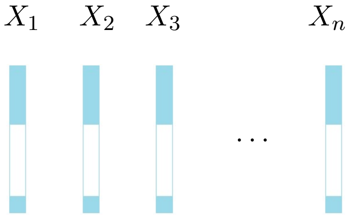
Sample-Measurement Tradeoff in Support Recovery Under a Subgaussian Prior

Lekshmi Ramesh   Chandra R. Murthy   Himanshu Tyagi

Indian Institute of Science
Bangalore

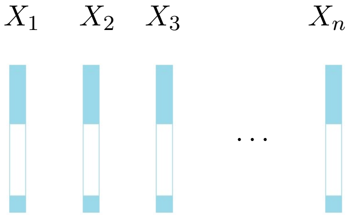
Support recovery from multiple samples

- Samples $X_1, \ldots, X_n$ from $\mathbb{R}^d$ with a common support S of size k

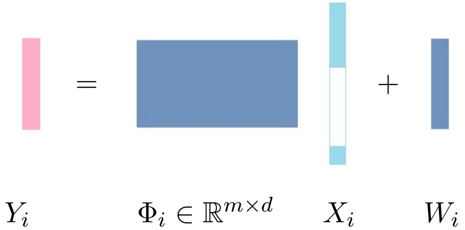


Support recovery from multiple samples

- Samples $X_1, \ldots, X_n$ from $\mathbb{R}^d$ with a common support S of size k



- Observe low dimensional projections of each sample



A generative model for the data

- Define $S_{k,d} = \{u \in \{0,1\}^d \ : \ |\texttt{supp}(u)| = k\}$

A generative model for the data

- Define $S_{k,d} = \{u \in \{0,1\}^d \ : \ |\mathtt{supp}(u)| = k\}$
- A generative model for the data

$$X_i \overset{iid}{\sim} \mathcal{N}(0, K_\lambda), \quad i \in [n],$$

where $\lambda \in S_{k,d}$ and $K_\lambda = \mathtt{diag}(\lambda)$
Note that $\mathtt{supp}(X_i) = \mathtt{supp}(\lambda) \overset{\text{def}}{=} S, \quad \forall\, i \in [n]$ a.s.

A generative model for the data

- Define $S_{k,d} = \{u \in \{0,1\}^d \ : \ |\mathtt{supp}(u)| = k\}$
- A generative model for the data

$$X_i \overset{iid}{\sim} \mathcal{N}(0, K_\lambda), \quad i \in [n],$$

where $\lambda \in S_{k,d}$ and $K_\lambda = \mathtt{diag}(\lambda)$
Note that $\mathtt{supp}(X_i) = \mathtt{supp}(\lambda) \overset{\text{def}}{=} S, \quad \forall\, i \in [n]$ a.s.

- A single sample is sufficient to recover S if $m = O(k \log(d-k))$

A generative model for the data

- Define $S_{k,d} = \{u \in \{0,1\}^d \; : \; |\texttt{supp}(u)| = k\}$
- A generative model for the data

$$X_i \overset{iid}{\sim} \mathcal{N}(0, K_\lambda), \quad i \in [n],$$

  where $\lambda \in S_{k,d}$ and $K_\lambda = \texttt{diag}(\lambda)$
  Note that $\texttt{supp}(X_i) = \texttt{supp}(\lambda) \overset{\text{def}}{=} S, \quad \forall\, i \in [n]$ a.s.

- A single sample is sufficient to recover S if $m = O(k \log(d-k))$
- What happens when $m < k$? Can we still recover the support if we have access to multiple samples?

Sample complexity of support recovery

- Assumptions

Sample complexity of support recovery

- Assumptions
  - The entries of X_i, $i \in [n]$, are independent and zero-mean with $\mathbb{E}\left[X_{ij}^2\right] = \lambda_j$ for $\lambda \in S_{k,d}$ and $X_{ij} \sim \mathrm{subG}(c\lambda_j)$, where c is an absolute constant

Sample complexity of support recovery

- Assumptions
  - The entries of X_i, $i \in [n]$, are independent and zero-mean with $\mathbb{E}\left[X_{ij}^2\right] = \lambda_j$ for $\lambda \in S_{k,d}$ and $X_{ij} \sim \mathrm{subG}(c\lambda_j)$, where c is an absolute constant
  - The entries of Φ_i, $i \in [n]$, are independent and zero-mean with $\mathbb{E}\left[\Phi_i(u,v)^2\right] = 1/m$, $\Phi_i(u,v) \sim \mathrm{subG}(c'/m)$, and $\mathbb{E}\left[\Phi_i(u,v)^4\right] = c''/m^2$, where c' and c'' are absolute constants

Sample complexity of support recovery

- Assumptions
  - The entries of X_i, $i \in [n]$, are independent and zero-mean with $\mathbb{E}\left[X_{ij}^2\right] = \lambda_j$ for $\lambda \in S_{k,d}$ and $X_{ij} \sim \mathrm{subG}(c\lambda_j)$, where c is an absolute constant
  - The entries of Φ_i, $i \in [n]$, are independent and zero-mean with $\mathbb{E}\left[\Phi_i(u,v)^2\right] = 1/m$, $\Phi_i(u,v) \sim \mathrm{subG}(c'/m)$, and $\mathbb{E}\left[\Phi_i(u,v)^4\right] = c''/m^2$, where c' and c'' are absolute constants
- For $m < k/2$ and $k < d-1$, the sample complexity of support recovery under the asuumptions above is

$$n^*(m,k,d) = \frac{k^2}{m^2}\log(k(d-k)).$$

Previous work

- Work in the multiple sample setting has mostly focused on the $m > k$ case

[1] Piya Pal and P. P. Vaidyanathan. "Pushing the Limits of Sparse Support Recovery Using Correlation Information". In: *IEEE Trans. on Sig. Proc.* 63.3 (2015), pp. 711–726.

[2] M. Azizyan, A. Krishnamurthy, and A. Singh. "Extreme Compressive Sampling for Covariance Estimation". In: 64.12 (Dec. 2018), pp. 7613–7635.

[3] T. Tony Cai and Anru Zhang. "ROP: Matrix recovery via rank-one projections". In: *The Annal of Statistics* 43.1 (2015), pp. 102–138.

Previous work

- Work in the multiple sample setting has mostly focused on the $m > k$ case
- Recent work[1] showed possibity of operating in $m < k$ regime when multiple samples available, however sample complxity not fully characterized

[1] Piya Pal and P. P. Vaidyanathan. "Pushing the Limits of Sparse Support Recovery Using Correlation Information". In: *IEEE Trans. on Sig. Proc.* 63.3 (2015), pp. 711–726.

[2] M. Azizyan, A. Krishnamurthy, and A. Singh. "Extreme Compressive Sampling for Covariance Estimation". In: 64.12 (Dec. 2018), pp. 7613–7635.

[3] T. Tony Cai and Anru Zhang. "ROP: Matrix recovery via rank-one projections". In: *The Annal of Statistics* 43.1 (2015), pp. 102–138.

Previous work

- Work in the multiple sample setting has mostly focused on the $m > k$ case
- Recent work[1] showed possibity of operating in $m < k$ regime when multiple samples available, however sample complxity not fully characterized
- Also connections to literature on covariance estimation[2,3]

---

[1] Piya Pal and P. P. Vaidyanathan. "Pushing the Limits of Sparse Support Recovery Using Correlation Information". In: *IEEE Trans. on Sig. Proc.* 63.3 (2015), pp. 711–726.

[2] M. Azizyan, A. Krishnamurthy, and A. Singh. "Extreme Compressive Sampling for Covariance Estimation". In: 64.12 (Dec. 2018), pp. 7613–7635.

[3] T. Tony Cai and Anru Zhang. "ROP: Matrix recovery via rank-one projections". In: *The Annal of Statistics* 43.1 (2015), pp. 102–138.

The estimator

- Based on estimating variance along each coordinate

The estimator

- Based on estimating variance along each coordinate
    - Compute for all $i \in [d]$

$$\tilde{\lambda}_i = \frac{1}{n}\sum_{j=1}^{n}(\Phi_{ji}^\top Y_j)^2,$$

where Φ_{ji} is the ith column of Φ_j

The estimator

- Based on estimating variance along each coordinate
  - Compute for all $i \in [d]$

$$\tilde{\lambda}_i = \frac{1}{n} \sum_{j=1}^{n} (\Phi_{ji}^\top Y_j)^2,$$

    where Φ_{ji} is the ith column of Φ_j

  - Sort the entries of $\tilde{\lambda}$: $\tilde{\lambda}_{(1)} \geq \cdots \geq \tilde{\lambda}_{(d)}$

    Output $\tilde{S} = \{(1), \ldots, (k)\}$

Performance of the estimator

- Hard to analyze $\tilde{S}$, we analyze the folowing threshold-based estimator:
$$\hat{S} = \texttt{supp}\left(\mathbb{1}_{\{\tilde{\lambda} > \tau\}}\right),$$
since $\mathrm{P}(\tilde{S} \neq S) \leq \mathrm{P}(\hat{S} \neq S)$

Performance of the estimator

- Hard to analyze $\tilde{S}$, we analyze the folowing threshold-based estimator:

$$\hat{S} = \texttt{supp}\Big(\mathbb{1}_{\{\tilde{\lambda} > \tau\}}\Big),$$

  since $\mathrm{P}(\tilde{S} \neq S) \leq \mathrm{P}(\hat{S} \neq S)$

- Here τ is an appropriately chosen threshold (and can depend on S)

Performance of the estimator

- Hard to analyze $\tilde{S}$, we analyze the folowing threshold-based estimator:

$$\hat{S} = \texttt{supp}\Big(\mathbb{1}_{\{\tilde{\lambda} > \tau\}}\Big),$$

since $\mathrm{P}(\tilde{S} \neq S) \leq \mathrm{P}(\hat{S} \neq S)$

- Here τ is an appropriately chosen threshold (and can depend on S)

- Probability of error

$$P_e \stackrel{\text{def}}{=} \mathrm{P}(\hat{S} \neq S) \leq \sum_{i \in S} \mathrm{P}(\tilde{\lambda} < \tau) + \sum_{i \in S^c} \mathrm{P}(\tilde{\lambda} \geq \tau)$$

Analysis based on tail bounds for $\tilde{\lambda}$ based on subgaussian/subexponential concentration inequalities

Performance of the estimator

- Key step in the analysis: P_e can be made small if the following separation condition holds for all $(i, i') \in S \times S^c$

$$\mu_i - \nu_i \geq \mu_{i'} + \nu_{i'}$$

where
$\mu_i,\ \mu_{i'}$: mean of the estimator conditioned on Φ_1^n
$\nu_i,\ \nu_{i'}$: deviation terms arising from concentration bounds

Performance of the estimator

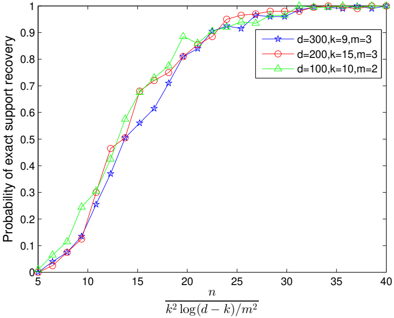
- Key step in the analysis: P_e can be made small if the following separation condition holds for all $(i, i') \in S \times S^c$

$$\mu_i - \nu_i \geq \mu_{i'} + \nu_{i'}$$

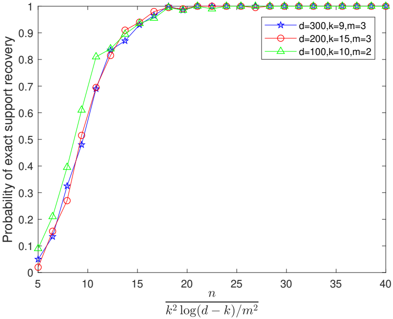
  where
  $\mu_i,\ \mu_{i'}$: mean of the estimator conditioned on Φ_1^n
  $\nu_i,\ \nu_{i'}$: deviation terms arising from concentration bounds

- Condition fails to hold for $n = 1$, recovery requires $n > 1$ when $m < k$

Phase transition



(a) Gaussian prior



(b) Rademacher prior

Figure 1: Phase transition of the closed-form estimator.

Lower bound

- Construct a set of $k(d-k)$ supports that are difficult to distinguish and use Fano's method

Lower bound

- Construct a set of $k(d-k)$ supports that are difficult to distinguish and use Fano's method
- Construction: Start with $S_0 = \{1, \ldots, k\}$, create a set of supports G by removing $i \in \{1, \ldots, k\}$ and adding $j \in \{k+1, \ldots, d\}$

Lower bound

- Construct a set of $k(d-k)$ supports that are difficult to distinguish and use Fano's method

- Construction: Start with $S_0 = \{1, \ldots, k\}$, create a set of supports G by removing $i \in \{1, \ldots, k\}$ and adding $j \in \{k+1, \ldots, d\}$

- Let U be uniformly distributed over G. By Fano's inequality

$$\begin{aligned} \mathrm{P}(\hat{S} \neq U) &\geq 1 - \frac{I(Y_1^n; U) + 1}{\log(k(d-k))} \\ &\geq 1 - \frac{\max_{S \in G} D(\mathrm{P}_{Y^n|S} \| \mathrm{P}_{Y^n|S_0}) + 1}{\log(k(d-k))} \end{aligned}$$

Lower bound

- For fixed Φ_1^n, the divergence term depends on the eigenvalues $\{a_i\}$ and $\{b_i\}$ of $\Phi_S \Phi_S^\top$ and $\Phi_{S_0} \Phi_{S_0}^\top$

$$D\big(\mathrm{P}_{Y^n|S,\Phi^n} \| \mathrm{P}_{Y^n|S_0,\Phi^n}\big) \leq \frac{n}{2} \sum_{i=1}^{m} \frac{(a_i - b_i)^2}{a_i b_i}$$

Lower bound

- For fixed Φ_1^n, the divergence term depends on the eigenvalues $\{a_i\}$ and $\{b_i\}$ of $\Phi_S \Phi_S^\top$ and $\Phi_{S_0} \Phi_{S_0}^\top$

$$D\big(\mathrm{P}_{Y^n|S,\Phi^n} \| \mathrm{P}_{Y^n|S_0,\Phi^n}\big) \leq \frac{n}{2} \sum_{i=1}^{m} \frac{(a_i - b_i)^2}{a_i b_i}$$

- Using results on the spectra of Gaussian random matrices and a few other tools, it can be shown that

$$D\big(\mathrm{P}_{Y^n|S} \| \mathrm{P}_{Y^n|S_0}\big) \leq \frac{cnm^2}{k^2(1 - m/k)^4}$$

This gives the required scaling of n.

Summary and extensions

- Multiple samples make support recovery possible with very few measurements per sample ($m < k$)

Summary and extensions

- Multiple samples make support recovery possible with very few measurements per sample ($m < k$)
- Can look at more general settings with nonbinary variance and changing support

Summary and extensions

- Multiple samples make support recovery possible with very few measurements per sample ($m < k$)
- Can look at more general settings with nonbinary variance and changing support
- Current estimator requires knowledge of an upper bound on k

Summary and extensions

- Multiple samples make support recovery possible with very few measurements per sample ($m < k$)
- Can look at more general settings with nonbinary variance and changing support
- Current estimator requires knowledge of an upper bound on k
  - Can first try to estimate k using observations, and then use our estimator

Summary and extensions

- Multiple samples make support recovery possible with very few measurements per sample ($m < k$)
- Can look at more general settings with nonbinary variance and changing support
- Current estimator requires knowledge of an upper bound on k
  - Can first try to estimate k using observations, and then use our estimator
  - Other estimators with similar sample complexity