

Probabilistic Algorithms for Finding Matrix Decompositions

Sparsity and Compressed Sensing
MVA 2017/18 - ENS Cachan

Alex Nowak

Advised by: Gabriel Peyré

Problem: Low rank approximation of a matrix

$$\begin{matrix} \mathbf{A} & \approx & \mathbf{B} & \mathbf{C}, \\ m \times n & & m \times k & k \times n. \end{matrix}$$

- **Standard Decompositions:**

- ① SVD:

$$\mathbf{A} = \left(\mathbf{U} \mathbf{\Sigma}^{1/2} \right) \left(\mathbf{V} \mathbf{\Sigma}^{1/2} \right)^*$$

- ② QR:

$$\mathbf{A} = \mathbf{Q} \mathbf{R}$$

- **Classical Algorithms:**

- ① Computationally expensive: $\mathcal{O}(mnk)$.
 - ② Need $\mathcal{O}(k)$ passes over data.
 - ③ Can't deal with inaccurate matrices.

\implies Not adequate to deal with massive datasets!

Two Stages Solution

- **Stage A:** (*Randomized*) Find $m \times k$ orthonormal \mathbf{Q} whose columns approximate the range of \mathbf{A} :

$$\mathbf{A} \approx \mathbf{Q}\mathbf{Q}^* \mathbf{A}$$

- **Stage B:** (*Deterministic*) Use \mathbf{Q} to find the desired decomposition.
E.g, set $\mathbf{B} = \mathbf{Q}$ and $\mathbf{C} = \mathbf{Q}^* \mathbf{A}$

Stage A: Randomize!

RANDOMIZED RANGE FINDER

- 1 Draw an $n \times \ell$ Gaussian random matrix $\mathbf{\Omega}$.
- 2 Form the $m \times \ell$ matrix $\mathbf{Y} = \mathbf{A}\mathbf{\Omega}$.
- 3 Construct an $m \times \ell$ matrix \mathbf{Q} whose columns form an orthonormal basis for the range of \mathbf{Y} , e.g., using the QR factorization $\mathbf{Y} = \mathbf{Q}\mathbf{R}$.

- If we set $\ell = k + p$ with $p > 0$ ($p \ll k$) the *oversampling parameter*, we can control the error $\|\mathbf{A} - \mathbf{Q}\mathbf{Q}^* \mathbf{A}\|$ with arbitrary precision! 😊

$$\mathbb{E} \|\mathbf{A} - \mathbf{Q}\mathbf{Q}^* \mathbf{A}\| \leq \left(1 + \sqrt{\frac{k}{p-1}}\right) \sigma_{k+1} + \frac{e\sqrt{k+p}}{p} \left(\sum_{j>k} \sigma_j^2\right)^{1/2}.$$

- But, error large if singular values $(\sigma_i)_{i=1}^r$ decay slowly... 😞
- But, product $\mathbf{A}\mathbf{\Omega}$ too expensive... $\mathcal{O}(mn\ell)$ 😞

Stage A: Solving the two issues.

- If \mathbf{A} has slow decaying singular values $(\sigma_i)_{i=1}^r$, take powers! 😊

$$\mathbf{Y} = \mathbf{B}\mathbf{\Omega} \quad \mathbf{B} = (\mathbf{A}\mathbf{A}^*)^q \mathbf{A} \quad \sigma_j(\mathbf{B}) = \sigma_j(\mathbf{A})^{2q+1}, \quad j = 1, 2, 3, \dots$$

$$\frac{\mathbb{E} \|\mathbf{A} - \mathbf{Q}\mathbf{Q}^* \mathbf{A}\|}{\sigma_{k+1}} \leq \left[1 + \sqrt{\frac{k}{p-1}} + \frac{e\sqrt{k+p}}{p} \cdot \sqrt{\min\{m, n\} - k} \right]^{1/(2q+1)}$$

- Make the product $\mathbf{A}\mathbf{\Omega}$ cheaper by using a *structured random matrix* instead of Gaussian.

Use *Subsampled Random Fourier Transform* (SRFT)

$$\mathbf{\Omega} = \sqrt{n/\ell} \cdot \mathbf{DFR}^*$$

Can compute the sample matrix $\mathbf{Y} = \mathbf{A}\mathbf{\Omega}$ in $\mathcal{O}(mn \log(\ell))$ operations via a *subsampled FFT*. 😊

Stage B: Construct $A \approx BC$ decomposition from Q .

Obs: Deterministic stage.

- **Direct SVD:** Set $B = Q$ and $C = Q^*A$.

\implies Product Q^*A costs $\mathcal{O}(mnk)$, too expensive!

- *Solution:* Can be reduced to $\mathcal{O}(k^2(m+n))$ via *row extraction*, but an additive error term comes up.

Theory: Expectation and high probability error bounds.

Obs: Deterministic stage.

- **Expectation bound:**

$$\mathbb{E} \|\mathbf{A} - \mathbf{Q}\mathbf{Q}^* \mathbf{A}\| \leq \left(1 + \sqrt{\frac{k}{p-1}}\right) \sigma_{k+1} + \frac{e\sqrt{k+p}}{p} \left(\sum_{j>k} \sigma_j^2\right)^{1/2}.$$

with probability

- **Tail bound:**

$$\begin{aligned} & \|(\mathbf{I} - \mathbf{P}_Y) \mathbf{A}\| \leq \\ & \leq \left(1 + 8\sqrt{(k+p) \cdot p \log p}\right) \sigma_{k+1} + 3\sqrt{k+p} \left(\sum_{j>k} \sigma_j^2\right)^{1/2}, \end{aligned}$$

with failure probability at most $6p^{-p}$.

Remark: Very fast decay with *oversampling parameter* p .

Experiment: Compare theoretical bound with numerical error for powers of normalized gaussian matrices \mathbf{A} .

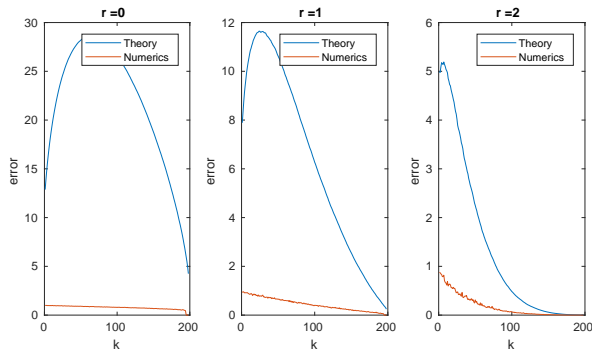


Figure: Comparison between theoretical mean bound and numerical error produced by *Randomized Range Finder* for $(\mathbf{A}\mathbf{A}^*)^r \mathbf{A}$ with $r = 0, 1, 2$.

Experiment: Compare theoretical bound with numerical error for powers of normalized gaussian matrices \mathbf{A} .

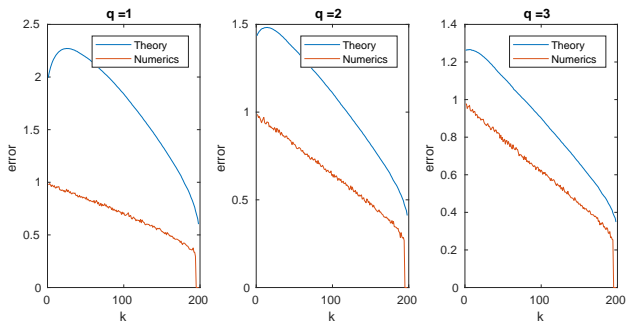


Figure: Comparison between theoretical mean bound and numerical error produced by *Randomized Power Iteration* for \mathbf{A} and $q = 1, 2, 3$

Experiment: MNIST and Laplacian eigenvectors of image patches.

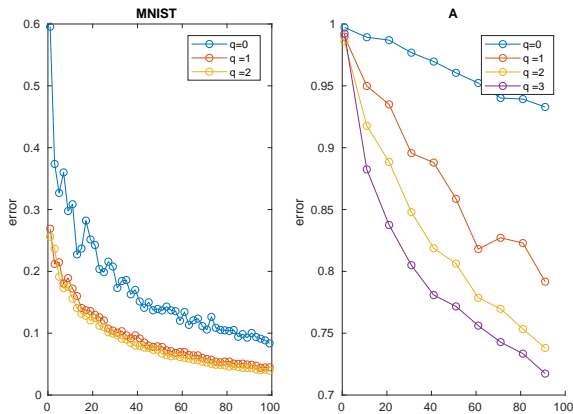


Figure: **Left:** Fast decaying singular values. **Right:** Slow decaying singular values.

- Necessary tool for Data Scientists working with massive or inaccurate datasets.
- Methods with strong experimental evidence backed up with theory.
- Possible future directions: Improve error bounds sharpness under additional hypothesis of matrix \mathbf{A} .