Lecture 2: Singular value decomposition (SVD)

School of Mathematical Sciences, Xiamen University

1. Singular value decomposition

- Definition: Let m and n be arbitrary positive integers ($m \geq n$ or $m < n$). Given $\mathbf{A} \in \mathbb{C}^{m\times n}$, not necessarily of full rank, a *singular value decomposition* (SVD) of $\mathbf{A}$ is a factorization

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*,$$

where $\mathbf{U} \in \mathbb{C}^{m\times m}$ is unitary, $\mathbf{V} \in \mathbb{C}^{n\times n}$ is unitary, and $\mathbf{\Sigma} \in \mathbb{R}^{m\times n}$ is diagonal. In addition, it is assumed that the diagonal entries σ_i of $\mathbf{\Sigma}$ are nonnegative and in nonincreasing order; that is

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0,$$

where $p = \min\{m, n\}$.

Theorem 1 (Existence of SVD)

Every matrix $\mathbf{A} \in \mathbb{C}^{m\times n}$ has a singular value decomposition.

Proof. Assume $\mathbf{A} \neq \mathbf{0}$; otherwise we can take $\mathbf{\Sigma} = \mathbf{0}$ and let $\mathbf{U}$ and $\mathbf{V}$ be arbitrary unitary matrices. Next, we use induction on m and n to prove the existence of SVD for the case $m \geq n$ (consider $\mathbf{A}^*$ if $m < n$): Assume that an SVD exists for any $(m-1) \times (n-1)$ matrix and prove it for any $m \times n$ matrix.

(i) The basic step: $m \geq n = 1$.

Write $\mathbf{A} = \mathbf{u}_1 \mathbf{\Sigma}_1 \mathbf{V}^*$ with $\mathbf{u}_1 = \mathbf{A}/\|\mathbf{A}\|_2, \mathbf{\Sigma}_1 = \|\mathbf{A}\|_2$ and $\mathbf{V} = 1$. Choose $\mathbf{U}_{\mathrm{c}}$ such that $\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{U}_{\mathrm{c}} \end{bmatrix} \in \mathbb{C}^{m \times m}$ is unitary. Let $\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \end{bmatrix}^\top \in \mathbb{R}^{m \times 1}$. Then $\mathbf{A}$ has an SVD $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$.

(ii) The induction step: $m \geq n > 1$.

Let $\widetilde{\mathbf{v}}_1 \in \mathbb{C}^n$ be a unit (i.e., $\|\widetilde{\mathbf{v}}_1\|_2 = 1$) eigenvector corresponding to the eigenvalue $\lambda_{\max}(\mathbf{A}^*\mathbf{A})$. Then we have $\|\mathbf{A}\widetilde{\mathbf{v}}_1\|_2 = \|\mathbf{A}\|_2 > 0$. Let $\widetilde{\mathbf{u}}_1 = \mathbf{A}\widetilde{\mathbf{v}}_1/\|\mathbf{A}\widetilde{\mathbf{v}}_1\|_2$, which is a unit vector. Choose $\widetilde{\mathbf{U}}_{\mathrm{c}}$ and $\widetilde{\mathbf{V}}_{\mathrm{c}}$ such that $\widetilde{\mathbf{U}} = \begin{bmatrix} \widetilde{\mathbf{u}}_1 & \widetilde{\mathbf{U}}_{\mathrm{c}} \end{bmatrix} \in \mathbb{C}^{m \times m}$ and $\widetilde{\mathbf{V}} = \begin{bmatrix} \widetilde{\mathbf{v}}_1 & \widetilde{\mathbf{V}}_{\mathrm{c}} \end{bmatrix} \in \mathbb{C}^{n \times n}$ are unitary.

Now we have

$$\widetilde{\mathbf{U}}^*\mathbf{A}\widetilde{\mathbf{V}} = \begin{bmatrix} \widetilde{\mathbf{u}}_1^* \\ \widetilde{\mathbf{U}}_c^* \end{bmatrix} \mathbf{A} \begin{bmatrix} \widetilde{\mathbf{v}}_1 & \widetilde{\mathbf{V}}_c \end{bmatrix} = \begin{bmatrix} \widetilde{\mathbf{u}}_1^*\mathbf{A}\widetilde{\mathbf{v}}_1 & \widetilde{\mathbf{u}}_1^*\mathbf{A}\widetilde{\mathbf{V}}_c \\ \widetilde{\mathbf{U}}_c^*\mathbf{A}\widetilde{\mathbf{v}}_1 & \widetilde{\mathbf{U}}_c^*\mathbf{A}\widetilde{\mathbf{V}}_c \end{bmatrix}.$$

We note that

$$\widetilde{\mathbf{u}}_1^*\mathbf{A}\widetilde{\mathbf{v}}_1 = \frac{(\mathbf{A}\widetilde{\mathbf{v}}_1)^*(\mathbf{A}\widetilde{\mathbf{v}}_1)}{\|\mathbf{A}\widetilde{\mathbf{v}}_1\|_2} = \|\mathbf{A}\widetilde{\mathbf{v}}_1\|_2 = \|\mathbf{A}\|_2,$$

$$\widetilde{\mathbf{U}}_c^*\mathbf{A}\widetilde{\mathbf{v}}_1 = \widetilde{\mathbf{U}}_c^*\widetilde{\mathbf{u}}_1\|\mathbf{A}\widetilde{\mathbf{v}}_1\|_2 = \mathbf{0},$$

and

$$\widetilde{\mathbf{u}}_1^*\mathbf{A}\widetilde{\mathbf{V}}_c = \frac{(\mathbf{A}\widetilde{\mathbf{v}}_1)^*}{\|\mathbf{A}\widetilde{\mathbf{v}}_1\|_2}\mathbf{A}\widetilde{\mathbf{V}}_c = \frac{\widetilde{\mathbf{v}}_1^*\mathbf{A}^*\mathbf{A}\widetilde{\mathbf{V}}_c}{\|\mathbf{A}\widetilde{\mathbf{v}}_1\|_2} = \frac{\lambda_{\max}(\mathbf{A}^*\mathbf{A})\widetilde{\mathbf{v}}_1^*\widetilde{\mathbf{V}}_c}{\|\mathbf{A}\widetilde{\mathbf{v}}_1\|_2} = \mathbf{0}.$$

Let

$$\sigma_1 := \|\mathbf{A}\|_2.$$

Then we have

$$\widetilde{\mathbf{U}}^*\mathbf{A}\widetilde{\mathbf{V}} = \begin{bmatrix} \sigma_1 & \mathbf{0} \\ \mathbf{0} & \widetilde{\mathbf{U}}_{\mathrm{c}}^*\mathbf{A}\widetilde{\mathbf{V}}_{\mathrm{c}} \end{bmatrix}.$$

By the induction hypothesis, we know that the $(m-1)\times(n-1)$ matrix $\widetilde{\mathbf{U}}_{\mathrm{c}}^*\mathbf{A}\widetilde{\mathbf{V}}_{\mathrm{c}}$ has an SVD:

$$\widetilde{\mathbf{U}}_{\mathrm{c}}^*\mathbf{A}\widetilde{\mathbf{V}}_{\mathrm{c}} = \widehat{\mathbf{U}}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{V}}^*.$$

It follows from $\sigma_1 = \|\mathbf{A}\|_2$, unitary invariance of $\|\cdot\|_2$, and

$$\widetilde{\mathbf{U}}^*\mathbf{A}\widetilde{\mathbf{V}} = \begin{bmatrix} \sigma_1 & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{U}}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{V}}^* \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{U}} \end{bmatrix} \begin{bmatrix} \sigma_1 & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{\Sigma}} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{V}} \end{bmatrix}^*$$

that $\sigma_1 \geq \|\widehat{\mathbf{\Sigma}}\|_2$. Now it is straightforward to show that

$$\mathbf{A} = \widetilde{\mathbf{U}} \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{U}} \end{bmatrix} \begin{bmatrix} \sigma_1 & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{\Sigma}} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{V}} \end{bmatrix}^* \widetilde{\mathbf{V}}^* =: \mathbf{U} \begin{bmatrix} \sigma_1 & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{\Sigma}} \end{bmatrix} \mathbf{V}^*$$

is an SVD of $\mathbf{A}$. □

- Full SVD:

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$$

- Reduced SVD (the case $m \geq n$):

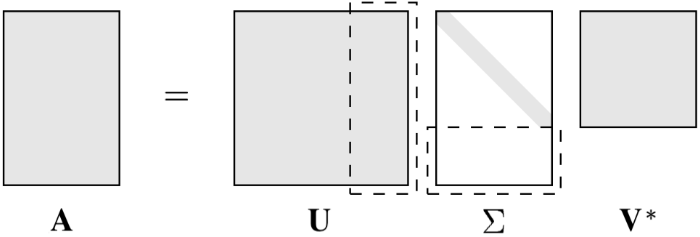
$$\mathbf{A} = \mathbf{U}_n\mathbf{\Sigma}_n\mathbf{V}^*$$

where

$$\mathbf{U}_n = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix},$$

and

$$\mathbf{\Sigma}_n = \mathrm{diag}\{\sigma_1, \sigma_2, \ldots, \sigma_n\}.$$



- Rank SVD or compact SVD or condensed SVD:

$$\mathbf{A} = \begin{bmatrix} \mathbf{U}_r & \mathbf{U}_\mathrm{c} \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_r^* \\ \mathbf{V}_\mathrm{c}^* \end{bmatrix} = \mathbf{U}_r \mathbf{\Sigma}_r \mathbf{V}_r^* = \sum\nolimits_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^*$$

  where $r = \operatorname{rank}(\mathbf{A})$,

$$\mathbf{U}_r = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_r \end{bmatrix}, \quad \mathbf{U}_\mathrm{c} = \begin{bmatrix} \mathbf{u}_{r+1} & \mathbf{u}_{r+2} & \cdots & \mathbf{u}_m \end{bmatrix},$$

$$\mathbf{V}_r = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_r \end{bmatrix}, \quad \mathbf{V}_\mathrm{c} = \begin{bmatrix} \mathbf{v}_{r+1} & \mathbf{v}_{r+2} & \cdots & \mathbf{v}_n \end{bmatrix},$$

  and

$$\mathbf{\Sigma}_r = \operatorname{diag}\{\sigma_1, \sigma_2, \ldots, \sigma_r\}.$$

- $\{\sigma_i^2, \mathbf{u}_i\}$ are eigenvalue-eigenvector pairs of $\mathbf{A}\mathbf{A}^*$, and $\{\sigma_i^2, \mathbf{v}_i\}$ are eigenvalue-eigenvector pairs of $\mathbf{A}^*\mathbf{A}$:

$$\mathbf{A}\mathbf{A}^*\mathbf{u}_i = \sigma_i^2 \mathbf{u}_i, \quad \mathbf{A}^*\mathbf{A}\mathbf{v}_i = \sigma_i^2 \mathbf{v}_i, \quad i = 1, 2, \ldots, p$$

- $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p$ are called the *singular values* of $\mathbf{A}$.

- the *left singular vectors* $\{\mathbf{u}_i\}$ and the *right singular vectors* $\{\mathbf{v}_i\}$:
  $\mathbf{u}_i^*\mathbf{A} = \sigma_i \mathbf{v}_i^*, \quad \mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i, \quad i = 1, 2, \ldots, p$

Theorem 2

The set of singular values $\{\sigma_i\}$ is uniquely determined and invariant under unitary multiplication.

Theorem 3

If $\mathbf{A}$ is square and all the σ_i are distinct, the left and right singular vectors are uniquely determined up to complex signs (i.e., complex scalar factors of absolute value 1).

Hint: There exists only one linearly independent eigenvector for each eigenvalue of $\mathbf{A}^*\mathbf{A}$ or $\mathbf{A}\mathbf{A}^*$ if the eigenvalues are distinct.

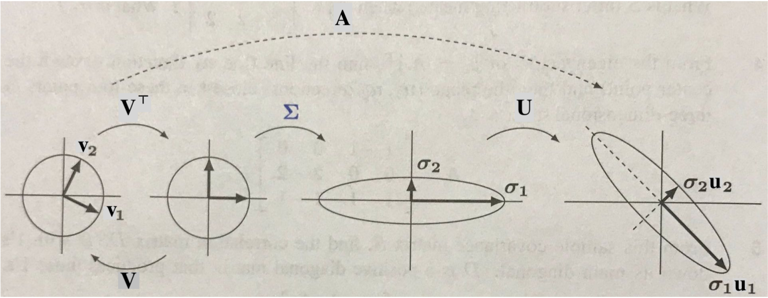
Theorem 4 (Real SVD)

Every matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ has a real singular value decomposition.

1.1. Geometric observation

- The image of the unit sphere (in the 2-norm) of $\mathbb{C}^n$ under any $m \times n$ matrix is a hyperellipse of $\mathbb{C}^m$.

  For example, 2×2 real matrix $\mathbf{A}$



SVD of a matrix can not be emphasized too much!

2. Matrix properties via SVD: $\mathbf{A} = \mathbf{U\Sigma V}^*$

- 2-norm

$$\|\mathbf{A}\|_2 = \sigma_1 = \|\mathbf{A}^*\|_2 = \|\mathbf{A}^\top\|_2 = \|\overline{\mathbf{A}}\|_2$$

- F-norm

$$\|\mathbf{A}\|_{\mathrm{F}} = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_r^2} = \|\mathbf{A}^*\|_{\mathrm{F}} = \|\mathbf{A}^\top\|_{\mathrm{F}} = \|\overline{\mathbf{A}}\|_{\mathrm{F}}$$

- *range* or *column space* of $\mathbf{A} \in \mathbb{C}^{m\times n}$, spanned by the columns of $\mathbf{A}$

$$\begin{aligned}\mathrm{range}(\mathbf{A}) := &\{\mathbf{y} \in \mathbb{C}^m \mid \exists\ \mathbf{x} \in \mathbb{C}^n \quad s.t. \quad \mathbf{y} = \mathbf{A}\mathbf{x}\}\\ = &\,\mathrm{span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}\end{aligned}$$

- *kernel* or *nullspace* of $\mathbf{A} \in \mathbb{C}^{m\times n}$

$$\mathrm{null}(\mathbf{A}) := \{\mathbf{x} \in \mathbb{C}^n \mid \mathbf{A}\mathbf{x} = \mathbf{0}\} = \mathrm{span}\{\mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \ldots, \mathbf{v}_n\}$$

- Range and nullspace of $\mathbf{A}^*$:

$$\mathrm{range}(\mathbf{A}^*) = \mathrm{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\} = \mathrm{null}(\mathbf{A})^\perp$$

$$\mathrm{null}(\mathbf{A}^*) = \mathrm{span}\{\mathbf{u}_{r+1}, \mathbf{u}_{r+2}, \ldots, \mathbf{u}_m\} = \mathrm{range}(\mathbf{A})^\perp$$

- Relations between the four subspaces

$$\mathrm{range}(\mathbf{A}^*) \perp \mathrm{null}(\mathbf{A}), \quad \mathrm{range}(\mathbf{A}^*) + \mathrm{null}(\mathbf{A}) = \mathbb{C}^n$$

$$\mathrm{range}(\mathbf{A}) \perp \mathrm{null}(\mathbf{A}^*), \quad \mathrm{range}(\mathbf{A}) + \mathrm{null}(\mathbf{A}^*) = \mathbb{C}^m$$

- Absolute value of determinant of $\mathbf{A} \in \mathbb{C}^{m\times m}$:

$$|\det(\mathbf{A})| = \prod_{i=1}^{m} \sigma_i$$

- If $\mathbf{A}$ is Hermitian, i.e., $\mathbf{A} = \mathbf{A}^*$, then

singular values are absolute values of eigenvalues.

- An eigendecomposition via SVD:

$$\begin{bmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{A}^* & \mathbf{0} \end{bmatrix} = \mathbf{Q} \begin{bmatrix} \mathbf{\Sigma}_r & & & \\ & -\mathbf{\Sigma}_r & & \\ & & \mathbf{0} & \\ & & & \mathbf{0} \end{bmatrix} \mathbf{Q}^*, \mathbf{Q} = \begin{bmatrix} \frac{\mathbf{U}_r}{\sqrt{2}} & \frac{\mathbf{U}_r}{\sqrt{2}} & \mathbf{U}_c & \mathbf{0} \\ \frac{\mathbf{V}_r}{\sqrt{2}} & \frac{-\mathbf{V}_r}{\sqrt{2}} & \mathbf{0} & \mathbf{V}_c \end{bmatrix}.$$

- A random square matrix is "always" nonsingular. Or more general, a random rectangular matrix is "always" of full rank. Why?

2.1. Low-rank approximation (LRA)

Theorem 5 (Eckart-Young-Mirski)

For any integer k with $1 \le k < r = \operatorname{rank}(\mathbf{A})$, define

$$\mathbf{A}_k = \sum\nolimits_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^*.$$

Then

$$\|\mathbf{A} - \mathbf{A}_k\|_2 = \sigma_{k+1} = \min_{\substack{\mathbf{B} \in \mathbb{C}^{m \times n}, \\ \operatorname{rank}(\mathbf{B}) \le k}} \|\mathbf{A} - \mathbf{B}\|_2,$$

and

$$\|\mathbf{A} - \mathbf{A}_k\|_{\mathrm{F}} = \sqrt{\sigma_{k+1}^2 + \cdots + \sigma_r^2} = \min_{\substack{\mathbf{B} \in \mathbb{C}^{m \times n}, \\ \operatorname{rank}(\mathbf{B}) \le k}} \|\mathbf{A} - \mathbf{B}\|_{\mathrm{F}}.$$

- Discussion: Is the minimizer in Theorem 5 unique?

Proof of Theorem 5.

- Suppose there is some $\mathbf{B} \in \mathbb{C}^{m \times n}$ with $\text{rank}(\mathbf{B}) \le k < r$ such that
$$\|\mathbf{A} - \mathbf{B}\|_2 < \sigma_{k+1}.$$
It follows that $\dim(\text{null}(\mathbf{B})) = n - \text{rank}(\mathbf{B}) \ge n - k$. Thus there exists an $(n-k)$-dimensional subspace $\mathcal{W} \subseteq \text{null}(\mathbf{B})$. For any nonzero $\mathbf{x} \in \mathcal{W}$, we have
$$\|\mathbf{A}\mathbf{x}\|_2 = \|(\mathbf{A} - \mathbf{B})\mathbf{x}\|_2 \le \|\mathbf{A} - \mathbf{B}\|_2 \|\mathbf{x}\|_2 < \sigma_{k+1}\|\mathbf{x}\|_2.$$
Let $\mathcal{V} = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k+1}\}$. For any $\mathbf{x} \in \mathcal{V}$, we have
$$\|\mathbf{A}\mathbf{x}\|_2 = \|\mathbf{A}\mathbf{V}_{k+1}\mathbf{y}\|_2 = \|\mathbf{U}_{k+1}\mathbf{\Sigma}_{k+1}\mathbf{y}\|_2 = \|\mathbf{\Sigma}_{k+1}\mathbf{y}\|_2 \ge \sigma_{k+1}\|\mathbf{x}\|_2.$$
Since $\dim\mathcal{W} + \dim\mathcal{V} = (n-k) + (k+1) > n$, there must be a nonzero vector lying in both, and this is a contradiction.
- Case $\|\cdot\|_F$: Generalized Inverses: Theory and Applications, 2nd edition, Adi Ben-Israel and Thomas N.E. Greville, Page 213. □

Application of low-rank approximation: image compression

- An image can be represented as a matrix. For example, typical grayscale images consist of a rectangular array of pixels, m in the vertical direction, n in the horizontal direction. The color of each of those pixels is denoted by a single number, an integer between 0 (black) and 255 (white). (This gives $2^8 = 256$ different shades of gray for each pixel. Color images are represented by three such matrices: one for red, one for green, and one for blue. Thus each pixel in a typical color image takes $(2^8)^3 = 2^{24}$ shades.)
- The objective of image compression is to reduce irrelevance and redundancy of the image data in order to be able to store or transmit data in an efficient form.
- Low-rank SVD approximation is a good candidate. (Note: `jpeg` compression algorithm uses similar idea, on subimages)

  Image Compression with Singular Value Decomposition Demo

3. Moore–Penrose pseudoinverse

- Let $\mathbf{A} \in \mathbb{C}^{m\times n}$ have an SVD (rank form) $\mathbf{A} = \mathbf{U}_r \mathbf{\Sigma}_r \mathbf{V}_r^*$. The *Moore–Penrose pseudoinverse* of $\mathbf{A}$, denoted by $\mathbf{A}^\dagger$:

$$\mathbf{A}^\dagger := \mathbf{V}_r \mathbf{\Sigma}_r^{-1} \mathbf{U}_r^* = \sum\nolimits_{i=1}^{r} \frac{1}{\sigma_i} \mathbf{v}_i \mathbf{u}_i^*.$$

- The matrix $\mathbf{A}^\dagger$ is the *unique* matrix satisfying the four equations

$$\mathbf{AXA} = \mathbf{A}, \quad \mathbf{XAX} = \mathbf{X}, \quad (\mathbf{AX})^* = \mathbf{AX}, \quad (\mathbf{XA})^* = \mathbf{XA}.$$

  For a proof, see Page 122 of Numerical Linear Algebra (in Chinese) by Zhihao Cao.

- If $\mathbf{A}$ has full column rank, then

$$\mathbf{A}^\dagger = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*.$$

  If $\mathbf{A}$ has full row rank, then

$$\mathbf{A}^\dagger = \mathbf{A}^*(\mathbf{A}\mathbf{A}^*)^{-1}.$$

4. A wonderful reference

- Zhihua Zhang (arXiv:1510.08532)

  The singular value decomposition, applications and beyond

5. Another proof of Theorem 5

- Holger Wendland

  Numerical Linear Algebra An Introduction

  Cambridge University Press, 2018.

  See Page 295, Theorem 7.41.

6. Computationally more feasible methods for LRA

- Adaptive cross approximation (ACA)

  See Page 297 of Numerical Linear Algebra An Introduction.

- Joel A. Tropp and Robert J. Webber (arXiv:2306.12418)

  Randomized algorithms for low-rank matrix approximation: Design, analysis, and applications