Entrywise Estimation of Singular Vectors of Low-Rank Matrices with Heteroskedasticity and Dependence

Joshua Agterberg

Johns Hopkins University

JSM 2022
Based on a paper with Zachary Lubberts and Carey Priebe
Outline

1. The Problem
2. Theoretical Results
3. Numerical Example
4. Conclusion
# Outline

1. The Problem
2. Theoretical Results
3. Numerical Example
4. Conclusion
Motivation: Spectral Methods

Spectral methods are ubiquitous in machine learning and statistics.

- Spectral Clustering
- Principal Components Analysis
- Nonconvex algorithm initializations (tensor SVD, phase retrieval, blind deconvolution)
Motivation: Spectral Methods

Spectral methods are ubiquitous in machine learning and statistics.

- Spectral Clustering
- Principal Components Analysis
- Nonconvex algorithm initializations (tensor SVD, phase retrieval, blind deconvolution)

Problem

Lots is known about convergence, but less is known about uncertainty quantification.
Motivation: Spectral Methods

Spectral methods are ubiquitous in machine learning and statistics.

- Spectral Clustering
- Principal Components Analysis
- Nonconvex algorithm initializations (tensor SVD, phase retrieval, blind deconvolution)

Problem

Lots is known about convergence, but less is known about uncertainty quantification.

Goal

Develop fine-grained statistical theory for spectral methods.
Signal Plus Noise Model

\[
\text{Signal} + \text{Noise} = \text{Observation}
\]
Signal Plus Noise Model

Goal: More Precise

Develop *fine-grained statistical theory* for an estimator of the left singular subspace of the *signal matrix*. 
Motivation: Spectral Methods

In many problems there is *heteroskedasticity* and *dependence* within each row of the noise.
Motivation: Spectral Methods

In many problems there is heteroskedasticity and dependence within each row of the noise.
This Talk

Goal: Even More Precise

Develop *fine-grained statistical theory* for an estimator of the left singular subspace of the *signal matrix* in the presence of heteroskedasticity and dependence within each row of the *noise matrix*. 
Goal: Even More Precise

Develop \textit{fine-grained statistical theory} for an estimator of the left singular subspace of the \textit{signal matrix} \textit{in the presence of heteroskedasticity} and \textit{dependence within each row of the noise matrix}.

“...the geometric relationship between the \textit{signal matrix}, the \textit{covariance structure of the noise}, and the distribution of the errors...”
We observe a **low-rank signal matrix** corrupted by **additive noise**:

\[ \hat{M} = M + E. \]
We observe a low-rank signal matrix corrupted by additive noise:

\[
\hat{M} = M + E.
\]

The signal matrix \( M \) is assumed to be (low) rank \( r \) with (thin or compact) singular value decomposition (SVD)

\[
M = U\Lambda V^\top
\]

- \( U \in \mathbb{O}(n, r) \) is matrix of leading left singular vectors (its columns \( U_j \) are orthonormal unit vectors)
- \( \Lambda \) is a diagonal \( r \times r \) matrix of singular values \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0 \)
- \( V \in \mathbb{O}(d, r) \) is matrix of leading right singular vectors
General Model

We observe a low-rank signal matrix corrupted by additive noise:

\[ \hat{M} = M \underbrace{\text{signal}}_{\text{signal}} + E \underbrace{\text{noise}}_{\text{noise}}. \]

The noise matrix \( E \) has

- independent, mean-zero rows of the form \( E_i = \Sigma_i^{1/2} Y_i \)
- \( \Sigma_i \in \mathbb{R}^{d \times d} \) is a positive semidefinite matrix
- \( Y_i \in \mathbb{R}^d \) is a vector with independent (sub)gaussian components with variance one
General Model

Goal: Most Precise

Develop *fine-grained statistical theory* for an estimator $\hat{U}$ of the $n \times r$ matrix $U$ of leading left singular vectors of $M$ upon observing $M + E$. 

When rows of $E$ have different covariances (heteroskedasticity), the left singular vectors of $M + E$ can be biased! 

Solution: Use HeteroPCA algorithm of Zhang et al. (2022) to debias the estimated singular vectors.
General Model

Goal: Most Precise

Develop *fine-grained statistical theory* for an estimator $\hat{U}$ of the $n \times r$ matrix $U$ of leading left singular vectors of $M$ upon observing $M + E$.

Problem

When rows of $E$ have different covariances (heteroskedasticity), the left singular vectors of $M + E$ can be biased!
General Model

**Goal: Most Precise**

Develop *fine-grained statistical theory* for an estimator $\hat{U}$ of the $n \times r$ matrix $U$ of leading left singular vectors of $M$ upon observing $M + E$.

**Problem**

When rows of $E$ have different covariances (heteroskedasticity), the left singular vectors of $M + E$ can be biased!

**Solution**

Use *HeteroPCA* algorithm of Zhang et al. (2022) to *debias* the estimated singular vectors.
**Algorithm 1:** HeteroPCA Algorithm of Zhang et al. (2022)

**Input:** \( N_0 = \hat{\mathbf{M}}\hat{\mathbf{M}}^\top - \text{diag}(\hat{\mathbf{M}}\hat{\mathbf{M}}^\top) \), max number of iterations \( T_{\text{max}} \)

**while** \( T \leq T_{\text{max}} \) **do**

\[
\tilde{N}_T := \text{SVD}_r(N_T), \text{ the best rank } r \text{ approximation to } N_T;
\]

\[
N_{T+1} := N_T - \text{diag}(N_T) + \text{diag}(\tilde{N}_T);
\]

**end**

**Return:** \( \hat{U} = \text{Left } r \text{ singular vectors of } N_{T_{\text{max}}} \)
Outline

1. The Problem
2. Theoretical Results
3. Numerical Example
4. Conclusion
Asymptotic Normality: $r$ fixed

Theorem (Agterberg et al. (2022))

Suppose some technical and regularity conditions hold, and suppose the signal-to-noise ratio is sufficiently large. Define

$$S^{(i)} := \Lambda^{-1} V^\top \Sigma_i V \Lambda^{-1}.$$  

Then as $n, d \to \infty$, with $d \geq n \geq \log(d)$, there exists a sequence of $r \times r$ orthogonal matrices $O_*$ such that

$$(S^{(i)})^{-1/2}(\hat{U}O_* - U)_i \to N(0, I_r).$$
Asymptotic Normality: $r$ fixed

Theorem (Agterberg et al. (2022))

Suppose some technical and regularity conditions hold, and suppose the signal-to-noise ratio is sufficiently large. Define

$$S^{(i)} := \Lambda^{-1} V^\top \Sigma_i V \Lambda^{-1}.$$

Then as $n, d \to \infty$, with $d \geq n \geq \log(d)$, there exists a sequence of $r \times r$ orthogonal matrices $O_*$ such that

$$(S^{(i)})^{-1/2}(\hat{U}O_* - U)_i. \to N(0, I_r).$$

Asymptotic covariance of $i$'th row of $\hat{U}$ depends on how $i$'th row of noise matrix $E$ interacts with $\Lambda$ and $V$. 
Corollary (Agterberg et al. (2022))

Under the conditions of Theorem 1, suppose further that $\Sigma_i = \sigma_i^2 I_d$ (independent noise with equal variance within each row). Then

$$(S^{(i)})_{jj} := \left\| \Sigma_i^{1/2} V \cdot j \right\|^2 / \lambda_j^2 = \frac{\sigma_i^2}{\lambda_j^2}.$$ 

Then there exists a sequence of orthogonal matrices $O_\ast$ such that

$$\frac{\lambda_j}{\sigma_j} (\hat{U} O_\ast - U)_{ij} \to N(0, 1).$$
Corollary (Agterberg et al. (2022))

Under the conditions of Theorem 1, suppose further that $\Sigma_i = \sigma_i^2 I_d$ (independent noise with equal variance within each row). Then

$$(S^{(i)})_{jj} := \left\| \Sigma_j^{1/2} V_{:j} \right\|^2 = \frac{\sigma_i^2}{\lambda_j^2}.$$ 

Then there exists a sequence of orthogonal matrices $O_*$ such that

$$\frac{\lambda_j}{\sigma_i} (\hat{U} O_* - U)_{ij} \to N(0, 1).$$

Asymptotic variance of entries of $j$’th estimated singular vector is proportional to $j$’th singular value.
Recall that $S^{(i)} = \Lambda^{-1} V^\top \Sigma_i V \Lambda^{-1}$
Recall that $S^{(i)} = \Lambda^{-1} V^\top \Sigma_i V \Lambda^{-1}$

Consider

$$
\Sigma_1 := 5 V_1 V_1^\top + 5 V_\theta V_\theta^\top + .1 I_d
$$

where $V_\theta$ satisfies $V_\theta^\top V_2 = \theta$ and is orthogonal to $V_1$
Recall that $S^{(i)} = \Lambda^{-1} V^\top \Sigma_i V \Lambda^{-1}$

Consider

$$
\Sigma_1 := 5 V_1 V_1^\top + 5 V_\theta V_\theta^\top + .1 I_d
$$

where $V_\theta$ satisfies $V_\theta^\top V_2 = \theta$ and is orthogonal to $V_1$

Theory suggests that $\Lambda(\hat{U} O \ast - U)_V \approx N(0, V^\top \Sigma_1 V)$ and hence

$$
V^\top \Sigma_1 V = V^\top \left( 5 V_1 V_1^\top + 5 V_\theta V_\theta^\top + .1 I_d \right) V
$$

$$
= \begin{pmatrix}
5 & 0 \\
0 & 5\theta
\end{pmatrix} + \begin{pmatrix}
.1 & 0 \\
0 & .1
\end{pmatrix}
$$
Recall that $S^{(i)} = \Lambda^{-1} V^\top \Sigma_i V \Lambda^{-1}$

Consider

$$ \Sigma_1 := 5 V_1 V_1^\top + 5 V_\theta V_\theta^\top + .1 I_d $$

where $V_\theta$ satisfies $V_\theta^\top V_2 = \theta$ and is orthogonal to $V_1$

Theory suggests that $\Lambda(\hat{U} A_\star - U)_1. \approx N(0, V^\top \Sigma_1 V)$ and hence

$$ V^\top \Sigma_1 V = V^\top \left( 5 V_1 V_1^\top + 5 V_\theta V_\theta^\top + .1 I_d \right) V $$

$$ = \begin{pmatrix} 5 & 0 \\ 0 & 5\theta \end{pmatrix} + \begin{pmatrix} .1 & 0 \\ 0 & .1 \end{pmatrix} $$

So decreasing $\theta$ decreases the limiting variance along the second dimension
Figure: 1000 MonteCarlo iterations of the first row of $\Lambda(\widehat{U}O_* - U)$ with $n = d = 1800$, where the covariance changes according to previous slide.
Conclusion

Wanted to develop fine-grained statistical theory for an estimator of the left singular vectors of $M = U \Lambda V^T$ in the presence of heteroskedasticity and dependence.
Conclusion

- Wanted to develop *fine-grained statistical theory* for an estimator of the left singular vectors of $M = UV^\top$ in the presence of *heteroskedasticity and dependence*.

- Our estimator is based on applying the HeteroPCA algorithm of Zhang et al. (2022) to the sample gram matrix $\hat{MM}^\top$. 
Conclusion

- Wanted to develop *fine-grained statistical theory* for an estimator of the left singular vectors of $M = U\Lambda V^\top$ in the presence of *heteroskedasticity and dependence*.

- Our estimator is based on applying the *HeteroPCA* algorithm of Zhang et al. (2022) to the sample gram matrix $\hat{M}\hat{M}^\top$.

- We prove limiting entrywise asymptotic normality results for our estimator in a high-dimensional regime showcasing the geometric relationship between the signal matrix, the covariance structure of the noise, and the limiting distribution of the errors via the limiting covariance matrix

$$S^{(i)} := \Lambda^{-1} V^\top \Sigma_i V \Lambda^{-1}.$$
Conclusion

- Wanted to develop fine-grained statistical theory for an estimator of the left singular vectors of $M = U\Lambda V^\top$ in the presence of heteroskedasticity and dependence.

- Our estimator is based on applying the HeteroPCA algorithm of Zhang et al. (2022) to the sample gram matrix $\hat{MM}^\top$.

- We prove limiting entrywise asymptotic normality results for our estimator in a high-dimensional regime showcasing the geometric relationship between the signal matrix, the covariance structure of the noise, and the limiting distribution of the errors via the limiting covariance matrix

$$S^{(i)} := \Lambda^{-1} V^\top \Sigma_i V \Lambda^{-1}.$$  

- Results in paper stated as Berry-Esseen Theorems (with $r$ allowed to grow), and we show we can estimate limiting covariance in high-dimensional mixture models, yielding asymptotically valid confidence intervals.

Thank you!

@JAgterberger
Asymptotic Normality: \( r \) fixed

**Theorem (Agterberg et al. (2022))**

Suppose some technical and regularity conditions hold. Suppose that

\[
\max \left\{ \frac{\log(d)}{\text{SNR}}, \max_j \frac{\| \Sigma_i^{1/2} V_{.j} \|^3}{\| \Sigma_i^{1/2} V_{.j} \|^3} \right\} \to 0
\]

as \( n, d \to \infty \), with \( d \geq n \geq \log(d) \). Define

\[
S^{(i)} := \Lambda^{-1} V^\top \Sigma_i V \Lambda^{-1}.
\]

Then there exists a sequence of \( r \times r \) orthogonal matrices \( \mathcal{O}_* \) such that

\[
(S^{(i)})^{-1/2}(\widehat{U}\mathcal{O}_* - U)_i. \to N(0, I_r).
\]
Asymptotic Normality: \( r \) fixed

Theorem (Agterberg et al. (2022))

Suppose some technical and regularity conditions hold. Suppose that

\[
\max \left\{ \frac{\log(d)}{\text{SNR}}, \max_j \frac{\| \Sigma_i^{1/2} V_j \|_3^3}{\| \Sigma_i^{1/2} V_j \|_3^3} \right\} \to 0
\]

as \( n, d \to \infty \), with \( d \geq n \geq \log(d) \). Define

\[
S^{(i)} := \Lambda^{-1} V^\top \Sigma_i V \Lambda^{-1}.
\]

Then there exists a sequence of \( r \times r \) orthogonal matrices \( O_* \) such that

\[
(S^{(i)})^{-1/2} (\widehat{U} O_* - U)_i \to N(0, I_r).
\]

Asymptotic covariance of \( i \)'th row of \( \widehat{U} \) depends on how \( \Sigma_i \) interacts with \( \Lambda \) and \( V \).
More Explanation

We require that

$$\max \left\{ \frac{\log(d')}{\text{SNR}}, \max_j \frac{\| \Sigma_j^{1/2} V_j \|_3^3}{\| \Sigma_j^{1/2} V_j \|_3} \right\} \to 0$$
More Explanation

- We require that

\[
\max \left\{ \frac{\log(d')}{\text{SNR}} \right\}, \max_j \left\{ \frac{\|\Sigma_j^{1/2} V_j\|_3^3}{\|\Sigma_j^{1/2} V_j\|_3} \right\} \to 0
\]

condition on SNR

Interaction between noise and signal
More Explanation

- We require that

\[
\max \left\{ \frac{\log(d')}{\text{SNR}}, \max_j \frac{\|\Sigma_i^{1/2} V_j\|_3^3}{\|\Sigma_i^{1/2} V_j\|_3^3} \right\} \to 0
\]

Interaction between noise and signal

- Special case: \(\Sigma_i \equiv I_d, V_j = \frac{\pm 1}{\sqrt{d}}\) (most incoherent vector). Then

\[
\frac{\|\Sigma_i^{1/2} V_j\|_3^3}{\|\Sigma_i^{1/2} V_j\|_3^3} = \frac{\|V_j\|_3^3}{\|V_j\|_3^3} = \frac{\sum_{l=1}^d \left(\frac{1}{\sqrt{d}}\right)^3}{1} = \frac{1}{\sqrt{d}}
\]
Asymptotic Normality: $r$ growing

**Theorem (Agterberg et al. (2022))**

Suppose some technical and regularity conditions hold. Suppose that

$$\max \left\{ \frac{r \log(d)}{\sqrt{n}}, \frac{r \log(d)}{\text{SNR}}, \frac{\|\Sigma_i^{1/2} V_j\|^3}{\|\Sigma_i^{1/2} V_j\|^3} \right\} \to 0$$

as $n, d \to \infty$, with $d \geq n \geq \log(d)$. Define

$$\sigma_{ij}^2 := \frac{\|\Sigma_i^{1/2} V_j\|^2}{\lambda_j^2}.$$

Then there exists a sequence of orthogonal matrices $O_*$ such that

$$\frac{1}{\sigma_{ij}} (\hat{U}O_* - U)_{ij} \to N(0, 1).$$
Singular vectors of $\hat{M} = \text{Eigenvectors of } \hat{M}\hat{M}^\top$

$\approx \text{Eigenvectors of } \mathbb{E}(\hat{M}\hat{M}^\top)$

$= \text{Eigenvectors of } MM^\top + D,$

where $D_{ii} = \text{Trace}(\Sigma_i).$
**Bias**

Singular vectors of $\hat{M} = \text{Eigenvectors of } \hat{M}\hat{M}^T$

$\approx \text{Eigenvectors of } E(\hat{M}\hat{M}^T)$

$= \text{Eigenvectors of } MM^T + D,$

where $D_{ii} = \text{Trace}(\Sigma_i)$.

**Problem**

If $\Sigma_i$’s are different (i.e. *heteroskedastic*), then the singular vectors of $\hat{M}$ are approximating a *deterministic diagonal perturbation* of $MM^T$. 
Correcting the Bias

- Could delete the diagonal of $\hat{M}\hat{M}^\top$ and take eigenvectors of that
Correcting the Bias

- Could delete the diagonal of $\hat{M}\hat{M}^\top$ and take eigenvectors of that
- Still biased! Then this approximates the eigenvectors of the matrix

$$MM^\top - \text{diag}(MM^\top) \neq MM^\top$$
Correcting the Bias

- Could delete the diagonal of $\hat{M}\hat{M}^\top$ and take eigenvectors of that
- Still biased! Then this approximates the eigenvectors of the matrix

$$MM^\top - \text{diag}(MM^\top) \neq MM^\top$$

- Just deleting the diagonal results in an error that does not scale with the noise
- Our idea: use existing HeteroPCA algorithm of Zhang et al. (2022) to impute the diagonals
Important parameters:

- Measure of noise: \( \sigma^2 := \max_i \| \Sigma_i \| \)
- Measure of signal: \( \lambda_r = \) smallest nonzero singular value of \( M \)
Notation

Important parameters:

- Measure of noise: \( \sigma^2 := \max_i \| \Sigma_i \| \)
- Measure of signal: \( \lambda_r = \) smallest nonzero singular value of \( M \)
- Define the signal-to-noise ratio:

\[
\text{SNR} := \frac{\lambda_r}{\sigma \sqrt{rd}}.
\]
Important parameters:

- Measure of noise: \( \sigma^2 := \max_i \| \Sigma_i \| \)
- Measure of signal: \( \lambda_r \) = smallest nonzero singular value of \( M \)
- Define the signal-to-noise ratio:

\[
\text{SNR} := \frac{\lambda_r}{\sigma \sqrt{rd}}.
\]

In the homoskedastic setting, \( \text{SNR} \to \infty \) is required for consistency when \( d \approx n \) with \( n, d \to \infty \).
Important parameters:

- Measure of noise: $\sigma^2 := \max_i \| \Sigma_i \|
- Measure of signal: $\lambda_r =$ smallest nonzero singular value of $\mathbf{M}$
- Define the signal-to-noise ratio:

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

In the homoskedastic setting, $\text{SNR} \to \infty$ is required for consistency when $d \approx n$ with $n, d \to \infty$.

- Condition number of $\mathbf{M}$, $\kappa := \frac{\lambda_1}{\lambda_r}$
New concept:

\textit{Covariance Condition Number}: \[ \kappa_\sigma := \max_{i,j} \frac{\sigma}{\| \Sigma_{i}^{1/2} V.j \|} \]

Quantifies the geometric relationship with the covariance structure of the noise on the right singular subspace.

Consider the following special case: \( \Sigma_i \equiv I_d \) for all \( i \) (or any multiple). Then \( \kappa_\sigma \equiv 1 \) only blows up when \( \| \Sigma_{i}^{1/2} V.j \| \) is very small relative to the overall noise (nondegeneracy condition).
Notation

New concept:

- **Covariance Condition Number**:

\[ \kappa_\sigma := \max_{i,j} \frac{\sigma}{\| \Sigma^{1/2} V_i \cdot V_j \|} \]

Quantifies the geometric relationship with the covariance structure of the noise on the right singular subspace \( V \).
Notation

New concept:

- **Covariance Condition Number**:

  \[ \kappa_\sigma := \max_{i,j} \frac{\sigma}{\| \Sigma_i^{1/2} V_j \|} \]

  Quantifies the geometric relationship with the covariance structure of the noise on the right singular subspace \( V \)

- Consider the following special case:
New concept:

- **Covariance Condition Number.**

\[
\kappa_\sigma := \max_{i,j} \frac{\sigma}{\| \Sigma_i^{1/2} V_j \|}
\]

Quantifies the geometric relationship with the covariance structure of the noise on the right singular subspace \( V \)

Consider the following special case:

- \( \Sigma_i \equiv I_d \) for all \( i \) (or any multiple)
- Then \( \kappa_\sigma \equiv 1 \)
Notation

New concept:

- **Covariance Condition Number**:

\[
\kappa_\sigma := \max_{i,j} \frac{\sigma}{\| \Sigma_i^{1/2} V_j \|}
\]

Quantifies the geometric relationship with the covariance structure of the noise on the right singular subspace \( V \).

Consider the following special case:

- \( \Sigma_i \equiv l_d \) for all \( i \) (or any multiple)
- Then \( \kappa_\sigma \equiv 1 \)

\( \kappa_\sigma \) only blows up when \( \| \Sigma_i^{1/2} V_j \| \) is very small relative to the overall noise \( \sigma \) (nondegeneracy condition)
Notation

Incoherence parameter:

- Incoherence parameter $\mu_0$ of the matrix $M$:

$$\max_i \|U_i\|, \|V_i\| \leq \mu_0 \sqrt{\frac{r}{n}}$$

Measures “spikiness” of $M$
Notation

Incoherence parameter:

- Incoherence parameter $\mu_0$ of the matrix $M$:

$$\max_i \|U_i\|, \|V_i\| \leq \mu_0 \sqrt{\frac{r}{n}}$$

Measures “spikiness” of $M$

Examples (consider $n = d$ for simplicity):

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

Incoherence parameter:

- Incoherence parameter $\mu_0$ of the matrix $M$:

$$\max_i \|U_i\|, \|V_i\| \leq \mu_0 \sqrt{\frac{r}{n}}$$

Measures “spikiness” of $M$

Examples (consider $n = d$ for simplicity):

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

$\mu_0 = \sqrt{\frac{n}{r}}$ vs. $\mu_0 = 1$
Asymptotic Normality: \( r \) fixed

**Theorem (Agterberg et al. (2022))**

Suppose that \( \kappa, \mu_0, \) and \( \kappa_\sigma \) are bounded, and that \( r \) is fixed. Suppose that

\[
\max \left\{ \frac{\log(d)}{\text{SNR}}, \max_j \frac{\|\Sigma_i^{1/2} V_j\|_3^3}{\|\Sigma_i^{1/2} V_j\|_3} \right\} \to 0
\]

as \( n, d \to \infty \), with \( d \geq n \geq \log(d) \). Define

\[
S^{(i)} := \Lambda^{-1} V^\top \Sigma_i V \Lambda^{-1}.
\]

Then there exists a sequence of orthogonal matrices \( \mathcal{O}_* \) such that

\[
(S^{(i)})^{-1/2}(\hat{U}\mathcal{O}_* - U)_i. \to N(0, I_r).
\]
The Problem

Asymptotic Normality: \( r \) growing

Theoretical Results

Theorem (Agterberg et al. (2022))

Suppose that \( \kappa, \mu_0, \) and \( \kappa_\sigma \) are bounded. Suppose that

\[
\max \left\{ \frac{r \log(d)}{\sqrt{n}}, \frac{r \log(d)}{\text{SNR}}, \frac{\|\Sigma_i^{1/2} V_j\|^3_3}{\|\Sigma_i^{1/2} V_j\|^3} \right\} \to 0
\]

as \( n, d \to \infty, \) with \( d \geq n \geq \log(d) \). Define

\[
\sigma_{ij}^2 := \frac{\|\Sigma_i^{1/2} V_j\|^2_2}{\lambda_j^2}.
\]

Then there exists a sequence of orthogonal matrices \( \mathcal{O}_* \) such that

\[
\frac{1}{\sigma_{ij}} (\widehat{U} \mathcal{O}_* - U)_{ij} \to N(0, 1).
\]
Figure: 1000 MonteCarlo iterations of the first row of $\hat{U}O_\ast - U$ with $n = d = 1800$, under a three component mixture model with spherical (identity) covariances within each component.
Figure: 1000 MonteCarlo iterations of the first row of $\Lambda(\hat{U}O_* - U)$ with $n = d = 1800$, under a three component mixture model with both spherical and elliptical covariances within the first component.