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

Signal \text{ Plus Noise} = \text{Observation}
Signal Plus Noise Model

Signal: $\mathbf{z}$

Noise: $\mathbf{z}$

Observation: $\mathbf{z}$

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.
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*. 

...the geometric relationship between the signal matrix, the covariance structure of the noise, and the distribution of the errors...
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*.

“...the geometric relationship between the *signal matrix*, the *covariance structure of the noise*, and the distribution of the errors...”
General Model

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

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

*signal* *noise*
General Model

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:

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

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 $M + E$. 
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.
**HeteroPCA Algorithm (Zhang et al., 2022)**

**Algorithm 1:** HeteroPCA Algorithm of Zhang et al. (2022)

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

**while** $T \leq T_{\text{max}}$ **do**

- $\tilde{N}_T := \text{SVD}_r(N_T)$, the best rank $r$ 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 \( V \).
Understanding the Limiting Variance

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

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

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

$$\frac{\lambda_j}{\sigma_j} (\widehat{U} O_* - U)_{ij} \to N(0, 1).$$
Understanding the Limiting Variance

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

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

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

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

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

- Recall that $S^{(i)} = \Lambda^{-1} V^\top \sum_i V \Lambda^{-1}$
- 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 $\Lambda(\hat{U}O_* - U)_1 \approx N(0, V^\top \Sigma_1 V)$ and hence

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

$$= \begin{pmatrix} 5 & 0 \\ 0 & 5\theta \end{pmatrix} + \begin{pmatrix} .1 & 0 \\ 0 & .1 \end{pmatrix}$$
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 $\Lambda(\hat{U}O_\star - U)_1 \approx N(0, V^\top \Sigma_1 V)$ 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 $\Lambda(\hat{U}\mathcal{O}_* - U)$ with $n = d = 1800$, where the covariance changes according to previous slide.
Outline

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

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

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

- 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

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

- 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

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

- 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.
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 2022. ISSN 1557-9654. doi: 10.1109/TIT.2022.3159085.

Anru R. Zhang, T. Tony Cai, and Yihong Wu. Heteroskedastic PCA: Algorithm, optimality, and applications. *The Annals of Statistics*, 50(1):53–80, February 2022. ISSN 0090-5364, 2168-8966. doi: 10.1214/21-AOS2074.
Thank you!

🐦: @JAgterberger
Theorem (Agterberg et al. (2022))

Suppose some technical and regularity conditions hold. Suppose that

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

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

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

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

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

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

Suppose some technical and regularity conditions hold. Suppose that

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

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 \( O_* \) 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 \( \Sigma_i \) interacts with \( \Lambda \) and \( V \).
More Explanation

We require that

$$\max \left\{ \frac{\log(d)}{\text{SNR}}, \max_j \frac{\| \Sigma_j^{1/2} V_j \|^3_3}{\| \Sigma_j^{1/2} V_j \|^3} \right\} \rightarrow 0$$
We require that

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

condition on SNR

Interaction between noise and signal
More Explanation

- We require that

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

- Interaction between noise and signal

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

Then

\[
\frac{\| \Sigma_j^{1/2} V_j \|_3^3}{\| \Sigma_j^{1/2} V_j \|_3^3} = \frac{\| V_j \|_3^3}{\| V_j \|_3^3} = \frac{\sum_{i=1}^{d} \left( \frac{1}{\sqrt{d}} \right)^3}{1} = \frac{1}{\sqrt{d}}
\]
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

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

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 $\hat{M} = \text{Eigenvectors of } \hat{M}\hat{M}^\top$

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

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

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

Problem

If $\Sigma_i$'s are different (i.e. heteroskedastic), then the singular vectors of $\hat{M}$ are approximating a deterministic diagonal perturbation of $MM^\top$. 
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 $M$
Important parameters:

- Measure of noise: \( \sigma^2 := \max_i \| \Sigma_i \| \)
- Measure of signal: \( \lambda_r = \text{smallest nonzero singular value of } M \)
- 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: $\lambda_r = \text{smallest nonzero singular value of } M$
- 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 $n, d \to \infty$. 
Notation

Important parameters:

- Measure of noise: $\sigma^2 := \max_i \| \Sigma_i \|
- Measure of signal: $\lambda_r = \text{smallest nonzero singular value of } M$
- 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 $n, d \to \infty$.
- Condition number of $M$, $\kappa := \frac{\lambda_1}{\lambda_r}$
New concept:

- *Covariance Condition Number*:

\[
\kappa_{\sigma} := \max_{i,j} \frac{\sigma}{\| \sum_{i}^{1/2} V_{i,j} \|}
\]
New concept:  

- **Covariance Condition Number**:

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

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

\textit{Covariance Condition Number.}

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

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**:

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

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

- Consider the following special case:
  - \( \Sigma_i \equiv I_d \) for all \( i \) (or any multiple)
  - Then \( \kappa_\sigma \equiv 1 \)
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 \( \kappa_\sigma \equiv 1 \)

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

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

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \quad \mu_0 = \sqrt{\frac{n}{r}}$$

versus

$$\begin{pmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & \ddots & \vdots \\ \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix} \quad \mu_0 = 1$$
Asymptotic Normality: \( r \) fixed

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

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

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

as \( n, d \rightarrow \infty \), 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 \( O_* \) such that

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

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

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

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

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 $O_*$ such that

$$\frac{1}{\sigma_{ij}} (\hat{U} O_* - U)_{ij} \to N(0, 1).$$
Figure: 1000 Monte Carlo iterations of the first row of $\hat{U}O_* - U$ with $n = d = 1800$, under a three component mixture model with spherical (identity) covariances within each component.
Simulation II

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.