Multiple Support Recovery Using Very Few Measurements Per Sample

Lekshmi Ramesh [1]

Joint work with: Chandra R. Murthy [2]   Himanshu Tyagi [2]

[1]Columbia University

[2]Indian Institue of Science, Bangalore

Algorithmic Structures for Uncoordinated Communications
and Statistical Inference in Exceedingly Large Spaces
Banff, March 2024

Outline

- Multiple support recovery
  - Setup and background
  - The case of very few measurements
- A spectral algorithm
- Sample complexity upper bound
- Discussion and Open problems

Problem Setting

The multiple support recovery problem

- Samples $X_1, \ldots, X_n$ in $\mathbb{R}^d$, each sample has k nonzero entries

The multiple support recovery problem

- Samples $X_1, \ldots, X_n$ in $\mathbb{R}^d$, each sample has k nonzero entries
- For each X_i, the location of the nonzero entries is called the support of X_i, denoted $\texttt{supp}(X_i)$

The multiple support recovery problem

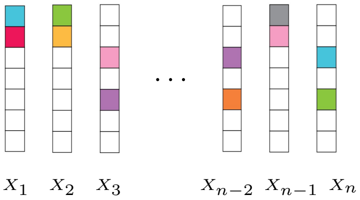
- Samples $X_1, \ldots, X_n$ in $\mathbb{R}^d$, each sample has k nonzero entries
- For each X_i, the location of the nonzero entries is called the support of X_i, denoted $\mathtt{supp}(X_i)$
- The support of each sample is drawn from a small set of allowed supports

$$\mathtt{supp}(X_i) \in \{\mathcal{S}_1, \ldots, \mathcal{S}_\ell\}$$

where $\mathcal{S}_i$ are subsets of $[d]$ of cardinality k

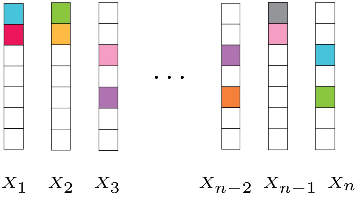
The multiple support recovery problem

- Example: Two allowed supports $\mathcal{S}_1 = \{1, 2\}$ and $\mathcal{S}_2 = \{3, 5\}$



The multiple support recovery problem

- Example: Two allowed supports $\mathcal{S}_1 = \{1, 2\}$ and $\mathcal{S}_2 = \{3, 5\}$



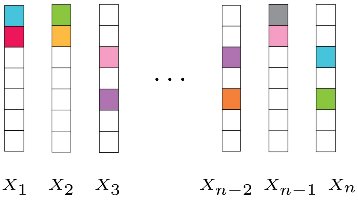
- We only observe low-dimensional linear projections

$$Y_i = \Phi_i X_i,\ i \in [n],$$

where $\Phi_i \in \mathbb{R}^{m \times d}$ with $m < d$

The multiple support recovery problem

- Example: Two allowed supports $\mathcal{S}_1 = \{1, 2\}$ and $\mathcal{S}_2 = \{3, 5\}$



- We only observe low-dimensional linear projections

$$Y_i = \Phi_i X_i, \; i \in [n],$$

where $\Phi_i \in \mathbb{R}^{m \times d}$ with $m < d$

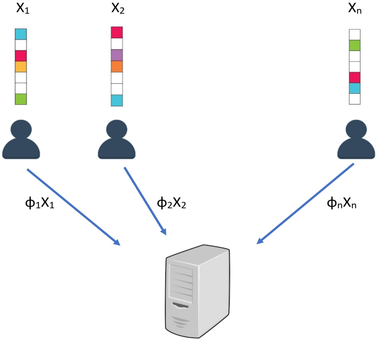
- Given $\{\Phi_i, Y_i\}_{i=1}^{n}$, recover $\{\mathcal{S}_1, \ldots, \mathcal{S}_\ell\}$

Application to feature clustering

- Coordinate clustering/feature clustering problems can be understood using our formulation

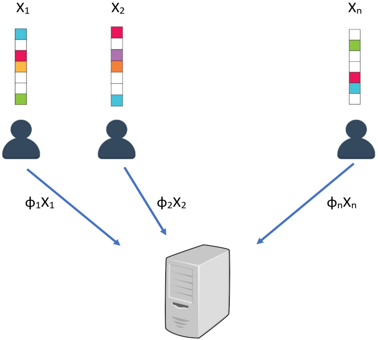
Application to feature clustering

- Coordinate clustering/feature clustering problems can be understood using our formulation



Application to feature clustering

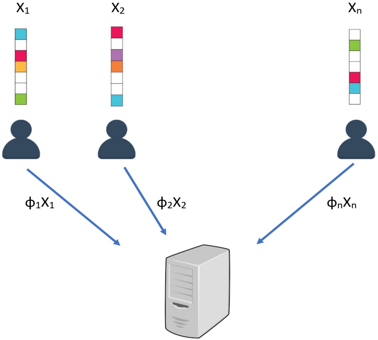
- Coordinate clustering/feature clustering problems can be understood using our formulation



- Distributed user profiling

Application to feature clustering

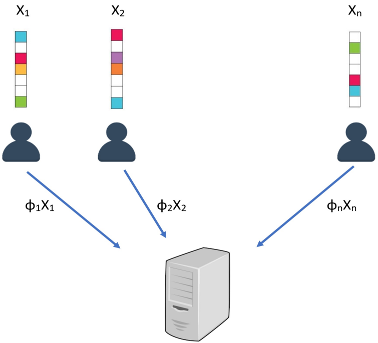
- Coordinate clustering/feature clustering problems can be understood using our formulation



- Distributed user profiling
  - Users have profile vectors indicating ratings/preferences for features (e.g. type of website visited)

Application to feature clustering

- Coordinate clustering/feature clustering problems can be understood using our formulation



- Distributed user profiling
  - Users have profile vectors indicating ratings/preferences for features (e.g. type of website visited)
  - For a given population, center wants to find features that occur together

Background

Related work

Mixed linear models

- Used to model heterogeneous data: population can be divided into groups, linear model within each group

Related work

Mixed linear models

- Used to model heterogeneous data: population can be divided into groups, linear model within each group
- Similar setting:
  Mixture of linear regressions [De Veaux 1989; Chen 2013; Chaganty 2013; Yin 2019; Li 2020]
  Phase retrieval [Candès 2013; Netrapalli 2014; Eldar 2014; Jaganathan 2015]
  Learning mixtures of low-rank models [Chen 2021]

Related work

Mixed linear models

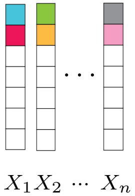
- Used to model heterogeneous data: population can be divided into groups, linear model within each group
- Similar setting:
  Mixture of linear regressions [De Veaux 1989; Chen 2013; Chaganty 2013; Yin 2019; Li 2020]
  Phase retrieval [Candès 2013; Netrapalli 2014; Eldar 2014; Jaganathan 2015]
  Learning mixtures of low-rank models [Chen 2021]
- Usually focus either on worst-case formulation or on recovering data vectors

Related work

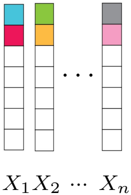
Mixed linear models

- Used to model heterogeneous data: population can be divided into groups, linear model within each group
- Similar setting:
  Mixture of linear regressions [De Veaux 1989; Chen 2013; Chaganty 2013; Yin 2019; Li 2020]
  Phase retrieval [Candès 2013; Netrapalli 2014; Eldar 2014; Jaganathan 2015]
  Learning mixtures of low-rank models [Chen 2021]
- Usually focus either on worst-case formulation or on recovering data vectors
- Current algorithms require at least roughly k measurements per sample – can this be reduced?

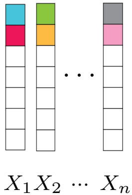
The common support case ($\ell = 1$)

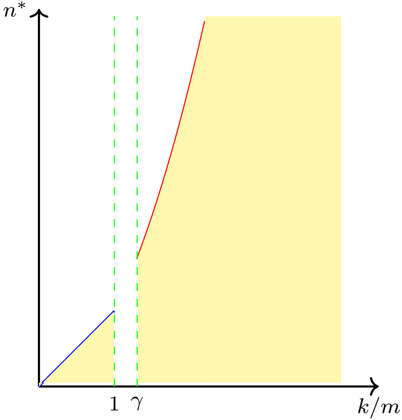


The common support case ($\ell = 1$)

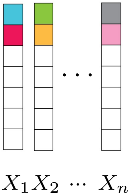


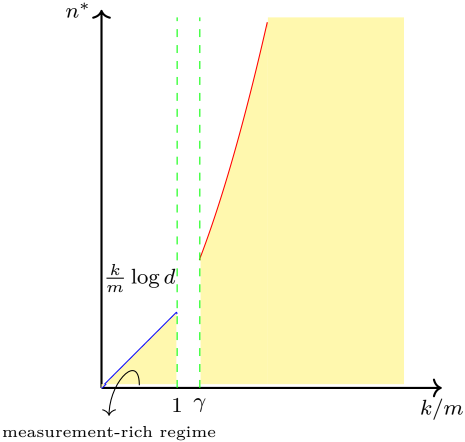
The common support case ($\ell = 1$)



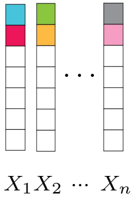


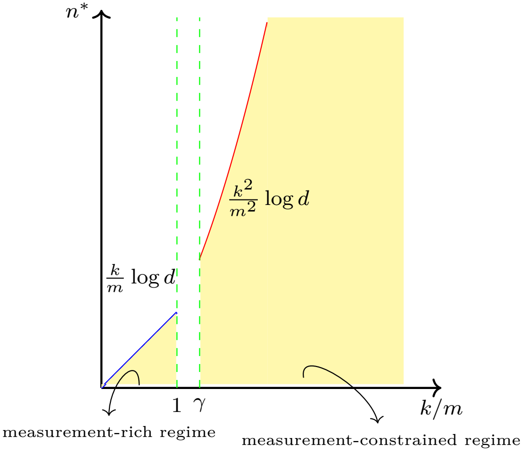
The common support case ($\ell = 1$)



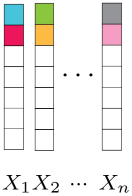


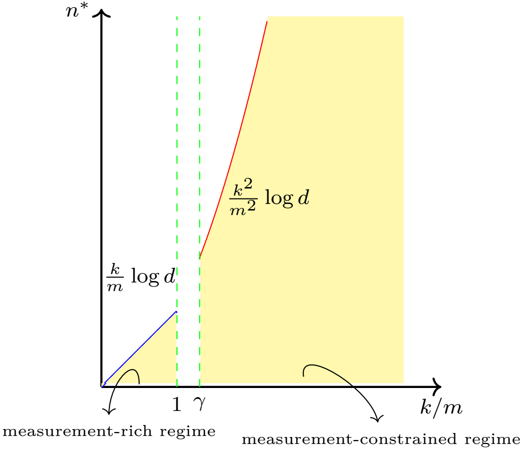
The common support case ($\ell = 1$)





The common support case ($\ell = 1$)





- Can operate with $m < k$ measurements per sample unlike conventional algorithms, but require more samples

---

L. Ramesh, C. R. Murthy, and H. Tyagi. "Phase Transitions for Support Recovery from Gaussian Linear Measurements", ISIT 2021

Multiple supports ($\ell > 1$)

- Two sets of unknowns: labels associating measurements to supports, and the underlying supports

Multiple supports ($\ell > 1$)

- Two sets of unknowns: labels associating measurements to supports, and the underlying supports
- Given knowledge of the labels, can group together measurements corresponding to same support and recover each support

Multiple supports ($\ell > 1$)

- Two sets of unknowns: labels associating measurements to supports, and the underlying supports
- Given knowledge of the labels, can group together measurements corresponding to same support and recover each support
- Alternatively, if we have roughly k measurements per sample, can use standard algorithms on each sample and recover each support

Multiple supports ($\ell > 1$)

- Two sets of unknowns: labels associating measurements to supports, and the underlying supports
- Given knowledge of the labels, can group together measurements corresponding to same support and recover each support
- Alternatively, if we have roughly k measurements per sample, can use standard algorithms on each sample and recover each support
- When labeling not known and $m < k$?

Multiple supports ($\ell > 1$)

- Is it possible to recover $(\mathcal{S}_1, \ldots, \mathcal{S}_\ell)$ when labeling not known and $m < k$?

Multiple supports ($\ell > 1$)

- Is it possible to recover $(\mathcal{S}_1, \ldots, \mathcal{S}_\ell)$ when labeling not known and $m < k$?

  Yes, we will see an efficient algorithm for multiple support recovery from very few measurements

Multiple supports ($\ell > 1$)

- Is it possible to recover $(\mathcal{S}_1, \ldots, \mathcal{S}_\ell)$ when labeling not known and $m < k$?

  Yes, we will see an efficient algorithm for multiple support recovery from very few measurements

- How many samples are required for recovery?

Multiple supports ($\ell > 1$)

- Is it possible to recover $(\mathcal{S}_1, \ldots, \mathcal{S}_\ell)$ when labeling not known and $m < k$?

  Yes, we will see an efficient algorithm for multiple support recovery from very few measurements

- How many samples are required for recovery?

  We can approximately recover all the supports using roughly $(k\ell/m)^4$ samples

A Spectral Algorithm

From fully observed data

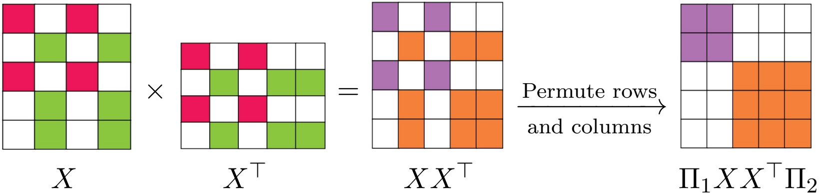
- We will assume supports to be disjoint, and a constant fraction of samples corresponding to each support

From fully observed data

- We will assume supports to be disjoint, and a constant fraction of samples corresponding to each support
- The sample covariance matrix has a block structure (under permutation of its rows and columns)

From fully observed data

- We will assume supports to be disjoint, and a constant fraction of samples corresponding to each support
- The sample covariance matrix has a block structure (under permutation of its rows and columns)

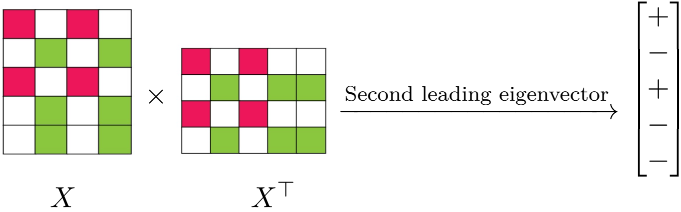


From fully observed data

- Unknown permutation can be found using eigenvectors of sample covariance matrix, after normalizing each row by its row sum

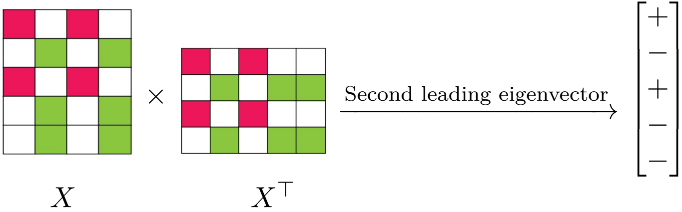
From fully observed data

- Unknown permutation can be found using eigenvectors of sample covariance matrix, after normalizing each row by its row sum



From fully observed data

- Unknown permutation can be found using eigenvectors of sample covariance matrix, after normalizing each row by its row sum



- When there are ℓ blocks (supports), use the top-ℓ eigenvectors and a nearest neighbor step

[F. McSherry, 2001]; [Ng et al., 2002]; [Newman, 2006].

From linear measurements

- Can a similar approach be used when we only have linear measurements of the samples?

From linear measurements

- Can a similar approach be used when we only have linear measurements of the samples?

  Yes, we will run spectral clustering using coordinate-wise variance estimate $a_i = (\Phi_i^\top Y_i) \circ (\Phi_i^\top Y_i)$

From linear measurements

- Can a similar approach be used when we only have linear measurements of the samples?

  Yes, we will run spectral clustering using coordinate-wise variance estimate $a_i = (\Phi_i^\top Y_i) \circ (\Phi_i^\top Y_i)$

$$a_{iu} a_{iv} = \begin{cases} \text{large, when } (u, v) \text{ in same support,} \\ \text{small, when } (u, v) \text{ in different supports.} \end{cases}$$

From linear measurements

- Can a similar approach be used when we only have linear measurements of the samples?

  Yes, we will run spectral clustering using coordinate-wise variance estimate $a_i = (\Phi_i^\top Y_i) \circ (\Phi_i^\top Y_i)$

$$a_{iu}a_{iv} = \begin{cases} \text{large, when } (u,v) \text{ in same support,} \\ \text{small, when } (u,v) \text{ in different supports.} \end{cases}$$

- Performing spectral clustering on $\sum_{i=1}^n a_i a_i^\top \in \mathbb{R}^{d\times d}$ is computationally intensive

From linear measurements

- Can a similar approach be used when we only have linear measurements of the samples?

  Yes, we will run spectral clustering using coordinate-wise variance estimate $a_i = (\Phi_i^\top Y_i) \circ (\Phi_i^\top Y_i)$

$$a_{iu} a_{iv} = \begin{cases} \text{large, when } (u, v) \text{ in same support,} \\ \text{small, when } (u, v) \text{ in different supports.} \end{cases}$$

- Performing spectral clustering on $\sum_{i=1}^{n} a_i a_i^\top \in \mathbb{R}^{d \times d}$ is computationally intensive

  We will first estimate the union $\cup_{i=1}^{\ell} \mathcal{S}_i$, and run spectral clustering restricted to the union

The algorithm

- Step 1. Compute variance estimates $a_i = (\Phi_i^\top Y_i) \circ (\Phi_i^\top Y_i) \in \mathbb{R}^d$ for each $i \in [n]$

The algorithm

- Step 1. Compute variance estimates $a_i = (\Phi_i^\top Y_i) \circ (\Phi_i^\top Y_i) \in \mathbb{R}^d$ for each $i \in [n]$
- Step 2. Compute sample mean $(1/n) \sum_{i=1}^n a_i$, top $k\ell$ coordinates give estimate $\hat{\mathcal{S}}_{\mathrm{un}}$ for the union

The algorithm

- Step 1. Compute variance estimates $a_i = (\Phi_i^\top Y_i) \circ (\Phi_i^\top Y_i) \in \mathbb{R}^d$ for each $i \in [n]$
- Step 2. Compute sample mean $(1/n) \sum_{i=1}^n a_i$, top $k\ell$ coordinates give estimate $\hat{\mathcal{S}}_{\text{un}}$ for the union
- Step 3. Perform spectral clustering on sample covariance matrix $T = (1/n) \sum_{i=1}^n (a_i)_{\hat{\mathcal{S}}_{\text{un}}} (a_i)_{\hat{\mathcal{S}}_{\text{un}}}^\top$ to partition the union into ℓ supports

The algorithm

- Step 1. Compute variance estimates $a_i = (\Phi_i^\top Y_i) \circ (\Phi_i^\top Y_i) \in \mathbb{R}^d$ for each $i \in [n]$
- Step 2. Compute sample mean $(1/n) \sum_{i=1}^n a_i$, top $k\ell$ coordinates give estimate $\hat{\mathcal{S}}_{\text{un}}$ for the union
- Step 3. Perform spectral clustering on sample covariance matrix $T = (1/n) \sum_{i=1}^n (a_i)_{\hat{\mathcal{S}}_{\text{un}}} (a_i)_{\hat{\mathcal{S}}_{\text{un}}}^\top$ to partition the union into ℓ supports
- Second order statistic recovers the union, fourth order statistic required to partition the union

Analyzing the Algorithm

Sample complexity of multiple support recovery

- Our analysis assumes a subgaussian generative model for the samples and measurement matrices

Sample complexity of multiple support recovery

- Our analysis assumes a subgaussian generative model for the samples and measurement matrices
- We seek approximate recovery of the supports up to permutation of the support labels

Sample complexity of multiple support recovery

- Our analysis assumes a subgaussian generative model for the samples and measurement matrices
- We seek approximate recovery of the supports up to permutation of the support labels
- The smallest n for which an estimator satisfying the recovery criterion exists is the sample complexity n^*

Sample complexity of multiple support recovery

- Our analysis assumes a subgaussian generative model for the samples and measurement matrices
- We seek approximate recovery of the supports up to permutation of the support labels
- The smallest n for which an estimator satisfying the recovery criterion exists is the sample complexity n^*

Sample complexity of multiple support recovery

- Our analysis assumes a subgaussian generative model for the samples and measurement matrices
- We seek approximate recovery of the supports up to permutation of the support labels
- The smallest n for which an estimator satisfying the recovery criterion exists is the sample complexity n^*

Theorem

Let $(\log k\ell)^2 \leq m < k$. Then,

$$n^* = \tilde{O}\left(\left(\frac{k\ell}{m}\right)^4\right).$$

Proof Sketch

Analyzing the two steps

- Recovery of the union. Can recover the union with roughly $k^2\ell^2 \log(d/m)$ samples[1]

[1] L. Ramesh, C. R. Murthy, and H. Tyagi "Sample-Measurement Tradeoff for Support Recovery under a Subgaussian Prior", ISIT 2019.

Analyzing the two steps

- Recovery of the union. Can recover the union with roughly $k^2\ell^2 \log(d/m)$ samples[1]
- Recovering each support. The expected value of the clustering matrix T has a block structure (under permutation of rows and columns)

$$\mathbb{E}[T] = \begin{bmatrix} \mu_0 & \mu^s & \mu^d & \mu^d \\ \mu^s & \mu_0 & \mu^d & \mu^d \\ \mu^d & \mu^d & \mu_0 & \mu^s \\ \mu^d & \mu^d & \mu^s & \mu_0 \end{bmatrix} \begin{matrix} \left.\begin{matrix} \\ \\ \end{matrix}\right\} \mathcal{S}_1 \\ \left.\begin{matrix} \\ \\ \end{matrix}\right\} \mathcal{S}_2 \end{matrix}$$

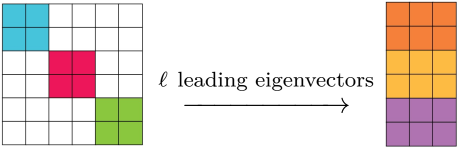
[1] L. Ramesh, C. R. Murthy, and H. Tyagi "Sample-Measurement Tradeoff for Support Recovery under a Subgaussian Prior", ISIT 2019.

Properties of the eigenvectors

- Eigenvectors of the expected clustering matrix $\mathbb{E}[T]$ has a repeating rows structure

Properties of the eigenvectors

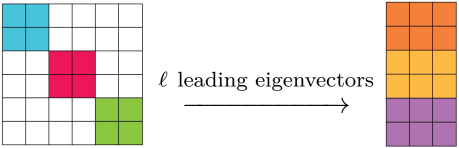
- Eigenvectors of the expected clustering matrix $\mathbb{E}[T]$ has a repeating rows structure



$\mathbb{E}[T] \in \mathbb{R}^{k\ell \times k\ell}$ $\quad$ $V \in \mathbb{R}^{k\ell \times \ell}$

Properties of the eigenvectors

- Eigenvectors of the expected clustering matrix $\mathbb{E}[T]$ has a repeating rows structure



$\mathbb{E}[T] \in \mathbb{R}^{k\ell \times k\ell}$ $\qquad$ $V \in \mathbb{R}^{k\ell \times \ell}$

- A nearest neighbor step can then partition the union estimate into ℓ subsets

Extending to sample-based statistic

- Can show that the sample version of the clustering matrix T suffices when we have roughly $k^4\ell^4/m^4$ samples

Extending to sample-based statistic

- Can show that the sample version of the clustering matrix T suffices when we have roughly $k^4\ell^4/m^4$ samples

- Eigenvectors of T and $\mathbb{E}[T]$ are close when $\|T - \mathbb{E}[T]\|_{op}$ is small (Davis-Kahan theorem)

Extending to sample-based statistic

- Can show that the sample version of the clustering matrix T suffices when we have roughly $k^4\ell^4/m^4$ samples

- Eigenvectors of T and $\mathbb{E}[T]$ are close when $\|T - \mathbb{E}[T]\|_{op}$ is small (Davis-Kahan theorem)

- Showing $\|T - \mathbb{E}[T]\|_{op}$ is small:

Extending to sample-based statistic

- Can show that the sample version of the clustering matrix T suffices when we have roughly $k^4\ell^4/m^4$ samples

- Eigenvectors of T and $\mathbb{E}[T]$ are close when $\|T - \mathbb{E}[T]\|_{op}$ is small (Davis-Kahan theorem)

- Showing $\|T - \mathbb{E}[T]\|_{op}$ is small:
  - T is a sum of rank-one matrices with heavy-tailed entries

Extending to sample-based statistic

- Can show that the sample version of the clustering matrix T suffices when we have roughly $k^4\ell^4/m^4$ samples

- Eigenvectors of T and $\mathbb{E}[T]$ are close when $\|T - \mathbb{E}[T]\|_{op}$ is small (Davis-Kahan theorem)

- Showing $\|T - \mathbb{E}[T]\|_{op}$ is small:
  - T is a sum of rank-one matrices with heavy-tailed entries
  - Standard methods difficult to adapt to this setting

Extending to sample-based statistic

- Can show that the sample version of the clustering matrix T suffices when we have roughly $k^4\ell^4/m^4$ samples
- Eigenvectors of T and $\mathbb{E}[T]$ are close when $\|T - \mathbb{E}[T]\|_{op}$ is small (Davis-Kahan theorem)
- Showing $\|T - \mathbb{E}[T]\|_{op}$ is small:
  - T is a sum of rank-one matrices with heavy-tailed entries
  - Standard methods difficult to adapt to this setting
  - We use a result by Rudelson[2] to bound $\|T - \mathbb{E}[T]\|_{op}$ under relaxed assumptions on moments

[2] M. Rudelson. Random vectors in the isotropic position, JFA 1999.

Conclusion

- We developed a spectral algorithm that can recover multiple supports from linear observations

Conclusion

- We developed a spectral algorithm that can recover multiple supports from linear observations
- Works with fewer than k measurements per sample, requires roughly $k^4\ell^4/m^4$ samples

Conclusion

- We developed a spectral algorithm that can recover multiple supports from linear observations
- Works with fewer than k measurements per sample, requires roughly $k^4\ell^4/m^4$ samples
- Open questions: overlapping supports; imbalanced groups; lower bounds

Conclusion

- We developed a spectral algorithm that can recover multiple supports from linear observations
- Works with fewer than k measurements per sample, requires roughly $k^4\ell^4/m^4$ samples
- Open questions: overlapping supports; imbalanced groups; lower bounds

Thank you

For more details: "Multiple Support Recovery Using Very Few Measurements Per Sample", IEEE Transactions on Signal Processing, May 2022 and ISIT 2021.