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

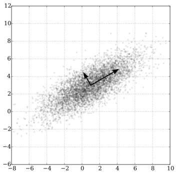
Outline

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

Motivation: Spectral Methods

Spectral methods are ubiquitous in machine learning and statistics.

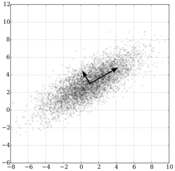
- Spectral Clustering
- Principal Components Analysis
- Nonconvex algorithm initializations (tensor SVD, phase retrieval, blind deconvolution)



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)



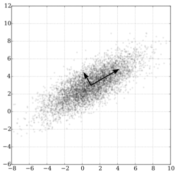
Problem

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)



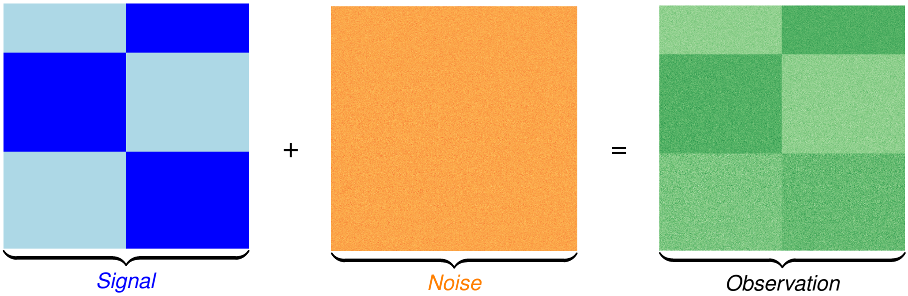
Problem

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

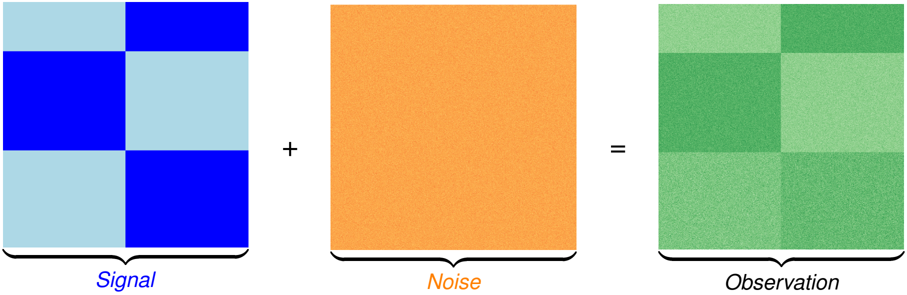
Goal

Develop *fine-grained statistical theory* for spectral methods.

Signal Plus Noise Model



Signal Plus Noise Model



Goal: More Precise

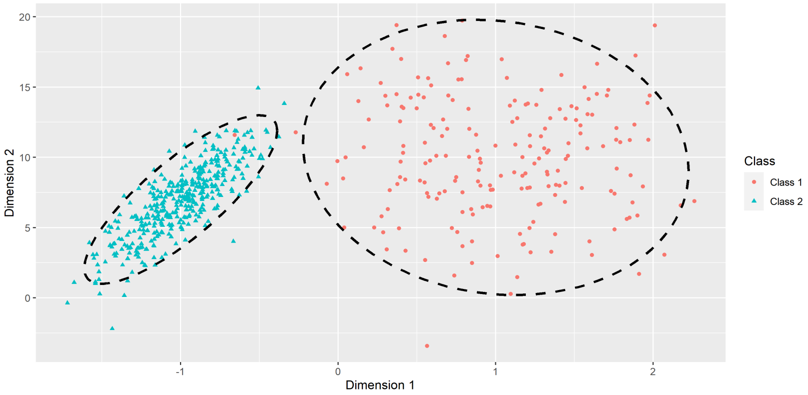
Develop *fine-grained statistical theory* for **an estimator of the left singular subspace of the signal matrix.**

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.**

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.**

"...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:

$$\widehat{M} = \underbrace{M}_{\text{signal}} + \underbrace{E}_{\text{noise}}.$$

General Model

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

$$\widehat{M} = \underbrace{M}_{\text{signal}} + \underbrace{E}_{\text{noise}}.$$

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 $U_{\cdot j}$ are orthonormal unit vectors)
- Λ 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:

$$\widehat{M} = \underbrace{M}_{\text{signal}} + \underbrace{E}_{\text{noise}}.$$

The noise matrix E has

- independent, mean-zero rows of the form $E_{i\cdot} = \Sigma_i^{1/2} Y_i$
- $\Sigma_i \in \mathbb{R}^{d \times d}$ 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 $\widehat{U}$ 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 $\widehat{U}$ 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 $\widehat{U}$ 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 : $N_0 = \widehat{M}\widehat{M}^\top - \mathrm{diag}(\widehat{M}\widehat{M}^\top)$, max number of iterations $T_{\max}$

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

  $\widetilde{N}_T := \mathrm{SVD}_r(N_T)$, the best rank r approximation to N_T;

  $N_{T+1} := N_T - \mathrm{diag}(N_T) + \mathrm{diag}(\widetilde{N}_T)$;

end

Return: $\widehat{U}$ = Left r singular vectors of $N_{T_{\max}}$

Outline

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 $\mathcal{O}_*$ such that

$$(S^{(i)})^{-1/2} (\widehat{U}\mathcal{O}_* - U)_{i\cdot} \to N(0, I_r).$$

Asymptotic Normality: r fixed

Theorem (Agterberg et al. (2022))

Suppose some technical and regularity conditions hold, 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 $\mathcal{O}_*$ such that

$$(S^{(i)})^{-1/2} (\widehat{U}\mathcal{O}_* - U)_{i\cdot} \to N(0, I_r).$$

Asymptotic covariance of i'th row of $\widehat{U}$ depends on how i'th row of noise matrix E interacts with Λ 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

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

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

$$\frac{\lambda_j}{\sigma_i}\left(\widehat{U}\mathcal{O}_* - U\right)_{ij} \to N(0,1).$$

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

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

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

$$\frac{\lambda_j}{\sigma_i}\left(\widehat{U}\mathcal{O}_* - U\right)_{ij} \to N(0,1).$$

Asymptotic variance of entries of j'th estimated singular vector is proportional to j'th singular value.

Outline

Setup

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

Setup

- Recall that $S^{(i)} = \Lambda^{-1} V^{\top} \Sigma_i V \Lambda^{-1}$
- Consider

$$\Sigma_1 := 5 V_{\cdot 1} V_{\cdot 1}{}^{\top} + 5 V_{\theta} V_{\theta}^{\top} + .1 I_d$$

where V_{θ} satisfies $V_{\theta}^{\top} V_{\cdot 2} = \theta$ and is orthogonal to $V_{\cdot 1}$

Setup

- Recall that $S^{(i)} = \Lambda^{-1} V^\top \Sigma_i V \Lambda^{-1}$
- Consider

$$\Sigma_1 := 5 V_{\cdot 1} V_{\cdot 1}{}^\top + 5 V_\theta V_\theta^\top + .1 I_d$$

where V_θ satisfies $V_\theta^\top V_{\cdot 2} = \theta$ and is orthogonal to $V_{\cdot 1}$

- Theory suggests that $\Lambda(\widehat{U}\mathcal{O}_* - U)_{1\cdot} \approx N(0, V^\top \Sigma_1 V)$ and hence

$$V^\top \Sigma_1 V = V^\top \left(5 V_{\cdot 1} V_{\cdot 1}{}^\top + 5 V_\theta V_\theta^\top + .1 I_d \right) V = \begin{pmatrix} 5 & 0 \\ 0 & 5\theta \end{pmatrix} + \begin{pmatrix} .1 & 0 \\ 0 & .1 \end{pmatrix}$$

Setup

- Recall that $S^{(i)} = \Lambda^{-1} V^\top \Sigma_i V \Lambda^{-1}$
- Consider

$$\Sigma_1 := 5 V_{\cdot 1} V_{\cdot 1}{}^\top + 5 V_\theta V_\theta^\top + .1 I_d$$

  where V_θ satisfies $V_\theta^\top V_{\cdot 2} = \theta$ and is orthogonal to $V_{\cdot 1}$

- Theory suggests that $\Lambda(\widehat{U}\mathcal{O}_* - U)_{1\cdot} \approx N(0, V^\top \Sigma_1 V)$ and hence

$$V^\top \Sigma_1 V = V^\top \left(5 V_{\cdot 1} V_{\cdot 1}{}^\top + 5 V_\theta V_\theta^\top + .1 I_d \right) V$$
$$= \begin{pmatrix} 5 & 0 \\ 0 & 5\theta \end{pmatrix} + \begin{pmatrix} .1 & 0 \\ 0 & .1 \end{pmatrix}$$

- So decreasing θ decreases the limiting variance along the second dimension

Simulation Result

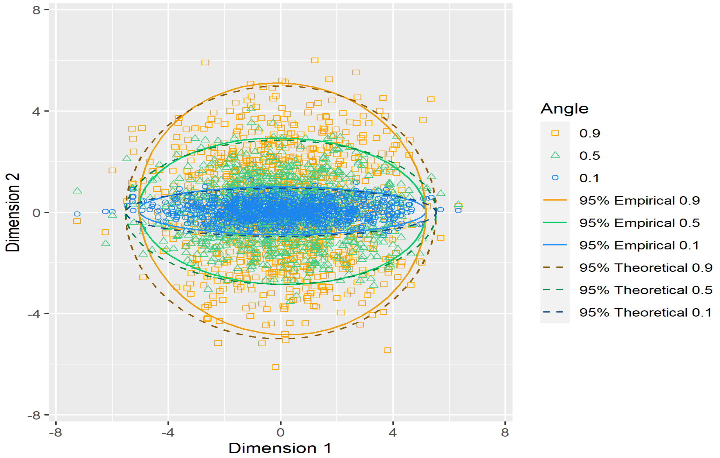


Figure: 1000 MonteCarlo iterations of the first row of $\Lambda(\widehat{U}\mathcal{O}_* - U)$ with $n = d = 1800$, where the covariance changes according to previous slide.

Outline

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*

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 $\widehat{M}\widehat{M}^{\top}$

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 $\widehat{M}\widehat{M}^\top$
- 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 $\widehat{M}\widehat{M}^\top$
- 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.

References I

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

Asymptotic Normality: r fixed

Theorem (Agterberg et al. (2022))

Suppose some technical and regularity conditions hold. Suppose that

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

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*

$$(S^{(i)})^{-1/2}(\widehat{U}\mathcal{O}_* - U)_{i\cdot} \to N(0, I_r).$$

Asymptotic Normality: r fixed

Theorem (Agterberg et al. (2022))

Suppose some technical and regularity conditions hold. Suppose that

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

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

$$(S^{(i)})^{-1/2} (\widehat{U}\mathcal{O}_* - U)_{i\cdot} \to N(0, I_r).$$

Asymptotic covariance of i'th row of $\widehat{U}$ depends on how Σ_i interacts with Λ and V.

More Explanation

- We require that

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

More Explanation

- We require that

$$\max\left\{ \underbrace{\frac{\log(d)}{\mathrm{SNR}}}_{\text{condition on SNR}}, \underbrace{\max_j \frac{\|\Sigma_i^{1/2} V_{\cdot j}\|_3^3}{\|\Sigma_i^{1/2} V_{\cdot j}\|^3}}_{\text{Interaction between noise and signal}} \right\} \to 0$$

More Explanation

- We require that

$$\max\left\{ \underbrace{\frac{\log(d)}{\mathrm{SNR}}}_{\text{condition on SNR}}, \underbrace{\max_j \frac{\|\Sigma_i^{1/2} V_{\cdot j}\|_3^3}{\|\Sigma_i^{1/2} V_{\cdot j}\|^3}}_{\substack{\text{Interaction between} \\ \text{noise and signal}}} \right\} \to 0$$

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

$$\frac{\|\Sigma_i^{1/2} V_{\cdot j}\|_3^3}{\|\Sigma_i^{1/2} V_{\cdot j}\|^3} = \frac{\|V_{\cdot j}\|_3^3}{\|V_{\cdot j}\|^3} = \frac{\sum_{l=1}^d \left(\frac{1}{\sqrt{d}}\right)^3}{1} = \frac{1}{\sqrt{d}}$$

Asymptotic Normality: r growing

Theorem (Agterberg et al. (2022))

Suppose some technical and regularity conditions hold. Suppose that

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

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

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

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

$$\frac{1}{\sigma_{ij}}\left(\widehat{U}\mathcal{O}_* - U\right)_{ij} \to N(0,1).$$

Bias

$$
\begin{aligned}
\text{Singular vectors of } \widehat{M} &= \text{Eigenvectors of } \widehat{M}\widehat{M}^\top \\
&\approx \text{Eigenvectors of } \mathbb{E}(\widehat{M}\widehat{M}^\top) \\
&= \text{Eigenvectors of } MM^\top + D, \\
&\quad \text{where } D_{ii} = \mathrm{Trace}(\Sigma_i).
\end{aligned}
$$

Bias

$$
\begin{aligned}
\text{Singular vectors of } \widehat{M} &= \text{Eigenvectors of } \widehat{M}\widehat{M}^\top \\
&\approx \text{Eigenvectors of } \mathbb{E}(\widehat{M}\widehat{M}^\top) \\
&= \text{Eigenvectors of } MM^\top + D, \\
&\quad \text{where } D_{ii} = \mathrm{Trace}(\Sigma_i).
\end{aligned}
$$

Problem

If Σ_i's are different (i.e. *heteroskedastic*), then the singular vectors of $\widehat{M}$ are approximating a *deterministic diagonal perturbation* of $MM^\top$.

Correcting the Bias

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

Correcting the Bias

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

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

Correcting the Bias

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

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

- 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

Notation

Important parameters:

- Measure of noise: $\sigma^2 := \max_i \|\Sigma_i\|$
- Measure of signal: λ_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:

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

Notation

Important parameters:

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

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

In the homoskedastic setting, $\mathrm{SNR} \to \infty$ 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: $\lambda_r =$ smallest nonzero singular value of M
- Define the signal-to-noise ratio:

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

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

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

Notation

New concept:

- *Covariance Condition Number*:

$$\kappa_\sigma := \max_{i,j} \frac{\sigma}{\|\Sigma_i^{1/2} V_{\cdot j}\|}$$

Notation

New concept:

- *Covariance Condition Number*:

$$\kappa_\sigma := \max_{i,j} \frac{\sigma}{\|\Sigma_i^{1/2} V_{\cdot j}\|}$$

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

Notation

New concept:

- *Covariance Condition Number*:

$$\kappa_\sigma := \max_{i,j} \frac{\sigma}{\|\Sigma_i^{1/2} V_{\cdot j}\|}$$

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

- Consider the following special case:

Notation

New concept:

- *Covariance Condition Number*:

$$\kappa_\sigma := \max_{i,j} \frac{\sigma}{\|\Sigma_i^{1/2} V_{\cdot j}\|}$$

  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*:

$$\kappa_\sigma := \max_{i,j} \frac{\sigma}{\|\Sigma_i^{1/2} V_{\cdot j}\|}$$

  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 $\kappa_\sigma \equiv 1$
- κ_σ only blows up when $\|\Sigma_i^{1/2} V_{\cdot j}\|$ is very small relative to the overall noise σ (nondegeneracy condition)

Notation

Incoherence parameter:

- Incoherence parameter μ_0 of the matrix M:

$$\max_i \|U_{i\cdot}\|, \|V_{i\cdot}\| \leq \mu_0 \sqrt{\frac{r}{n}}$$

Measures “spikiness” of M

Notation

Incoherence parameter:

- Incoherence parameter μ_0 of the matrix M:

$$\max_i \|U_{i\cdot}\|, \|V_{i\cdot}\| \leq \mu_0 \sqrt{\frac{r}{n}}$$

  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 μ_0 of the matrix M:

$$\max_i \|U_{i\cdot}\|, \|V_{i\cdot}\| \le \mu_0 \sqrt{\frac{r}{n}}$$

Measures “spikiness” of M

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

$$\underbrace{\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}}_{\mu_0 = \sqrt{\frac{n}{r}}} \text{ versus } \underbrace{\begin{pmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & \ddots & \vdots \\ \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}}_{\mu_0 = 1}$$

Asymptotic Normality: r fixed

Theorem (Agterberg et al. (2022))

Suppose that κ, μ_0, and κ_σ are bounded, and that r is fixed. Suppose that

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

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

$$(S^{(i)})^{-1/2}(\widehat{U}\mathcal{O}_* - U)_{i\cdot} \to N(0, I_r).$$

Asymptotic Normality: r growing

Theorem (Agterberg et al. (2022))

Suppose that κ, μ_0, and κ_σ are bounded. Suppose that

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

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

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

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

$$\frac{1}{\sigma_{ij}}\left(\widehat{U}\mathcal{O}_* - U\right)_{ij} \to N(0,1).$$

Simulation I

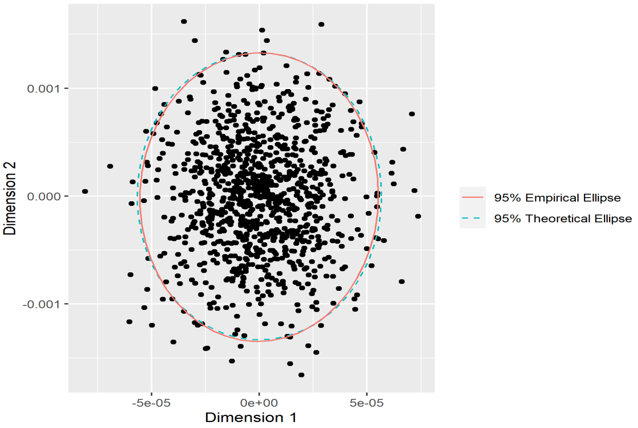


Figure: 1000 MonteCarlo iterations of the first row of $\widehat{U}\mathcal{O}_* - U$ with $n = d = 1800$, under a three component mixture model with spherical (identity) covariances within each component.

Simulation II

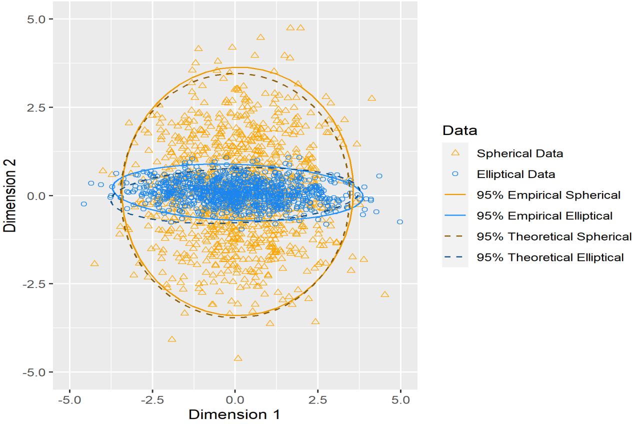


Figure: 1000 MonteCarlo iterations of the first row of $\Lambda(\widehat{U}\mathcal{O}_* - U)$ with $n = d = 1800$, under a three component mixture model with both spherical and elliptical covariances within the first component.