP - d’Aspremont et al. (07), MDP - B. and R. (12), other heuristics.

<table>
<thead>
<tr>
<th>$\theta^*$</th>
<th>$\tilde{\theta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>no detection</td>
<td>combinatorial method</td>
</tr>
<tr>
<td>polynomial-time method</td>
<td></td>
</tr>
</tbody>
</table>

$$\sqrt{\frac{k \log(d)}{n}} \quad \sqrt{\frac{k^2 \log(d)}{n}}$$

Q.Berthet - Complexity Theoretic Lower Bounds for Sparse Principal Component Detection
Situation suggested by those results

\[ \sqrt{\frac{k \log(d)}{n}} \quad \sqrt{\frac{k^2 \log(d)}{n}} \]

\( \theta^* \)  

\( \tilde{\theta} \)

- no detection
- combinatorial method
  - no polynomial-time detection
- polynomial-time method
Detection rates

So far, only upper bounds, suggestions

\[ \theta^* \quad \theta^\alpha \quad \tilde{\theta} \]

- no detection
- combinatorial method
- polynomial-time method

\[ \sqrt{\frac{k \log(d)}{n}} \quad \sqrt{\frac{k^\alpha}{n}} \]

Situation could be very different.

Need for **Complexity Theoretic Lower Bounds**
Planted clique problem
Erdős-Rényi graphs

$G(m, 1/2)$: Each edge is randomly connected, with probability $1/2$, independently.
Erdős-Rényi graphs

The expectation of the adjacency matrix is constant: pure noise setting.

Random instance

Expectation
Erdős-Rényi graphs

$G(m, 1/2, \kappa)$: A clique of size $\kappa$ is planted in a graph from $G(m, 1/2)$. 

Graph

Adjacency matrix
Erdős-Rényi graphs

$G(m, 1/2, \kappa)$: A clique of size $\kappa$ is planted in a graph from $G(m, 1/2)$. 

Graph

Adjacency matrix
Erdős-Rényi graphs

The expectation of the adjacency matrix has a sparse signal structure

Random instance

Expectation
Planted clique problem

Detection of a structured signal of sparsity $\kappa$ in a random graph of size $m$.

$G(m, 1/2) = P_0^{(G)}$

$G(m, 1/2, \kappa) = P_1^{(G)}$
Planted clique problem

Detection of a structured signal of sparsity $\kappa$ in a random graph of size $m$.

$G(m, 1/2) = P_0^{(G)}$  

$G(m, 1/2, \kappa) = P_1^{(G)}$
Distinguishing those two distributions is called the Planted Clique problem

\[
\begin{aligned}
H_0^{PC} : G &\sim \mathcal{G}(m, 1/2) = P_0^{(G)} \\
H_1^{PC} : G &\sim \mathcal{G}(m, 1/2, \kappa) = P_1^{(G)}.
\end{aligned}
\]

- Detection for \( \kappa > 2 \log_2(m) \), NP-hard method (Max-clique) Pencer (94).
- Polynomial time detection for \( \kappa = O(\sqrt{m}) \) Alon et al. (98).
- Strong reasons to believe impossible detection in polynomial time for \( \kappa = O(m^c), c < 1/2 \).

Ames and Vavasis (11), Dekel et al. (10); Feige and Krauthgamer (00); Feige and Ron (10)

Jerrum (92), Feige and Krauthgamer (92), Rossman (10), Feldman et al. (12)

Juels and Peinado (00), Alon et al. (07), Hazan and Krauthgamer (11), Alon et al. (2011)
**Hypothesis** $A_{PC}$

For any $a, b \in (0, 1)$, all randomized $P$-time tests $\xi$, there exists $\Gamma > 0$ such that

\[
P_0^{(G)}(\xi_{m, \kappa}(G) = 1) \lor P_1^{(G)}(\xi_{m, \kappa}(G) = 0) \geq 1.2\delta, \quad \forall \ m^{\frac{a}{2}} < \Gamma \kappa < m^{\frac{b}{2}}.
\]

**Remarks**

- Formalization of computational hardness of the statistical problem.
- Constant $\Gamma$ can depend on $\xi, a, b$: Asymptotic nature of the class $P$.

Create a randomized $P$-time function from graphs to random vectors.

\[G(m, 1/2) = P_0^{(G)}\]

\[G(m, 1/2, \kappa) = P_1^{(G)}\]
$G(2m, 1/2)$

$G(2m, 1/2, \kappa)$

$P_{bl(G)} = P_0^{\otimes n}$

$\kappa \leq (2m)^{b/2}$

$\theta \leq \sqrt{\frac{k^a}{n}}$

$\mathcal{D}_0$

$\mathcal{D}_1^k(\theta)$
Reduction description

Function based on the bottom-left corner of the adjacency matrix.

- $G(2m, 1/2) \rightarrow n$ vectors i.i.d. with independent $\{-1, 1\}$ coefficients $\in \mathcal{D}_0$.
- $G(2m, 1/2, \kappa_b) \rightarrow n$ vectors with distribution close (in TV) to i.i.d. $\in \mathcal{D}^k_1(\theta^\alpha)$.

Key observation: Sampling without replacement close to with replacement
Rates

- Detection at the rate $\theta^\alpha$ would contradict Hypothesis $A_{PC}$.
- SDP, MDP optimal among $\mathcal{P}$-time methods.

\[ \sqrt{\frac{k \log(d)}{n}} \quad \text{no detection} \quad \text{combinatorial method} \quad \sqrt{\frac{k^\alpha}{n}} \quad \text{polynomial-time methods} \]

Take-home message: gap of $\sqrt{k}$ for methods in $\mathcal{P}$-time
Conclusion

In this work

- Theoretical formulation of computational lower bounds.
- Link between Sparse PCA and Planted Clique problem.
- Optimal rates in \( P \)-time for SDP, MDP.

Future work

- Computational Lower bounds for other problems
- Strengthening the complexity assumption, random 3-SAT, NP-hardness...

Reference

Computational Lower Bounds for Sparse PCA.  
[arxiv.org/abs/1304.0828](http://arxiv.org/abs/1304.0828)  
supported by CAREER award DMS-1053987.