Part 1b: Singular value decomposition (SVD)

September 28, 2018

1. Singular value decomposition

Theorem 1 (SVD, case $m \geq n$)

Every matrix $\mathbf{A} \in \mathbb{C}^{m\times n}$ has a (full, reduced, and rank) singular value decomposition:

$$\begin{aligned}\mathbf{A} \quad &= \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^* = \mathbf{U}_n\boldsymbol{\Sigma}_n\mathbf{V}^* = \mathbf{U}_r\boldsymbol{\Sigma}_r\mathbf{V}_r^* = \sum\nolimits_{j=1}^{r} \sigma_j \mathbf{u}_j \mathbf{v}_j^* \\ &= \begin{bmatrix}\mathbf{U}_r & \mathbf{U}_\mathrm{c}\end{bmatrix} \begin{bmatrix}\boldsymbol{\Sigma}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix} \begin{bmatrix}\mathbf{V}_r^* \\ \mathbf{V}_\mathrm{c}^*\end{bmatrix},\end{aligned}$$

where $\boldsymbol{\Sigma} = \begin{bmatrix}\boldsymbol{\Sigma}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix} \in \mathbb{R}^{m\times n}$, $\boldsymbol{\Sigma}_r = \mathrm{diag}\{\sigma_1, \sigma_2, \cdots, \sigma_r\}$, $r = \mathrm{rank}(\mathbf{A})$,

$$\begin{aligned}&\mathbf{U} \in \mathbb{C}^{m\times m}, \quad \mathbf{U}^*\mathbf{U} = \mathbf{I}_m, \quad \mathbf{V} \in \mathbb{C}^{n\times n}, \quad \mathbf{V}^*\mathbf{V} = \mathbf{I}_n, \\ &\mathbf{U}_r = \begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_r\end{bmatrix}, \quad \mathbf{U}_\mathrm{c} = \begin{bmatrix}\mathbf{u}_{r+1} & \mathbf{u}_{r+2} & \cdots & \mathbf{u}_m\end{bmatrix}, \\ &\mathbf{V}_r = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_r\end{bmatrix}, \quad \mathbf{V}_\mathrm{c} = \begin{bmatrix}\mathbf{v}_{r+1} & \mathbf{v}_{r+2} & \cdots & \mathbf{v}_n\end{bmatrix}, \\ &\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_n = 0.\end{aligned}$$

Proof. We use induction on m and n.

- Assume that the SVD exists for $(m-1) \times (n-1)$ matrices and prove it for $m \times n$ matrices.

  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.

- The basic step occurs when $n = 1$ (since $m \geq n$).

  We 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 $\mathbf{V} = 1$.

- For the induction step, choose $\mathbf{v}$ so that $\|\mathbf{v}\|_2 = 1$ and $\|\mathbf{A}\|_2 = \|\mathbf{A}\mathbf{v}\|_2 > 0$. Such a vector $\mathbf{v}$ exists by

$$\|\mathbf{A}\|_2 = \max_{\|\mathbf{v}\|_2=1} \|\mathbf{A}\mathbf{v}\|_2 = \max_{\|\mathbf{v}\|_2=1} \sqrt{\mathbf{v}^*\mathbf{A}^*\mathbf{A}\mathbf{v}}.$$

  Let $\mathbf{u} = \mathbf{A}\mathbf{v}/\|\mathbf{A}\mathbf{v}\|_2$, which is a unit vector. Choose $\widehat{\mathbf{U}}$ and $\widehat{\mathbf{V}}$ so that $\widetilde{\mathbf{U}} = \begin{bmatrix} \mathbf{u} & \widehat{\mathbf{U}} \end{bmatrix}$ and $\widetilde{\mathbf{V}} = \begin{bmatrix} \mathbf{v} & \widehat{\mathbf{V}} \end{bmatrix}$ are $m \times m$ and $n \times n$ unitary matrices, respectively.

Now we have

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

We note that

$$\mathbf{u}^*\mathbf{A}\mathbf{v} = \frac{(\mathbf{A}\mathbf{v})^*(\mathbf{A}\mathbf{v})}{\|\mathbf{A}\mathbf{v}\|_2} = \|\mathbf{A}\mathbf{v}\|_2 = \|\mathbf{A}\|_2 \equiv \sigma_1,$$

and

$$\widehat{\mathbf{U}}^*\mathbf{A}\mathbf{v} = \widehat{\mathbf{U}}^*\mathbf{u}\|\mathbf{A}\mathbf{v}\|_2 = \mathbf{0}.$$

We claim $\mathbf{u}^*\mathbf{A}\widehat{\mathbf{V}} = \mathbf{0}$ too because otherwise

$$\begin{aligned} \sigma_1 &= \|\mathbf{A}\|_2 = \|\widetilde{\mathbf{U}}^*\mathbf{A}\widetilde{\mathbf{V}}\|_2 \\ &= \|\begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}\|_2 \cdot \|\widetilde{\mathbf{U}}^*\mathbf{A}\widetilde{\mathbf{V}}\|_2 \\ &\geq \|\begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}\widetilde{\mathbf{U}}^*\mathbf{A}\widetilde{\mathbf{V}}\|_2 = \|[\sigma_1 \ \mathbf{u}^*\mathbf{A}\widehat{\mathbf{V}}]\|_2 > \sigma_1, \end{aligned}$$

which is a contradiction.

Therefore,

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

Apply the induction hypothesis to the $(m-1)\times(n-1)$ matrix $\widehat{\mathbf{U}}^*\mathbf{A}\widehat{\mathbf{V}}$ to get an SVD:

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

It follows from

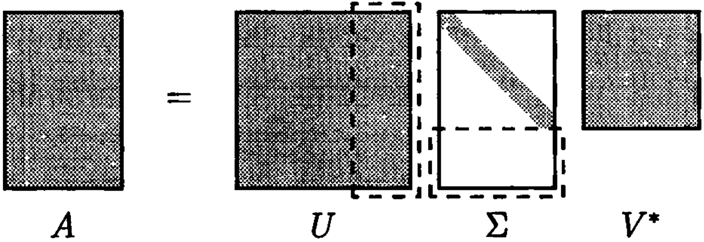
$$\widetilde{\mathbf{U}}^*\mathbf{A}\widetilde{\mathbf{V}} = \begin{bmatrix} \sigma_1 & \mathbf{0} \\ \mathbf{0} & \check{\mathbf{U}}\check{\mathbf{\Sigma}}\check{\mathbf{V}}^* \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \check{\mathbf{U}} \end{bmatrix} \begin{bmatrix} \sigma_1 & \mathbf{0}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

that

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

It is easy to show that this is an SVD of $\mathbf{A}$. $\square$

- Full SVD (case $m \geq n$) and Reduced SVD: Matlab `svd`



- σ_i^2 are eigenvalues of $\mathbf{AA}^*$ or $\mathbf{A}^*\mathbf{A}$, $\mathbf{u}_i$ and $\mathbf{v}_i$ are corresponding eigenvectors
- *singular values* $\{\sigma_i\}$: uniquely determined, invariant under unitary multiplication
- *left singular vectors* $\{\mathbf{u}_i\}$, *right singular vectors* $\{\mathbf{v}_i\}$:

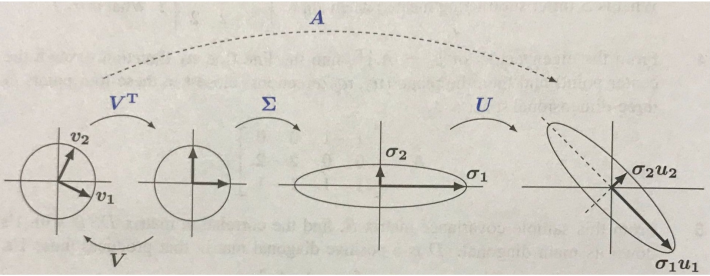
$$\mathbf{u}_i^*\mathbf{A} = \sigma_i \mathbf{v}_i^*, \quad \mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i, \quad i = 1, 2, \dots, n$$

- 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).

2. Geometric observation

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

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



The SVD of a matrix cannot be emphasized too much!

3. Matrix properties via the SVD

- 2-norm

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

- F-norm

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

- range($\mathbf{A}$): *column space* of $\mathbf{A}$, spanned by the columns of $\mathbf{A}$

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

- null($\mathbf{A}$): *kernel* or *null space* of $\mathbf{A}$

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

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

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

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

- Relations between the four subspaces

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

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

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

  singular values are absolute values of eigenvalues

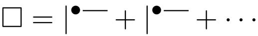
- Determinant of $\mathbf{A} \in \mathbb{C}^{m \times m}$

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

4. Low-rank approximation

- For any k with $0 \le k \le r$, define

$$\mathbf{A}_k = \sum_{j=1}^{k} \sigma_j \mathbf{u}_j \mathbf{v}_j^*.$$



Theorem 2

If $k = \min\{m, n\}$, define $\sigma_{k+1} = 0$. For $1 \le k \le \min\{m, n\}$, we have

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

and

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

Discussion: Is the minimizer unique?

Proof.

- Suppose there is some $\mathbf{B}$ with $\operatorname{rank}(\mathbf{B}) \leq k$ such that

$$\|\mathbf{A} - \mathbf{B}\|_2 < \sigma_{k+1} = \|\mathbf{A} - \mathbf{A}_k\|_2.$$

  Then there is an $(n-k)$-dimensional subspace $\mathcal{W} \subseteq \operatorname{null}(\mathbf{B})$. For any $\mathbf{x} \in \mathcal{W}$, we have

$$\|\mathbf{A}\mathbf{x}\|_2 = \|(\mathbf{A} - \mathbf{B})\mathbf{x}\|_2 \leq \|\mathbf{A} - \mathbf{B}\|_2 \|\mathbf{x}\|_2 < \sigma_{k+1} \|\mathbf{x}\|_2.$$

  Let $\mathcal{V} = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_{k+1}\}$. For any $\mathbf{x} \in \mathcal{V}$, we have

$$\|\mathbf{A}\mathbf{x}\|_2 = \|\mathbf{A}\mathbf{V}_{k+1}\mathbf{y}\|_2 = \|\mathbf{U}_{k+1}\mathbf{\Sigma}_{k+1}\mathbf{y}\|_2 = \|\mathbf{\Sigma}_{k+1}\mathbf{y}\|_2 \geq \sigma_{k+1}\|\mathbf{x}\|_2.$$

  Since $\dim\mathcal{W} + \dim\mathcal{V} > n$, there must be a nonzero vector lying in both, and this is a contradiction.

- The case for $\|\cdot\|_{\mathrm{F}}$, see Page 213 of Generalized Inverses: Theory and Applications, Adi Ben-Israel and Thomas N.E. Greville. □

5. Application: 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. See the Matlab codes in the website. (Note: `jpeg` compression algorithm uses similar idea, on subimages)

6. Discussion

- A random $n \times n$ matrix is "always" nonsingular. Why?

7. 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 $\mathbf{A}^\dagger$:

$$\mathbf{A}^\dagger = \mathbf{V}_r \mathbf{\Sigma}_r^{-1} \mathbf{U}_r^* = \sum_{j=1}^{r} \frac{1}{\sigma_j} \mathbf{v}_j \mathbf{u}_j^*.$$

- The matrix $\mathbf{A}^\dagger$ is the unique matrix satisfying the four Moore-Penrose 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 $\mathbf{A}$ is of full column rank, then $\mathbf{A}^\dagger = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*$.

8. A wonderful reference

- Zhihua Zhang

  The singular value decomposition, applications and beyond

  arXiv:1510.08532