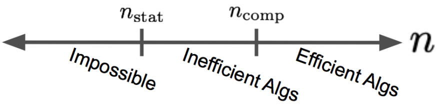
Reductions and the Complexity of Statistical Problems

Matthew Brennan and Guy Bresler

MIT EECS, LIDS & IDSS

Motivation: Statistical-Computational Gaps

A gap in the sample complexity, estimation rate or level of signal needed by efficient vs. inefficient algorithms



Sparse PCA Detection: Decide if n samples are from $H_0 : \mathcal{N}(0, I_d)$ or $H_1 : \mathcal{N}\left(0, I_d + \theta v v^\top\right)$ where v is a k-sparse unit vector

$$n_{\text{stat}} \asymp \frac{k \log d}{\theta^2} \quad \text{and} \quad n_{\text{comp}} \asymp \frac{k^2 \wedge d}{\theta^2}$$

Statistical-Computational Gaps: Approaches

1. **Failure of Classes of Algorithms:** Showing that classes of efficient algorithms fail up to conjectured computational limits e.g. AMP, local search algs, the SOS hierarchy, statistical query algs, etc.

2. **Average-Case Reductions:** Complexity-theoretic approach giving poly-time reductions directly relating different problems and their gaps

Because of complexity-theoretic barriers to basing average-case lower bounds on $\mathsf{P} \neq \mathsf{NP}$, the reductions approach typically is to map *between* statistical problems

Average-Case Reductions to Statistical Problems

Since the first reduction of [BR13] to sparse PCA, there have been many reductions from planted clique (PC) to

- Sparse PCA [BR13, WBS16, GMZ17, BB19]
- Planted dense subgraph [HWX15, BBH18]
- Gaussian biclustering and recovery [MW15, CLR15, CW18, BBH18]
- Incoherence in matrix completion [Che15]
- RIP certification [KZ14, WBP16]
- Testing k-wise independence [AAK^{+}07]
- Universal submatrix detection [BBH19]
- Web of reductions among several problems with sparsity [BBH18]
- Larger web of reductions from variants of PC [BB20]

An Expanded Family of Reductions from Variants of PC

Our Focus: Recent web of reductions from variants of PC [BB20] which breaks out of sparse submatrix plus independent noise matrix structure

(1) Produce negative correlations with inverted Wishart

(2) Dense Bernoulli rotations with $K_{2,t}$

(3) Dense Bernoulli rotations with $K_{3,t}$

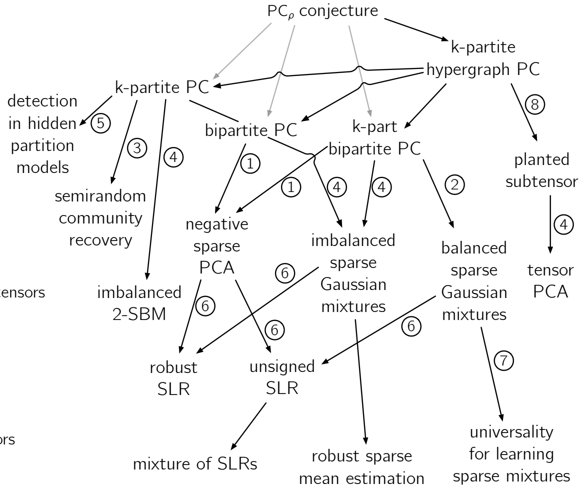
(4) Dense Bernoulli rotations with $K_{r,t}$

(5) Dense Bernoulli rotations with design tensors

(6) LR decomposition and label generation

(7) Symmetric 3-ary rejection kernels

(8) Multi-query reduction completing tensors from hypergraphs

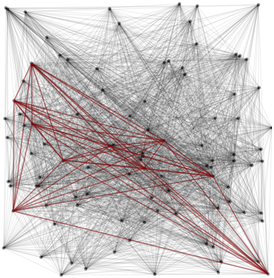


The Planted Clique Conjecture

Planted Clique (PC): Given a graph G with N nodes decide if

$$H_0 : G \sim \mathcal{G}(N, 1/2)$$
$$H_1 : G \sim \mathcal{G}(N, 1/2) \text{ with u.a.r. added } K\text{-clique}$$

PC Conjecture: If $K \ll \sqrt{N}$, then any poly-time algorithm for PC has Type I+II error $1 - o(1)$



k-Part Bipartite PC

Bipartite PC (BPC): Given a bipartite graph G with sides of size M and N vertices, decide if

$$H_0 : G \sim \mathcal{G}_B(M, N, 1/2) \text{ i.e. a u.a.r. } M \times N \text{ bipartite graph}$$
$$H_1 : G \sim \mathcal{G}_B(M, N, 1/2) \text{ with u.a.r. added } K_M \times K_N \text{ biclique}$$

BPC Conjecture: If $K_N \ll \sqrt{N}$ and $K_M \ll \sqrt{M}$ and $M = \text{poly}(N)$, then any poly-time algorithm for BPC has Type I+II error $1 - o(1)$

k-Part Bipartite PC

***k*-Part Bipartite PC (*k*-BPC):** Given a bipartite graph G, decide if

$$H_0 : G \sim \mathcal{G}_B(M, N, 1/2)$$
$$H_1 : G \sim \mathcal{G}_B(M, N, 1/2) \text{ with u.a.r. added } K_M \times K_N \text{ biclique}$$

where the K_N right vertices are chosen u.a.r. to have one one vertex per part of a given partition of $[N]$ into K_N parts of size N/K_N

***k*-BPC Conjecture:** If $K_N \ll \sqrt{N}$ and $K_M \ll \sqrt{M}$ and $M = \text{poly}(N)$, then any poly-time algorithm for BPC has Type I+II error $1 - o(1)$

Remark: The k-BPC conjecture also is implied by a "k-part" extension of the PC conjecture to hypergraphs (k-HPC conjecture)

Robust Sparse Mean Estimation (RSME)

Sparse Mean Estimation: Estimate a k-sparse $\mu \in \mathbb{R}^d$ within ℓ_2 error γ from $X_1, X_2, \dots, X_n \sim_{\text{i.i.d.}} \mathcal{N}(\mu, I_d)$

$$n_{\text{stat}} \asymp n_{\text{comp}} \asymp \frac{k \log d}{\gamma^2}$$

Robust Sparse Mean Estimation (RSME)

Robust Sparse Mean Estimation: Estimate a k-sparse $\mu \in \mathbb{R}^d$ within ℓ_2 error $O(\epsilon)$ from $X_1, X_2, \dots, X_n \sim_{\text{i.i.d.}} \mathcal{N}(\mu, I_d)$, ϵn of which are corrupted

$$n_{\text{stat}} \asymp \frac{k \log d}{\epsilon^2} \quad \text{and} \quad n_{\text{comp}} \lesssim \frac{k^2 \log d}{\epsilon^2}$$

through convex programming and SDPs [Li17, BDLS17]

Robust Sparse Mean Estimation (RSME)

Robust Sparse Mean Estimation: Estimate a k-sparse $\mu \in \mathbb{R}^d$ within ℓ_2 error $O(\epsilon)$ from $X_1, X_2, \dots, X_n \sim_{\text{i.i.d.}} \mathcal{N}(\mu, I_d)$, ϵn of which are corrupted

$$n_{\text{stat}} \asymp \frac{k \log d}{\epsilon^2} \quad \text{and} \quad n_{\text{comp}} \lesssim \frac{k^2 \log d}{\epsilon^2}$$

through convex programming and SDPs [Li17, BDLS17]

Theorem (Lower Bounds for RSME)

The k-BPC conjecture implies that estimating within ℓ_2 error γ requires

$$n_{comp} \gtrsim \frac{k^2 \epsilon^2}{\gamma^4}$$

i.e. any poly-time alg for RSME outputting $\hat{\mu}$ with $\|\mu - \hat{\mu}\|_2 \leq \gamma$ w.p. at least $2/3$ requires this sample complexity.

Proof Plan

Reduction in TV: Construct a poly-time reduction mapping

1. H_0 of k-BPC to within $o(1)$ total variation (TV) of $\mathcal{N}(0, I_d)^{\otimes n}$
2. H_1 of k-BPC to within $o(1)$ TV of n i.i.d. samples from the mixture

$$\left(1 - \frac{\epsilon}{2}\right) \cdot \mathcal{N}(2\gamma\mu, I_d) + \frac{\epsilon}{2} \cdot \mathcal{N}\left(-2\gamma(2\epsilon^{-1} - 1)\mu, I_d\right)$$

where μ is u.a.r. from $\{0, 1/\sqrt{k}\}^d \cap \mathbb{S}^{d-1}$

Why does this imply the Theorem? Composing the reduction with an alg in the theorem has Type I+II error $2/3 + o(1)$ on k-BPC

Reduction Plan:

1. Introduce general technique dense Bernoulli rotations (DBR)
2. Apply DBR locally to subvectors of the k-BPC adjacency matrix
3. Choose the "output means" of DBR carefully to produce (1) and (2)

Total Variation

Simplifying Notation: Throughout this talk, we will abbreviate:

- $d_{\text{TV}}(\mathcal{P}, \mathcal{Q}) \le \epsilon$ as

$$\mathcal{P} \approx_{\epsilon} \mathcal{Q}$$

- $d_{\text{TV}}(\mathcal{A}(X), \mathcal{Q}) \le \epsilon$ where $X \sim \mathcal{P}$ and $\mathcal{A}$ is a (random) function as

$$\mathcal{P} \xrightarrow{\mathcal{A}}_{\epsilon} \mathcal{Q}$$

Data-Processing Inequality: If $\mathcal{P} \xrightarrow{\mathcal{A}_1}_{\epsilon_1} \mathcal{Q}$ and $\mathcal{Q} \xrightarrow{\mathcal{A}_1}_{\epsilon_2} \mathcal{R}$, then

$$\mathcal{P} \xrightarrow{\mathcal{A}_2 \circ \mathcal{A}_1}_{\epsilon_1 + \epsilon_2} \mathcal{R}$$

Union Bounds: If $\mathcal{P}_i \approx_{\epsilon} \mathcal{Q}_i$, then

$$\mathcal{P}_1 \otimes \mathcal{P}_2 \otimes \cdots \otimes \mathcal{P}_n \approx_{n\epsilon} \mathcal{Q}_1 \otimes \mathcal{Q}_2 \otimes \cdots \otimes \mathcal{Q}_n$$

Rejection Kernels: Gaussian Example [MW15, BBH18]

Goal: A *computational change of measure* i.e. an efficient map RK with

$$1 \xrightarrow{\mathrm{RK}}_{o(N^{-3})} \mathcal{N}(\nu, 1) \quad \text{where } \nu = \tilde{\Theta}(1)$$

$$\mathrm{Bern}(1/2) \xrightarrow{\mathrm{RK}}_{o(N^{-3})} \mathcal{N}(0, 1)$$

Idea: If φ_ν is the PDF of $\mathcal{N}(\nu, 1)$, then

- if input $= 1$, sample φ_ν
- if input $= 0$, sample $2 \cdot \varphi_0 - \varphi_\nu$

An Issue: $2 \cdot \varphi_0 - \varphi_\nu$ is not a valid PDF!

Can truncate to x s.t. $2 \cdot \varphi_0(x) \geq \varphi_\nu(x)$ which is the bulk if $\nu \asymp \frac{1}{\sqrt{\log N}}$

Implementation: Can sample $2 \cdot \varphi_0 - \varphi_\nu$ with rejection sampling

Dense Bernoulli Rotations

Parameters: $A \in \mathbb{R}^{n \times m}$ with $\sigma_{\max}(A) \leq 1$ and $\tau \lesssim 1/\sqrt{\log n}$

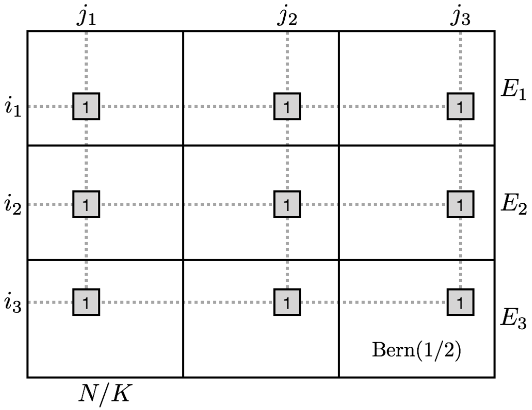
Guarantee: Transforms a vector of n i.i.d. Bern$(1/2)$ with an unknown bit i fixed to 1 into an approx sample from $\mathcal{N}(\tau A_i, I_m)$ in TV, for each $i \leq n$

1. Let $V \in \{0,1\}^n$ be the input vector with an unknown planted ith bit
2. Form V' by applying Gaussian rejection kernels entrywise to V, mapping approx to $\mathcal{N}(\tau \cdot \mathbf{1}_i, I_m)$
3. Sample a vector $U \sim \mathcal{N}(0,1)^{\otimes n}$ and output

$$X = \underbrace{AV'}_{\text{correct mean } A_i} + \underbrace{\left(I_n - AA^\top\right)^{1/2} U}_{\text{cancels induced correlations}}$$

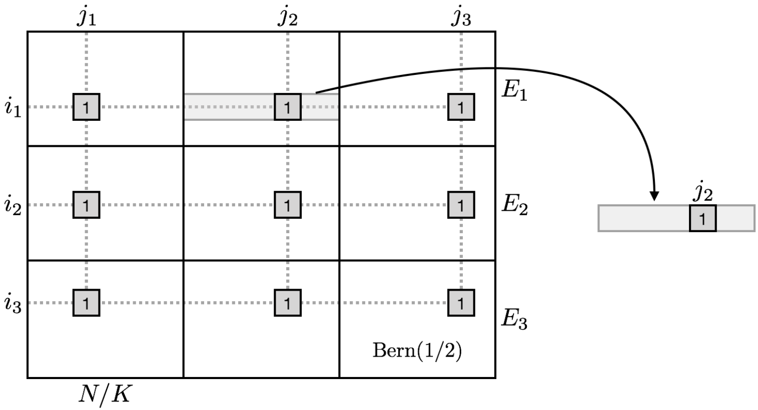
Reduction Sketch

Apply Bernoulli rotations to each row locally in each block of k-BPC



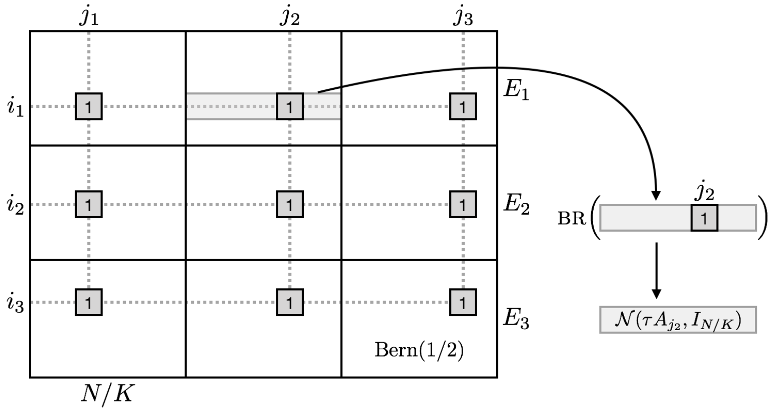
Reduction Sketch

Apply Bernoulli rotations to each row locally in each block of k-BPC



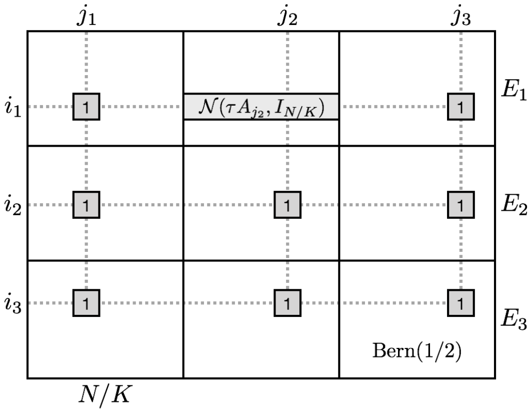
Reduction Sketch

Apply Bernoulli rotations to each row locally in each block of k-BPC



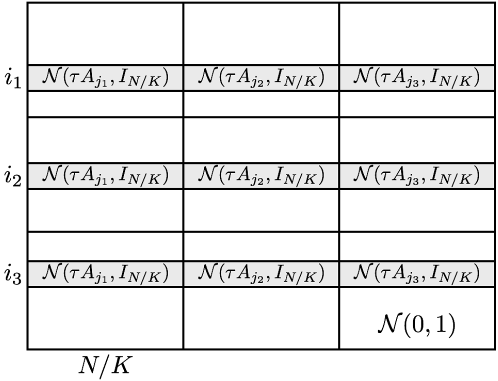
Reduction Sketch

Apply Bernoulli rotations to each row locally in each block of k-BPC



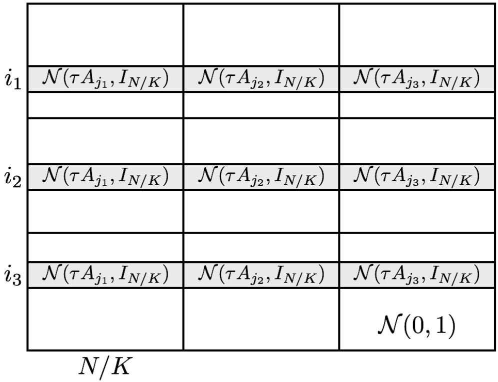
Reduction Sketch

Apply Bernoulli rotations to each row locally in each block of k-BPC



Reduction Sketch

Apply Bernoulli rotations to each row locally in each block of k-BPC



What Remains? Choosing the output mean vectors $A_1, A_2, \dots, A_n$

What do we want from $A_1, A_2, \ldots, A_n$?

- The ith row of output is $\approx_{o(N^{-2})}$ distributed as
  1. $\mathcal{N}(X, I_N)$ where $X = \tau \cdot [A_{j_1}, \ldots, A_{j_k}]$ if $i \in$ left clique
  2. $\mathcal{N}(0, I_N)$ if $i \notin$ left clique
- This is right instance of robust sparse mean estimation as long as X always contains an $(1 - \epsilon/2)$-fraction of its entries equal to $2\gamma/\sqrt{k}$!
- Suppose $A_1, A_2, \ldots, A_m$ have the following properties:
  1. A_j is zero-sum N/K-dimensional unit vector
  2. $A_j \in \{x, y\}^{N/K}$ contains a $(1 - \epsilon/2)$-fraction of a x
  3. $\sigma_{\max}(A) = \Theta(1)$
- Key Question: What lower bound would this show?

$$\frac{2\gamma}{\sqrt{k}} \asymp \tau \cdot x \approx x$$

What do we want from $A_1, A_2, \ldots, A_n$?

- The ith row of output is $\approx_{o(N^{-2})}$ distributed as
  1. $\mathcal{N}(X, I_N)$ where $X = \tau \cdot [A_{j_1}, \ldots, A_{j_k}]$ if $i \in$ left clique
  2. $\mathcal{N}(0, I_N)$ if $i \notin$ left clique
- This is right instance of robust sparse mean estimation as long as X always contains an $(1-\epsilon)$-fraction of its entries equal to $2\gamma/\sqrt{k}$!
- Suppose $A_1, A_2, \ldots, A_m$ have the following properties:
  1. A_j is zero-sum N/K-dimensional unit vector
  2. $A_j \in \{x, y\}^{N/K}$ contains a $(1-\epsilon)$-fraction of a x
  3. $\sigma_{\max}(A) = \Theta(1)$
- Key Question: What lower bound would this show?

$$\frac{2\gamma}{\sqrt{k}} \asymp \tau \cdot x \approx x = \sqrt{\frac{\epsilon}{(1-\epsilon)N/K}} \approx \frac{\sqrt{\epsilon}}{N^{1/4}}$$

What do we want from $A_1, A_2, \ldots, A_n$?

- The ith row of output is $\approx_{o(N^{-2})}$ distributed as
  1. $\mathcal{N}(X, I_N)$ where $X = \tau \cdot [A_{j_1}, \ldots, A_{j_k}]$ if $i \in$ left clique
  2. $\mathcal{N}(0, I_N)$ if $i \notin$ left clique
- This is right instance of robust sparse mean estimation as long as X always contains an $(1-\epsilon)$-fraction of its entries equal to $2\gamma/\sqrt{k}$!
- Suppose $A_1, A_2, \ldots, A_m$ have the following properties:
  1. A_j is zero-sum N/K-dimensional unit vector
  2. $A_j \in \{x, y\}^{N/K}$ contains a $(1-\epsilon)$-fraction of a x
  3. $\sigma_{\max}(A) = \Theta(1)$
- Key Question: What lower bound would this show?

$$\frac{2\gamma}{\sqrt{k}} \approx \frac{\sqrt{\epsilon}}{N^{1/4}} \quad \Longleftrightarrow \quad N \approx \frac{k^2\epsilon^2}{\gamma^4}$$

$\mathbb{F}_r^t$ Design Matrices

Goal: Construct $A_1, A_2, \ldots, A_m$ such that

1. A_j is zero-sum N/K-dimensional unit vector
2. $A_j \in \{x, y\}^{N/K}$ contains a $(1 - \epsilon/2)$-fraction of a x
3. $\sigma_{\max}(A) = \Theta(1)$

Construction: Let r be a prime, $\mathbb{F}_r^t = \{P_1, \ldots, P_{r^t}\}$ and $V_1, \ldots, V_\ell$ be all affine shifts of hyperplanes in $\mathbb{F}_r^t$ where $\ell = \frac{r(r^t - 1)}{r - 1}$

$$A_{ji} = \frac{1}{\sqrt{r^t(r-1)}} \cdot \begin{cases} 1 & \text{if } P_i \not\in V_j \\ 1 - r & \text{if } P_i \in V_j \end{cases}$$

satisfies 1-3 with $\epsilon/2 = 1/r$ and $\sigma_{\max}(A) = \sqrt{1 + (r-1)^{-1}}$

Conclusions

1. We gave an example reduction to robust sparse mean estimation
2. This is one of many reductions beginning with a variant of the PC conjecture and mapping to problems with different hidden structures
3. Many open problems about reduction techniques, reductions to negative SPCA and reductions to sparse generalized linear models