Efficient Bounds and Estimates for Canonical Angles in Randomized Subspace Approximations

Yijun Dong*, Per-Gunnar Martinsson*, Yuji Nakatsukasa†

*Oden Institute, University of Texas at Austin

†Mathematical Institute, University of Oxford

Aug 24, 2023, ICIAM 2023, Tokyo, Japan

This presentation is based on the following arXiv paper: Dong, Yijun, Per-Gunnar Martinsson, and Yuji Nakatsukasa. "Efficient Bounds and Estimates for Canonical Angles in Randomized Subspace Approximations." arXiv preprint arXiv:2211.04676 (2022).

Outline

- **Problem setup: randomized subspace approximations & canonical angles**
- Prior probabilistic bounds/estimates & posterior residual-based guarantees
- Numerical comparisons: effectiveness of canonical angle bounds & estimates in practice

Leading Singular Subspaces

- Singular value decomposition (SVD)

  Given $\mathbf{A} \in \mathbb{C}^{m \times n}$, $1 \leq k \leq r = \mathrm{rank}(\mathbf{A})$, rank-k truncated SVD:

$$\mathbf{A}_k = \underset{m\times k}{\mathbf{U}_k} \ \underset{k\times k}{\mathbf{\Sigma}_k} \ \underset{k\times n}{\mathbf{V}_k^*}$$

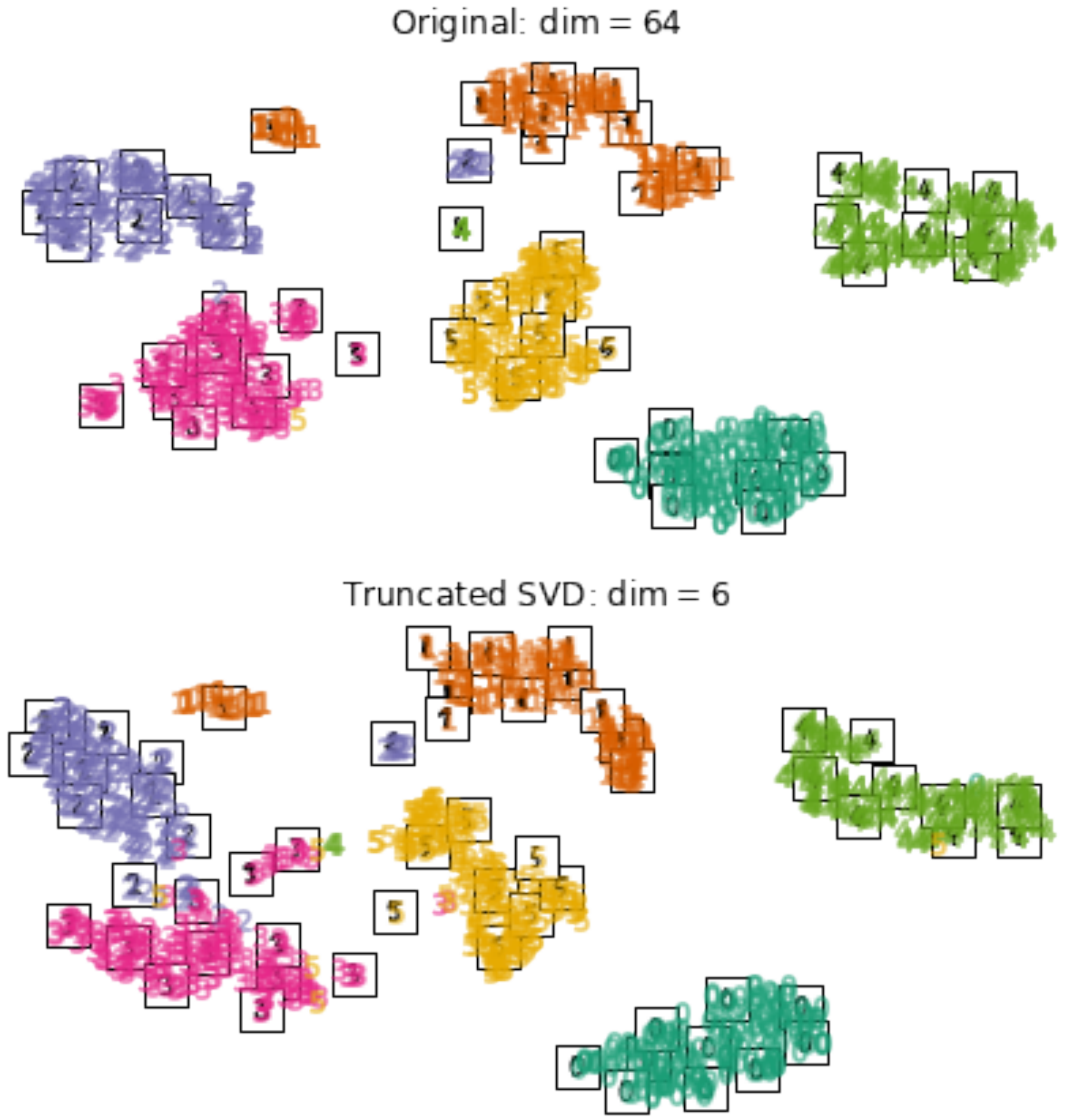
  - Maximum-k singular values: $\mathbf{\Sigma}_k = \mathrm{diag}(\sigma_1, \ldots, \sigma_k)$
  - **Leading-k singular subspaces**: $\mathbf{U}_k^* \mathbf{U}_k = \mathbf{V}_k^* \mathbf{V}_k = \mathbf{I}_k$

- Eckart–Young–Mirsky theorem

$$\mathbf{A}_k = \min_{\mathrm{rank}(\widehat{\mathbf{A}}) \leq k} \|\mathbf{A} - \widehat{\mathbf{A}}\|_F$$

  - Truncated SVD provides the optimal rank-k approximation

- Broad Applications
  - Low-rank approximations, PCA, CCA, spectral clustering, leverage score sampling, etc.



Spectral clustering on the dimension-6 leading singular subspace of a mini-MNIST dataset (8×8 images of digits 0-5)

Sketching: Approximate leading singular subspaces efficiently for large matrices

Questions: How accurate are these approximations? Tight & efficiently computable error bounds & estimates?

Randomized Subspace Approximations with Sketching

- Inputs: $\mathbf{A} \in \mathbb{C}^{m\times n}$, sample size l with $k < l \leq r = \mathrm{rank}(\mathbf{A})$ (e.g., $l = 2k \ll r$), number of power iterations $q \in \{0,1,2,\cdots\}$ ($q \leq 2$ usually)
- Outputs: $\mathrm{RSVD}(\mathbf{A}, l, q) = (\widehat{\mathbf{U}}_l \in \mathbb{C}^{m\times l}, \widehat{\boldsymbol{\Sigma}}_l \in \mathbb{C}^{l\times l}, \widehat{\mathbf{V}}_l \in \mathbb{C}^{n\times l})$ such that $\widehat{\mathbf{A}}_l = \widehat{\mathbf{U}}_l \widehat{\boldsymbol{\Sigma}}_l \widehat{\mathbf{V}}_l^* \approx \mathbf{A}$

1. **Randomized linear embedding** (Johnson-Lindenstrauss transforms, etc.)
   - Draw $\boldsymbol{\Omega} \sim P(\mathbb{C}^{n\times l})$ with i.i.d. entries $\Omega_{ij} \sim \mathcal{N}(0, l^{-1})$ such that $\mathbb{E}[\boldsymbol{\Omega}\boldsymbol{\Omega}^*] = \mathbf{I}_n$

   Isotropic embedding
2. **Sketching** with power iterations
   - Randomized **power** iterations (unstable): $\mathbf{X}^{(q)} = (\mathbf{A}\mathbf{A}^*)^q \mathbf{A}\boldsymbol{\Omega}$
   - Randomized **subspace** iterations (stable): $\mathbf{X}^{(0)} = \mathrm{ortho}(\mathbf{A}\boldsymbol{\Omega})$, $\mathbf{X}^{(i)} = \mathrm{ortho}(\mathbf{A}\ \mathrm{ortho}(\mathbf{A}^*\mathbf{X}^{(i-1)}))\ \forall\ i \in [q]$
3. $\mathbf{Q}_X = \mathrm{ortho}(\mathbf{X}^{(q)})$
4. $[\widetilde{\mathbf{U}}_l, \widehat{\boldsymbol{\Sigma}}_l, \widehat{\mathbf{V}}_l] = \mathrm{svd}(\mathbf{A}^*\mathbf{Q}_X)$
5. $\widehat{\mathbf{U}}_l = \mathbf{Q}_X \widetilde{\mathbf{U}}_l$

Key observations: with $\boldsymbol{\Sigma}$ being the spectrum of $\mathbf{A}$

- For any $q \in \mathbb{N}$, q power iterations correspond to $\boldsymbol{\Sigma}^{2q+1}$
- Compared to $\widehat{\mathbf{U}}_l$, $\widehat{\mathbf{V}}_l$ enjoys half more power iterations (i.e., $\boldsymbol{\Sigma}^{2q+2}$)

Canonical Angles: Alignment between Subspaces

- Canonical angles $\angle(\mathscr{U}, \mathscr{V}) = (\theta_1, \cdots, \theta_k)$ measure the alignment between two subspaces $\mathscr{U}, \mathscr{V} \subseteq \mathbb{C}^d$ with dimensions $k, l \leq d$ respectively ($k < l$ w.l.o.g), e.g.,
  - True leading singular subspace: $\mathscr{U} = \mathrm{range}(\mathbf{U}_k)$
  - Approximated leading singular subspace: $\mathscr{V} = \mathrm{range}(\widehat{\mathbf{U}}_l)$
- Left & right **canonical angles** of $\mathrm{RSVD}(\mathbf{A}, l, q) = (\widehat{\mathbf{U}}_l, \widehat{\boldsymbol{\Sigma}}_l, \widehat{\mathbf{V}}_l)$: $\forall\ i \in [k]$,

$$\sin\angle_i(\mathbf{U}_k, \widehat{\mathbf{U}}_l) = \sigma_{k-i+1}((\mathbf{I}_m - \widehat{\mathbf{U}}_l \widehat{\mathbf{U}}_l^*)\mathbf{U}_k), \quad \cos\angle_i(\mathbf{U}_k, \widehat{\mathbf{U}}_l) = \sigma_i(\widehat{\mathbf{U}}_l^* \mathbf{U}_k)$$

$$\sin\angle_i(\mathbf{V}_k, \widehat{\mathbf{V}}_l) = \sigma_{k-i+1}((\mathbf{I}_m - \widehat{\mathbf{V}}_l \widehat{\mathbf{V}}_l^*)\mathbf{V}_k), \quad \cos\angle_i(\mathbf{V}_k, \widehat{\mathbf{V}}_l) = \sigma_i(\widehat{\mathbf{V}}_l^* \mathbf{V}_k)$$

Prior v.s. **posterior** guarantees: computed without v.s. with the outputs $(\widehat{\mathbf{U}}_l, \widehat{\boldsymbol{\Sigma}}_l, \widehat{\mathbf{V}}_l)$

- Prior guarantees are probabilistic, with randomness from $\boldsymbol{\Omega} \sim P(\mathbb{C}^{n \times l})$
- Posterior guarantees are deterministic with given $(\widehat{\mathbf{U}}_l, \widehat{\boldsymbol{\Sigma}}_l, \widehat{\mathbf{V}}_l)$

Outline

- Problem setup: randomized subspace approximations & canonical angles
- **Prior probabilistic bounds/estimates & posterior residual-based guarantees**
- Numerical comparisons: effectiveness of canonical angle bounds & estimates in practice

Space-agnostic Prior Probabilistic Bounds

Theorem 1. (Space-agnostic bounds under multiplicative oversampling. (D., Martinsson, Nakatsukasa, 2022))

- With Gaussian embedding; small $q \in \mathbb{N}$ such that $\eta \triangleq \left(\sum_{j=k+1}^{r} \sigma_j^{4q+4} \right)^2 \Big/ \sum_{j=k+1}^{r} \sigma_j^{2(4q+4)} = \Omega(l)$; oversampling $l = \Omega(k)$
  - Notice that $1 < \eta \leq r - k$ and usually $r - k \gg l$. $\eta = \Omega(l)$ refers to a realistic case with non-negligible approximation error: when the tail of the spectrum $\{\sigma_j\}_{j=k+1}^{r}$ remains non-trivial after q power iterations
- With high probability (at least $1 - e^{-\Theta(k)} - e^{-\Theta(l)}$), there exist $\epsilon_1 = \Theta(\sqrt{k/l})$, $\epsilon_2 = \Theta(\sqrt{l/\eta})$, $\epsilon_1, \epsilon_2 \in (0,1)$ such that, $\forall\ i \in [k]$

$$\left(1 + O_{\epsilon_1,\epsilon_2}\left(\frac{l \cdot \sigma_i^{4q+2}}{\sum_{j=k+1}^{r} \sigma_j^{4q+2}}\right)\right)^{-\frac{1}{2}} \leq \sin\angle_i(\mathbf{U}_k, \widehat{\mathbf{U}}_l) \leq \left(1 + \frac{1-\epsilon_1}{1+\epsilon_2} \cdot \frac{l \cdot \sigma_i^{4q+2}}{\sum_{j=k+1}^{r} \sigma_j^{4q+2}}\right)^{-\frac{1}{2}}$$

$$\left(1 + O_{\epsilon_1,\epsilon_2}\left(\frac{l \cdot \sigma_i^{4q+4}}{\sum_{j=k+1}^{r} \sigma_j^{4q+4}}\right)\right)^{-\frac{1}{2}} \leq \sin\angle_i(\mathbf{V}_k, \widehat{\mathbf{V}}_l) \leq \left(1 + \frac{1-\epsilon_1}{1+\epsilon_2} \cdot \frac{l \cdot \sigma_i^{4q+4}}{\sum_{j=k+1}^{r} \sigma_j^{4q+4}}\right)^{-\frac{1}{2}}$$

- **In practice**, taking $\epsilon_1 = \sqrt{k/l}$, $\epsilon_2 = \sqrt{l/(r-k)}$ is sufficient for upper bounds when $l \geq 1.6k$ and $q \leq 10$

Comparison with Existing Prior Probabilistic Guarantees

- Given $\mathbf{\Omega} \sim P(\mathbb{C}^{n\times l})$, let $\mathbf{\Omega}_1 \triangleq \mathbf{V}_k^* \mathbf{\Omega}$ and $\mathbf{\Omega}_2 \triangleq \mathbf{V}_{r\backslash k}^* \mathbf{\Omega}$. Then, $\mathbf{\Omega}_1 \sim P(\mathbb{C}^{k\times l})$ and $\mathbf{\Omega}_2 \sim P(\mathbb{C}^{(r-k)\times l})$

Isotropic embedding: $\mathbf{\Omega}_1$, $\mathbf{\Omega}_2$ are agnostic of $\mathbf{V}_k, \mathbf{V}_{r\backslash k}$

- **Prior work (Saibaba, 2018)**[1]:

$$\sin\angle_i(\mathbf{U}_k, \widehat{\mathbf{U}}_l) \leq \left(1 + \frac{\sigma_i^{4q+2}}{\sigma_{k+1}^{4q+2}\|\mathbf{\Omega}_2\mathbf{\Omega}_1^\dagger\|_2^2}\right)^{-\frac{1}{2}}, \quad \sin\angle_i(\mathbf{V}_k, \widehat{\mathbf{V}}_l) \leq \left(1 + \frac{\sigma_i^{4q+4}}{\sigma_{k+1}^{4q+4}\|\mathbf{\Omega}_2\mathbf{\Omega}_1^\dagger\|_2^2}\right)^{-\frac{1}{2}}$$

where for $l \geq k+2$, given any $\delta \in (0,1)$, with probability at least $1-\delta$,

$$\|\mathbf{\Omega}_2\mathbf{\Omega}_1^\dagger\|_2 \leq \frac{e\sqrt{l}}{l-k+1}\left(\frac{2}{\delta}\right)^{\frac{1}{l-k+1}}\left(\sqrt{n-k} + \sqrt{l} + \sqrt{2\log\frac{2}{\delta}}\right) = \Omega\left(\sqrt{\frac{n-k}{l}}\right)$$

Recall the correspondence in Theorem 1:

$$\frac{1}{l}\sum_{j=k+1}^{r}\sigma_j^{4q+2} \leq \frac{n-k}{l}\sigma_{k+1}^{4q+2}$$

where the smaller values lead to the tighter upper bounds

- Theorem 1 is **space-agnostic** since the randomized linear embedding $\mathbf{\Omega} \sim P(\mathbb{C}^{n\times l})$ is **isotropic**
  - Only depends on the spectrum $\{\sigma_j\}_{j=1}^r$, but not on the singular subspaces $(\mathbf{U}_k, \mathbf{U}_{r\backslash k})$ or $(\mathbf{V}_k, \mathbf{V}_{r\backslash k})$
  - In proof, we took an integrated view on the concentration of $\mathbf{\Sigma}_{r\backslash k}^{2q+1}\mathbf{\Omega}_2$

1. Saibaba, Arvind K. "Randomized subspace iteration: Analysis of canonical angles and unitarily invariant norms." *SIAM Journal on Matrix Analysis and Applications* 40.1 (2019): 23-48.

Unbiased Space-agnostic Estimates

- Draw independent Gaussian random matrices $\left\{\mathbf{\Omega}_1^{(j)} \sim P(\mathbb{C}^{k\times l}) \middle| j \in [N]\right\}$ and $\left\{\mathbf{\Omega}_2^{(j)} \sim P(\mathbb{C}^{(r-k)\times l}) \middle| j \in [N]\right\}$
- Unbiased canonical angle estimates $\alpha_i = \mathbb{E}\left[\sin\angle_i(\mathbf{U}_k, \widehat{\mathbf{U}}_l)\right]$, $\beta_i = \mathbb{E}\left[\sin\angle_i(\mathbf{V}_k, \widehat{\mathbf{V}}_l)\right]$ $\forall\ i \in [k]$ such that

$$\sin\angle_i(\mathbf{U}_k, \widehat{\mathbf{U}}_l) \approx \alpha_i = \frac{1}{N}\sum_{j=1}^{N}\left(1 + \sigma_i^2\left(\mathbf{\Sigma}_k^{2q+1}\mathbf{\Omega}_1^{(j)}\left(\mathbf{\Sigma}_{r\backslash k}^{2q+1}\mathbf{\Omega}_2^{(j)}\right)^{\dagger}\right)\right)^{-\frac{1}{2}}$$

Corresponds to $\dfrac{1 \mp \epsilon_1}{1 \pm \epsilon_2} \cdot \dfrac{l \cdot \sigma_i^{4q+2}}{\sum_{j=k+1}^{r}\sigma_j^{4q+2}}$ in the upper/lower bounds of Theorem 1

$$\sin\angle_i(\mathbf{V}_k, \widehat{\mathbf{V}}_l) \approx \beta_i = \frac{1}{N}\sum_{j=1}^{N}\left(1 + \sigma_i^2\left(\mathbf{\Sigma}_k^{2q+2}\mathbf{\Omega}_1^{(j)}\left(\mathbf{\Sigma}_{r\backslash k}^{2q+2}\mathbf{\Omega}_2^{(j)}\right)^{\dagger}\right)\right)^{-\frac{1}{2}}$$

- **Low variance** in practice (i.e., negligible when $N \geq 3$)
- **Can be <u>computed</u> efficiently** with $O(Nrl^2)$ operations (for a given spectrum $\mathbf{\Sigma}$)
- **For any** $k \leq l \leq r$, without further assumptions on the sample size (e.g., $\eta = \Omega(l), l = \Omega(k)$)

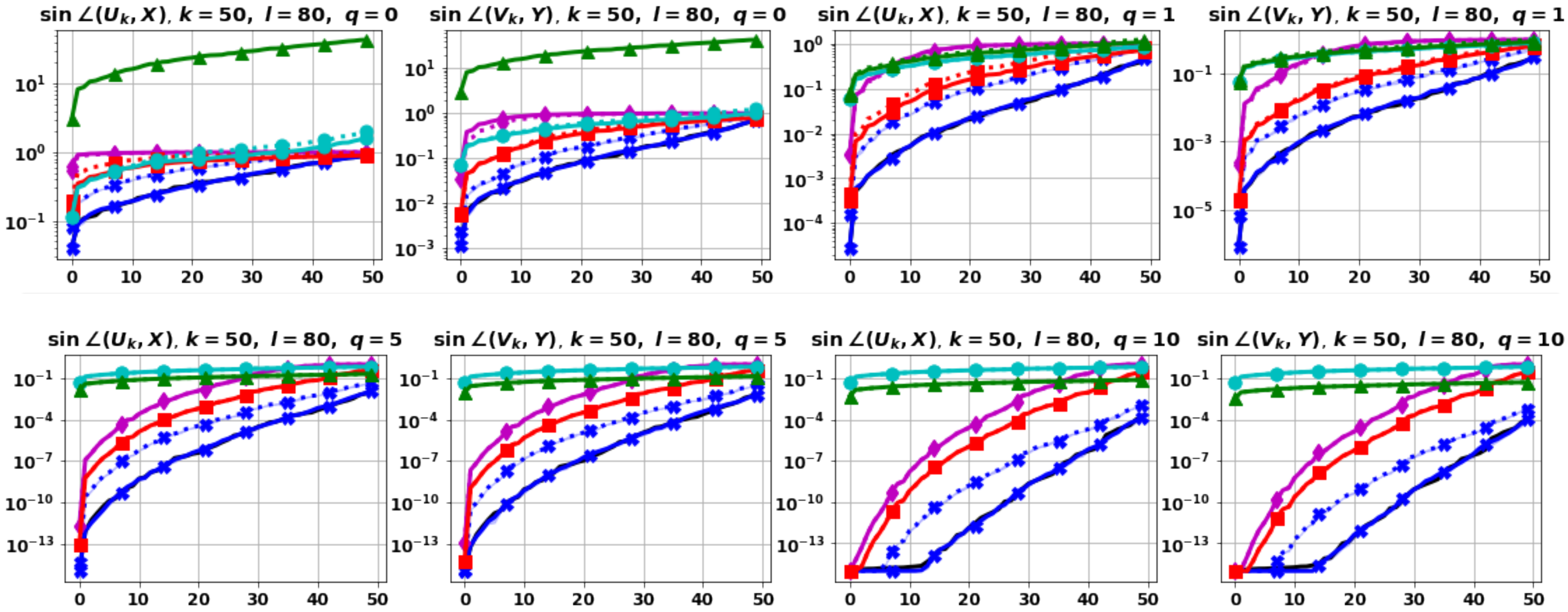
Posterior Residual-based Guarantees

1. Posterior bounds based on full residuals: Theorem 2. (D., Martinsson, Nakatsukasa, 2022)
   - $\sin\angle_i(\mathbf{U}_k, \widehat{\mathbf{U}}_l) \leq \dfrac{\sigma_{k-i+1}\left(\left(\mathbf{I}_m - \widehat{\mathbf{U}}_l\widehat{\mathbf{U}}_l^*\right)\mathbf{A}\right)}{\sigma_k} \wedge \dfrac{\sigma_1\left(\left(\mathbf{I}_m - \widehat{\mathbf{U}}_l\widehat{\mathbf{U}}_l^*\right)\mathbf{A}\right)}{\sigma_i}$
   - Deterministic and **algorithm-independent** (e.g., holds for any $k \leq l \leq r$, and any embedding $\mathbf{\Omega}$)
   - Can be approximated with $O(mnl)$ operations
2. Posterior bounds based on sub-residuals: Theorem 3.
   - Let $\mathbf{E}_{31} \triangleq \widehat{\mathbf{U}}^*_{m\setminus l}\mathbf{A}\widehat{\mathbf{V}}_k$, $\mathbf{E}_{32} \triangleq \widehat{\mathbf{U}}^*_{m\setminus l}\mathbf{A}\widehat{\mathbf{V}}_{l\setminus k}$, $\mathbf{E}_{33} \triangleq \widehat{\mathbf{U}}^*_{m\setminus l}\mathbf{A}\widehat{\mathbf{V}}_{n\setminus l}$, $\Gamma_1 \triangleq \dfrac{\sigma_k^2 - \|\mathbf{E}_{33}\|_2^2}{\sigma_k}$, $\Gamma_2 \triangleq \dfrac{\sigma_k^2 - \|\mathbf{E}_{33}\|_2^2}{\|\mathbf{E}_{33}\|_2}$.
     Assume $\sigma_k > \|\mathbf{E}_{33}\|_2$. Then, for any unitary invariant norm $\|\cdot\|$, $\|\sin\angle(\mathbf{U}_k, \widehat{\mathbf{U}}_l)\| \leq \|[\mathbf{E}_{31}, \mathbf{E}_{32}]\| / \Gamma_1$
   - Deterministic and holds for any $k \leq l \leq r$, and any embedding $\mathbf{\Omega}$
   - Can be approximated with $O(mnl)$ operations

Outline

- Problem setup: randomized subspace approximations & canonical angles
- Prior probabilistic bounds/estimates & posterior residual-based guarantees
- **Numerical comparisons: effectiveness of canonical angle bounds & estimates in practice**

Space-agnostic bounds & estimates win on MNIST: Polynomial spectral decay



$l = 1.6k$

$q \in \{0,1\}$

$l = 1.6k$

$q \in \{5,10\}$

Black line: true canonical angles

Blue lines/dashes (with shade): unbiased space-agnostic estimates computed with true/approximated singular values

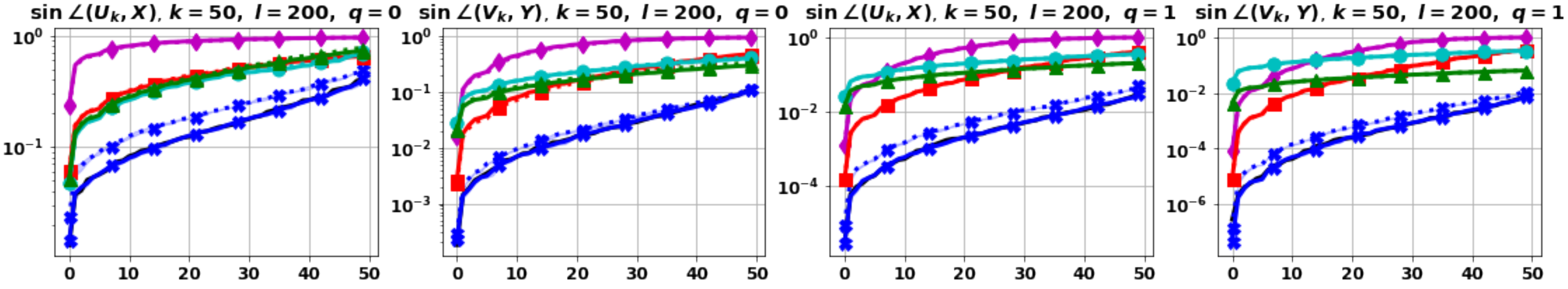
Red lines/dashes: space-agnostic upper bounds with true/approximated singular values, $\epsilon_1 = \sqrt{k/l}, \epsilon_2 = \sqrt{l/(r-k)}$

Meganta lines/dashes: (Saibaba, 2018) bounds with true/approximated singular values and the true singular subspaces

Cyan & green lines/dashes: Posterior residual-based bounds in Theorem 2 & 3 with true/approximated singular values

shade = min/max in $N = 3$ samples $\Rightarrow$ negligible variance!

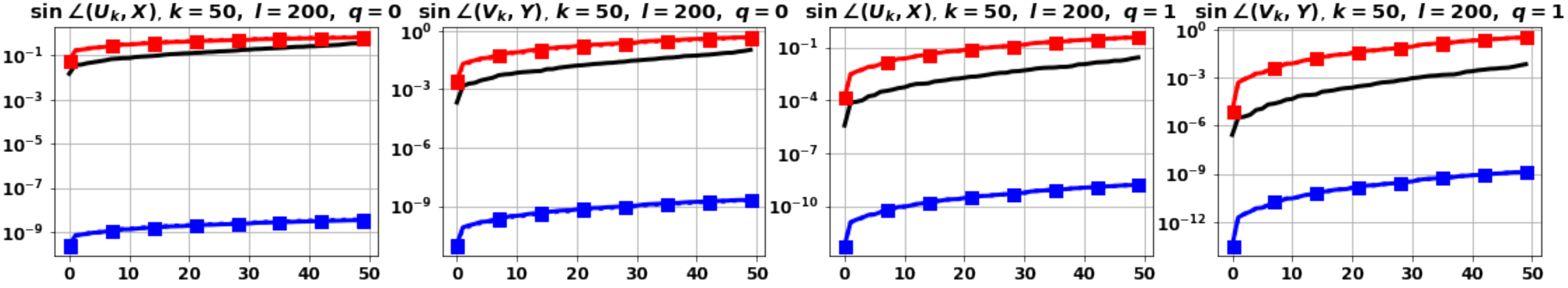
How about space-agnostic lower bounds in practice: MNIST



$l = 4k$

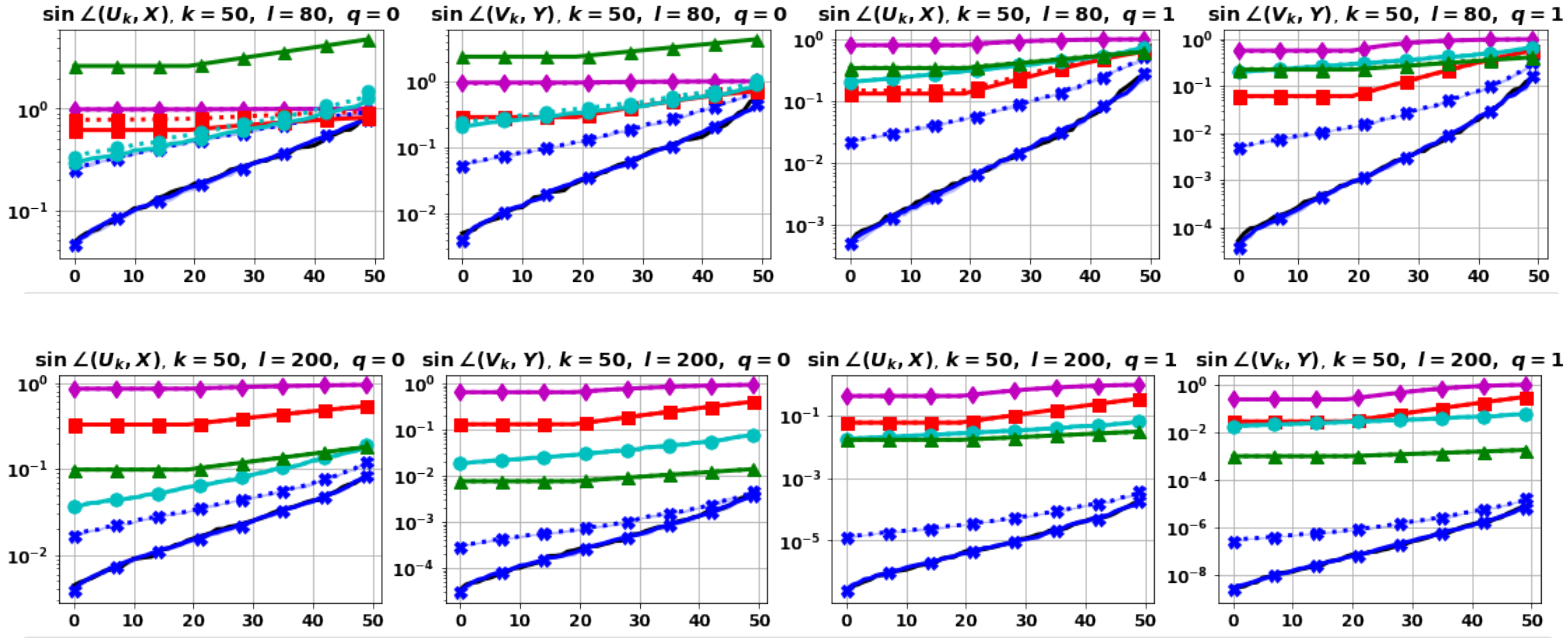
$q \in \{0,1\}$

Unbiased space-agnostic estimates, space-agnostic upper bounds, (Saibaba, 2018) bounds, Posterior residual-based bounds in Theorem 2 & 3 (with true/approximated singular values), and true canonical angles



Space-agnostic upper bounds and lower bounds with true singular values and $\epsilon_1 = \sqrt{k/l}, \epsilon_2 = \sqrt{l/(r-k)}$

When are posterior bounds more effective: Exponential spectral decay + low-error regimes



$l = 1.6k$

$q \in \{0,1\}$

$l = 4k$

$q \in \{0,1\}$

Unbiased space-agnostic estimates, space-agnostic upper bounds, (Saibaba, 2018) bounds, Posterior residual-based bounds in Theorem 2 & 3 (with true/approximated singular values), and true canonical angles

Thank You! Happy to take any questions

arXiv: https://arxiv.org/abs/2211.04676

GitHub: https://github.com/dyjdongyijun/Randomized_Subspace_Approximation