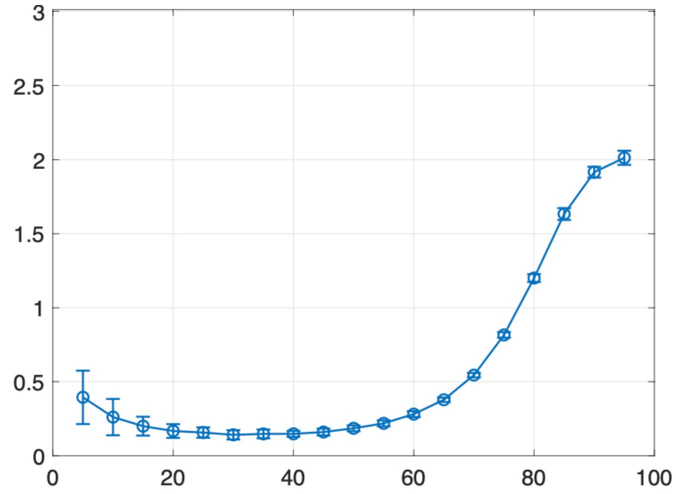
Singular Subspace Perturbation Bounds via Rectangular Random Matrix Diffusions

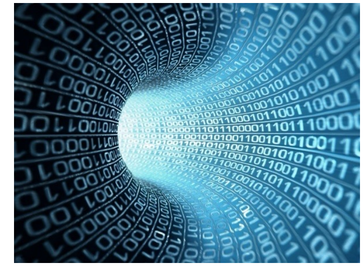


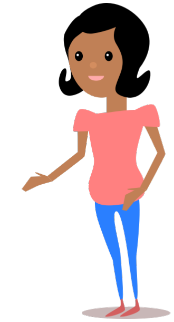
Peiyao Lai
WPI

Oren Mangoubi
WPI

ICLR 2025

Matrix Approximation





Matrix Approximation Problem:

Given $\gamma_1 \geq \cdots \geq \gamma_d$, $\Gamma \coloneqq \mathrm{diag}(\gamma_1, \cdots, \gamma_d)$,

Rectangular matrix $A \in \mathbb{R}^{m \times d}$ with singular values $\sigma_1 \geq \cdots \geq \sigma_d$ and diagonalization $M = \mathrm{U}\Sigma V^\top$.

$$\min_{\hat{V} \in \mathrm{O}(d)} ||V\Gamma^2 V^\top - \hat{V}\Gamma^2 \hat{V}^\top||_F$$

Special cases:

- Singular Subspace Recovery: $\gamma_1 = \cdots = \gamma_k = 1, \quad \gamma_{k+1} = \cdots = \gamma_d = 0$
- Rank-k covariance approximation: $\gamma_i = \sigma_i, \quad i \leq k$

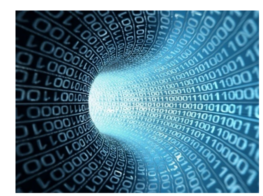
- Many applications to ML, statistics, medicine, engineering, etc.

Rectangular Matrix Perturbations

Given:
- $\gamma_1 \geq \cdots \geq \gamma_d$, $\Gamma := \mathrm{diag}(\gamma_1, \cdots, \gamma_d)$,
- A (deterministic) rectangular $A \in R^{m\times d}$ with singular values $\sigma_1 \geq \cdots \geq \sigma_d$ and singular value decomposition $\mathrm{A} = \mathrm{U}\Sigma V^\top$.
- A random rectangular matrix $G \in \mathbb{R}^{m\times d}$ with iid $N(0,1)$ entries

Let $\hat{A} = A + G$, and let $\hat{A} = \hat{U}\hat{\Sigma}\hat{V}^\top$ be its singular value decomposition

Goal: find an upper bound on $\left\|\hat{V}\Gamma^2\hat{V}^\top - V\Gamma^2 V^\top\right\|_F \leq ?$



Some Applications:
- Statistics under noise-corrupted data
- Differential privacy (releasing an entire privatized *dataset*)
- Randomized numerical linear algebra

Previous work: Singular Vector Perturbation

- Previous works give perturbation bounds for worst-case (deterministic) perturbations E: $\hat{A} = A + E, \quad \hat{A} =: \hat{U}\,\hat{\Sigma}\,\hat{V}^\top$

$$|||V_k V_k^\top - \hat{V}_k \hat{V}_k^\top||| \leq \frac{|||E|||}{\sigma_k - \sigma_{k+1}}$$ [Wedin, '72]

For Gaussian E, implies $\left\| V_k V_k^\top - \hat{V}_k\,\hat{V}_k^\top \right\|_F \leq O\left(\frac{\sqrt{m}\sqrt{k}}{\sigma_k - \sigma_{k+1}}\right)$ w.h.p.

- More recent works give improved bounds when perturbation is, e.g., a Gaussian random matrix: ($r := \mathrm{rank}(A)$)

$$\max\left(\|\hat{U}_k\hat{U}_k^\top - U_k U_k^\top\|_F,\ \|\hat{V}_k\hat{V}_k^\top - V_k V_k^\top\|_F\right) \leq O\left(r\sqrt{k}\sqrt{\sum_{j=1}^{k}\frac{1}{(\sigma_j - \sigma_{k+1})^2}} + \frac{\sqrt{m}\sqrt{k}}{\sigma_k}\right)$$ w.h.p. [O'Rourke,Vu, Wang, '23]

- Many works also give bounds specialized to symmetric/Hermitian case: [Davis, Kahan '70] (deterministic perturb.), e.g. [Dwork, Talwar, Thakurta, Zhang '14] , [Mangoubi, Vishnoi '22, '23, '25] (Gaussian perturbations), & more general random perturbations e.g. [O'Rourke, Vu, Wang,'18]

Bounds of [O'Rourke,Vu,Wang,23] are tight for ***left*** singular subspace $U_k \subseteq \mathbb{R}^{m\times k}$, $m \geq d$

Current bounds on **right** singular subspace $V_k \subseteq \mathbb{R}^{d\times k}$ are not tight for all σ when perturbation is Gaussian, and **depend on m**.

In many applications, $m = \#$ of datapoints, $d =$ number of features, with $m \gg d$.

Can one obtain *right* singular subspace bounds independent of m, even when $m \gg d$?

Main result [L-M '25] : Rectangular Gaussian perturbations

Assumption(A): The top-k singular values of A satisfy $\sigma_i - \sigma_{i+1} \geq \widetilde{\Omega}(\sqrt{dT})\ \forall\, i \leq k$

Theorem: Given $T > 0$, $k \leq d$, $\gamma_1 \geq \cdots \geq \gamma_d$ with $\gamma_i = 0$ for $i > k$, $A \in \mathbb{R}^{m \times d}$ with singular values $\sigma_1 \geq \cdots \geq \sigma_d$. Let $\hat{A} = A + \sqrt{T}G$ and $G, V, \hat{V}$ as above. Then

$$\mathbb{E}[\|\hat{V}\Gamma^2\hat{V}^\top - V\Gamma^2 V^\top\|_F^2] \leq \tilde{O}\left(\sum_{i=1}^{k}\sum_{j=i+1}^{d}\frac{(\gamma_i^2-\gamma_j^2)}{(\sigma_i-\sigma_j)^2}\right)T.$$

Corollary (right singular subspace recovery of Gaussian-perturbed *rectangular* matrix):

$$\mathbf{E}\,[||V_k V_k^\top - \widehat{V}_k \widehat{V}_k^\top||_F] \leq \tilde{O}(\frac{\sqrt{d}\sqrt{k}}{\sigma_k - \sigma_{k+1}})$$

$$\mathbf{E}\,[||V_k V_k^\top - \hat{V}_k \hat{V}_k^\top||_F] \leq \tilde{O}(\frac{\sqrt{d}}{\sigma_k - \sigma_{k+1}})$$ (if we also have $\sigma_i - \sigma_{i+1} \geq \sigma_k - \sigma_{k+1} \forall\, i \leq k$)

- Improves by $\sqrt{k}\frac{\sqrt{m}}{\sqrt{d}}$ (in expectation) on bound $||V_k V_k^\top - \hat{V}_k \hat{V}_k^\top||_F \leq O\left(\frac{\sqrt{m}\sqrt{k}}{\sigma_k - \sigma_{k+1}}\right)$ w.h.p. implied by [Davis, Kahan '70], [Wedin, '72], *eliminating dependence on m*
- Similar improvement on right sing. sub. bound of [O'Rourke,Vu,Wang,23] if e.g. $\sigma_k - \sigma_{k+1} = \Omega(\sigma_k)$

Corollary (covariance approximation of Gaussian-perturbed *rectangular* matrix):

$$\mathbb{E}\left[\|\hat{V}\hat{\Sigma}_k^\top\hat{\Sigma}_k\hat{V}^\top - V\Sigma_k^\top\Sigma_k V^\top\|_F\right] \leq O\left(\sqrt{k}\sqrt{d}\left(\sigma_1 + \sigma_k\frac{\sigma_k}{\sigma_k - \sigma_{k+1}}\right)\right)$$

- Improves by $\sqrt{k}\frac{\sqrt{m}}{\sqrt{d}}$ (in expectation) on bound implied by [O'Rourke, Vu, Wang '18] for Gaussian perturbations if e.g. $\sigma_k - \sigma_{k+1} = \Omega(\sigma_k)$, *eliminating dependence on m*

Singular Vector Flow: The Dyson Bessel Process

View addition of Gaussian noise as *rectangular* matrix-valued diffusion

$$\hat{A}(t) = A + W(t), \qquad t > 0$$

- Each entry of $W(t)$ is a standard Brownian Motion (BM)

Take SVD of $\hat{A}(t) = U(t)\Sigma(t)V^\top(t)$

$$U(t) = \begin{pmatrix} u_1(t), & \cdots & , u_d(t) \end{pmatrix} \quad \Sigma(t) = \begin{pmatrix} \sigma_1(t) & & \\ & \ddots & \\ & & \sigma_d(t) \end{pmatrix} \quad V(t) = \begin{pmatrix} v_1(t), & \cdots & , v_d(t) \end{pmatrix}$$

Eigenvalues and eigenvectors evolve according to Dyson Bessel Process SDEs (see e.g. [Bru, '89]):

$$\mathrm{d}\sigma_i(t) = \mathrm{d}\beta_{ii}(t) + \left(\frac{1}{2\sigma_i(t)} \sum_{j \neq i} \frac{(\sigma_i(t))^2 + (\sigma_j(t))^2}{(\sigma_i(t))^2 - (\sigma_j(t))^2} + \frac{m-1}{2\sigma_i(t)} \right) \mathrm{d}t, 1 \leq i \leq d,$$

$$\mathrm{d}v_i(t) = \sum_{j \neq i} v_j(t) \sqrt{\frac{(\sigma_j(t))^2 + (\sigma_i(t))^2}{((\sigma_j(t))^2 - (\sigma_i(t))^2)^2}} \mathrm{d}\beta_{ji}(t) - \frac{1}{2} v_i(t) \sum_{j \neq i} \frac{(\sigma_j(t))^2 + (\sigma_i(t))^2}{((\sigma_j(t))^2 - (\sigma_i(t))^2)^2} \mathrm{d}t$$

$$=: \sum_{j \neq i} v_j(t) c_{ij}(t) \mathrm{d}\beta_{ji}(t) - \frac{1}{2} v_i(t) \sum_{j \neq i} c_{ij}^2(t) \mathrm{d}t,$$

Here $\beta_{ij}(t),\ 1 \leq i < j \leq d$ is a family of iid standard Brownian motions

(analogous SDEs for $du_i(t)$)

Using SDEs to bound the error:

- Define projected process: $\Psi(t) = V(t)\Gamma^2 V^\top(t)$
- We want to bound $||\Psi(T) - \Psi(0)||_F^2 = ||\int_0^T \mathrm{d}\Psi(t)\,||_F^2$

Use singular vector evolution equations to derive SDE for $\Psi(t)$:

$$\mathrm{d}\Psi(t) = \sum_{i=1}^{d}\sum_{j\neq i}(\gamma_i^2 - \gamma_j^2)\Bigg[\underbrace{\frac{1}{2}\sqrt{\frac{\sigma_j^2(t)+\sigma_i^2(t)}{(\sigma_j^2(t)-\sigma_i^2(t))^2}}\mathrm{d}\beta_{ji}(t)(v_i(t)v_j^\top(t) + v_j(t)v_i^\top(t))}_{\mathrm{d}\omega_{ij}(t)} - \frac{\sigma_j^2(t)+\sigma_i^2(t)}{(\sigma_j^2(t)-\sigma_i^2(t))^2}\mathrm{d}t(v_i(t)v_i^\top(t))\Bigg]$$

- $\mathrm{d}\Psi(t)$ is a sum of independent random terms $\mathrm{d}\omega_{ij}(t)$:
  - $\mathrm{d}\omega_{ij}(t)$ independent for all i, j, and independent of *past* times t

Next, use independence to "add up" Frobenius norms of $\mathrm{d}\omega_{ij}(t)$ as sum-of-squares. Then use Ito's Lemma to integrate Frobenius norm over time:

$$\mathbb{E}\left[\|\Psi(T) - \Psi(0)\|_F^2\right] \leq \int_0^T \mathbb{E}\left[\sum_{i=1}^{d}\sum_{j\neq i}(\gamma_i^2 - \gamma_j^2)^2\frac{\sigma_j^2(t)+\sigma_i^2(t)}{(\sigma_j^2(t)-\sigma_i^2(t))^2}\right]\mathrm{d}t + \cdots$$

Finally, use Weyl's inequality to bound singular value gaps:

$$\sigma_i(t) - \sigma_j(t) \geq \sigma_i - \sigma_j - \|W(t)\|_2 \geq \frac{1}{2}(\sigma_i - \sigma_j) \quad \text{w.h.p. ,}$$

as long as $\sigma_i - \sigma_{i+1} \geq \sqrt{d} \quad \forall\, i \leq k.$

Conclusion

Introduced diffusion-based techniques for bounding singular vectors of *rectangular* matrices perturbed by Gaussian noise

Singular vector perturbation bound: $\mathbf{E}\,[||V_k V_k^\top - \hat{V}_k \hat{V}_k^\top||_F] \leq \tilde{O}(\frac{\sqrt{d}}{\sigma_k - \sigma_{k+1}})$

Improves by $\sqrt{k}\frac{\sqrt{m}}{\sqrt{d}}$ when noise is Gaussian, $\sigma_i - \sigma_{i+1} \geq \sigma_k - \sigma_{k+1},\ i \leq k$

Replaces dependence on m with dependence on d!

Covariance approximation bound:

$$\mathbb{E}\left[\|\hat{V}\hat{\Sigma}_k^\top \hat{\Sigma}_k \hat{V}^\top - V\Sigma_k^\top \Sigma_k V^\top\|_F\right] \leq O\left(\sqrt{k}\sqrt{d}\left(\sigma_1 + \sigma_k \frac{\sigma_k}{\sigma_k - \sigma_{k+1}}\right)\right)$$

Improves by $\sqrt{k}\frac{\sqrt{m}}{\sqrt{d}}$ when noise is Gaussian, $\sigma_i - \sigma_{i+1} \geq \sigma_k - \sigma_{k+1},\ i \leq k$

Replaces dependence on m with dependence on d!

Can one obtain stronger bounds for inputs A with additional structure?

Can one extend diffusion techniques to non-Gaussian noise?

Thanks!