Asymptotics and Statistical Inference in High-Dimensional Low-Rank Matrix Models

Joshua Agterberg

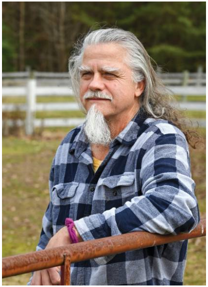
JOHNS HOPKINS MATHEMATICAL INSTITUTE for DATA SCIENCE

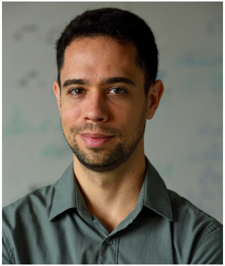
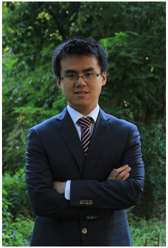
JOHNS HOPKINS WHITING SCHOOL of ENGINEERING

February 2023


Take 2

Collaborators




Outline

1. High-Dimensional Low-Rank Matrix Models
2. Asymptotics
3. Statistical Inference
4. Contributions

Outline

High-Dimensional Models

$$\underbrace{x_i}_{observation} = \underbrace{\mu}_{signal} + \underbrace{\sigma\varepsilon_i}_{noise}; \qquad i = 1, \ldots, n;$$

$$\mu \in \mathbb{R}^d \qquad \sigma > 0$$

ε_i is mean-zero isotropic noise.

High-Dimensional Models

$$\underbrace{x_i}_{observation} = \underbrace{\mu}_{signal} + \underbrace{\sigma\varepsilon_i}_{noise}; \qquad i = 1, \ldots, n;$$

$$\mu \in \mathbb{R}^d \qquad \sigma > 0$$

ε_i is mean-zero isotropic noise.

- Easy to check:

$$\mathbb{E}\|\bar{x} - \mu\|^2 = \sigma^2 \frac{d}{n}.$$

High-Dimensional Models

$$\underbrace{x_i}_{observation} = \underbrace{\mu}_{signal} + \underbrace{\sigma\varepsilon_i}_{noise}; \qquad i = 1, \ldots, n;$$

$$\mu \in \mathbb{R}^d \qquad \sigma > 0$$

ε_i is mean-zero isotropic noise.

- Easy to check:

$$\mathbb{E}\|\bar{x} - \mu\|^2 = \sigma^2 \frac{d}{n}.$$

- Can show there are matching *lower bounds* over all possible estimators of μ

High-Dimensional Models

$$\underbrace{x_i}_{observation} = \underbrace{\mu}_{signal} + \underbrace{\sigma\varepsilon_i}_{noise}; \qquad i = 1, \ldots, n;$$

$$\mu \in \mathbb{R}^d \qquad \sigma > 0$$

ε_i is mean-zero isotropic noise.

- Easy to check:

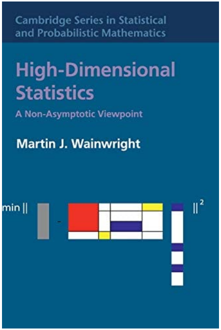
$$\mathbb{E}\|\bar{x} - \mu\|^2 = \sigma^2 \frac{d}{n}.$$

- Can show there are matching *lower bounds* over all possible estimators of μ

Problem

With fixed σ, we need to have $\frac{d}{n} \to 0$ for consistency.

High-Dimensional Models



Given these "no free lunch" guarantees, what can help us in the high-dimensional setting? Essentially, our only hope is that the data is endowed with some form of *low-dimensional structure,* one which makes it simpler than the high-dimensional view might suggest. Much of high-dimensional statistics involves constructing models of high-dimensional phenomena that involve some implicit form of low-dimensional structure, and then studying the statistical and computational gains afforded by exploiting this structure. In order to illustrate, let us

Low-Dimensional Structure via Sparsity

Assume that μ only has s nonzero entries, with $s \ll d$.

Low-Dimensional Structure via Sparsity

Assume that μ only has s nonzero entries, with $s \ll d$.

Low-Dimensional Structure via Sparsity

Assume that μ only has s nonzero entries, with $s \ll d$.

- Can show there is an estimator $\bar{x}^{(\mathrm{sparsity})}$ satisfying

$$\mathbb{E}\|\bar{x}^{(\mathrm{sparsity})} - \mu\|^2 \lesssim \sigma^2 \frac{s \log(ed/s)}{n}$$

Low-Dimensional Structure via Sparsity

Assume that μ only has s nonzero entries, with $s \ll d$.

- Can show there is an estimator $\bar{x}^{(\text{sparsity})}$ satisfying

$$\mathbb{E}\|\bar{x}^{(\text{sparsity})} - \mu\|^2 \lesssim \sigma^2 \frac{s \log(ed/s)}{n}$$

- Similar lower bound over all s-sparse vectors μ.

Low-Dimensional Structure via Sparsity

Assume that μ only has s nonzero entries, with $s \ll d$.

- Can show there is an estimator $\bar{x}^{(\mathrm{sparsity})}$ satisfying

$$\mathbb{E}\|\bar{x}^{(\mathrm{sparsity})} - \mu\|^2 \lesssim \sigma^2 \frac{s \log(ed/s)}{n}$$

- Similar lower bound over all s-sparse vectors μ.

Key Takeaway

Imposing *low-dimensional structural assumptions* can maintain consistency in high dimensions.

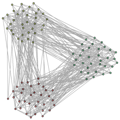
High-Dimensional Matrix Models

Data doesn't have to be Euclidean!

High-Dimensional Matrix Models

Data doesn't have to be Euclidean!

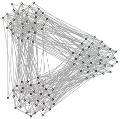
- Network data



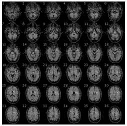
High-Dimensional Matrix Models

Data doesn't have to be Euclidean!

- Network data



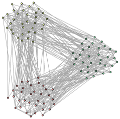
- Brain image data



High-Dimensional Matrix Models

Data doesn't have to be Euclidean!

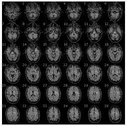
- Network data



- Matrix time series

$$
\begin{array}{l} \\ Revenue_t \\ Assets_t \\ Dividends\ per\ share_t \\ \vdots \end{array}
\begin{array}{c} \begin{array}{cccc} Apple & Twitter & Tesla & \cdots \end{array} \\ \begin{pmatrix} X_{11}^{(t)} & X_{12}^{(t)} & X_{13}^{(t)} & \cdots \\ X_{21}^{(t)} & X_{22}^{(t)} & X_{23}^{(t)} & \cdots \\ X_{31}^{(t)} & X_{32}^{(t)} & X_{33}^{(t)} & \cdots \\ \cdots & \cdots & \cdots & \ddots \end{pmatrix} \end{array}
$$

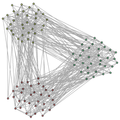
- Brain image data



High-Dimensional Matrix Models

Data doesn't have to be Euclidean!

- Network data

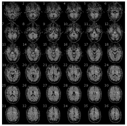


- Matrix time series

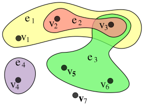
$$
\begin{array}{l}
\\
Revenue_t \\
Assets_t \\
Dividends\ per\ share_t \\
\vdots
\end{array}
\quad
\begin{array}{cccc}
Apple & Twitter & Tesla & \cdots \\
\end{array}
$$

$$
\begin{pmatrix}
X_{11}^{(t)} & X_{12}^{(t)} & X_{13}^{(t)} & \cdots \\
X_{21}^{(t)} & X_{22}^{(t)} & X_{23}^{(t)} & \cdots \\
X_{31}^{(t)} & X_{32}^{(t)} & X_{33}^{(t)} & \cdots \\
\cdots & \cdots & \cdots & \ddots
\end{pmatrix}
$$

- Brain image data



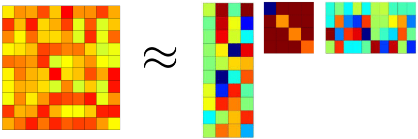
- Hypergraph data



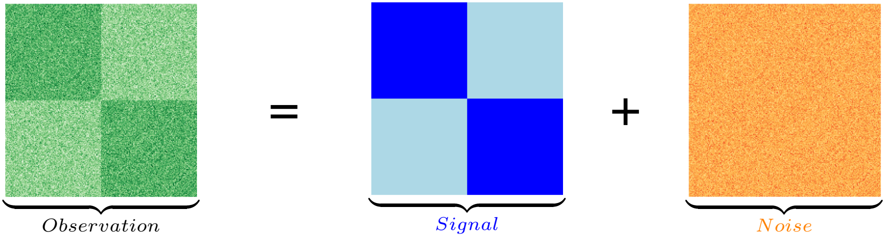
High-Dimensional Low-Rank Matrix Models

Ansatz

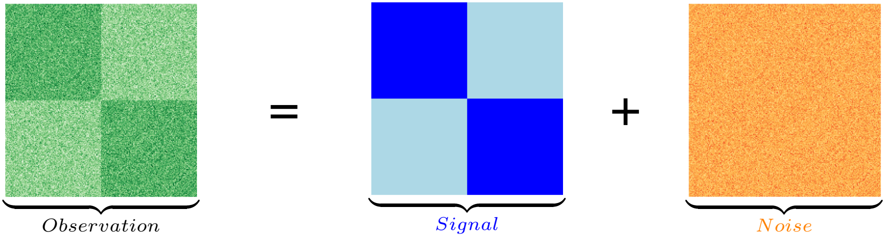
By imposing *low-dimensional structural assumptions* (low-rankedess), we can maintain consistency *and perform valid inference* in high dimensions



A Canonical Model



A Canonical Model



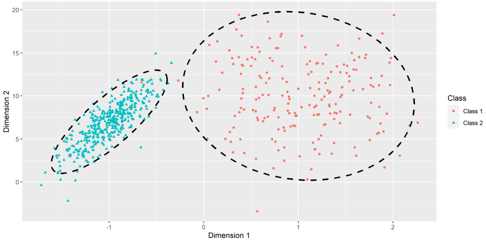
Canonical Matrix Denoising Model

$$\underbrace{\widehat{\mathbf{S}}}_{observation} = \underbrace{\mathbf{S}}_{signal} + \underbrace{\mathbf{N}}_{noise} \in \mathbb{R}^{n\times n};$$

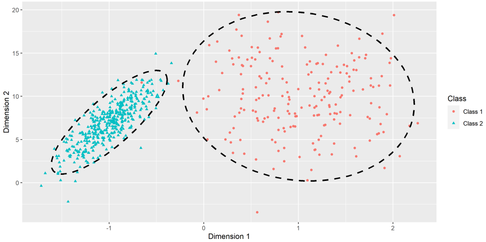
$\mathbf{S}$ is low-rank and symmetric;

$\mathbf{N}$ satisfies $\mathbb{E}\mathbf{N}_{ij}^2 = \sigma^2$.

High-Dimensional Mixture Model



High-Dimensional Mixture Model



High-Dimensional Mixture Model

$$X_i = \mu_{z(i)} + Y_i \in \mathbb{R}^d, \qquad 1 \le i \le n;$$

$z(i)$ is the membership of the i'th observation;

μ_k are the K different means;

$\mathbb{E}Y_i = 0.$

High-Dimensional Mixture Model

High-Dimensional Mixture Model

$$X_i = \mu_{z(i)} + Y_i \qquad \in \mathbb{R}^d, \qquad 1 \leq i \leq n;$$

$z(i)$ is the membership of the i'th observation;

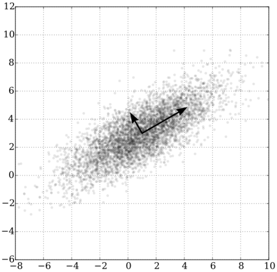
μ_k are the K different means;

$\mathbb{E}Y_i = 0.$

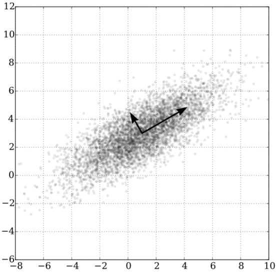
Corresponding Low-Rank Matrix Model

$$\mathbf{X} = \underbrace{\begin{pmatrix} \mu_{z(1)} \\ \vdots \\ \mu_{z(n)} \end{pmatrix}}_{Low-Rank\ Matrix} + \mathbf{Y}$$

Principal Component Analysis



Principal Component Analysis



Principal Component Analysis

Estimate leading eigenvectors of covariance:

$$X_i = \mathbf{\Sigma}^{1/2} Y_i \qquad \in \mathbb{R}^d, \qquad 1 \le i \le n;$$

$$\mathbb{E} Y_i = 0; \qquad \mathbb{E} Y_i Y_i^\top = \mathbf{I}_d$$

Principal Component Analysis

Principal Component Analysis

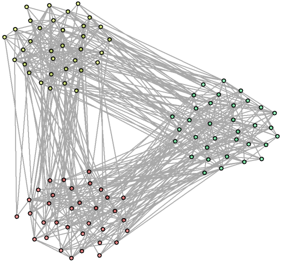
Estimate leading eigenvectors of covariance:

$$X_i = \mathbf{\Sigma}^{1/2} Y_i \qquad \in \mathbb{R}^d, \qquad 1 \le i \le n;$$
$$\mathbb{E} Y_i = 0; \qquad \mathbb{E} Y_i Y_i^\top = \mathbf{I}_d$$

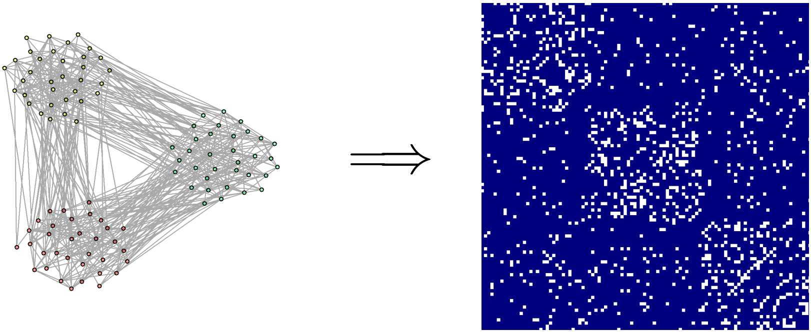
Corresponding Low-Rank Matrix Model

$$\widehat{\mathbf{\Sigma}} = \frac{1}{n} \sum_{i=1}^{n} X_i X_i^\top \qquad = \underbrace{\mathbf{\Sigma}_0}_{Low-rank\ matrix} + \underbrace{\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}_0}_{\text{“}noise\text{”}}$$

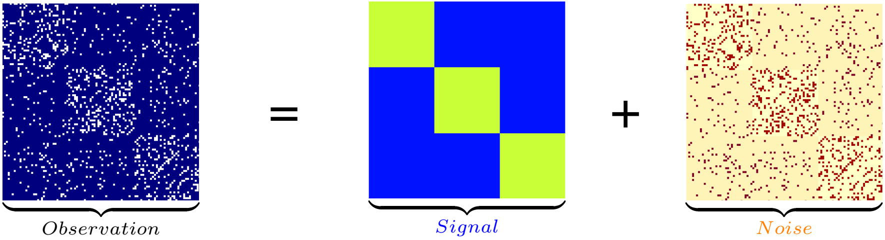
Network Analysis



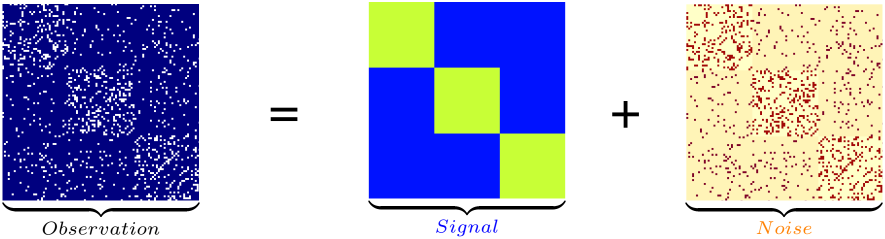
Network Analysis



Network Analysis



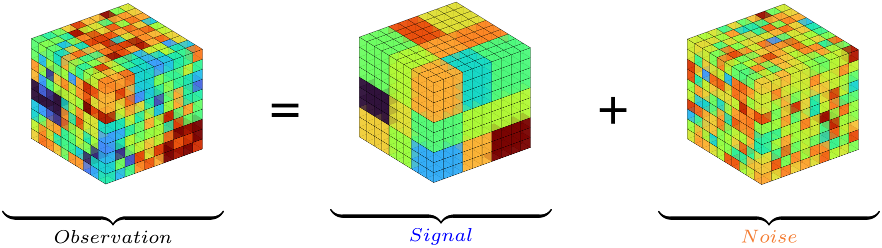
Network Analysis



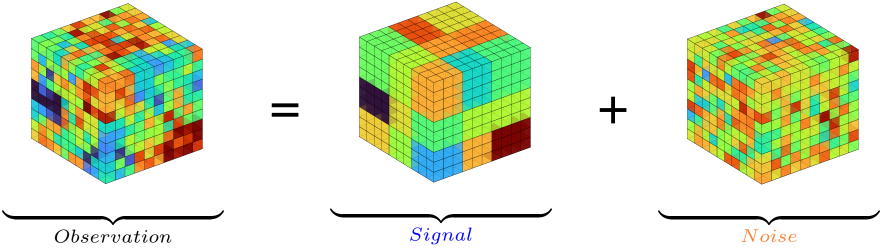
Statistical Network Analysis

$$\mathbf{A} = \underbrace{\mathbf{P}}_{\textit{Probability Matrix}} + \underbrace{\mathbf{E}}_{\textit{Bernoulli noise}} \in \{0,1\}^{n\times n};$$

Tensor Data Analysis



Tensor Data Analysis



Tensor Data Analysis

$$\underbrace{\widehat{\mathcal{T}}}_{observation} = \underbrace{\mathcal{T}}_{signal} + \underbrace{\mathcal{Z}}_{noise};$$

$\mathcal{T}$ is Tucker low-rank;

$$\mathbb{E}\mathcal{Z}_{ijk} = 0; \qquad \mathbb{E}\mathcal{Z}_{ijk}^2 \le \sigma^2$$

Eigenvector/Singular Vector Estimation

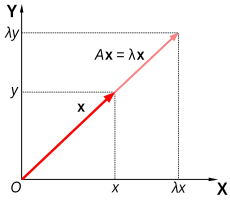
Fundamental Observation

In all of the previous models, it is the *eigenvectors, singular vectors*, or *related quantities* that contain important information.

Eigenvector/Singular Vector Estimation

Fundamental Observation

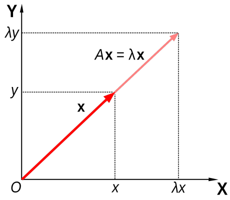
In all of the previous models, it is the *eigenvectors, singular vectors*, or *related quantities* that contain important information.



Eigenvector/Singular Vector Estimation

Fundamental Observation

In all of the previous models, it is the *eigenvectors, singular vectors*, or *related quantities* that contain important information.

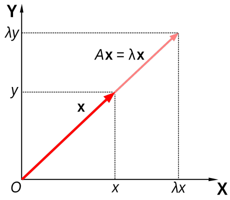


This dissertation:

Eigenvector/Singular Vector Estimation

Fundamental Observation

In all of the previous models, it is the *eigenvectors, singular vectors*, or *related quantities* that contain important information.



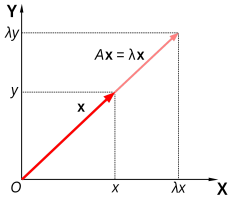
This dissertation:

- Study eigenvector/singular vector *rates of convergence* via fine-grained bounds;

Eigenvector/Singular Vector Estimation

Fundamental Observation

In all of the previous models, it is the *eigenvectors, singular vectors*, or *related quantities* that contain important information.



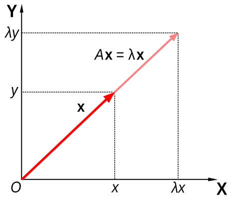
This dissertation:

- Study eigenvector/singular vector *rates of convergence* via fine-grained bounds;
- Establish *asymptotic expansions* and *limit theorems*;

Eigenvector/Singular Vector Estimation

Fundamental Observation

In all of the previous models, it is the *eigenvectors, singular vectors*, or *related quantities* that contain important information.

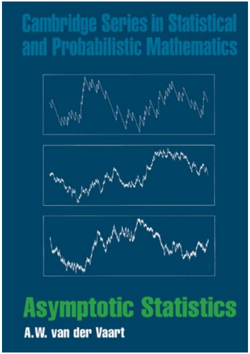


This dissertation:

- Study eigenvector/singular vector *rates of convergence* via fine-grained bounds;
- Establish *asymptotic expansions* and *limit theorems*;
- Design *principled inference procedures*.

Outline

Asymptotics



Why asymptotic statistics? The use of asymptotic approximations is two-fold. First, they enable us to find approximate tests and confidence regions. Second, approximations can be used theoretically to study the quality (efficiency) of statistical procedures.

Consistency

5.2 CONSISTENCY

5.2.1 Plug-In Estimates and MLEs in Exponential Family Models

Suppose that we have a sample $X_1, \ldots, X_n$ from $P_{\boldsymbol{\theta}}$ where $\boldsymbol{\theta} \in \Theta$ and want to estimate a real or vector $q(\boldsymbol{\theta})$. The least we can ask of our estimate $\widehat{q}_n(X_1, \ldots, X_n)$ is that as $n \to \infty$, $\widehat{q}_n \xrightarrow{P_{\boldsymbol{\theta}}} q(\boldsymbol{\theta})$ for all $\boldsymbol{\theta}$. That is, in accordance with (A.14.1) and (B.7.1), for all $\boldsymbol{\theta} \in \Theta$, $\epsilon > 0$,

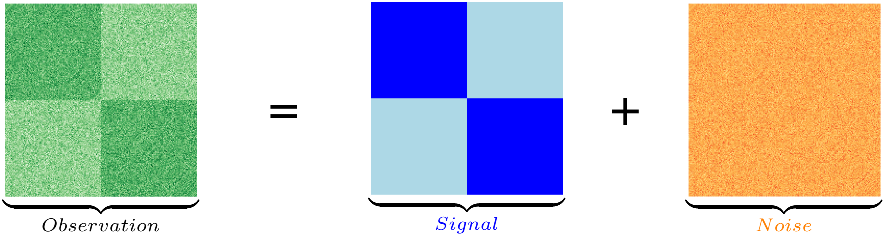
$$P_{\boldsymbol{\theta}}[|\widehat{q}_n(X_1, \ldots, X_n) - q(\boldsymbol{\theta})| \geq \epsilon] \to 0. \tag{5.2.1}$$

where $|\cdot|$ denotes Euclidean distance. A stronger requirement is

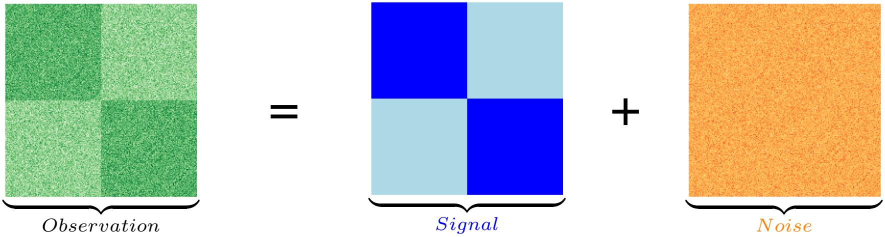
$$\sup_{\boldsymbol{\theta}} \left\{ P_{\boldsymbol{\theta}} \left[|\widehat{q}_n(X_1, \ldots, X_n) - q(\boldsymbol{\theta})| \geq \epsilon \right] : \boldsymbol{\theta} \in \Theta \right\} \to 0. \tag{5.2.2}$$

Bounds $b(n, \epsilon)$ for $\sup_{\boldsymbol{\theta}} P_{\boldsymbol{\theta}} [|\widehat{q}_n - q(\boldsymbol{\theta})| \geq \epsilon]$ that yield (5.2.2) are preferable and we shall indicate some of qualitative interest when we can. But, with all the caveats of Section 5.1, (5.2.1), which is called *consistency* of $\widehat{q}_n$ and can be thought of as 0'th order asymptotics, remains central to all asymptotic theory. The stronger statement (5.2.2) is called *uniform consistency*. If Θ is replaced by a smaller set K, we talk of uniform consistency over K.

Matrix Denoising Consistency



Matrix Denoising Consistency



Canonical Matrix Denoising Model

$$\underbrace{\widehat{\mathbf{S}}}_{observation} = \underbrace{\mathbf{S}}_{signal} + \underbrace{\mathbf{N}}_{noise} \in \mathbb{R}^{n\times n};$$

$\mathbf{S}$ is low-rank and symmetric;

$\mathbf{N}$ satisfies $\mathbb{E}\mathbf{N}_{ij}^2 = \sigma^2$.

Matrix Denoising Consistency

$$\underbrace{\widehat{\mathbf{S}}}_{observation} = \underbrace{\mathbf{S}}_{signal} + \underbrace{\mathbf{N}}_{noise} \in \mathbb{R}^{n\times n};$$

$$\mathbf{S} = \mathbf{U}\Lambda\mathbf{U}^\top;$$

$$\mathbf{N} \text{ satisfies } \mathbb{E}\mathbf{N}_{ij}^2 = \sigma^2.$$

Matrix Denoising Consistency

$$\underbrace{\widehat{\mathbf{S}}}_{observation} = \underbrace{\mathbf{S}}_{signal} + \underbrace{\mathbf{N}}_{noise} \in \mathbb{R}^{n\times n};$$

$$\mathbf{S} = \mathbf{U}\Lambda\mathbf{U}^\top;$$

$$\mathbf{N} \text{ satisfies } \mathbb{E}\mathbf{N}_{ij}^2 = \sigma^2.$$

- $\mathbf{U} \in \mathbb{O}(n, r)$ (resp. $\widehat{\mathbf{U}}$) is matrix of leading eigenvectors of $\mathbf{S}$ (resp. $\widehat{\mathbf{S}}$)

Matrix Denoising Consistency

$$\underbrace{\widehat{\mathbf{S}}}_{observation} = \underbrace{\mathbf{S}}_{signal} + \underbrace{\mathbf{N}}_{noise} \in \mathbb{R}^{n\times n};$$

$$\mathbf{S} = \mathbf{U}\Lambda\mathbf{U}^\top;$$

$$\mathbf{N} \text{ satisfies } \mathbb{E}\mathbf{N}_{ij}^2 = \sigma^2.$$

- $\mathbf{U} \in \mathbb{O}(n, r)$ (resp. $\widehat{\mathbf{U}}$) is matrix of leading eigenvectors of $\mathbf{S}$ (resp. $\widehat{\mathbf{S}}$)
- Λ is a diagonal $r \times r$ matrix of singular values $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$

Matrix Denoising Consistency

$$\underbrace{\widehat{\mathbf{S}}}_{observation} = \underbrace{\mathbf{S}}_{signal} + \underbrace{\mathbf{N}}_{noise} \in \mathbb{R}^{n\times n};$$
$$\mathbf{S} = \mathbf{U}\Lambda\mathbf{U}^\top;$$
$$\mathbf{N} \text{ satisfies } \mathbb{E}\mathbf{N}_{ij}^2 = \sigma^2.$$

- $\mathbf{U} \in \mathbb{O}(n, r)$ (resp. $\widehat{\mathbf{U}}$) is matrix of leading eigenvectors of $\mathbf{S}$ (resp. $\widehat{\mathbf{S}}$)
- Λ is a diagonal $r \times r$ matrix of singular values $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$
- Define the signal-to-noise ratio:

$$\text{SNR} := \frac{\lambda_r}{\sigma}.$$

Matrix Denoising Consistency

$$\underbrace{\widehat{\mathbf{S}}}_{observation} = \underbrace{\mathbf{S}}_{signal} + \underbrace{\mathbf{N}}_{noise} \in \mathbb{R}^{n\times n};$$

$$\mathbf{S} = \mathbf{U}\Lambda\mathbf{U}^\top;$$

$$\mathbf{N} \text{ satisfies } \mathbb{E}\mathbf{N}_{ij}^2 = \sigma^2.$$

- Eigenvector matrix $\widehat{\mathbf{U}}$ is *highly nonlinear function* of noise $\mathbf{N}$

Matrix Denoising Consistency

$$\underbrace{\widehat{\mathbf{S}}}_{observation} = \underbrace{\mathbf{S}}_{signal} + \underbrace{\mathbf{N}}_{noise} \in \mathbb{R}^{n\times n};$$

$$\mathbf{S} = \mathbf{U}\Lambda\mathbf{U}^\top;$$

$$\mathbf{N} \text{ satisfies } \mathbb{E}\mathbf{N}_{ij}^2 = \sigma^2.$$

- Eigenvector matrix $\widehat{\mathbf{U}}$ is *highly nonlinear function* of noise $\mathbf{N}$
- Davis-Kahan Theorem:

$$\inf_{\mathcal{O}:\mathcal{O}\mathcal{O}^\top=\mathbf{I}_r} \|\widehat{\mathbf{U}}\mathcal{O} - \mathbf{U}\| \le C\frac{\sqrt{n}}{\mathrm{SNR}}.$$

Matrix Denoising Consistency

$$\underbrace{\widehat{\mathbf{S}}}_{observation} = \underbrace{\mathbf{S}}_{signal} + \underbrace{\mathbf{N}}_{noise} \in \mathbb{R}^{n \times n};$$

$$\mathbf{S} = \mathbf{U}\Lambda\mathbf{U}^\top;$$

$$\mathbf{N} \text{ satisfies } \mathbb{E}\mathbf{N}_{ij}^2 = \sigma^2.$$

- Eigenvector matrix $\widehat{\mathbf{U}}$ is *highly nonlinear function* of noise $\mathbf{N}$
- Davis-Kahan Theorem:

$$\inf_{\mathcal{O}:\mathcal{O}\mathcal{O}^\top = \mathbf{I}_r} \|\widehat{\mathbf{U}}\mathcal{O} - \mathbf{U}\| \leq C\frac{\sqrt{n}}{\mathrm{SNR}}.$$

- Means that $\mathrm{SNR} \gg \sqrt{n}$ is *sufficient* for consistency!

Matrix Denoising Consistency

$$\underbrace{\widehat{\mathbf{S}}}_{observation} = \underbrace{\mathbf{S}}_{signal} + \underbrace{\mathbf{N}}_{noise} \in \mathbb{R}^{n\times n};$$

$$\mathbf{S} = \mathbf{U}\Lambda\mathbf{U}^\top;$$

$\mathbf{N}$ satisfies $\mathbb{E}\mathbf{N}_{ij}^2 = \sigma^2$.

- Eigenvector matrix $\widehat{\mathbf{U}}$ is *highly nonlinear function* of noise $\mathbf{N}$
- Davis-Kahan Theorem:

$$\inf_{\mathcal{O}:\mathcal{O}\mathcal{O}^\top=\mathbf{I}_r} \|\widehat{\mathbf{U}}\mathcal{O} - \mathbf{U}\| \le C\frac{\sqrt{n}}{\mathrm{SNR}}.$$

- Means that $\mathrm{SNR} \gg \sqrt{n}$ is *sufficient* for consistency!
- Can show it is also *necessary* under Gaussian noise.

Matrix Denoising Consistency: Finer Grained Bounds

Problem

Results of the form $\|\widehat{\mathbf{U}}\mathcal{O} - \mathbf{U}\| \to 0$ are often *too weak* to to guarantee anything besides consistency.

Matrix Denoising Consistency: Finer Grained Bounds

Problem

Results of the form $\|\widehat{\mathbf{U}}\mathcal{O} - \mathbf{U}\| \to 0$ are often *too weak* to to guarantee anything besides consistency.

A First Solution

Study the $\ell_{2,\infty}$ perturbation of the form:

$$\|\widehat{\mathbf{U}}\mathcal{O} - \mathbf{U}\|_{2,\infty} := \max_{1 \leq i \leq n} \|(\widehat{\mathbf{U}}\mathcal{O} - \mathbf{U})_{i\cdot}\| \leq ???$$

Matrix Denoising Consistency: Finer Grained Bounds

Problem

Results of the form $\|\widehat{\mathbf{U}}\mathcal{O} - \mathbf{U}\| \to 0$ are often *too weak* to to guarantee anything besides consistency.

A First Solution

Study the $\ell_{2,\infty}$ perturbation of the form:

$$\|\widehat{\mathbf{U}}\mathcal{O} - \mathbf{U}\|_{2,\infty} := \max_{1 \leq i \leq n} \|(\widehat{\mathbf{U}}\mathcal{O} - \mathbf{U})_{i\cdot}\| \leq ???$$

- Can be used to study *implicit regularization* and nonconvex optimization

Matrix Denoising Consistency: Finer Grained Bounds

Problem

Results of the form $\|\widehat{\mathbf{U}}\mathcal{O} - \mathbf{U}\| \to 0$ are often *too weak* to to guarantee anything besides consistency.

A First Solution

Study the $\ell_{2,\infty}$ perturbation of the form:

$$\|\widehat{\mathbf{U}}\mathcal{O} - \mathbf{U}\|_{2,\infty} := \max_{1 \leq i \leq n} \|(\widehat{\mathbf{U}}\mathcal{O} - \mathbf{U})_{i\cdot}\| \leq ???$$

- Can be used to study *implicit regularization* and nonconvex optimization
- Can be used to obtain *perfect clustering* in mixture models

Matrix Denoising Consistency: Finer Grained Bounds

Problem

Results of the form $\|\widehat{\mathbf{U}}\mathcal{O} - \mathbf{U}\| \to 0$ are often *too weak* to to guarantee anything besides consistency.

A First Solution

Study the $\ell_{2,\infty}$ perturbation of the form:

$$\|\widehat{\mathbf{U}}\mathcal{O} - \mathbf{U}\|_{2,\infty} := \max_{1 \leq i \leq n} \|(\widehat{\mathbf{U}}\mathcal{O} - \mathbf{U})_{i\cdot}\| \leq ???$$

- Can be used to study *implicit regularization* and nonconvex optimization
- Can be used to obtain *perfect clustering* in mixture models
- Often a precursor to limit theory

Incoherence Parameter

Definition

The *incoherence parameter* of a symmetric rank r matrix $\mathbf{S}$ with eigendecomposition $\mathbf{U}\Lambda\mathbf{U}^\top$ is defined as the smallest number μ_0 such that

$$\|\mathbf{U}\|_{2,\infty} = \max_i \|\mathbf{U}_{i\cdot}\| \leq \mu_0 \sqrt{\frac{r}{n}}.$$

Incoherence Parameter

Definition

The *incoherence parameter* of a symmetric rank r matrix $\mathbf{S}$ with eigendecomposition $\mathbf{U}\Lambda\mathbf{U}^\top$ is defined as the smallest number μ_0 such that

$$\|\mathbf{U}\|_{2,\infty} = \max_i \|\mathbf{U}_{i\cdot}\| \le \mu_0 \sqrt{\frac{r}{n}}.$$

- Examples:

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \text{ versus } \begin{pmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & \ddots & \vdots \\ \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}$$

Incoherence Parameter

Definition

The *incoherence parameter* of a symmetric rank r matrix $\mathbf{S}$ with eigendecomposition $\mathbf{U}\Lambda\mathbf{U}^\top$ is defined as the smallest number μ_0 such that

$$\|\mathbf{U}\|_{2,\infty} = \max_i \|\mathbf{U}_{i\cdot}\| \le \mu_0 \sqrt{\frac{r}{n}}.$$

- Examples:

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \text{ versus } \begin{pmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & \ddots & \vdots \\ \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}$$

Larger values of μ_0 means “more spiky” $\mathbf{S}$!

$\ell_{2,\infty}$ Perturbation in Matrix Denoising

Theorem

Suppose that $\mathrm{SNR} \geq C\sqrt{n\log(n)}$ *for some sufficiently large constant* C.

$\ell_{2,\infty}$ Perturbation in Matrix Denoising

Theorem

Suppose that $\mathrm{SNR} \geq C\sqrt{n\log(n)}$ *for some sufficiently large constant* C*. Suppose* $\mathbf{S}$ *is incoherent with incoherence parameter* μ_0*.*

$\ell_{2,\infty}$ Perturbation in Matrix Denoising

Theorem

Suppose that $\mathrm{SNR} \geq C\sqrt{n\log(n)}$ *for some sufficiently large constant* C*. Suppose* $\mathbf{S}$ *is incoherent with incoherence parameter* μ_0*. Then there exists a universal constant* C' *such that with probability at least* $1 - O(n^{-20})$

$$\|\widehat{\mathbf{U}} - \mathbf{U}\mathcal{O}_*\|_{2,\infty} \leq C' \frac{\mu_0\sqrt{r\log(n)}}{\mathrm{SNR}}.$$

$\ell_{2,\infty}$ Perturbation in Matrix Denoising

(Davis-Kahan Bound) $$\|\widehat{\mathbf{U}} - \mathbf{U}\mathcal{O}\|_F \lesssim \frac{\sqrt{nr}}{\mathrm{SNR}};$$

($\ell_{2,\infty}$ Bound) $$\max_i \left\|\left(\widehat{\mathbf{U}} - \mathbf{U}\mathcal{O}\right)_{i\cdot}\right\| \lesssim \frac{\mu_0\sqrt{r\log(n)}}{\mathrm{SNR}}$$

$\ell_{2,\infty}$ Perturbation in Matrix Denoising

$$\text{(Davis-Kahan Bound)} \qquad \|\widehat{\mathbf{U}} - \mathbf{U}\mathcal{O}\|_F \lesssim \frac{\sqrt{nr}}{\text{SNR}};$$

$$(\ell_{2,\infty} \text{ Bound}) \qquad \max_i \left\| \left(\widehat{\mathbf{U}} - \mathbf{U}\mathcal{O}\right)_{i\cdot} \right\| \lesssim \frac{\mu_0 \sqrt{r \log(n)}}{\text{SNR}}$$

Key Takeaway

Errors are spread out amongst the rows when $\mathbf{S}$ is not too spiky!

Beyond Perturbation Bounds

- $\ell_{2,\infty}$ bounds can reveal new information about how signal and noise interact (e.g., through incoherence μ_0).

Beyond Perturbation Bounds

- $\ell_{2,\infty}$ bounds can reveal new information about how signal and noise interact (e.g., through incoherence μ_0).
- Still not enough to develop optimal clustering error rates or to be used in inference problems.

Beyond Perturbation Bounds

- $\ell_{2,\infty}$ bounds can reveal new information about how signal and noise interact (e.g., through incoherence μ_0).
- Still not enough to develop optimal clustering error rates or to be used in inference problems.
- Develop *asymptotic expansions* that are amenable to analysis.

Beyond Perturbation Bounds

- $\ell_{2,\infty}$ bounds can reveal new information about how signal and noise interact (e.g., through incoherence μ_0).
- Still not enough to develop optimal clustering error rates or to be used in inference problems.
- Develop *asymptotic expansions* that are amenable to analysis.
- Develop distributional theory (e.g., limit theorems) that are first step to obtaining principled inference tools.

Beyond Perturbation Bounds

- $\ell_{2,\infty}$ bounds can reveal new information about how signal and noise interact (e.g., through incoherence μ_0).
- Still not enough to develop optimal clustering error rates or to be used in inference problems.
- Develop *asymptotic expansions* that are amenable to analysis.
- Develop distributional theory (e.g., limit theorems) that are first step to obtaining principled inference tools.

Asymptotic Expansion for Matrix Denoising

Theorem

Suppose that $\mathrm{SNR} \geq C\sqrt{n\log(n)}$.

Asymptotic Expansion for Matrix Denoising

Theorem

Suppose that $\mathrm{SNR} \geq C\sqrt{n\log(n)}$. *Then there is an event* $\mathcal{E}$ *satisfying* $\mathbb{P}(\mathcal{E}) \geq 1 - n^{-10}$ *such that on this event,*

Asymptotic Expansion for Matrix Denoising

Theorem

Suppose that $\mathrm{SNR} \geq C\sqrt{n\log(n)}$. *Then there is an event* $\mathcal{E}$ *satisfying* $\mathbb{P}(\mathcal{E}) \geq 1 - n^{-10}$ *such that on this event,*

$$\widehat{\mathbf{U}}\mathcal{O}_*^\top - \mathbf{U} = \underbrace{\mathbf{N}\mathbf{U}\mathbf{\Lambda}^{-1}}_{\text{Linear in } \mathbf{N}} + \underbrace{\Gamma}_{\text{Small-Order Residual}}$$

Asymptotic Expansion for Matrix Denoising

Theorem

Suppose that $\mathrm{SNR} \geq C\sqrt{n\log(n)}$*. Then there is an event* $\mathcal{E}$ *satisfying* $\mathbb{P}(\mathcal{E}) \geq 1 - n^{-10}$ *such that on this event,*

$$\widehat{\mathbf{U}}\mathcal{O}_*^\top - \mathbf{U} = \underbrace{\mathbf{N}\mathbf{U}\Lambda^{-1}}_{\text{Linear in } \mathbf{N}} + \underbrace{\Gamma}_{\text{Small-Order Residual}}$$

where

$$\|\Gamma\|_{2,\infty} \lesssim \frac{\mu_0\left(r + \sqrt{r\log(n)}\right)}{\sqrt{n} \times \mathrm{SNR}} + \frac{\mu_0\sqrt{rn}\log(n)}{\mathrm{SNR}^2}.$$

Asymptotic Expansion for Matrix Denoising

(Previous result) $$\max_i \left\| \left(\widehat{\mathbf{U}} \mathcal{O}_*^\top - \mathbf{U} \right)_{i\cdot} \right\| = \widetilde{O}\left(\frac{1}{\mathrm{SNR}} \right);$$

(This result) $$\max_i \left\| \left(\widehat{\mathbf{U}} \mathcal{O}_*^\top - \mathbf{U} - \mathbf{N}\mathbf{U}\Lambda^{-1} \right)_{i\cdot} \right\| = \widetilde{o}\left(\frac{1}{\mathrm{SNR}} \right).$$

Asymptotic Expansion for Matrix Denoising

(Previous result) $$\max_i \|(\widehat{\mathbf{U}}\mathcal{O}_*^\top - \mathbf{U})_{i\cdot}\| = \widetilde{O}\left(\frac{1}{\mathrm{SNR}}\right);$$

(This result) $$\max_i \|(\widehat{\mathbf{U}}\mathcal{O}_*^\top - \mathbf{U} - \mathbf{N}\mathbf{U}\Lambda^{-1})_{i\cdot}\| = \widetilde{o}\left(\frac{1}{\mathrm{SNR}}\right).$$

Key Takeaway

$\widehat{\mathbf{U}}$ is approximately a linear function of noise matrix $\mathbf{N}$, population eigenvector matrix $\mathbf{U}$, and population eigenvalue matrix Λ!

Distributional Theory for Matrix Denoising

Theorem

Suppose that r is fixed, that the condition number κ of $\mathbf{S}$ is bounded, and that μ_0 is bounded.

Distributional Theory for Matrix Denoising

Theorem

Suppose that r is fixed, that the condition number κ of $\mathbf{S}$ is bounded, and that μ_0 is bounded. Suppose further that $\mathrm{SNR} \gg \sqrt{n}\log(n)$.

Distributional Theory for Matrix Denoising

Theorem

Suppose that r is fixed, that the condition number κ of $\mathbf{S}$ is bounded, and that μ_0 is bounded. Suppose further that $\mathrm{SNR} \gg \sqrt{n}\log(n)$. Then it holds that

$$\frac{\Lambda}{\sigma}\left(\widehat{\mathbf{U}}\mathcal{O}_*^\top - \mathbf{U}\right)_{i\cdot} \to \mathcal{N}(0, \mathbf{I}_r)$$

in distribution as $n \to \infty$.

Distributional Theory for Matrix Denoising

Theorem

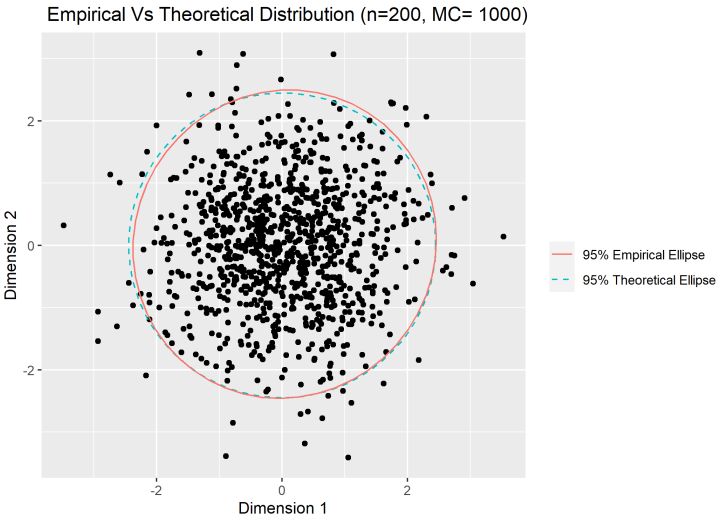
Suppose that r is fixed, that the condition number κ of $\mathbf{S}$ is bounded, and that μ_0 is bounded. Suppose further that $\mathrm{SNR} \gg \sqrt{n}\log(n)$. Then it holds that

$$\frac{\Lambda}{\sigma}\left(\widehat{\mathbf{U}}\mathcal{O}_*^\top - \mathbf{U}\right)_{i\cdot} \to \mathcal{N}(0, \mathbf{I}_r)$$

in distribution as $n \to \infty$.

Limiting variance of the entries of l'th eigenvector is $\frac{\sigma^2}{\lambda_l{}^2}$!

Distributional Theory for Matrix Denoising



Outline

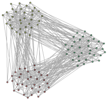
Statistical Inference

Inference problems of interest:

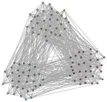
Statistical Inference

Inference problems of interest:

- Two-sample network testing



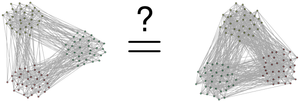
$\overset{?}{=}$



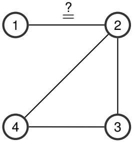
Statistical Inference

Inference problems of interest:

- Two-sample network testing



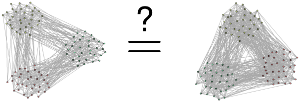
- Testing vertex memberships



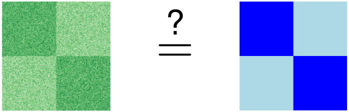
Statistical Inference

Inference problems of interest:

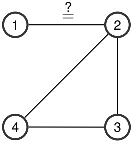
- Two-sample network testing



- Goodness of fit testing



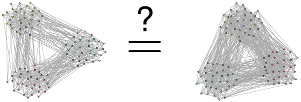
- Testing vertex memberships



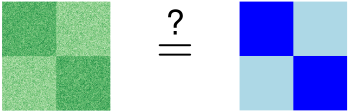
Statistical Inference

Inference problems of interest:

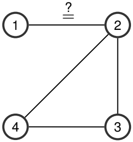
- Two-sample network testing



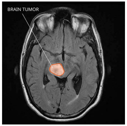
- Goodness of fit testing



- Testing vertex memberships



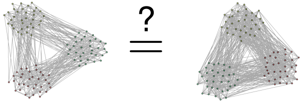
- Simultaneous confidence intervals



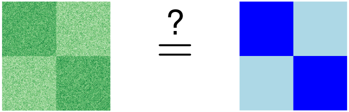
Statistical Inference

Inference problems of interest:

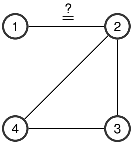
- Two-sample network testing



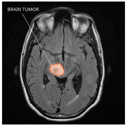
- Goodness of fit testing



- Testing vertex memberships



- Simultaneous confidence intervals



Main Idea

Use the previous results to justify subsequent inference with eigenvectors, singular vectors, or related quantities.

A Simple Testing Problem

- Consider the null hypothesis:

$$H_0 : \mathbf{S}_{i\cdot} = \mathbf{S}_{j\cdot}$$

A Simple Testing Problem

- Consider the null hypothesis:

$$H_0 : \mathbf{S}_{i\cdot} = \mathbf{S}_{j\cdot}$$

- Since $\mathbf{S}$ is low-rank, H_0 holds if and only if

$$\|(\mathbf{U}\Lambda)_{i\cdot} - (\mathbf{U}\Lambda)_{j\cdot}\|^2 = 0.$$

A Simple Testing Problem

- Consider the null hypothesis:

$$H_0 : \mathbf{S}_{i\cdot} = \mathbf{S}_{j\cdot}$$

- Since $\mathbf{S}$ is low-rank, H_0 holds if and only if

$$\|(\mathbf{U}\Lambda)_{i\cdot} - (\mathbf{U}\Lambda)_{j\cdot}\|^2 = 0.$$

- Don't have access to $\mathbf{U}$ and Λ!

A Simple Testing Problem

- Consider the null hypothesis:

$$H_0 : \mathbf{S}_{i\cdot} = \mathbf{S}_{j\cdot}$$

- Since $\mathbf{S}$ is low-rank, H_0 holds if and only if

$$\|(\mathbf{U}\Lambda)_{i\cdot} - (\mathbf{U}\Lambda)_{j\cdot}\|^2 = 0.$$

- Don't have access to $\mathbf{U}$ and Λ!

Test Statistic

Define

$$T_{ij}^2 := \frac{1}{2\sigma^2}\|(\widehat{\mathbf{U}}\widehat{\Lambda})_{i\cdot} - (\widehat{\mathbf{U}}\widehat{\Lambda})_{j\cdot}\|^2,$$

(we assume that σ is known for convenience).

A Simple Testing Problem

Theorem

Suppose that r *is fixed, and* μ_0, κ *are bounded. Suppose that* $\mathrm{SNR} \gg \sqrt{n}\log(n)$. *Then:*

A Simple Testing Problem

Theorem

Suppose that r *is fixed, and* μ_0, κ *are bounded. Suppose that* $\mathrm{SNR} \gg \sqrt{n}\log(n)$*. Then:*

- *(Consistency under the null) If* $\mathbf{S}_{i\cdot} = \mathbf{S}_{j\cdot}$*, it holds that* $T_{ij}^2 \to \chi_r^2$*.*

A Simple Testing Problem

Theorem

Suppose that r *is fixed, and* μ_0, κ *are bounded. Suppose that* $\mathrm{SNR} \gg \sqrt{n}\log(n)$. *Then:*

- *(Consistency under the null) If* $\mathbf{S}_{i\cdot} = \mathbf{S}_{j\cdot}$, *it holds that* $T_{ij}^2 \to \chi_r^2$.
- *(Consistency under local alternatives) If it holds that* $\|\mathbf{S}_{i\cdot} - \mathbf{S}_{j\cdot}\| \gg \sigma$, *then it holds that* $\mathbb{P}(T_{ij}^2 > C) \to 1$ *for any* $C > 0$.

A Simple Testing Problem

Theorem

Suppose that r *is fixed, and* μ_0, κ *are bounded. Suppose that* $\mathrm{SNR} \gg \sqrt{n}\log(n)$*. Then:*

- *(Consistency under the null) If* $\mathbf{S}_{i\cdot} = \mathbf{S}_{j\cdot}$*, it holds that* $T_{ij}^2 \to \chi_r^2$*.*
- *(Consistency under local alternatives) If it holds that* $\|\mathbf{S}_{i\cdot} - \mathbf{S}_{j\cdot}\| \gg \sigma$*, then it holds that* $\mathbb{P}(T_{ij}^2 > C) \to 1$ *for any* $C > 0$*. If instead it holds that* $\frac{1}{2\sigma^2}\|\mathbf{S}_{i\cdot} - \mathbf{S}_{j\cdot}\|^2 \to \mu > 0$*, then* $T_{ij}^2 \to \chi_r^2(\mu)$*.*

A Simple Testing Problem

Theorem

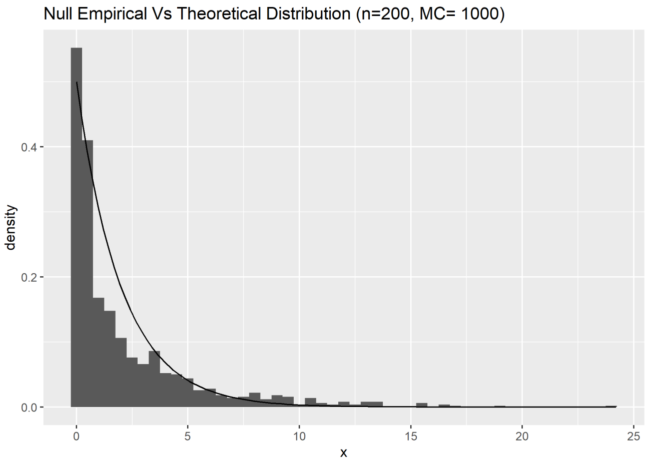
Suppose that r *is fixed, and* μ_0, κ *are bounded. Suppose that* $\mathrm{SNR} \gg \sqrt{n}\log(n)$. *Then:*

- *(Consistency under the null) If* $\mathbf{S}_{i\cdot} = \mathbf{S}_{j\cdot}$, *it holds that* $T_{ij}^2 \to \chi_r^2$.
- *(Consistency under local alternatives) If it holds that* $\|\mathbf{S}_{i\cdot} - \mathbf{S}_{j\cdot}\| \gg \sigma$, *then it holds that* $\mathbb{P}(T_{ij}^2 > C) \to 1$ *for any* $C > 0$. *If instead it holds that* $\frac{1}{2\sigma^2}\|\mathbf{S}_{i\cdot} - \mathbf{S}_{j\cdot}\|^2 \to \mu > 0$, *then* $T_{ij}^2 \to \chi_r^2(\mu)$.

Key Takeaway

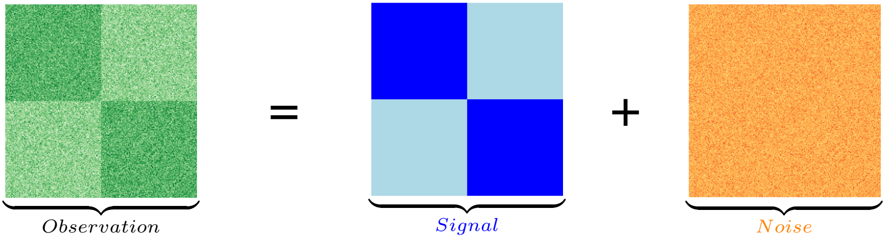
Consistent testing is possible in high dimensions given knowledge of the underlying low-rank structure!

A Simple Testing Problem

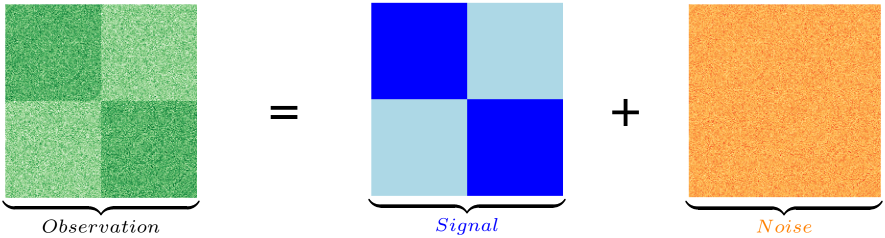


Outline

Chapter 1: Introduction

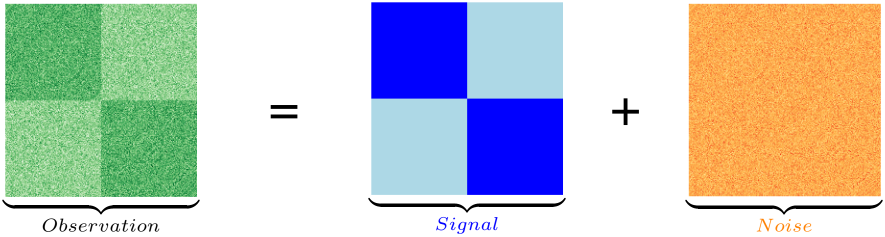


Chapter 1: Introduction



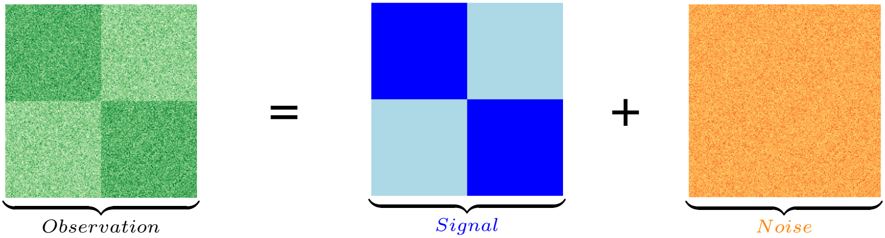
- Develop *framework for statistical inference*

Chapter 1: Introduction



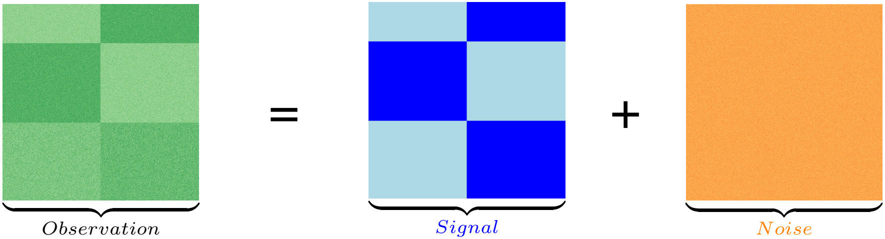
- Develop *framework for statistical inference*
- Examples with matrix denoising model

Chapter 1: Introduction

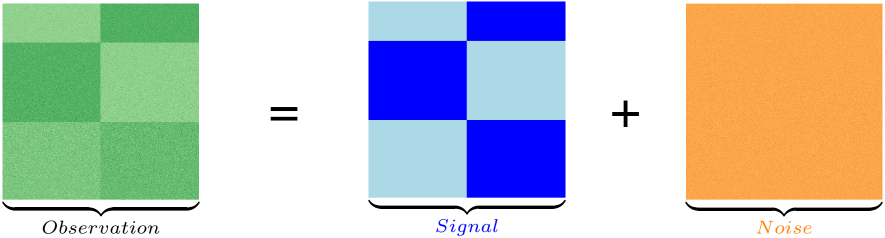


- Develop *framework for statistical inference*
- Examples with matrix denoising model
- A few novel results you have seen today

Chapter 2: Rectangular Signal Plus Noise Model

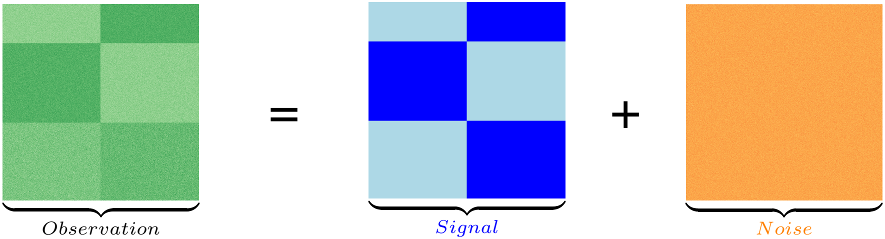


Chapter 2: Rectangular Signal Plus Noise Model



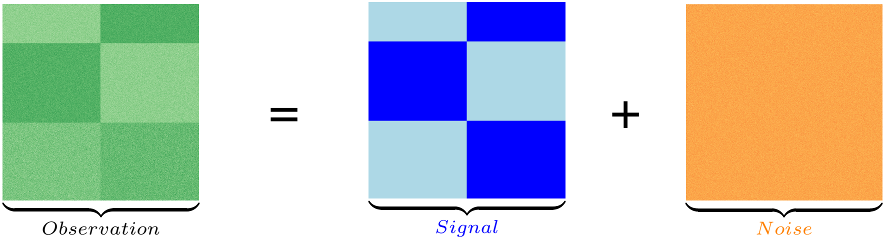
- Distributional theory and inference in the rectangular signal plus noise model

Chapter 2: Rectangular Signal Plus Noise Model



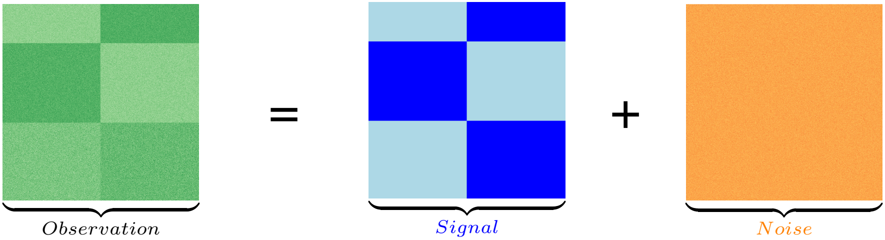
- Distributional theory and inference in the rectangular signal plus noise model
- Allow for *heteroskedasticity* and *dependence* in the noise

Chapter 2: Rectangular Signal Plus Noise Model



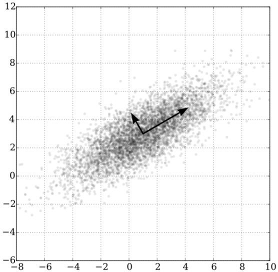
- Distributional theory and inference in the rectangular signal plus noise model
- Allow for *heteroskedasticity* and *dependence* in the noise
- Apply to high-dimensional mixture models

Chapter 2: Rectangular Signal Plus Noise Model

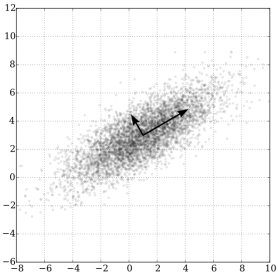


- Distributional theory and inference in the rectangular signal plus noise model
- Allow for *heteroskedasticity* and *dependence* in the noise
- Apply to high-dimensional mixture models
- Based on "Entrywise Estimation of Singular Vectors with Heteroskedasticity and Dependence" (Agterberg et al., 2022b)

Chapter 3: Sparse PCA

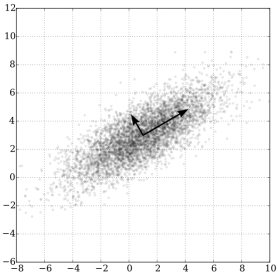


Chapter 3: Sparse PCA



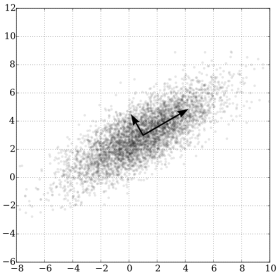
- $\ell_{2,\infty}$ rates of convergence in Sparse PCA model

Chapter 3: Sparse PCA



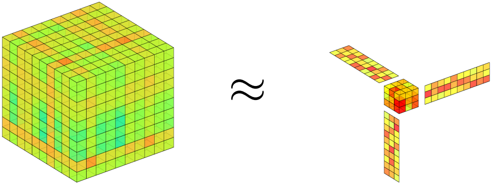
- $\ell_{2,\infty}$ rates of convergence in Sparse PCA model
- Results reveal new phenomena with respect to sparsity and estimation

Chapter 3: Sparse PCA

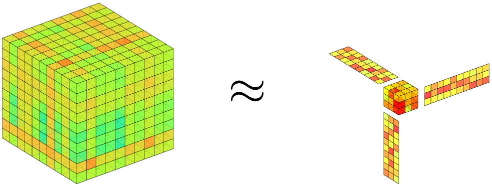


- $\ell_{2,\infty}$ rates of convergence in Sparse PCA model
- Results reveal new phenomena with respect to sparsity and estimation
- Based on "Entrywise Recovery Guarantees for Sparse PCA via Sparsistent Algorithms" (Agterberg and Sulam, 2022)

Chapter 4: Tensor Data Analysis

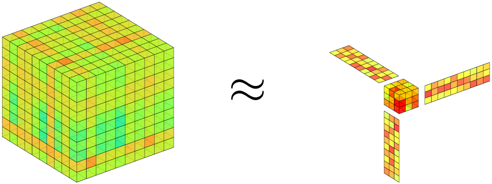


Chapter 4: Tensor Data Analysis



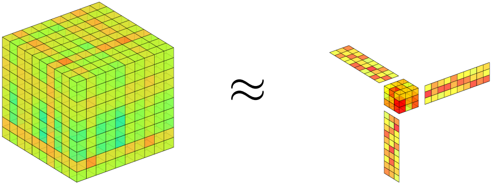
- $\ell_{2,\infty}$ rates of estimation in tensor mixed-membership blockmodel and more general tensor denoising model

Chapter 4: Tensor Data Analysis



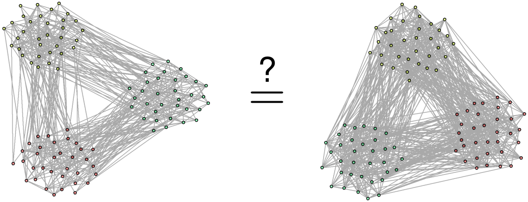
- $\ell_{2,\infty}$ rates of estimation in tensor mixed-membership blockmodel and more general tensor denoising model
- Results reveal new phenomena for Tensor SVD that are different from Matrix SVD

Chapter 4: Tensor Data Analysis

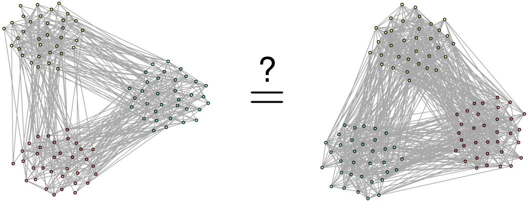


- $\ell_{2,\infty}$ rates of estimation in tensor mixed-membership blockmodel and more general tensor denoising model
- Results reveal new phenomena for Tensor SVD that are different from Matrix SVD
- Based on "Estimating Higher-Order mixed Memberships via the $\ell_{2,\infty}$ Tensor Perturbation Bound" (Agterberg and Zhang, 2022)

Chapter 5: Two-Sample Network Testing

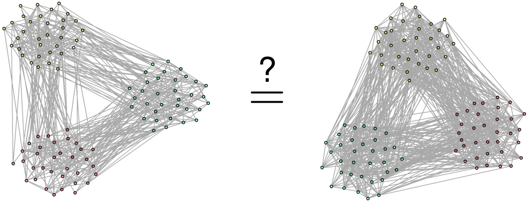


Chapter 5: Two-Sample Network Testing



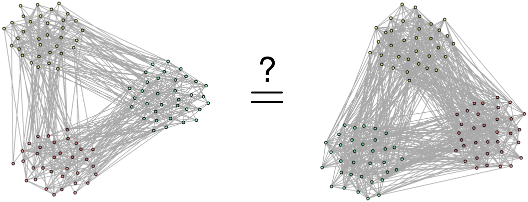
- Provide a testing procedure for a hypothesis test for two networks and show that it is consistent under null and alternative hypotheses

Chapter 5: Two-Sample Network Testing



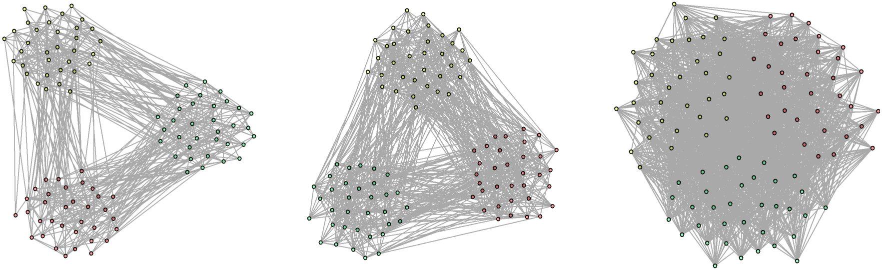
- Provide a testing procedure for a hypothesis test for two networks and show that it is consistent under null and alternative hypotheses
- Results based on asymptotic expansions developed in Rubin-Delanchy et al. (2020) as well as results in Agterberg et al. (2020b)

Chapter 5: Two-Sample Network Testing

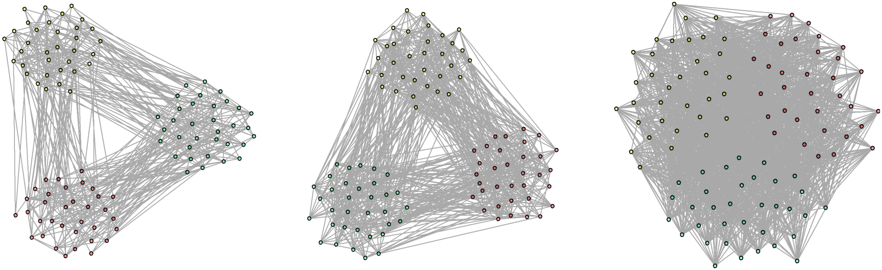


- Provide a testing procedure for a hypothesis test for two networks and show that it is consistent under null and alternative hypotheses
- Results based on asymptotic expansions developed in Rubin-Delanchy et al. (2020) as well as results in Agterberg et al. (2020b)
- Based on “Nonparametric Two-Sample Hypothesis Testing for Random Graphs with Negative and Repeated Eigenvalues” (Agterberg et al., 2020a)

Chapter 6: Clustering in Multilayer Networks

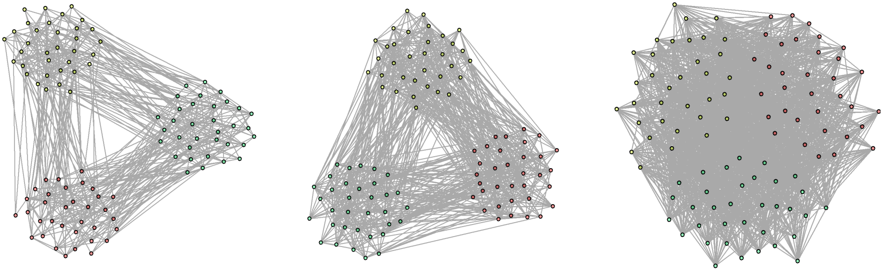


Chapter 6: Clustering in Multilayer Networks



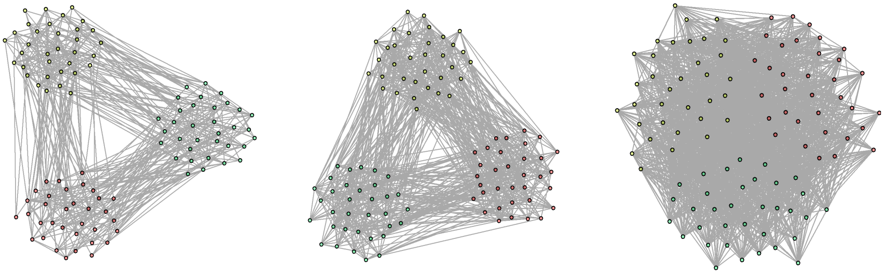
- Develop “exponential” misclustering error rates for a spectral algorithm for heterogeneous networks

Chapter 6: Clustering in Multilayer Networks



- Develop “exponential” misclustering error rates for a spectral algorithm for heterogeneous networks
- Results based on two novel asymptotic expansions, and results reveal how error rates improve with more networks

Chapter 6: Clustering in Multilayer Networks



- Develop “exponential” misclustering error rates for a spectral algorithm for heterogeneous networks
- Results based on two novel asymptotic expansions, and results reveal how error rates improve with more networks
- Based on “Joint Spectral Clustering in Multilayer Degree-Corrected Stochastic Blockmodels” (Agterberg et al., 2022a)

Future and Ongoing Work

- Multilayer networks:
  - More general community models with estimation and testing guarantees with multilayer networks
  - Estimation accuracy in sparse network regimes
  - Network time series
  - Signed multilayer networks

Future and Ongoing Work

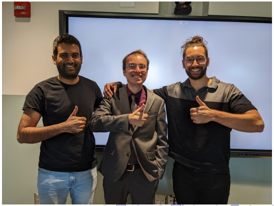
- Multilayer networks:
  - More general community models with estimation and testing guarantees with multilayer networks
  - Estimation accuracy in sparse network regimes
  - Network time series
  - Signed multilayer networks
- Tensor data analysis:
  - Statistical inference for low-rank tensors by building upon $\ell_{2,\infty}$ tensor perturbation bound (ongoing)
  - Other notions of low-rank
  - Robustness and sparsity

Future and Ongoing Work

- Multilayer networks:
  - More general community models with estimation and testing guarantees with multilayer networks
  - Estimation accuracy in sparse network regimes
  - Network time series
  - Signed multilayer networks
- Tensor data analysis:
  - Statistical inference for low-rank tensors by building upon $\ell_{2,\infty}$ tensor perturbation bound (ongoing)
  - Other notions of low-rank
  - Robustness and sparsity
- Spectral methods and nonconvex algorithms:
  - Entrywise guarantees for other nonconvex matrix and tensor algorithms under different noise models
  - Inference with the outputs of nonconvex procedures
  - Heterogeneous missingness mechanisms

Pictures

Pictures



References I

Joshua Agterberg and Jeremias Sulam. Entrywise Recovery Guarantees for Sparse PCA via Sparsistent Algorithms. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, pages 6591–6629. PMLR, May 2022. ISSN: 2640-3498.

Joshua Agterberg and Anru Zhang. Estimating Higher-Order Mixed Memberships via the $\ell_{2,\infty}$ Tensor Perturbation Bound, December 2022. arXiv:2212.08642 [math, stat].

Joshua Agterberg, Minh Tang, and Carey Priebe. Nonparametric Two-Sample Hypothesis Testing for Random Graphs with Negative and Repeated Eigenvalues. *arXiv:2012.09828 [math, stat]*, December 2020a.

Joshua Agterberg, Minh Tang, and Carey E. Priebe. On Two Distinct Sources of Nonidentifiability in Latent Position Random Graph Models. *arXiv:2003.14250 [math, stat]*, March 2020b.

References II

Joshua Agterberg, Zachary Lubberts, and Jesús Arroyo. Joint Spectral Clustering in Multilayer Degree-Corrected Stochastic Blockmodels, December 2022a. arXiv:2212.05053 [math, stat].

Joshua Agterberg, Zachary Lubberts, and Carey E. Priebe. Entrywise Estimation of Singular Vectors of Low-Rank Matrices With Heteroskedasticity and Dependence. *IEEE Transactions on Information Theory*, 68(7):4618–4650, July 2022b. ISSN 1557-9654. doi: 10.1109/TIT.2022.3159085.

Patrick Rubin-Delanchy, Joshua Cape, Minh Tang, and Carey E. Priebe. A statistical interpretation of spectral embedding: the generalised random dot product graph. *arXiv:1709.05506 [cs, stat]*, January 2020.

Thank you!

: @JAgterberger