Estimating Shared Subspace with AJIVE: Power and Limitation of Multiple Data Matrices

Cong Ma

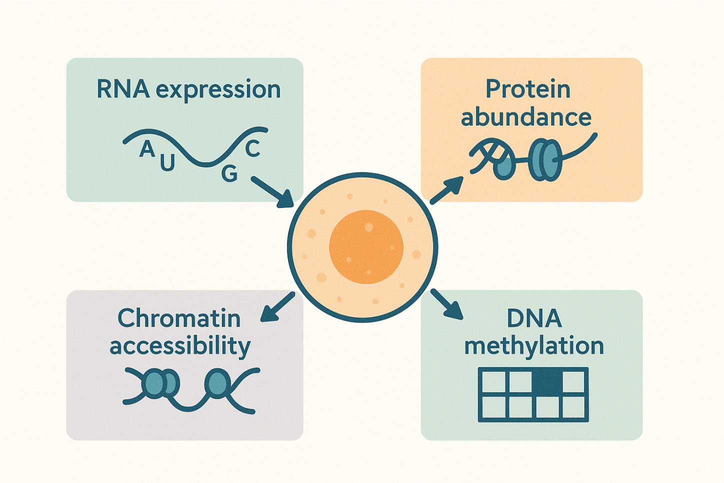
Department of Statistics, UChicago

Statistics Seminar, UC Davis, Apr. 2025

Yuepeng Yang

UChicago Statistics → Yale Statistics

Multimodal single-cell data



Each modality captures a different biological view

Multimodal data are ubiquitous

Examples with multiple high-dimensional data types

Field	**Object**	**Data types**
Computational biology	Tissue samples	Gene expression, microRNA, genotype, protein abundance/activity
Chemometrics	Chemicals	Mass spectra, NMR spectra, atomic composition
Atmospheric sciences	Locations	Temperature, humidity, particle concentrations over time
Internet traffic	Websites	Word frequencies, visitor demographics, linked pages

— from Lock et al. '13

Multimodal data are ubiquitous

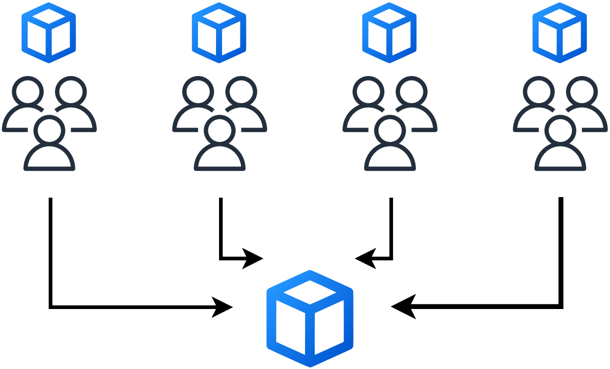
Examples with multiple high-dimensional data types

Field	**Object**	**Data types**
Computational biology	Tissue samples	Gene expression, microRNA, genotype, protein abundance/activity
Chemometrics	Chemicals	Mass spectra, NMR spectra, atomic composition
Atmospheric sciences	Locations	Temperature, humidity, particle concentrations over time
Internet traffic	Websites	Word frequencies, visitor demographics, linked pages

— from Lock et al. '13

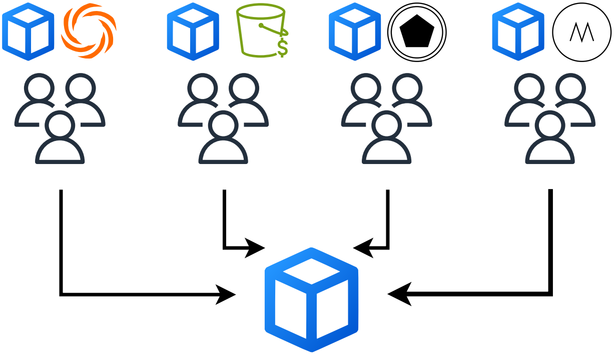
How to integrate information across different data types?

Learning shared structure



see e.g., Argelaguet et al. '18, Fan el al. '19, Arroyo et al. '22

Learning shared and unique structures



see e.g., Lock et al. '18, Lin and Zhang '23, Prothero et al. '24

Two key questions

- **Identification**: How to define shared and unique structures?
- **Estimation**: How to estimate shared and unique structures?

Joint and Individual Variation Explained (JIVE)

— Lock et el. '13

JIVE

We observe K matrices $\{\boldsymbol{A}_k\}_{1\leq k\leq K}$ with $\boldsymbol{A}_k \in \mathbb{R}^{n\times d_k}$ and

$$\boldsymbol{A}_k = \underbrace{\boldsymbol{U}^\star \boldsymbol{V}_k^{\star\top}}_{\substack{\text{rank}-r \\ \text{shared component}}} + \underbrace{\boldsymbol{U}_k^\star \boldsymbol{W}_k^{\star\top}}_{\substack{\text{rank}-r_k \\ \text{unique component}}} + \underbrace{\boldsymbol{E}_k}_{\text{Noise}}$$

- $\boldsymbol{U}^\star \in \mathbb{O}^{n\times r}$ is shared column space
- $\boldsymbol{U}_k^\star \in \mathbb{O}^{n\times r_k}$ are unique column spaces with $\boldsymbol{U}_k \perp \boldsymbol{U}$
- $\boldsymbol{V}_k^\star \in \mathbb{R}^{d_k\times r}$ and $\boldsymbol{W}_k^\star \in \mathbb{R}^{d_k\times r_k}$ are loading matrices

Two key questions

- **Identification**: How to define shared and unique structures?
- **Estimation**: How to estimate shared and unique structures?

Defining shared and unique structures in JIVE

JIVE

$$A_k^\star = \underbrace{U^\star V_k^{\star\top}}_{\substack{\text{rank}-r \\ \text{shared component}}} + \underbrace{U_k^\star W_k^{\star\top}}_{\substack{\text{rank}-r_k \\ \text{unique component}}}$$

JIVE defines shared information to be shared subspace $\cap_{k=1}^K \mathrm{col}(A_k^\star)$

Defining shared and unique structures in JIVE

JIVE

$$\boldsymbol{A}_k^\star = \underbrace{\boldsymbol{U}^\star \boldsymbol{V}_k^{\star\top}}_{\substack{\text{rank}-r\\ \text{shared component}}} + \underbrace{\boldsymbol{U}_k^\star \boldsymbol{W}_k^{\star\top}}_{\substack{\text{rank}-r_k\\ \text{unique component}}}$$

JIVE defines shared information to be shared subspace $\cap_{k=1}^K \mathrm{col}(\boldsymbol{A}_k^\star)$

But, does $\mathrm{col}(\boldsymbol{U}^\star) = \cap_{k=1}^K \mathrm{col}(\boldsymbol{A}_k^\star)$?

Faithfulness: $\mathrm{col}(\boldsymbol{U}^\star) \subset \cap_{k=1}^{K} \mathrm{col}(\boldsymbol{A}_k^\star)$

JIVE

$$\boldsymbol{A}_k^\star = \underbrace{\boldsymbol{U}^\star \boldsymbol{V}_k^{\star\top}}_{\substack{\text{rank}-r \\ \text{shared component}}} + \underbrace{\boldsymbol{U}_k^\star \boldsymbol{W}_k^{\star\top}}_{\substack{\text{rank}-r_k \\ \text{unique component}}}$$

- Counterexample: $\boldsymbol{V}_k^\star = \boldsymbol{W}_k^\star$

Faithfulness: $\mathrm{col}(\boldsymbol{U}^\star) \subset \cap_{k=1}^{K} \mathrm{col}(\boldsymbol{A}_k^\star)$

JIVE

$$\boldsymbol{A}_k^\star = \underbrace{\boldsymbol{U}^\star \boldsymbol{V}_k^{\star\top}}_{\substack{\text{rank}-r \\ \text{shared component}}} + \underbrace{\boldsymbol{U}_k^\star \boldsymbol{W}_k^{\star\top}}_{\substack{\text{rank}-r_k \\ \text{unique component}}}$$

- Counterexample: $\boldsymbol{V}_k^\star = \boldsymbol{W}_k^\star$
- Faithfulness is equivalent to assuming

$$\mathrm{rank}(\boldsymbol{A}_k^\star) = r + r_k$$

Exhaustiveness: $\cap_{k=1}^{K} \mathrm{col}(\boldsymbol{A}_k^\star) \subset \mathrm{col}(\boldsymbol{U}^\star)$

JIVE

$$\boldsymbol{A}_k^\star = \underbrace{\boldsymbol{U}^\star \boldsymbol{V}_k^{\star\top}}_{\substack{\text{rank}-r \\ \text{shared component}}} + \underbrace{\boldsymbol{U}_k^\star \boldsymbol{W}_k^{\star\top}}_{\substack{\text{rank}-r_k \\ \text{unique component}}}$$

- Counterexample: $\{\boldsymbol{U}_k^\star\}$ are identical

Exhaustiveness: $\cap_{k=1}^{K}\mathrm{col}(\boldsymbol{A}_k^\star) \subset \mathrm{col}(\boldsymbol{U}^\star)$

JIVE

$$\boldsymbol{A}_k^\star = \underbrace{\boldsymbol{U}^\star \boldsymbol{V}_k^{\star\top}}_{\substack{\text{rank}-r \\ \text{shared component}}} + \underbrace{\boldsymbol{U}_k^\star \boldsymbol{W}_k^{\star\top}}_{\substack{\text{rank}-r_k \\ \text{unique component}}}$$

- Counterexample: $\{\boldsymbol{U}_k^\star\}$ are identical
- Exhaustiveness is equivalent to assuming

$$\cap_{k=1}^{K}\mathrm{col}(\boldsymbol{U}_k^\star) = \varnothing$$

Exhaustiveness: $\cap_{k=1}^{K} \mathrm{col}(\boldsymbol{A}_k^\star) \subset \mathrm{col}(\boldsymbol{U}^\star)$

JIVE

$$\boldsymbol{A}_k^\star = \underbrace{\boldsymbol{U}^\star \boldsymbol{V}_k^{\star\top}}_{\substack{\text{rank}-r \\ \text{shared component}}} + \underbrace{\boldsymbol{U}_k^\star \boldsymbol{W}_k^{\star\top}}_{\substack{\text{rank}-r_k \\ \text{unique component}}}$$

- Counterexample: $\{\boldsymbol{U}_k^\star\}$ are identical
- Exhaustiveness is equivalent to assuming

$$\cap_{k=1}^{K} \mathrm{col}(\boldsymbol{U}_k^\star) = \varnothing$$

Now, $\boldsymbol{U}^\star$ is identifiable since $\mathrm{col}(\boldsymbol{U}^\star) = \cap_{k=1}^{K} \mathrm{col}(\boldsymbol{A}_k^\star)$

Two key questions

- **Identification**: How to define shared and unique structures?
- **Estimation**: How to estimate shared and unique structures?

Angle-based JIVE (AJIVE)

— Feng et al. '18

AJIVE: a two-step spectral method

1. Estimate shared + unique column space of each $\boldsymbol{A}_k$:
   Let $\widetilde{\boldsymbol{U}}_k$ be top-$(r + r_k)$ left singular vectors of $\boldsymbol{A}_k$
2. Estimate shared column space:
   Let $\widehat{\boldsymbol{U}}$ be the top-r eigenvectors of $\sum_{k=1}^{K} \widetilde{\boldsymbol{U}}_k \widetilde{\boldsymbol{U}}_k^\top$

Angle-based JIVE (AJIVE)

— Feng et al. '18

AJIVE: a two-step spectral method

1. Estimate shared + unique column space of each $\boldsymbol{A}_k$:
   Let $\widetilde{\boldsymbol{U}}_k$ be top-$(r + r_k)$ left singular vectors of $\boldsymbol{A}_k$

2. Estimate shared column space:
   Let $\widehat{\boldsymbol{U}}$ be the top-r eigenvectors of $\sum_{k=1}^{K} \widetilde{\boldsymbol{U}}_k \widetilde{\boldsymbol{U}}_k^\top$

- In the noiseless case, AJIVE recovers $\boldsymbol{U}^\star$ exactly

Angle-based JIVE (AJIVE)

— Feng et al. '18

AJIVE: a two-step spectral method

1. Estimate shared + unique column space of each $\boldsymbol{A}_k$:
   Let $\widetilde{\boldsymbol{U}}_k$ be top-$(r + r_k)$ left singular vectors of $\boldsymbol{A}_k$

2. Estimate shared column space:
   Let $\widehat{\boldsymbol{U}}$ be the top-r eigenvectors of $\sum_{k=1}^{K} \widetilde{\boldsymbol{U}}_k \widetilde{\boldsymbol{U}}_k^\top$

- In the noiseless case, AJIVE recovers $\boldsymbol{U}^\star$ exactly
- How does AJIVE perform with noisy observations?

Angle-based JIVE (AJIVE)

— Feng et al. '18

AJIVE: a two-step spectral method

1. Estimate shared + unique column space of each $\boldsymbol{A}_k$:
   Let $\widetilde{\boldsymbol{U}}_k$ be top-$(r + r_k)$ left singular vectors of $\boldsymbol{A}_k$
2. Estimate shared column space:
   Let $\widehat{\boldsymbol{U}}$ be the top-r eigenvectors of $\sum_{k=1}^{K} \widetilde{\boldsymbol{U}}_k \widetilde{\boldsymbol{U}}_k^\top$

- In the noiseless case, AJIVE recovers $\boldsymbol{U}^\star$ exactly
- How does AJIVE perform with noisy observations?

$\leftarrow$ focus of this talk

Theoretical results

Key problem parameters

Performance of AJIVE and hardness of shared subspace estimation depend on

- n, d, r, σ, and mininum singular value $\min_k \sigma_{r+r_k}(\boldsymbol{A}_k^{\star})$

Key problem parameters

Performance of AJIVE and hardness of shared subspace estimation depend on

- n, d, r, σ, and mininum singular value $\min_k \sigma_{r+r_k}(\boldsymbol{A}_k^\star)$
- K: number of matrices

— benefit of using multiple matrices?

Key problem parameters

Performance of AJIVE and hardness of shared subspace estimation depend on

- n, d, r, σ, and mininum singular value $\min_k \sigma_{r+r_k}(\boldsymbol{A}_k^\star)$
- K: number of matrices

  — benefit of using multiple matrices?
- θ: misalignment level of unique subspaces

Misalignment of unique subspaces

Recall for identifiability, we assume

$$\cap_{k=1}^{K}\mathrm{col}(\boldsymbol{U}_k^{\star}) = \varnothing$$

Misalignment of unique subspaces

Recall for identifiability, we assume

$$\cap_{k=1}^{K} \mathrm{col}(\boldsymbol{U}_k^\star) = \varnothing \quad \iff \quad \Bigg\| \underbrace{\frac{1}{K}\sum_{k=1}^{K} \boldsymbol{U}_k^\star {\boldsymbol{U}_k^\star}^\top}_{\text{Avg. Proj. Mat}} \Bigg\| < 1$$

Misalignment of unique subspaces

Recall for identifiability, we assume

$$\cap_{k=1}^{K} \mathrm{col}(\boldsymbol{U}_k^\star) = \varnothing \quad \iff \quad \Bigg\| \underbrace{\frac{1}{K}\sum_{k=1}^{K} \boldsymbol{U}_k^\star \boldsymbol{U}_k^{\star\top}}_{\text{Avg. Proj. Mat}} \Bigg\| < 1$$

Definition 1 (Misalignment)

We say collection of subspaces $\{\boldsymbol{U}_k^\star\}_{1\le k\le K}$ is θ-misaligned if

$$\Bigg\| \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{U}_k^\star \boldsymbol{U}_k^{\star\top} \Bigg\| \le 1-\theta$$

- θ tells us how misaligned unique subspaces are

Range of θ

θ-misalignment

$$\left\| \frac{1}{K} \sum_{k=1}^{K} \boldsymbol{U}_k^\star \boldsymbol{U}_k^{\star\top} \right\| \leq 1 - \theta$$

- Fully aligned: when $\cap_{k=1}^{K} \mathrm{col}(\boldsymbol{U}_k^\star) \neq \varnothing$, we have $\theta = 0$

Range of θ

θ-misalignment

$$\left\| \frac{1}{K} \sum_{k=1}^{K} \boldsymbol{U}_k^{\star} \boldsymbol{U}_k^{\star\top} \right\| \leq 1 - \theta$$

- Fully aligned: when $\cap_{k=1}^{K} \mathrm{col}(\boldsymbol{U}_k^{\star}) \neq \varnothing$, we have $\theta = 0$
- Fully misaligned: when $\{\boldsymbol{U}_k^{\star}\}_{1 \leq k \leq K}$ are orthonormal to each other, we have $\theta = 1 - 1/K$

Range of θ

θ-misalignment

$$\left\| \frac{1}{K} \sum_{k=1}^{K} \boldsymbol{U}_k^\star \boldsymbol{U}_k^{\star\top} \right\| \leq 1 - \theta$$

- Fully aligned: when $\cap_{k=1}^{K} \mathrm{col}(\boldsymbol{U}_k^\star) \neq \varnothing$, we have $\theta = 0$
- Fully misaligned: when $\{\boldsymbol{U}_k^\star\}_{1 \leq k \leq K}$ are orthonormal to each other, we have $\theta = 1 - 1/K$
- Any $\theta \in (0, 1 - 1/K]$ is realizable by some $\{\boldsymbol{U}_k^\star\}$

Performance guarantees of AJIVE

Let $\sigma_{\min} := \min_k \sigma_{r+r_k}(\boldsymbol{A}_k)$

For simplicity suppose $n = d_1 = \cdots = d_K$, $r = r_1 = \cdots = r_K \asymp 1$

Theorem 2 (Yang, Ma '25)

Assume $\frac{\sigma\sqrt{n}}{\sigma_{\min}} \ll \min\{\sqrt{\theta}, \sqrt{K}\theta\}$. *AJIVE obeys*

$$\left\| \widehat{\boldsymbol{U}}\widehat{\boldsymbol{U}}^\top - \boldsymbol{U}^\star \boldsymbol{U}^{\star\top} \right\| \lesssim \frac{\sigma}{\sigma_{\min}} \sqrt{\frac{n}{K} + \frac{r}{K\theta}} + \frac{1}{\theta(1 \wedge K\theta)} \cdot \frac{\sigma^2 n}{\sigma_{\min}^2}$$

Performance guarantees of AJIVE

Let $\sigma_{\min} := \min_k \sigma_{r+r_k}(\boldsymbol{A}_k)$

For simplicity suppose $n = d_1 = \cdots = d_K$, $r = r_1 = \cdots = r_K \asymp 1$

Theorem 2 (Yang, Ma '25)

Assume $\frac{\sigma\sqrt{n}}{\sigma_{\min}} \ll \min\{\sqrt{\theta}, \sqrt{K}\theta\}$. *AJIVE obeys*

$$\left\|\widehat{\boldsymbol{U}}\widehat{\boldsymbol{U}}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\right\| \lesssim \frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{K} + \frac{r}{K\theta}} + \frac{1}{\theta(1 \wedge K\theta)} \cdot \frac{\sigma^2 n}{\sigma_{\min}^2}$$

- $\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{K} + \frac{r}{K\theta}}$: first-order error in high-SNR regime
- $\frac{1}{\theta(1 \wedge K\theta)} \cdot \frac{\sigma^2 n}{\sigma_{\min}^2}$: second-order error in low-SNR regime

High-SNR regime

Minimax lower bounds for estimating $\mathrm{col}(\boldsymbol{U}^\star)$

Theorem 3 (Yang, Ma '25)

$$\inf_{\widetilde{\boldsymbol{U}}} \sup_{\{\boldsymbol{A}_k^\star\}} \mathbb{E}\left[\left\|\widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{U}}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\right\|\right] \gtrsim \frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{K} + \frac{r}{K\theta}}$$

Minimax lower bounds for estimating $\mathrm{col}(\boldsymbol{U}^\star)$

Theorem 3 (Yang, Ma '25)

$$\inf_{\widetilde{\boldsymbol{U}}} \sup_{\{\boldsymbol{A}_k^\star\}} \mathbb{E}\left[\left\|\widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{U}}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\right\|\right] \gtrsim \frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{K} + \frac{r}{K\theta}}$$

Recall upper bound of AJIVE when SNR is high:

$$\left\|\widehat{\boldsymbol{U}}\widehat{\boldsymbol{U}}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\right\| \lesssim \frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{K} + \frac{r}{K\theta}}$$

Minimax lower bounds for estimating $\mathrm{col}(\boldsymbol{U}^\star)$

Theorem 3 (Yang, Ma '25)

$$\inf_{\widetilde{\boldsymbol{U}}} \sup_{\{\boldsymbol{A}_k^\star\}} \mathbb{E}\left[\left\|\widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{U}}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\right\|\right] \gtrsim \frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{K} + \frac{r}{K\theta}}$$

Recall upper bound of AJIVE when SNR is high:

$$\left\|\widehat{\boldsymbol{U}}\widehat{\boldsymbol{U}}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\right\| \lesssim \frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{K} + \frac{r}{K\theta}}$$

AJIVE is minimax optimal in high-SNR regime

Understanding optimal rate under high SNR

JIVE

$$\boldsymbol{A}_k = \underbrace{\boldsymbol{U}^\star \boldsymbol{V}_k^{\star\top}}_{\substack{\text{rank}-r \\ \text{shared component}}} + \underbrace{\boldsymbol{U}_k^\star \boldsymbol{W}_k^{\star\top}}_{\substack{\text{rank}-r_k \\ \text{unique component}}} + \underbrace{\boldsymbol{E}_k}_{\text{Noise}}$$

- $\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{K}}$: optimal est. error when unique components are known

Understanding optimal rate under high SNR

JIVE

$$\boldsymbol{A}_k = \underbrace{\boldsymbol{U}^\star \boldsymbol{V}_k^{\star\top}}_{\substack{\text{rank}-r \\ \text{shared component}}} + \underbrace{\boldsymbol{U}_k^\star \boldsymbol{W}_k^{\star\top}}_{\substack{\text{rank}-r_k \\ \text{unique component}}} + \underbrace{\boldsymbol{E}_k}_{\text{Noise}}$$

- $\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{K}}$: optimal est. error when unique components are known
- $\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{r}{K\theta}}$: additional error due to unknown unique subspaces

Understanding optimal rate under high SNR

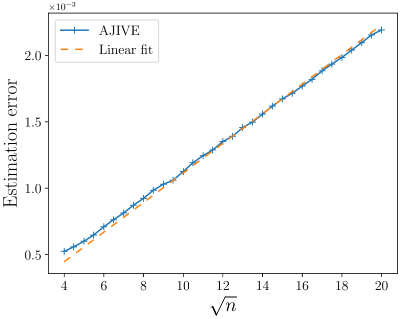
JIVE

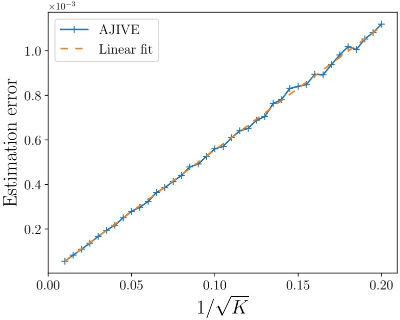
$$A_k = \underbrace{U^\star V_k^{\star\top}}_{\substack{\text{rank}-r \\ \text{shared component}}} + \underbrace{U_k^\star W_k^{\star\top}}_{\substack{\text{rank}-r_k \\ \text{unique component}}} + \underbrace{E_k}_{\text{Noise}}$$

- $\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{K}}$: optimal est. error when unique components are known
- $\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{r}{K\theta}}$: additional error due to unknown unique subspaces

Shared subspace estimation is harder as unique subspaces are more aligned, i.e., θ is smaller

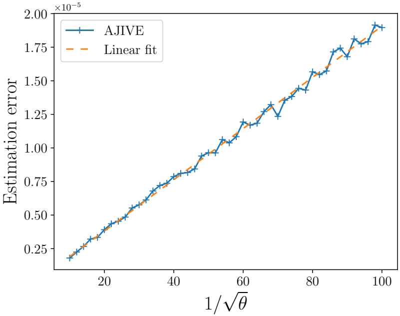
Empirical results: Large θ

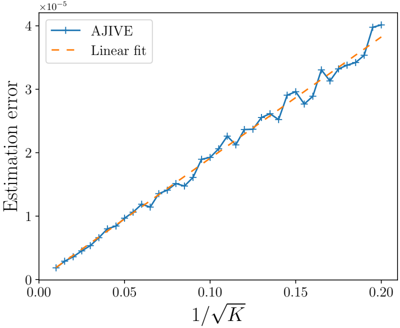




$$\theta = 0.5: \quad \|\widehat{\boldsymbol{U}\boldsymbol{U}}^\top - \boldsymbol{U}^\star \boldsymbol{U}^{\star\top}\| \propto \sqrt{n/K}$$

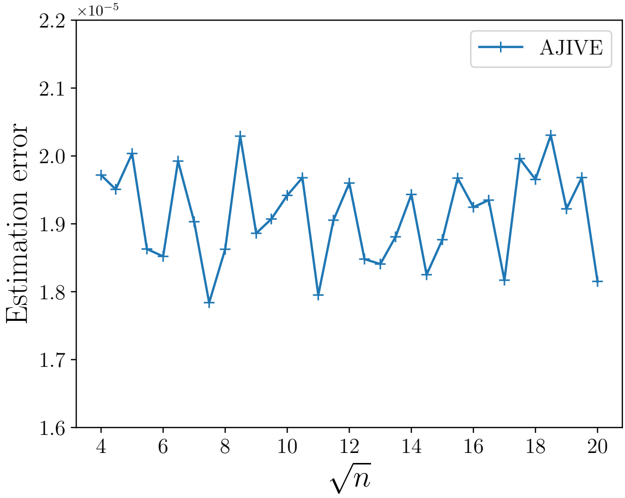
Empirical results: Small θ





$$\theta \in (0.0001, 0.01): \quad \|\widehat{\boldsymbol{U}}\widehat{\boldsymbol{U}}^\top - \boldsymbol{U}^\star \boldsymbol{U}^{\star\top}\| \propto \sqrt{1/K\theta}$$

Empirical results: Small θ



$$\theta = 0.0001$$

Low-SNR regime

Non-diminishing error in low-SNR regime

Revisiting upper bound of AJIVE when SNR is low:

$$\left\| \widehat{\boldsymbol{U}}\widehat{\boldsymbol{U}}^\top - \boldsymbol{U}^\star \boldsymbol{U}^{\star\top} \right\| \lesssim \frac{1}{\theta(1 \wedge K\theta)} \cdot \frac{\sigma^2 n}{\sigma_{\min}^2}$$

Non-diminishing error in low-SNR regime

Revisiting upper bound of AJIVE when SNR is low:

$$\left\| \widehat{\boldsymbol{U}}\widehat{\boldsymbol{U}}^\top - \boldsymbol{U}^\star \boldsymbol{U}^{\star\top} \right\| \lesssim \frac{1}{\theta(1 \wedge K\theta)} \cdot \frac{\sigma^2 n}{\sigma_{\min}^2}$$

Estimation error does NOT converge to 0 as K increases

Non-diminishing error in low-SNR regime

Revisiting upper bound of AJIVE when SNR is low:

$$\left\|\widehat{\boldsymbol{U}}\widehat{\boldsymbol{U}}^{\top}-\boldsymbol{U}^{\star}\boldsymbol{U}^{\star\top}\right\| \lesssim \frac{1}{\theta(1\wedge K\theta)}\cdot\frac{\sigma^2 n}{\sigma_{\min}^2}$$

Estimation error does NOT converge to 0 as K increases

Is this artifact in analysis or fundamental limitation?

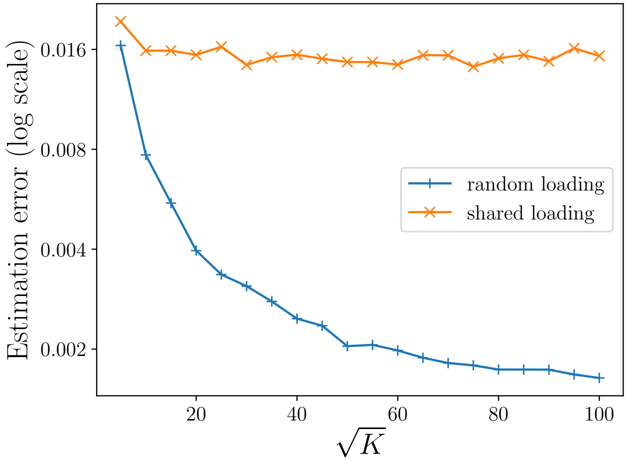
Two experimental settings

JIVE

$$\boldsymbol{A}_k = \underbrace{\boldsymbol{U}^\star \boldsymbol{V}_k^{\star\top}}_{\substack{\text{rank}-r \\ \text{shared component}}} + \underbrace{\boldsymbol{U}_k^\star \boldsymbol{W}_k^{\star\top}}_{\substack{\text{rank}-r_k \\ \text{unique component}}} + \underbrace{\boldsymbol{E}_k}_{\text{Noise}}$$

- Random loadings: $\boldsymbol{V}_k^\star$ and $\boldsymbol{W}_k^\star$ are independent random orthonormal matrices
- Shared loadings: Let $\boldsymbol{V}^\star$ and $\boldsymbol{W}^\star$ be random orthonormal matrices. Set $\boldsymbol{V}_k^\star = \boldsymbol{V}^\star$ and $\boldsymbol{W}_k^\star = \boldsymbol{W}^\star$ for all k

AJIVE has non-diminishing error



Shared vs random loadings on $\|\widehat{\boldsymbol{U}}\widehat{\boldsymbol{U}}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\|$ vs K

Intuition: Bias aligns

AJIVE

1. Let $\widetilde{\boldsymbol{U}}_k$ be top-$(r + r_k)$ left singular vectors of $\boldsymbol{A}_k$
2. Let $\widehat{\boldsymbol{U}}$ be top-r eigenvectors of $\sum_{k=1}^{K} \widetilde{\boldsymbol{U}}_k \widetilde{\boldsymbol{U}}_k^\top$

Intuition: Bias aligns

AJIVE

1. Let $\widetilde{\boldsymbol{U}}_k$ be top-$(r + r_k)$ left singular vectors of $\boldsymbol{A}_k$
2. Let $\widehat{\boldsymbol{U}}$ be top-r eigenvectors of $\sum_{k=1}^{K} \widetilde{\boldsymbol{U}}_k \widetilde{\boldsymbol{U}}_k^\top$

- SVD is biased: $\widetilde{\boldsymbol{U}}_k \widetilde{\boldsymbol{U}}_k^\top$ is biased estimate of column space of $\boldsymbol{A}_k^\star$

Intuition: Bias aligns

AJIVE

1. Let $\widetilde{\boldsymbol{U}}_k$ be top-$(r + r_k)$ left singular vectors of $\boldsymbol{A}_k$
2. Let $\widehat{\boldsymbol{U}}$ be top-r eigenvectors of $\sum_{k=1}^{K} \widetilde{\boldsymbol{U}}_k \widetilde{\boldsymbol{U}}_k^\top$

- SVD is biased: $\widetilde{\boldsymbol{U}}_k \widetilde{\boldsymbol{U}}_k^\top$ is biased estimate of column space of $\boldsymbol{A}_k^\star$
- Under shared loading, individual bias can be aligned, inducing a non-diminishing error in second step of AJIVE

Intuition: Bias aligns

AJIVE

1. Let $\widetilde{\boldsymbol{U}}_k$ be top-$(r + r_k)$ left singular vectors of $\boldsymbol{A}_k$
2. Let $\widehat{\boldsymbol{U}}$ be top-r eigenvectors of $\sum_{k=1}^{K} \widetilde{\boldsymbol{U}}_k \widetilde{\boldsymbol{U}}_k^\top$

- SVD is biased: $\widetilde{\boldsymbol{U}}_k \widetilde{\boldsymbol{U}}_k^\top$ is biased estimate of column space of $\boldsymbol{A}_k^\star$
- Under shared loading, individual bias can be aligned, inducing a non-diminishing error in second step of AJIVE

Is non-diminishing error fundamental to shared subspace estimation?

Oracle lower bound

Oracle spectral estimator

- Suppose unique components $\boldsymbol{U}_k^\star \boldsymbol{W}_k^{\star\top}$ are known, optimal estimator is top-r eigenspace of

$$\frac{1}{K}\sum_{k=1}^{K}\left(\boldsymbol{A}_k - \boldsymbol{U}_k^\star \boldsymbol{W}_k^{\star\top}\right)\left(\boldsymbol{A}_k - \boldsymbol{U}_k^\star \boldsymbol{W}_k^{\star\top}\right)^\top$$

Oracle spectral estimator

- Suppose unique components $\boldsymbol{U}_k^\star \boldsymbol{W}_k^{\star\top}$ are known, optimal estimator is top-r eigenspace of

$$\frac{1}{K}\sum_{k=1}^{K}\left(\boldsymbol{A}_k - \boldsymbol{U}_k^\star \boldsymbol{W}_k^{\star\top}\right)\left(\boldsymbol{A}_k - \boldsymbol{U}_k^\star \boldsymbol{W}_k^{\star\top}\right)^\top$$

- When unknown, replace $\boldsymbol{U}_k^\star \boldsymbol{W}_k^{\star\top}$ by estimate

Oracle spectral estimator

- Suppose unique components $\boldsymbol{U}_k^\star \boldsymbol{W}_k^{\star\top}$ are known, optimal estimator is top-r eigenspace of

$$\frac{1}{K}\sum_{k=1}^{K}\left(\boldsymbol{A}_k - \boldsymbol{U}_k^\star \boldsymbol{W}_k^{\star\top}\right)\left(\boldsymbol{A}_k - \boldsymbol{U}_k^\star \boldsymbol{W}_k^{\star\top}\right)^\top$$

- When unknown, replace $\boldsymbol{U}_k^\star \boldsymbol{W}_k^{\star\top}$ by estimate
  For instance, oracle-aided estimate

$$\text{top-}r_k \text{ SVD of} \quad \mathcal{P}_\star^\perp \boldsymbol{A}_k = \boldsymbol{U}_k^\star \boldsymbol{W}_k^{\star\top} + \mathcal{P}_\star^\perp \boldsymbol{E}_k,$$

  where $\mathcal{P}_\star^\perp := \boldsymbol{I} - \boldsymbol{U}^\star \boldsymbol{U}^{\star\top}$

Non-diminishing error of oracle estimator

Oracle spectral estimator

1. Let $\widehat{\boldsymbol{U}}_k \widehat{\boldsymbol{W}}_k^\top$ be top-r_k SVD of $\mathcal{P}_\star^\perp \boldsymbol{A}_k = \boldsymbol{U}_k^\star \boldsymbol{W}_k^{\star\top} + \mathcal{P}_\star^\perp \boldsymbol{E}_k$
2. Let $\widehat{\boldsymbol{U}}_{\text{oracle}}$ be top-r eigenspace of

$$\frac{1}{K}\sum_{k=1}^{K}\left(\boldsymbol{A}_k - \widehat{\boldsymbol{U}}_k \widehat{\boldsymbol{W}}_k^\top\right)\left(\boldsymbol{A}_k - \widehat{\boldsymbol{U}}_k \widehat{\boldsymbol{W}}_k^\top\right)^\top$$

Non-diminishing error of oracle estimator

Oracle spectral estimator

1. Let $\widehat{\boldsymbol{U}}_k \widehat{\boldsymbol{W}}_k^\top$ be top-r_k SVD of $\mathcal{P}_\star^\perp \boldsymbol{A}_k = \boldsymbol{U}_k^\star \boldsymbol{W}_k^{\star\top} + \mathcal{P}_\star^\perp \boldsymbol{E}_k$
2. Let $\widehat{\boldsymbol{U}}_{\text{oracle}}$ be top-r eigenspace of

$$\frac{1}{K}\sum_{k=1}^{K}\left(\boldsymbol{A}_k - \widehat{\boldsymbol{U}}_k \widehat{\boldsymbol{W}}_k^\top\right)\left(\boldsymbol{A}_k - \widehat{\boldsymbol{U}}_k \widehat{\boldsymbol{W}}_k^\top\right)^\top$$

Theorem 4 (Yang, Ma '25)

There exist $\boldsymbol{U}^\star$, $\{\boldsymbol{U}_k^\star\}_{k=1}^K$, $\{\boldsymbol{V}_k^\star\}_{k=1}^K$, $\{\boldsymbol{W}_k^\star\}_{k=1}^K$ *such that*

$$\left\|\widehat{\boldsymbol{U}}_{\text{oracle}} \widehat{\boldsymbol{U}}_{\text{oracle}}^\top - \boldsymbol{U}^\star \boldsymbol{U}^{\star\top}\right\| \geq C_2 \frac{\sigma^4 n^2}{\sigma_{\min}^4} - C_3 \frac{\log n}{\sqrt{K}} \cdot \frac{\sigma\sqrt{n}}{\sigma_{\min}}$$

Connection to nonconvex MLE

Maximum likelihood estimator

$$\min_{\boldsymbol{U},\boldsymbol{U}_k,\boldsymbol{V}_k,\boldsymbol{W}_k} \quad \sum_{k=1}^{K} \|\boldsymbol{U}\boldsymbol{V}_k^\top + \boldsymbol{U}_k\boldsymbol{W}_k^\top - \boldsymbol{A}_k\|_{\mathrm{F}}^2$$

$$\text{subject to} \quad \boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}_r, \qquad \boldsymbol{U}_k^\top\boldsymbol{U}_k = \boldsymbol{I}_{r_k}, \quad \boldsymbol{U}^\top\boldsymbol{U}_k = \boldsymbol{0}_{r\times r_k}$$

Connection to nonconvex MLE

Maximum likelihood estimator

$$\min_{\boldsymbol{U},\boldsymbol{U}_k,\boldsymbol{V}_k,\boldsymbol{W}_k} \quad \sum_{k=1}^{K} \|\boldsymbol{U}\boldsymbol{V}_k^\top + \boldsymbol{U}_k\boldsymbol{W}_k^\top - \boldsymbol{A}_k\|_{\mathrm{F}}^2$$

$$\text{subject to} \quad \boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}_r, \qquad \boldsymbol{U}_k^\top\boldsymbol{U}_k = \boldsymbol{I}_{r_k}, \quad \boldsymbol{U}^\top\boldsymbol{U}_k = \boldsymbol{0}_{r\times r_k}$$

Alternating minimization (AltMin):

- Fixing shared subspace $\boldsymbol{U}$, find unique components $\boldsymbol{U}_k\boldsymbol{W}_k^\top$

Connection to nonconvex MLE

Maximum likelihood estimator

$$\min_{\boldsymbol{U},\boldsymbol{U}_k,\boldsymbol{V}_k,\boldsymbol{W}_k} \quad \sum_{k=1}^{K} \|\boldsymbol{U}\boldsymbol{V}_k^\top + \boldsymbol{U}_k\boldsymbol{W}_k^\top - \boldsymbol{A}_k\|_{\mathrm{F}}^2$$

$$\text{subject to} \quad \boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}_r, \qquad \boldsymbol{U}_k^\top\boldsymbol{U}_k = \boldsymbol{I}_{r_k}, \quad \boldsymbol{U}^\top\boldsymbol{U}_k = \boldsymbol{0}_{r\times r_k}$$

Alternating minimization (AltMin):

- Fixing shared subspace $\boldsymbol{U}$, find unique components $\boldsymbol{U}_k\boldsymbol{W}_k^\top$
- Fixing unique components $\{\boldsymbol{U}_k\boldsymbol{W}_k^\top\}$, find shared component $\boldsymbol{U}$

Connection to nonconvex MLE

Maximum likelihood estimator

$$\min_{\boldsymbol{U},\boldsymbol{U}_k,\boldsymbol{V}_k,\boldsymbol{W}_k} \quad \sum_{k=1}^{K} \|\boldsymbol{U}\boldsymbol{V}_k^\top + \boldsymbol{U}_k\boldsymbol{W}_k^\top - \boldsymbol{A}_k\|_\mathrm{F}^2$$

$$\text{subject to} \quad \boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}_r, \quad \boldsymbol{U}_k^\top\boldsymbol{U}_k = \boldsymbol{I}_{r_k}, \quad \boldsymbol{U}^\top\boldsymbol{U}_k = \boldsymbol{0}_{r\times r_k}$$

Alternating minimization (AltMin):

- Fixing shared subspace $\boldsymbol{U}$, find unique components $\boldsymbol{U}_k\boldsymbol{W}_k^\top$
- Fixing unique components $\{\boldsymbol{U}_k\boldsymbol{W}_k^\top\}$, find shared component $\boldsymbol{U}$

oracle spectral estimator = one-step AltMin of MLE from $\boldsymbol{U}^\star$

Connection to nonconvex MLE

Maximum likelihood estimator

$$\min_{\boldsymbol{U},\boldsymbol{U}_k,\boldsymbol{V}_k,\boldsymbol{W}_k} \quad \sum_{k=1}^{K} \|\boldsymbol{U}\boldsymbol{V}_k^\top + \boldsymbol{U}_k\boldsymbol{W}_k^\top - \boldsymbol{A}_k\|_{\mathrm{F}}^2$$

$$\text{subject to} \quad \boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}_r, \qquad \boldsymbol{U}_k^\top\boldsymbol{U}_k = \boldsymbol{I}_{r_k}, \quad \boldsymbol{U}^\top\boldsymbol{U}_k = \boldsymbol{0}_{r\times r_k}$$

Alternating minimization (AltMin):

- Fixing shared subspace $\boldsymbol{U}$, find unique components $\boldsymbol{U}_k\boldsymbol{W}_k^\top$
- Fixing unique components $\{\boldsymbol{U}_k\boldsymbol{W}_k^\top\}$, find shared component $\boldsymbol{U}$

oracle spectral estimator = one-step AltMin of MLE from $\boldsymbol{U}^\star$

MLE is inconsistent as $K \to \infty$

Shared subspace estimation as an incidental parameter problem

— Neyman, Scott 1948

In incidental parameter problem, sequence of ind. observations $\{\boldsymbol{A}_k\}$ is governed by two sets of parameters:

Shared subspace estimation as an incidental parameter problem

— Neyman, Scott 1948

In incidental parameter problem, sequence of ind. observations $\{\boldsymbol{A}_k\}$ is governed by two sets of parameters:

- structural parameters $\boldsymbol{U}^\star$, which appears in the law of every observation

Shared subspace estimation as an incidental parameter problem

— Neyman, Scott 1948

In incidental parameter problem, sequence of ind. observations $\{\boldsymbol{A}_k\}$ is governed by two sets of parameters:

- structural parameters $\boldsymbol{U}^\star$, which appears in the law of every observation
- incidental parameters $(\{\boldsymbol{U}_k^\star\}, \{\boldsymbol{V}_k^\star\}, \{\boldsymbol{W}_k^\star\})$, which appears only in law of each individual observation

Shared subspace estimation as an incidental parameter problem

— Neyman, Scott 1948

In incidental parameter problem, sequence of ind. observations $\{\boldsymbol{A}_k\}$ is governed by two sets of parameters:

- structural parameters $\boldsymbol{U}^\star$, which appears in the law of every observation
- incidental parameters $(\{\boldsymbol{U}_k^\star\}, \{\boldsymbol{V}_k^\star\}, \{\boldsymbol{W}_k^\star\})$, which appears only in law of each individual observation

MLE can be inconsistent for estimating structural parameters

Classical example: Fixed-effect model

Setup:

$$(X_i, Y_i) \sim N(\alpha_i, \sigma^2), \quad i = 1, \ldots, n$$

Classical example: Fixed-effect model

Setup:

$$(X_i, Y_i) \sim N(\alpha_i, \sigma^2), \quad i = 1, \ldots, n$$

MLE:

$$\widehat{\alpha}_i = \frac{X_i + Y_i}{2}, \quad \widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} s_i^2, \quad s_i^2 = \frac{(X_i - Y_i)^2}{4}$$

Classical example: Fixed-effect model

Setup:

$$(X_i, Y_i) \sim N(\alpha_i, \sigma^2), \quad i = 1, \ldots, n$$

MLE:

$$\widehat{\alpha}_i = \frac{X_i + Y_i}{2}, \quad \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} s_i^2, \quad s_i^2 = \frac{(X_i - Y_i)^2}{4}$$

Neyman–Scott problem:

- Bias in estimating σ^2:

$$E(s_i^2) = \frac{\sigma^2}{2} \quad \text{(not equal to } \sigma^2\text{)}$$

- As $n \to \infty$, MLE is inconsistent

Another interesting example: Rasch Model

Setup: Probability of correct response of subject i to item j:

$$P(Y_{ij} = 1) = \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)}, \quad \text{for } 1 \le i \le n, 1 \le j \le m$$

- θ_i: Ability parameter of subject i (incidental parameters)
- β_j: Difficulty parameter of item j (main parameters of interest)

Another interesting example: Rasch Model

Setup: Probability of correct response of subject i to item j:

$$P(Y_{ij} = 1) = \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)}, \quad \text{for } 1 \leq i \leq n, 1 \leq j \leq m$$

- θ_i: Ability parameter of subject i (incidental parameters)
- β_j: Difficulty parameter of item j (main parameters of interest)

Neyman–Scott problem: Fixing m, as $n \to \infty$, MLE is inconsistent

Another interesting example: Rasch Model

Setup: Probability of correct response of subject i to item j:

$$P(Y_{ij} = 1) = \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)}, \quad \text{for } 1 \leq i \leq n, 1 \leq j \leq m$$

- θ_i: Ability parameter of subject i (incidental parameters)
- β_j: Difficulty parameter of item j (main parameters of interest)

Neyman–Scott problem: Fixing m, as $n \to \infty$, MLE is inconsistent

Y. Yang, and C. Ma, "Random pairing MLE for estimation of item parameters in Rasch model," arXiv:2406.13989, 2024

Neyman-Scott's problem in our case

JIVE

$$\boldsymbol{A}_k = \underbrace{\boldsymbol{U}^\star \boldsymbol{V}_k^{\star\top}}_{\substack{\text{rank}-r \\ \text{shared component}}} + \underbrace{\boldsymbol{U}_k^\star \boldsymbol{W}_k^{\star\top}}_{\substack{\text{rank}-r_k \\ \text{unique component}}} + \underbrace{\boldsymbol{E}_k}_{\text{Noise}}$$

Neyman-Scott's problem in our case

JIVE

$$\boldsymbol{A}_k = \underbrace{\boldsymbol{U}^\star \boldsymbol{V}_k^{\star\top}}_{\substack{\text{rank}-r \\ \text{shared component}}} + \underbrace{\boldsymbol{U}_k^\star \boldsymbol{W}_k^{\star\top}}_{\substack{\text{rank}-r_k \\ \text{unique component}}} + \underbrace{\boldsymbol{E}_k}_{\text{Noise}}$$

Maximum likelihood estimator

$$\min_{\boldsymbol{U},\boldsymbol{U}_k,\boldsymbol{V}_k,\boldsymbol{W}_k} \quad \sum_{k=1}^{K} \|\boldsymbol{U}\boldsymbol{V}_k^\top + \boldsymbol{U}_k\boldsymbol{W}_k^\top - \boldsymbol{A}_k\|_{\mathrm{F}}^2$$

$$\text{subject to} \quad \boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}_r, \quad \boldsymbol{U}_k^\top\boldsymbol{U}_k = \boldsymbol{I}_{r_k}, \quad \boldsymbol{U}^\top\boldsymbol{U}_k = \boldsymbol{0}_{r\times r_k}$$

Neyman-Scott's problem in our case

JIVE

$$\boldsymbol{A}_k = \underbrace{\boldsymbol{U}^\star \boldsymbol{V}_k^{\star\top}}_{\substack{\text{rank}-r \\ \text{shared component}}} + \underbrace{\boldsymbol{U}_k^\star \boldsymbol{W}_k^{\star\top}}_{\substack{\text{rank}-r_k \\ \text{unique component}}} + \underbrace{\boldsymbol{E}_k}_{\text{Noise}}$$

Maximum likelihood estimator

$$\min_{\boldsymbol{U},\boldsymbol{U}_k,\boldsymbol{V}_k,\boldsymbol{W}_k} \quad \sum_{k=1}^{K} \|\boldsymbol{U}\boldsymbol{V}_k^\top + \boldsymbol{U}_k\boldsymbol{W}_k^\top - \boldsymbol{A}_k\|_{\mathrm{F}}^2$$

$$\text{subject to} \quad \boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}_r, \quad \boldsymbol{U}_k^\top\boldsymbol{U}_k = \boldsymbol{I}_{r_k}, \quad \boldsymbol{U}^\top\boldsymbol{U}_k = \boldsymbol{0}_{r\times r_k}$$

MLE is inconsistent when $K \to \infty$

Conclusions

- Multimodal learning is ubiquitous and important

Conclusions

- Multimodal learning is ubiquitous and important
- JIVE and AJIVE are interesting model and method, respectively

Conclusions

- Multimodal learning is ubiquitous and important
- JIVE and AJIVE are interesting model and method, respectively
- When SNR is high, AJIVE is optimal

  $\rightarrow$ power of multiple matrices

Conclusions

- Multimodal learning is ubiquitous and important
- JIVE and AJIVE are interesting model and method, respectively
- When SNR is high, AJIVE is optimal
  $\rightarrow$ power of multiple matrices
- When SNR is low, AJIVE (and MLE) has non-diminishing error
  $\rightarrow$ potential limitation of multiple matrices

Conclusions

- Multimodal learning is ubiquitous and important
- JIVE and AJIVE are interesting model and method, respectively
- When SNR is high, AJIVE is optimal
  $\rightarrow$ power of multiple matrices
- When SNR is low, AJIVE (and MLE) has non-diminishing error
  $\rightarrow$ potential limitation of multiple matrices

Future directions:

- Information-theoretic lower bounds for non-diminishing error
- Missing data, outliers, etc.
- Adaptive rank estimation

Conclusions

- Multimodal learning is ubiquitous and important
- JIVE and AJIVE are interesting model and method, respectively
- When SNR is high, AJIVE is optimal
  $\rightarrow$ power of multiple matrices
- When SNR is low, AJIVE (and MLE) has non-diminishing error
  $\rightarrow$ potential limitation of multiple matrices

Future directions:

- Information-theoretic lower bounds for non-diminishing error
- Missing data, outliers, etc.
- Adaptive rank estimation

Y. Yang, C. Ma, "Estimating shared subspace with AJIVE: the power and limitation of multiple data matrices", arxiv:2501.093336, 2025