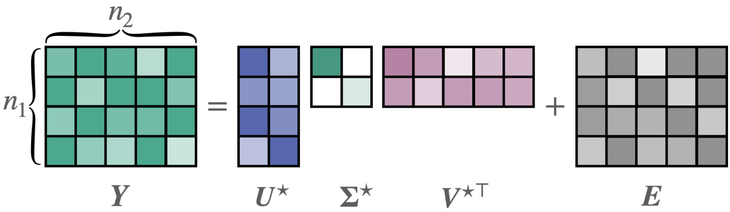
Deflated HeteroPCA: Overcoming the curse of ill-conditioning in heteroskedastic PCA

Yuxin Chen, Wharton Statistics & Data Science

Yuchen Zhou
Wharton Statistics & Data Science

A subspace estimation / model

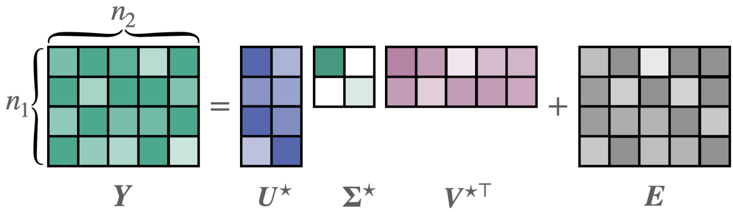


- **Ground truth:** rank-r matrix $\boldsymbol{X}^\star$ with SVD ($r \ll \min\{n_1, n_2\}$)

$$\boldsymbol{X}^\star = \boldsymbol{U}^\star \boldsymbol{\Sigma}^\star \boldsymbol{V}^{\star\top} = \sum_{i=1}^{r} \sigma_i^\star \boldsymbol{u}_i^\star \boldsymbol{v}_i^{\star\top} \in \mathbb{R}^{n_1 \times n_2}$$

where $\boldsymbol{U}^\star \in \mathbb{R}^{n_1 \times r}$, $\boldsymbol{\Sigma}^\star = \mathrm{diag}\{\sigma_1^\star, \cdots, \sigma_r^\star\}$, $\boldsymbol{V}^\star \in \mathbb{R}^{n_2 \times r}$

A subspace estimation / model

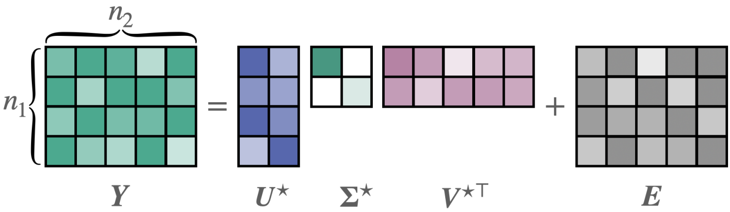


- **Ground truth:** rank-r matrix $\boldsymbol{X}^\star$ with SVD ($r \ll \min\{n_1, n_2\}$)

$$\boldsymbol{X}^\star = \boldsymbol{U}^\star \boldsymbol{\Sigma}^\star \boldsymbol{V}^{\star\top} = \sum_{i=1}^{r} \sigma_i^\star \boldsymbol{u}_i^\star \boldsymbol{v}_i^{\star\top} \in \mathbb{R}^{n_1 \times n_2}$$

- **Noisy observations:** $\boldsymbol{Y} = \boldsymbol{X}^\star + \underbrace{\boldsymbol{E}}_{\text{zero-mean ind. noise}}$

A subspace estimation / model



- **Ground truth:** rank-r matrix $\boldsymbol{X}^\star$ with SVD ($r \ll \min\{n_1, n_2\}$)

$$\boldsymbol{X}^\star = \boldsymbol{U}^\star \boldsymbol{\Sigma}^\star \boldsymbol{V}^{\star\top} = \sum_{i=1}^{r} \sigma_i^\star \boldsymbol{u}_i^\star \boldsymbol{v}_i^{\star\top} \in \mathbb{R}^{n_1 \times n_2}$$

- **Noisy observations:** $\boldsymbol{Y} = \boldsymbol{X}^\star + \underbrace{\boldsymbol{E}}_{\text{zero-mean ind. noise}}$

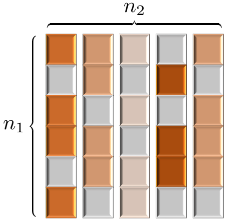
- **Goal**: estimate column subspace $\boldsymbol{U}^\star \in \mathbb{R}^{n_1 \times r}$ based on $\boldsymbol{Y}$

Two challenges

unbalanced dimensionality: $n_2 \gg n_1$ (a highly challenging regime)

Two challenges

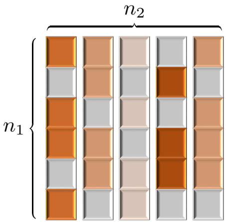
unbalanced dimensionality: $n_2 \gg n_1$ (a highly challenging regime)



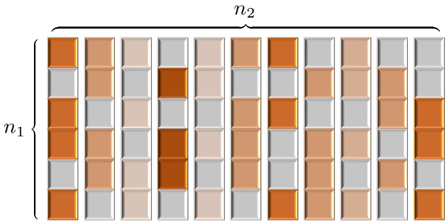
$n_2 \lesssim n_1$: solvable via matrix completion methods

Two challenges

unbalanced dimensionality: $n_2 \gg n_1$ (a highly challenging regime)



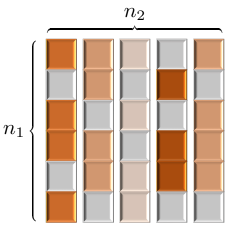
$n_2 \lesssim n_1$: solvable via matrix completion methods



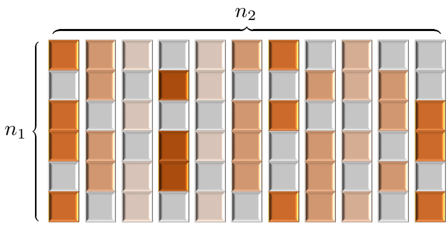
$n_2 \gg n_1$: sometimes it's only feasible to estimate col-space instead of whole matrix

Two challenges

unbalanced dimensionality: $n_2 \gg n_1$ (a highly challenging regime)



$n_2 \lesssim n_1$: solvable via matrix completion methods

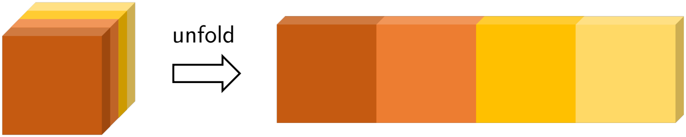


$n_2 \gg n_1$: sometimes it's only feasible to estimate col-space instead of whole matrix

heteroskedasticity: $\underbrace{\text{noise variances } \{\mathbb{E}[E_{i,j}^2]\}}_{\text{unknown } a\ priori}$ are location-varying

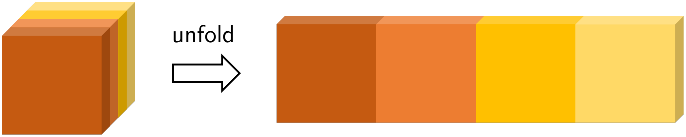
Applications beyond PCA

- Tensor completion

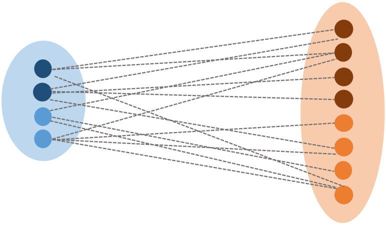


Applications beyond PCA

- Tensor completion

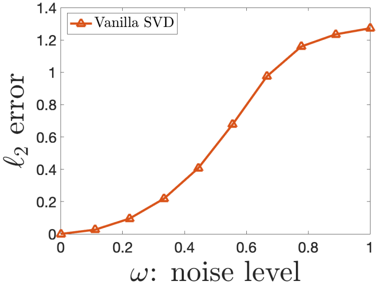


- One-sided community recovery in bipartite random graphs



Review of popular methods

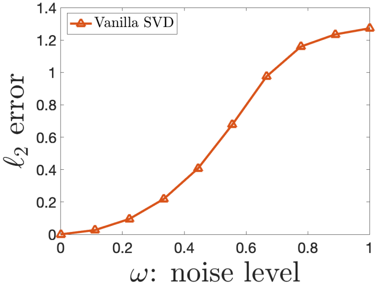
$$n_1 = 100, n_2 = 10,000$$
$$r = 2, \kappa = 2$$



vanilla SVD: $\boldsymbol{U} \longleftarrow$ rank-r left singular subspace of $\boldsymbol{Y} = \boldsymbol{X}^\star + \boldsymbol{E}$

Review of popular methods

$$n_1 = 100, n_2 = 10,000$$
$$r = 2, \kappa = 2$$



vanilla SVD: $\boldsymbol{U} \longleftarrow$ rank-r left singular subspace of $\boldsymbol{Y} = \boldsymbol{X}^\star + \boldsymbol{E}$

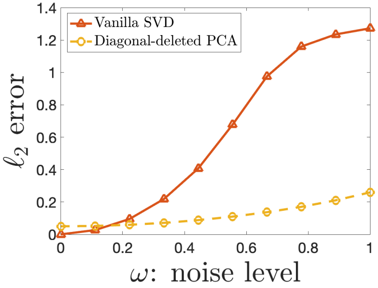
- often sub-optimal due to large bias in diagonal entries:

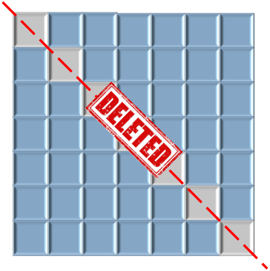
$$\mathbb{E}[\boldsymbol{Y}\boldsymbol{Y}^\top] = \underbrace{\boldsymbol{X}\boldsymbol{X}^\top}_{\checkmark} + \underbrace{\mathsf{diag}\Big\{\Big[\sum\nolimits_j \mathbb{E}[E_{i,j}^{\star 2}]\Big]_{1 \le i \le n_1}\Big\}}_{\text{potentially large diagonal matrix!}}$$

Review of popular methods

$n_1 = 100, n_2 = 10,000$

$r = 2, \kappa = 2$



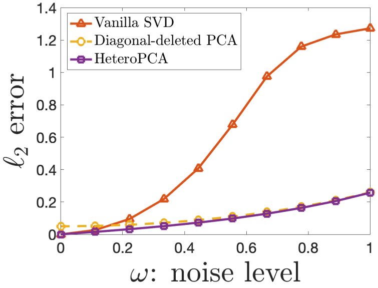


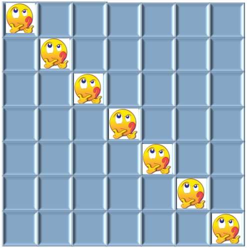
diagonal-deleted PCA:

- remove diag($\boldsymbol{Y}\boldsymbol{Y}^\top$)
- compute top-r eigen-space

Review of popular methods

$$n_1 = 100, n_2 = 10,000$$
$$r = 2, \kappa = 2$$



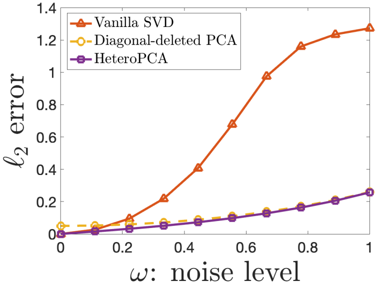


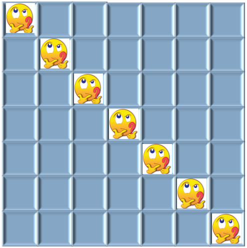
HeteroPCA (Zhang, Cai, Wu '22)

- iteratively estimate diag$(\boldsymbol{Y}\boldsymbol{Y}^\top)$
- compute top-r eigen-space

Review of popular methods

$$n_1 = 100, n_2 = 10,000$$
$$r = 2, \kappa = 2$$





HeteroPCA (Zhang, Cai, Wu '22)

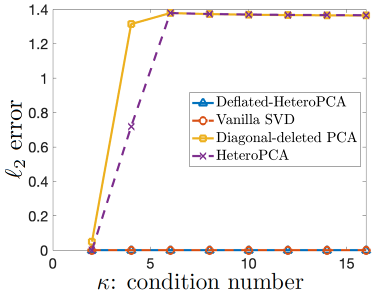
- **initialize:** $\boldsymbol{G}^0 = \mathcal{P}_{\text{off-diag}}(\boldsymbol{Y}\boldsymbol{Y}^\top)$
- for $t = 0, 1, \ldots$

  $(\boldsymbol{U}^t, \boldsymbol{\Lambda}^t) = \text{eigs}(\boldsymbol{G}^t, r)$

  $\boldsymbol{G}^{t+1} = \boldsymbol{G}^0 + \mathcal{P}_{\text{diag}}(\boldsymbol{U}^t\boldsymbol{\Lambda}^t\boldsymbol{U}^{t\top})$

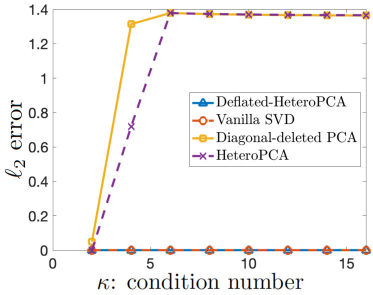
A curious phenomenon: curse of ill-conditioning

Somewhat surprising numerical example: $r = 2, n_1 = 200,\ n_2 = 40,000$

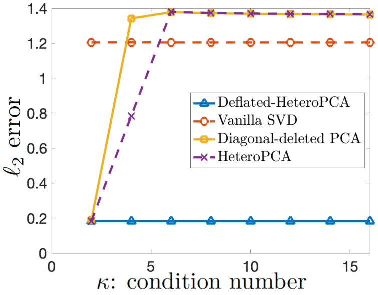


noiseless case

Somewhat surprising numerical example: $r = 2, n_1 = 200,\ n_2 = 40,000$

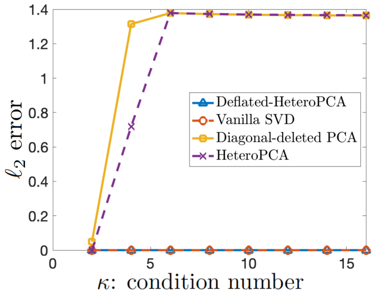


noiseless case

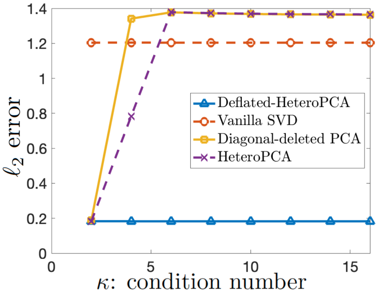


noisy case

Somewhat surprising numerical example: $r = 2, n_1 = 200,\ n_2 = 40,000$



noiseless case

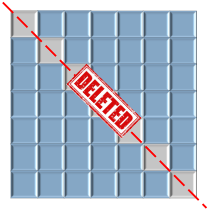


noisy case

Previous methods degrade as $\underbrace{\textit{condition number of } \boldsymbol{X}^\star \textit{ increases}}$!

but this actually makes problem info-theoretically easier . . .

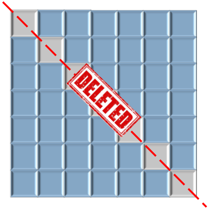
Diagonsis: influences of diagonal deletion



$$\mathbb{E}\Big[\mathcal{P}_{\text{off-diag}}\big(\boldsymbol{Y}\boldsymbol{Y}^\top\big)\Big] = \boldsymbol{X}^\star\boldsymbol{X}^{\star\top} - \underbrace{\mathcal{P}_{\text{diag}}\big(\boldsymbol{X}^\star\boldsymbol{X}^{\star\top}\big)}_{\text{ideally negligible compared to } \sigma_r^\star}$$

- ideally, we hope digonal deletion has negligible influences

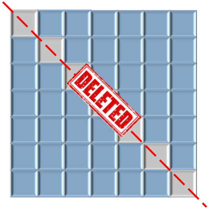
Diagonsis: influences of diagonal deletion



$$\mathbb{E}\Big[\mathcal{P}_{\text{off-diag}}(\boldsymbol{Y}\boldsymbol{Y}^\top)\Big] = \boldsymbol{X}^\star\boldsymbol{X}^{\star\top} - \underbrace{\mathcal{P}_{\text{diag}}(\boldsymbol{X}^\star\boldsymbol{X}^{\star\top})}_{\text{but larger than } \sigma_r^\star \text{ if } \sigma_1^\star/\sigma_r^\star \text{ is too large}}$$

- ideally, we hope digonal deletion has negligible influences
- non-negligible for ill-conditioned case though ...

Diagonsis: influences of diagonal deletion



$$\mathbb{E}\Big[\mathcal{P}_{\text{off-diag}}(\boldsymbol{Y}\boldsymbol{Y}^\top)\Big] = \boldsymbol{X}^\star\boldsymbol{X}^{\star\top} - \underbrace{\mathcal{P}_{\text{diag}}(\boldsymbol{X}^\star\boldsymbol{X}^{\star\top})}_{\text{but larger than } \sigma_r^\star \text{ if } \sigma_1^\star/\sigma_r^\star \text{ is too large}}$$

- ideally, we hope digonal deletion has negligible influences
- non-negligible for ill-conditioned case though ...

Both diagonal-deleted PCA & $\underbrace{\text{HeteroPCA}}_{\text{initialized by diagonal-deleted PCA}}$ become ineffective in the presence of ill-conditioning!

Limitations of prior art

	requirement on κ	$\ell_{2,\infty}$ error	SNR requirement
vanilla SVD Yan et al. '21	no requirement	sub-optimal	$\sqrt{n}_1 + \sqrt{n_2}$
HeteroPCA Yan et al. '21	$\kappa \lesssim n_1^{1/4}$	poly(κ) dependence	$\kappa\,(n_1 n_2)^{1/4} + \kappa^3 n_1^{1/2}$
diagonal-deleted PCA Cai et al. '21	$\kappa \lesssim n_1^{1/4}$	poly(κ) dependence	$\kappa\,(n_1 n_2)^{1/4} + \kappa^3 n_1^{1/2}$

Limitations of prior art

	requirement on κ	$\ell_{2,\infty}$ error	SNR requirement
vanilla SVD Yan et al. '21	no requirement	sub-optimal	$\sqrt{n}_1 + \sqrt{n_2}$
HeteroPCA Yan et al. '21	$\kappa \lesssim n_1^{1/4}$	poly(κ) dependence	$\kappa\,(n_1 n_2)^{1/4} + \kappa^3 n_1^{1/2}$
diagonal-deleted PCA Cai et al. '21	$\kappa \lesssim n_1^{1/4}$	poly(κ) dependence	$\kappa\,(n_1 n_2)^{1/4} + \kappa^3 n_1^{1/2}$

All prior theory suffers from at least one of the following issues:

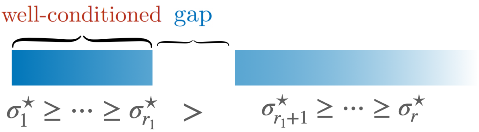
- sub-optimal statistical error bounds (decaying w/ cond. no. κ)
- sub-optimal SNR range

Can we break the curse of ill-conditioning while accommodating widest SNR range?

— *Zhang et al. '18, Yan et al. '21*

revisit HeteroPCA theory: works well if

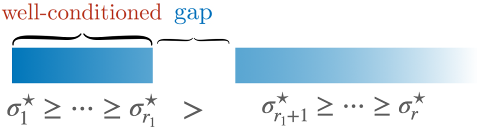
- $\boldsymbol{X}^\star$ is well-conditioned
- least singular value $\sigma_r^\star$ (or spectral gap) is not buried by noise



— *Zhang et al. '18, Yan et al. '21*

revisit HeteroPCA theory: works well if

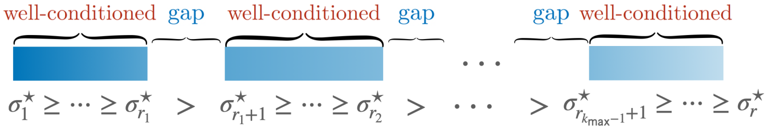
- $\boldsymbol{X}^\star$ is well-conditioned
- least singular value $\sigma_r^\star$ (or spectral gap) is not buried by noise



solution:

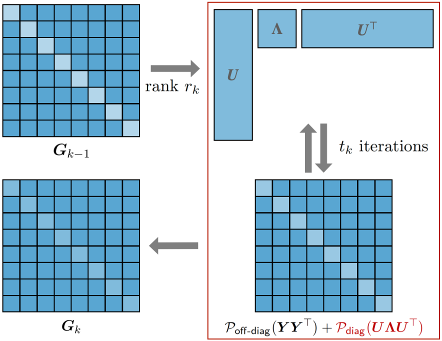
- divide eigenvalues into well-conditioned & well-separated subblocks
- estimate subblocks sequentially

Proposed algorithm: deflated-HeteroPCA



- sequentially choose ranks $r_0 = 0 < r_1 < \cdots < r_{k_{\mathsf{max}}} = r$ s.t.
  - $\sigma^\star_{r_{k-1}+1}/\sigma^\star_{r_k}$ is small
  - sufficient gap between $\sigma^\star_{r_k}$ and $\sigma^\star_{r_k+1}$

Proposed algorithm: deflated-HeteroPCA



- sequentially choose ranks $r_0 = 0 < r_1 < \cdots < r_{k_{\max}} = r$ s.t.
  - $\sigma^\star_{r_{k-1}+1}/\sigma^\star_{r_k}$ is small
  - sufficient gap between $\sigma^\star_{r_k}$ and $\sigma^\star_{r_k+1}$
- invoke HeteroPCA$(\underbrace{G_{k-1}}_{\text{input}}, \underbrace{r_k}_{\text{rank}})$ to impute diagonals & obtain G_k

Proposed algorithm: deflated-HeteroPCA

- **Initialize:** $\boldsymbol{G}^0 = \mathcal{P}_{\textsf{off-diag}}(\boldsymbol{Y}\boldsymbol{Y}^\top), k = 0, r_0 = 0$

- **Sequential updates:** while $r_k < r$

$$k = k + 1$$

select r_k in a data-driven manner

$$(\boldsymbol{G}_k, \boldsymbol{U}_k) = \textsf{HeteroPCA}(\underbrace{\boldsymbol{G}_{k-1}}_{\text{input}}, \underbrace{r_k}_{\text{rank}})$$

- **Output:** $\boldsymbol{U} \coloneqq \boldsymbol{U}_k \quad \longrightarrow \quad$ estimate of $\boldsymbol{U}^\star$

Proposed algorithm: deflated-HeteroPCA

- **Initialize:** $\boldsymbol{G}^0 = \mathcal{P}_{\text{off-diag}}(\boldsymbol{Y}\boldsymbol{Y}^\top), k = 0, r_0 = 0$

- **Sequential updates:** while $r_k < r$

$$k = k + 1$$

select r_k in a data-driven manner

$$(\boldsymbol{G}_k, \boldsymbol{U}_k) = \text{HeteroPCA}(\underbrace{\boldsymbol{G}_{k-1}}_{\text{input}}, \underbrace{r_k}_{\text{rank}})$$

- **Output:** $\boldsymbol{U} := \boldsymbol{U}_k \quad \longrightarrow \quad$ estimate of $\boldsymbol{U}^\star$

Select $r_k = \begin{cases} \max \mathcal{R}_k, & \text{if } \mathcal{R}_k \neq \emptyset, \\ r, & \text{otherwise}, \end{cases}$ where

$$\mathcal{R}_k := \{r' : r_{k-1} < r' \leq r, \overbrace{\sigma_{r_{k-1}+1}(\boldsymbol{G}_{k-1}) / \sigma_{r'}(\boldsymbol{G}_{k-1}) \leq 4}^{\text{well-conditioned}} \\ \& \underbrace{\sigma_{r'}(\boldsymbol{G}_{k-1}) - \sigma_{r'+1}(\boldsymbol{G}_{k-1}) \geq \sigma_{r'}(\boldsymbol{G}_{k-1}) / r}_{\text{gap}}\}.$$

Assumptions (ignoring log factors)

- **heteroskedasticity**: $E'_{i,j}s$ are indep. obeying
  - $\mathbb{E}[E_{i,j}] = 0, \qquad \mathsf{Var}\left[E_{i,j}\right] \le \omega^2$
  - $|E_{i,j}| \lesssim \omega_{\mathsf{max}} \min\left\{ (n_1 n_2)^{1/4}, \sqrt{n_2} \right\}$ with high prob.

Assumptions (ignoring log factors)

- **heteroskedasticity**: $E'_{i,j}s$ are indep. obeying
  - $\mathbb{E}[E_{i,j}] = 0, \qquad \mathsf{Var}\left[E_{i,j}\right] \le \omega^2$
  - $|E_{i,j}| \lesssim \omega_{\max} \min\left\{ (n_1 n_2)^{1/4}, \sqrt{n_2} \right\}$ with high prob.
- **signal-to-noise ratio (SNR)**:

$$\frac{\sigma_r^\star}{\omega} \gtrsim (n_1 n_2)^{1/4} + n_1^{1/2}$$

  - necessary for existence of consistent estimators (Cai et al. '21)

Assumptions (ignoring log factors)

- **heteroskedasticity**: $E'_{i,j}s$ are indep. obeying
  - $\mathbb{E}[E_{i,j}] = 0, \qquad \mathsf{Var}\,[E_{i,j}] \le \omega^2$
  - $|E_{i,j}| \lesssim \omega_{\mathsf{max}} \min\left\{ (n_1 n_2)^{1/4}, \sqrt{n_2} \right\}$ with high prob.
- **signal-to-noise ratio (SNR)**:

$$\frac{\sigma_r^\star}{\omega} \gtrsim (n_1 n_2)^{1/4} + n_1^{1/2}$$

  - necessary for existence of consistent estimators (Cai et al. '21)
- rank $r = O(1)$
- incoherence $\mu := \max\{\frac{n_1}{r} \|\boldsymbol{U}^\star\|_{2,\infty}^2, \frac{n_2}{r} \|\boldsymbol{V}^\star\|_{2,\infty}^2\} = O(1)$

Theoretical guarantees

Theorem 1 (Zhou, Chen '23)

With high prob., Deflated-HeteroPCA yields

$$\|\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}} - \boldsymbol{U}^{\star}\| \lesssim \zeta_{\mathsf{op}}$$

for some rotation matrix $\boldsymbol{R}_{\boldsymbol{U}}$, *where* $\zeta_{\mathsf{op}} = \frac{\sqrt{n_1 n_2}\,\omega^2}{\sigma_r^{\star 2}} + \frac{\sqrt{n_1}\,\omega}{\sigma_r^{\star}}$

- match minimax lower bounds in Zhang et al. '22 & Cai et al. '21
- condition-number-free

Theoretical guarantees

Theorem 1 (Zhou, Chen '23)

With high prob., Deflated-HeteroPCA yields

$$\|\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}} - \boldsymbol{U}^{\star}\| \lesssim \zeta_{\mathsf{op}}$$

$$\|\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}} - \boldsymbol{U}^{\star}\|_{2,\infty} \lesssim \frac{1}{\sqrt{n_1}}\zeta_{\mathsf{op}} \quad \textit{(fine-grained } \ell_{2,\infty}\textit{)}$$

for some rotation matrix $\boldsymbol{R}_{\boldsymbol{U}}$, *where* $\zeta_{\mathsf{op}} = \frac{\sqrt{n_1 n_2}\,\omega^2}{\sigma_r^{\star 2}} + \frac{\sqrt{n_1}\,\omega}{\sigma_r^{\star}}$

- match minimax lower bounds in Zhang et al. '22 & Cai et al. '21
- condition-number-free

Comparisons with prior theory

	requirement on κ	$\ell_{2,\infty}$ error	min. SNR threshold
Vanilla SVD Yan et al. '21	no requirement	$\sqrt{\frac{1}{n_1}}\mathcal{E}_{\mathsf{svd}}$	$\sqrt{n_1} + \sqrt{n_2}$
HeteroPCA Yan et al. '21	$\kappa \lesssim n_1^{1/4}$	$\kappa^2\sqrt{\frac{1}{n_1}}\mathcal{E}_{\mathsf{noise}}$	$\kappa\,(n_1 n_2)^{1/4} + \kappa^3 n_1^{1/2}$
Diagonal-deleted PCA Cai et al. '21	$\kappa \lesssim n_1^{1/4}$	$\kappa^2\sqrt{\frac{1}{n_1}}\left(\mathcal{E}_{\mathsf{noise}} + \mathcal{E}_{\mathsf{diag\text{-}del}}\right)$	$\kappa\,(n_1 n_2)^{1/4} + \kappa^3 n_1^{1/2}$
Deflated-HeteroPCA	no requirement	$\sqrt{\frac{1}{n_1}}\zeta_{\mathsf{op}}$	$(n_1 n_2)^{1/4} + n_1^{1/2}$

Comparisons with prior theory

	requirement on κ	$\ell_{2,\infty}$ error	min. SNR threshold
Vanilla SVD Yan et al. '21	no requirement	$\sqrt{\frac{1}{n_1}}\mathcal{E}_{\mathsf{svd}}$	$\sqrt{n_1}+\sqrt{n_2}$
HeteroPCA Yan et al. '21	$\kappa \lesssim n_1^{1/4}$	$\kappa^2\sqrt{\frac{1}{n_1}}\mathcal{E}_{\mathsf{noise}}$	$\kappa\,(n_1n_2)^{1/4}+\kappa^3 n_1^{1/2}$
Diagonal-deleted PCA Cai et al. '21	$\kappa \lesssim n_1^{1/4}$	$\kappa^2\sqrt{\frac{1}{n_1}}(\mathcal{E}_{\mathsf{noise}}+\mathcal{E}_{\mathsf{diag\text{-}del}})$	$\kappa\,(n_1n_2)^{1/4}+\kappa^3 n_1^{1/2}$
Deflated-HeteroPCA	no requirement	$\sqrt{\frac{1}{n_1}}\zeta_{\mathsf{op}}$	$(n_1n_2)^{1/4}+n_1^{1/2}$

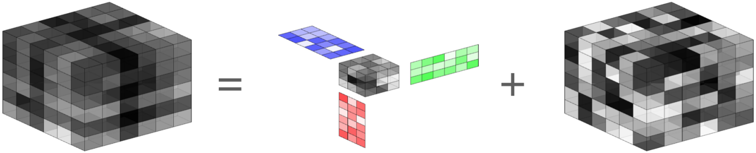
$$\zeta_{\mathsf{op}} = \frac{\sqrt{n_1 n_2}\,\omega^2}{\sigma_r^{\star 2}} + \frac{\sqrt{n_1}\,\omega}{\sigma_r^{\star}}$$

$$\mathcal{E}_{\mathsf{noise}} = \frac{\sqrt{n_1 n_2}\,\omega^2}{\sigma_r^{\star 2}} + \frac{\kappa\omega\,\sqrt{n_1}}{\sigma_r^{\star}} > \zeta_{\mathsf{op}}$$

$$\mathcal{E}_{\mathsf{svd}} = \frac{(n_1 \vee n_2)\omega^2}{\sigma_r^{\star 2}} + \frac{\sqrt{n_1}\,\omega}{\sigma_r^{\star}} > \zeta_{\mathsf{op}}$$

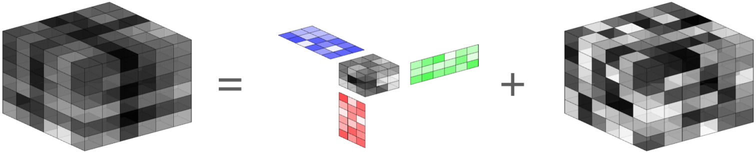
$$\mathcal{E}_{\mathsf{diag\text{-}del}} > 0$$

Application: tensor PCA



- Truth: $\boldsymbol{\mathcal{X}}^\star = \boldsymbol{\mathcal{S}}^\star \times_1 \boldsymbol{U}_1^\star \times_2 \boldsymbol{U}_2^\star \times_3 \boldsymbol{U}_3^\star$, where $\boldsymbol{\mathcal{S}}^\star \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ and $\boldsymbol{U}_i^\star \in \mathcal{O}^{n_i, r_i}$
- Noisy observation: $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}}^\star + \boldsymbol{\mathcal{E}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$
- **Goal**: estimate $\boldsymbol{\mathcal{X}}^\star$ and $\boldsymbol{U}_i^\star$

Application: tensor PCA



- Truth: $\boldsymbol{\mathcal{X}}^{\star} = \boldsymbol{\mathcal{S}}^{\star} \times_1 \boldsymbol{U}_1^{\star} \times_2 \boldsymbol{U}_2^{\star} \times_3 \boldsymbol{U}_3^{\star}$, where $\boldsymbol{\mathcal{S}}^{\star} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ and $\boldsymbol{U}_i^{\star} \in \mathcal{O}^{n_i, r_i}$
- Noisy observation: $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}}^{\star} + \boldsymbol{\mathcal{E}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$
- **Goal**: estimate $\boldsymbol{\mathcal{X}}^{\star}$ and $\boldsymbol{U}_i^{\star}$
- Assume $r = O(1), \mu = O(1), n_i \asymp n$, $E_{i,j,k}$'s are indep. zero-mean and ω-sub-Gaussian

Application: tensor PCA

Theorem 2

Assume that $\sigma^\star_{\mathsf{min}}/\omega \gtrsim n^{3/4}$. *Then with high prob., the outputs of Deflated-HeteroPCA + HOOI satisfy, for some rotation matrix* $\boldsymbol{R}_i$,

$$\|\widehat{\boldsymbol{U}}_i \boldsymbol{R}_i - \boldsymbol{U}_i^\star\| \lesssim \frac{\sqrt{n}\,\omega}{\sigma^\star_{\mathsf{min}}}$$

$$\|\widehat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}}^\star\|_{\mathrm{F}}^2 \lesssim n\omega^2$$

This is the first result that is simultaneously

- **rate-optimal**: match minimax lower bounds in Zhang et al. '18
- condition-number-free
- optimal in terms of SNR range (matching computation limit)

Concluding remarks

A new algorithm called Deflated-HeteroPCA that

- breaks curse of ill-conditioning w/o compromising SNR range
- achieves near-optimal statistical guarantees (ℓ_2 and $\ell_{2,\infty}$)

Concluding remarks

A new algorithm called Deflated-HeteroPCA that

- breaks curse of ill-conditioning w/o compromising SNR range
- achieves near-optimal statistical guarantees (ℓ_2 and $\ell_{2,\infty}$)

papers:

Y. Zhou, Y. Chen, "Deflated HeteroPCA: Overcoming the curse of ill-conditioning in heteroskedastic PCA," arxiv:2303.06198, 2023

Y. Yan, Y. Chen, J. Fan, "Inference for heteroskedastic PCA with missing data," arxiv:2107.12365, 2021

C. Cai, G. Li, Y. Chi, H. V. Poor, Y. Chen, "Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees," *Annals of Stats*, 2021