MA4230 Matrix Computation

Chapter 2: Singular Value Decomposition (SVD)

Notation: diagonal matrix

A matrix $A \in \mathbb{R}^{m\times n}$ is called **diagonal** iff $A = \mathrm{diag}_{m\times n}(\alpha_1, \dots, \alpha_p)$ for some $\alpha_1, \dots, \alpha_p \in \mathbb{R}$, where $p := \min(m, n)$.

(i) When $m \geq n$:

$$\mathrm{diag}_{m\times n}(\alpha_1, \dots, \alpha_n) := \begin{pmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_n \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \in \mathbb{R}^{m\times n}.$$

(ii) When $m < n$:

$$\mathrm{diag}_{m\times n}(\alpha_1, \dots, \alpha_m) := \begin{pmatrix} \alpha_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_m & 0 & \cdots & 0 \end{pmatrix} \in \mathbb{R}^{m\times n}.$$

Examples of diagonal matrices

- $\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} = \mathrm{diag}_{3\times 3}(a, b, c)$.
- $\begin{pmatrix} a & 0 \\ 0 & b \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} = \mathrm{diag}_{5\times 2}(a, b)$.
- $\begin{pmatrix} a & 0 & 0 & 0 & 0 & 0 \\ 0 & b & 0 & 0 & 0 & 0 \\ 0 & 0 & c & 0 & 0 & 0 \\ 0 & 0 & 0 & d & 0 & 0 \end{pmatrix} = \mathrm{diag}_{4\times 6}(a, b, c, d)$.
- $\begin{pmatrix} a \\ 0 \end{pmatrix} = \mathrm{diag}_{2\times 1}(a)$.
- $\begin{pmatrix} a & 0 \end{pmatrix} = \mathrm{diag}_{1\times 2}(a)$.

2.1 Definition and geometric interpretation

Definition (Singular Value Decomposition (SVD))

Let $A \in \mathbb{R}^{m\times n}$ and write $p := \min(m,n)$. If there exist

$$\begin{aligned} U &= (u_1|\cdots|u_m) && \in \mathbb{R}^{m\times m} && \text{orthogonal,} \\ V &= (v_1|\cdots|v_n) && \in \mathbb{R}^{n\times n} && \text{orthogonal,} \\ \Sigma &= \operatorname{diag}_{m\times n}(\sigma_1,\dots,\sigma_p) && \in \mathbb{R}^{m\times n} && \text{with } \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0, \end{aligned}$$

such that

$$A = U\Sigma V^{\mathrm{T}},$$

then we call this a (full) **singular value decomposition (SVD)** of A with

- **singular values** $\sigma_1,\dots,\sigma_p \in [0,\infty)$,
- **left singular vectors** $u_1,\dots,u_m \in \mathbb{R}^m$,
- **right singular vectors** $v_1,\dots,v_n \in \mathbb{R}^n$.

Note: Since U and V are orthogonal matrices ($U^{-1} = U^{\mathrm{T}}$, $V^{-1} = V^{\mathrm{T}}$):

$$A = U\Sigma V^{\mathrm{T}} \iff AV = U\Sigma \iff U^{\mathrm{T}}AV = \Sigma.$$

Reduced SVD

We can simplify the SVD of rectangular matrices ($m \neq n$):

(i) Suppose $A \in \mathbb{R}^{m\times n}$, $m > n$ (tall), and that it has a SVD

$$A = \underbrace{U}_{m\times m}\underbrace{\Sigma}_{m\times n}\underbrace{V^{\mathrm{T}}}_{n\times n} = (u_1|\cdots|u_n|u_{n+1}|\cdots|u_m)\begin{pmatrix}\mathrm{diag}_{n\times n}(\sigma_1,\ldots,\sigma_n)\\ 0_{(m-n)\times n}\end{pmatrix} V^{\mathrm{T}}.$$

Can simplify:

$$A = (u_1|\cdots|u_n)\ \mathrm{diag}_{n\times n}(\sigma_1,\ldots,\sigma_n)\ V^{\mathrm{T}} =: \hat{U}\hat{\Sigma}V^{\mathrm{T}}.$$

(ii) Suppose $A \in \mathbb{R}^{m\times n}$, $m < n$ (wide), and that it has a SVD

$$A{=}U\Sigma V^{\mathrm{T}}{=}U\left(\mathrm{diag}_{m\times m}(\sigma_1,\ldots,\sigma_m)\,\middle|\,0_{m\times(n-m)}\right)(v_1|\cdots|v_m|v_{m+1}|\cdots|v_n)^{\mathrm{T}}$$

Can simplify:

$$A = U\ \mathrm{diag}_{m\times m}(\sigma_1,\ldots,\sigma_m)\ (v_1|\cdots|v_m)^{\mathrm{T}} =: U\hat{\Sigma}\hat{V}^{\mathrm{T}}.$$

$\Longrightarrow$ The SVD can be simplified to

$$A = \hat{U}\hat{\Sigma}\hat{V}^{\mathrm{T}}$$

with (write $p := \min(m, n)$)

$$\begin{aligned} \hat{U} &= (u_1|\cdots|u_p) \in \mathbb{R}^{m\times p} \\ \hat{V} &= (v_1|\cdots|v_p) \in \mathbb{R}^{n\times p} \\ \hat{\Sigma} &= \mathrm{diag}_{p\times p}(\sigma_1, \ldots, \sigma_p). \end{aligned}$$

Note that,

- if $m > n$: $\hat{U} \in \mathbb{R}^{m\times n}$ is a rectangular tall matrix with orthonormal columns, and $\hat{V} = V \in \mathbb{R}^{n\times n}$ is orthogonal.
- if $m < n$: $\hat{U} = U \in \mathbb{R}^{m\times m}$ is orthogonal, and $\hat{V} \in \mathbb{R}^{n\times m}$ is a rectangular tall matrix with orthonormal columns.
- (if $m = n$: $\hat{U} = U$, $\hat{V} = V$ and $\hat{\Sigma} = \Sigma$.)

A decomposition $A = \hat{U}\hat{\Sigma}\hat{V}^{\mathrm{T}}$ with $\hat{U} \in \mathbb{R}^{m\times p}$ and $\hat{V} \in \mathbb{R}^{n\times p}$ having orthonormal columns, and $\hat{\Sigma} \in \mathbb{R}^{p\times p}$ being diagonal with non-negative and non-increasing diagonal entries, is called a **reduced SVD** of A.

Some fundamental questions

- *Existence*: Does every matrix have a SVD? If not, do certain special matrices have a SVD?
- *Uniqueness*: Is the SVD, if it exists, unique? I.e., are the matrices U, V, Σ unique, or equivalently, are the left singular vectors, the right singular vectors, and the singular values unique? If not, what can we say about them?
- *Computation*: If we know that there is a SVD for some matrix, how can we actually compute a SVD?
- *Geometric interpretation*: Is there a geometric motivation behind this decomposition?

Spoiler: **Every matrix has a SVD.** We start by discussing the geometric interpretation and then proceed to proving this existence result.

First of all, some explicit examples:

Examples of SVDs

(i)

$$\begin{pmatrix} 1 & 0 & -1 & 1 \\ -1 & 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 & 0 & 0 \\ 0 & \sqrt{3} & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & 0 \\ -\frac{1}{\sqrt{3}} & 0 & 0 & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{pmatrix}^{\mathrm{T}}.$$

Corresponding reduced SVD:

$$\begin{pmatrix} 1 & 0 & -1 & 1 \\ -1 & 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & \sqrt{3} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} \\ 0 & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix}^{\mathrm{T}}.$$

Examples of SVDs

(ii)

$$\begin{pmatrix} 1 & 1 \\ -1 & 1 \\ 1 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 & 0 \\ \frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}^{\mathrm{T}}.$$

Corresponding reduced SVD:

$$\begin{pmatrix} 1 & 1 \\ -1 & 1 \\ 1 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \\ \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & \sqrt{2} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}^{\mathrm{T}}.$$

(iii) $$\begin{pmatrix} 2 & 11 \\ 10 & -5 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 4\sqrt{10} & 0 \\ 0 & 3\sqrt{10} \end{pmatrix} \begin{pmatrix} -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}^{\mathrm{T}}.$$

Geometric interpretation of the SVD

The geometric interpretation of the SVD is that **the image of the 2-norm unit sphere $\{x \in \mathbb{R}^n | \, \|x\|_2 = 1\}$ under any $m \times n$ matrix is a hyperellipse**.

A hyperellipse in $\mathbb{R}^m$ is m-dimensional generalization of an ellipse: surface obtained by stretching the 2-norm unit sphere in $\mathbb{R}^m$ by factors $\sigma_1, \ldots, \sigma_m \geq 0$ in the directions of orthonormal vectors $u_1, \ldots, u_m \in \mathbb{R}^m$.

Indeed, observe that a SVD $A = U\Sigma V^{\mathrm{T}}$ of a matrix $A \in \mathbb{R}^{m\times n}$ yields

$$AV = U\Sigma \implies (a_1 | \cdots | a_n)(v_1 | \cdots | v_n) = (u_1 | \cdots | u_m)\mathrm{diag}_{m\times n}(\sigma_1, \ldots, \sigma_p)$$

(recall that $p := \min(m, n)$). Therefore,

$$\text{if } m \geq n : \quad Av_i = \sigma_i u_i \quad \forall 1 \leq i \leq n,$$

and

$$\text{if } m < n : \quad Av_i = \sigma_i u_i \quad \forall 1 \leq i \leq m, \quad Av_j = 0 \quad \forall m < j \leq n.$$

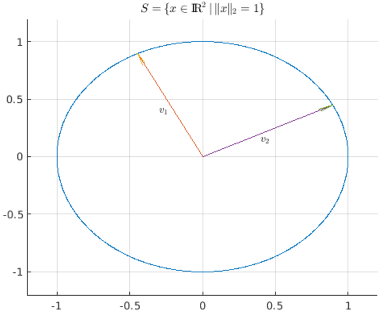
(Note if $m < n$ (wide): the j-th column of $U\Sigma$ is $0 \in \mathbb{R}^m$ for $j > m$)

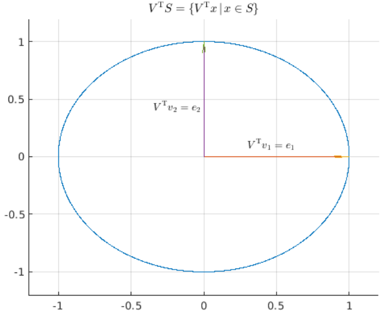
Example:

$$A := \begin{pmatrix} 2 & 11 \\ 10 & -5 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 4\sqrt{10} & 0 \\ 0 & 3\sqrt{10} \end{pmatrix} \begin{pmatrix} -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}^{\mathrm{T}} =: U\Sigma V^{\mathrm{T}}.$$

Have singular values $\sigma_1 = 4\sqrt{10}$, $\sigma_2 = 3\sqrt{10}$, left singular vectors $u_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$, $u_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$, right singular vectors $v_1 = \begin{pmatrix} -\frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix}$, $v_2 = \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}$.

Compute image of unit circle S under L_A: $L_A(S) = [L_U \circ L_\Sigma \circ L_{V^{\mathrm{T}}}](S)$.





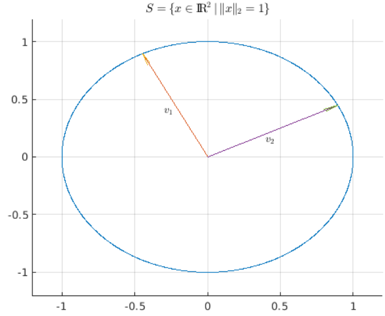
Step 1: $L_{V^{\mathrm{T}}}$ is reflection of plane across the line $y = \frac{1+\sqrt{5}}{2}x$.

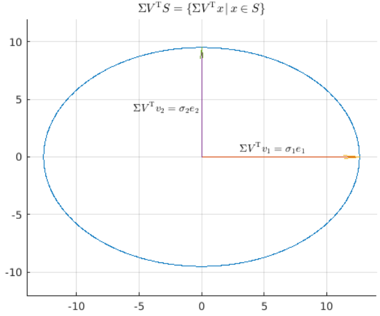
Example:

$$A := \begin{pmatrix} 2 & 11 \\ 10 & -5 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 4\sqrt{10} & 0 \\ 0 & 3\sqrt{10} \end{pmatrix} \begin{pmatrix} -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}^{\mathrm{T}} =: U\Sigma V^{\mathrm{T}}.$$

Have singular values $\sigma_1 = 4\sqrt{10}$, $\sigma_2 = 3\sqrt{10}$, left singular vectors $u_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$, $u_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$, right singular vectors $v_1 = \begin{pmatrix} -\frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix}$, $v_2 = \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}$.

Compute image of unit circle S under L_A: $L_A(S) = [L_U \circ L_\Sigma \circ L_{V^{\mathrm{T}}}](S)$.





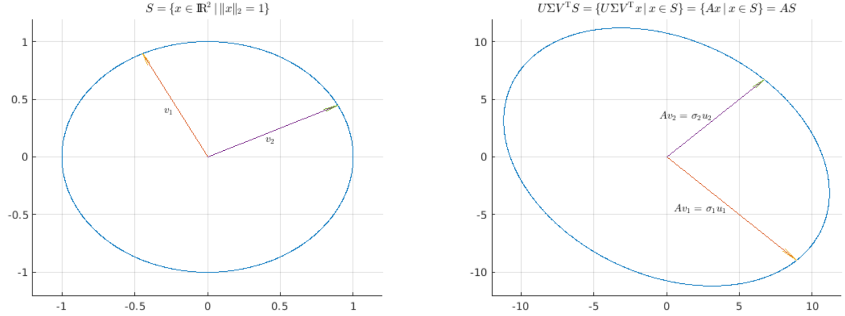
Step 2: L_Σ stretches scales x-coordinate by σ_1 and y-coordinate by σ_2.

Example:

$$A := \begin{pmatrix} 2 & 11 \\ 10 & -5 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 4\sqrt{10} & 0 \\ 0 & 3\sqrt{10} \end{pmatrix} \begin{pmatrix} -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}^{\mathrm{T}} =: U\Sigma V^{\mathrm{T}}.$$

Have singular values $\sigma_1 = 4\sqrt{10}$, $\sigma_2 = 3\sqrt{10}$, left singular vectors $u_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$, $u_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$, right singular vectors $v_1 = \begin{pmatrix} -\frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix}$, $v_2 = \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}$.

Compute image of unit circle S under L_A: $L_A(S) = [L_U \circ L_\Sigma \circ L_{V^{\mathrm{T}}}](S)$.



Step 3: L_U is rotation of plane clockwise by the angle $\frac{\pi}{4}$.

$\Longrightarrow L_A(S)$ is ellipse with principal semiaxes are $\sigma_1 u_1$ and $\sigma_2 u_2$.

For any $A \in \mathbb{R}^{2\times 2}$ with SVD $A = U\Sigma V^{\mathrm{T}}$:

$$L_A(S) = [L_U \circ L_\Sigma \circ L_{V^{\mathrm{T}}}](S)$$

1) $L_{V^{\mathrm{T}}}$ is either rotation or reflection of plane $\Longrightarrow L_{V^{\mathrm{T}}}(S) = S$.
2) L_Σ stretches x-coordinate by σ_1 and y-coordinate by $\sigma_2 \Longrightarrow$ $[L_\Sigma \circ L_{V^{\mathrm{T}}}](S) = L_\Sigma(S)$ is ellipse aligned with coordinate axes.
3) L_U is either rotation or reflection of plane $\Longrightarrow L_A(S) = L_U(L_\Sigma(S))$ is ellipse with principal semiaxes $\sigma_1 u_1$, $\sigma_2 u_2$.

2.2 Existence and uniqueness

Existence theorem for SVD

Theorem (Every matrix has a SVD)

Let $A \in \mathbb{R}^{m\times n}$ for some $m, n \in \mathbb{N}$. Then, there exists a SVD of A.

Proof: We need to show that $\exists$ orthogonal matrices $U \in \mathbb{R}^{m\times m}$, $V \in \mathbb{R}^{n\times n}$, and a matrix $\Sigma = \mathrm{diag}_{m\times n}(\sigma_1, \dots, \sigma_p)$ with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$ (recall $p = \min(m, n)$) such that

$$U^{\mathrm{T}} A V = \Sigma. \qquad (\Longrightarrow A = U\Sigma V^{\mathrm{T}} \text{ is SVD of } A.)$$

Note: $AV = U\Sigma$ requires $Av_i = \sigma_i u_i \; \forall\, 1 \leq i \leq p$ and thus, for $1 \leq i \leq p$:

$$\sigma_i = \|\sigma_i u_i\|_2 = \|Av_i\|_2 \leq \sup_{\substack{x\in\mathbb{R}^n \\ \|x\|_2=1}} \|Ax\|_2 = \|A\|_2.$$

Step 1. Construct σ_1, v_1, u_1.
Set $\sigma_1 := \|A\|_2 = \sup_{x\in S} \|Ax\|_2$, where $S := \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$.
S is compact subset of $\mathbb{R}^n$, the map $S \ni x \mapsto \|Ax\|_2 \in \mathbb{R}$ is continuous:

$\exists v_1 \in S : \|Av_1\|_2 = \sigma_1$, i.e., $Av_1 = \sigma_1 u_1$ for some unit vector $u_1 \in \mathbb{R}^m$.

Step 2. Construct orthogonal matrices $V_1 \in \mathbb{R}^{n\times n}$, $U_1 \in \mathbb{R}^{m\times m}$ such that

$$U_1^{\mathrm{T}} A V_1 = \left(\begin{array}{c|c} \sigma_1 & 0_{1\times(n-1)} \\ \hline 0_{(m-1)\times 1} & B \end{array}\right) \in \mathbb{R}^{m\times n} \quad \text{for some } B \in \mathbb{R}^{(m-1)\times(n-1)}.$$

Extend v_1 to orthonormal basis $\{v_1, \ldots, v_n\} \subseteq \mathbb{R}^n$ and u_1 to orthonormal basis $\{u_1, \ldots, u_m\} \subseteq \mathbb{R}^m$, and set

$$V_1 := (v_1|\cdots|v_n) \in \mathbb{R}^{n\times n}, \qquad U_1 := (u_1|\cdots|u_m) \in \mathbb{R}^{m\times m}.$$

Then, V_1 and U_1 are orthogonal matrices. By Step 1,

$$Av_1 = \sigma_1 u_1 \quad \Longrightarrow \quad Av_1 = U_1(\sigma_1 e_1) \quad \Longrightarrow \quad U_1^{\mathrm{T}} A v_1 = \sigma_1 e_1 \in \mathbb{R}^m.$$

Hence,

$$U_1^{\mathrm{T}} A V_1 = \left(\begin{array}{c|c} \sigma_1 & w^{\mathrm{T}} \\ \hline 0_{(m-1)\times 1} & B \end{array}\right) =: A_1 \in \mathbb{R}^{m\times n}$$

for some $w \in \mathbb{R}^{n-1}$ and $B \in \mathbb{R}^{(m-1)\times(n-1)}$.

We are done with Step 2 if we can show that $w = 0 \in \mathbb{R}^{n-1}$.

Recall from previous slide:

$$U_1^{\mathrm{T}} A V_1 = \left(\begin{array}{c|c} \sigma_1 & w^{\mathrm{T}} \\ \hline 0_{(m-1)\times 1} & B \end{array}\right) =: A_1 \in \mathbb{R}^{m\times n}$$

for some $w \in \mathbb{R}^{n-1}$ and $B \in \mathbb{R}^{(m-1)\times(n-1)}$. We will show $w = 0 \in \mathbb{R}^{n-1}$. Key is the observation

$$\|A_1\|_2 = \|U_1^{\mathrm{T}} A V_1\|_2 = \|A\|_2 = \sigma_1.$$

We claim that $\|A_1\|_2 \geq \sqrt{\sigma_1^2 + w^{\mathrm{T}} w}$. Once shown, this yields $w = 0$.

For $\tilde{w} := \begin{pmatrix} \sigma_1 \\ \hline w \end{pmatrix} \in \mathbb{R}^n$ we have

$$\|A_1 \tilde{w}\|_2^2 = \left\| \left(\begin{array}{c|c} \sigma_1 & w^{\mathrm{T}} \\ \hline 0_{(m-1)\times 1} & B \end{array}\right) \begin{pmatrix} \sigma_1 \\ \hline w \end{pmatrix} \right\|_2^2 = \left\| \begin{pmatrix} \sigma_1^2 + w^{\mathrm{T}} w \\ \hline Bw \end{pmatrix} \right\|_2^2$$
$$\geq (\sigma_1^2 + w^{\mathrm{T}} w)^2 = (\sigma_1^2 + w^{\mathrm{T}} w)\|\tilde{w}\|_2^2$$

$\Longrightarrow \|A_1\|_2 \geq \sqrt{\sigma_1^2 + w^{\mathrm{T}} w}$. Since $\|A_1\|_2 = \sigma_1$, we must have $w = 0$.

Step 2 is done: we have found orthogonal matrices $V_1 \in \mathbb{R}^{n\times n}$, $U_1 \in \mathbb{R}^{m\times m}$ such that

$$U_1^{\mathrm{T}} A V_1 = \left(\begin{array}{c|c} \sigma_1 & 0_{1\times(n-1)} \\ \hline 0_{(m-1)\times 1} & B \end{array}\right) \in \mathbb{R}^{m\times n} \quad \text{for some } B \in \mathbb{R}^{(m-1)\times(n-1)}.$$

Step 3. We conclude the proof using induction on the dimension of A.

Note

$$U_1^{\mathrm{T}} A V_1 = \left(\begin{array}{c} \sigma_1 \\ \hline O_{(m-1)\times 1} \end{array}\right) = \operatorname{diag}_{m\times 1}(\sigma_1) \quad \text{if } n = 1,$$
$$U_1^{\mathrm{T}} A V_1 = (\sigma_1 | \, 0_{1\times(n-1)}) = \operatorname{diag}_{1\times n}(\sigma_1) \quad \text{if } m = 1,$$

i.e., every $A \in \mathbb{R}^{m\times n}$ with $m = 1$ or $n = 1$ has a SVD. Now, assume $m, n \geq 2$ and that $B \in \mathbb{R}^{(m-1)\times(n-1)}$ has a SVD $B = U_2 \Sigma_2 V_2^{\mathrm{T}}$. Then,

$$U_1^{\mathrm{T}} A V_1 = \left(\begin{array}{c|c} \sigma_1 & 0_{1\times(n-1)} \\ \hline 0_{(m-1)\times 1} & U_2 \Sigma_2 V_2^{\mathrm{T}} \end{array}\right)$$
$$= \left(\begin{array}{c|c} 1 & 0_{1\times(m-1)} \\ \hline 0_{(m-1)\times 1} & U_2 \end{array}\right) \left(\begin{array}{c|c} \sigma_1 & 0_{1\times(n-1)} \\ \hline 0_{(m-1)\times 1} & \Sigma_2 \end{array}\right) \left(\begin{array}{c|c} 1 & 0_{1\times(n-1)} \\ \hline 0_{(n-1)\times 1} & V_2 \end{array}\right)^{\mathrm{T}}.$$

Bring first and third matrix from right-hand side to left-hand side:

$$U^{\mathrm{T}} A V = \left(\begin{array}{c|c} \sigma_1 & 0_{1\times(n-1)} \\ \hline 0_{(m-1)\times 1} & \Sigma_2 \end{array}\right) =: \Sigma \in \mathbb{R}^{m\times n},$$

where

$$U := U_1 \left(\begin{array}{c|c} 1 & 0_{1\times(m-1)} \\ \hline 0_{(m-1)\times 1} & U_2 \end{array}\right), \quad V := V_1 \left(\begin{array}{c|c} 1 & 0_{1\times(n-1)} \\ \hline 0_{(n-1)\times 1} & V_2 \end{array}\right).$$

Done:

- $\Sigma \in \mathbb{R}^{m\times n}$ is diagonal with non-increasing non-negative entries.
- $U \in \mathbb{R}^{m\times m}$, $V \in \mathbb{R}^{n\times n}$ are orthogonal matrices.

$\square$

Next question: Is the SVD unique?

Is the SVD unique?

No: let's consider the simplest possible case, $A = (a) \in \mathbb{R}^{1\times 1}$. Observe

$$Q = (q) \in \mathbb{R}^{1\times 1} \text{ is orthogonal} \quad \Longleftrightarrow \quad q \in \{-1, 1\}.$$

So, if we look for a SVD

$$A = (a) = U\Sigma V^{\mathrm{T}} = (u)(\sigma_1)(v)^{\mathrm{T}} \qquad [\, = (u\sigma_1 v)\,],$$

we need $u, v \in \{-1, 1\}$ and $\sigma_1 \geq 0$. Hence, must have

$$\sigma_1 = |a|.$$

We can list all SVDs of $A = (a)$:

- If $a > 0$: $(a) = (1)(\sigma_1)(1)^{\mathrm{T}}$ and $(a) = (-1)(\sigma_1)(-1)^{\mathrm{T}}$.
- If $a = 0$: $(a) = (1)(\sigma_1)(1)^{\mathrm{T}} = (-1)(\sigma_1)(-1)^{\mathrm{T}} = (-1)(\sigma_1)(1)^{\mathrm{T}} = (1)(\sigma_1)(-1)^{\mathrm{T}}$.
- If $a < 0$: $(a) = (-1)(\sigma_1)(1)^{\mathrm{T}}$ and $(a) = (1)(\sigma_1)(-1)^{\mathrm{T}}$.

$\Longrightarrow$ SVD not unique. However, in this example, the singular values are unique, and the left & right singular vectors are unique up to signs.

$\Longrightarrow$ There is hope: we might have uniqueness of singular values and/or uniqueness up to signs for left & right singular vectors.

A first uniqueness result: σ_1 is unique

Claim: For any matrix $A \in \mathbb{R}^{m\times n}$, its largest singular value σ_1 is uniquely determined and given by

$$\sigma_1 = \|A\|_2.$$

Proof: Suppose $A = U\Sigma V^{\mathrm{T}}$ is a SVD of A. Then, (recall $p := \min(m,n)$)

$$\|A\|_2 = \|U\Sigma V^{\mathrm{T}}\|_2 = \|\Sigma\|_2 = \|\mathrm{diag}_{m\times n}(\sigma_1,\ldots,\sigma_p)\|_2 = \max_{1\le i\le p}|\sigma_i| = \sigma_1.$$

$\square$

Uniqueness theorem for singular values and singular vectors

Theorem (Uniqueness result for SVD)

The singular values $\{\sigma_i\}$ of any given matrix $A \in \mathbb{R}^{m\times n}$ are unique and

$$\{\sigma_1^2, \ldots, \sigma_p^2\} = \begin{cases} \Lambda(A^{\mathrm{T}}A) & , \text{ if } m \geq n, \\ \Lambda(AA^{\mathrm{T}}) & , \text{ if } m < n. \end{cases} \tag{1}$$

If $A \in \mathbb{R}^{n\times n}$ is square and the singular values are positive and distinct, then the left singular vectors $\{u_i\}$ and right singular vectors $\{v_i\}$ are unique up to signs.

Proof: 1) We show (1). ($\Longrightarrow$ singular values are uniquely determined)

Let $A \in \mathbb{R}^{m\times n}$ and let $A = U\Sigma V^{\mathrm{T}}$ be a SVD of A. Then,

$A^{\mathrm{T}}A = V\Sigma^{\mathrm{T}}U^{\mathrm{T}}U\Sigma V^{\mathrm{T}} = V(\Sigma^{\mathrm{T}}\Sigma)V^{-1} \in \mathbb{R}^{n\times n} \implies A^{\mathrm{T}}A$ similar to $\Sigma^{\mathrm{T}}\Sigma$,

$AA^{\mathrm{T}} = U\Sigma V^{\mathrm{T}}V\Sigma^{\mathrm{T}}U^{\mathrm{T}} = U(\Sigma\Sigma^{\mathrm{T}})U^{-1} \in \mathbb{R}^{m\times m} \implies AA^{\mathrm{T}}$ similar to $\Sigma\Sigma^{\mathrm{T}}$.

$\Longrightarrow \Lambda(A^{\mathrm{T}}A) = \Lambda(\Sigma^{\mathrm{T}}\Sigma)$ and $\Lambda(AA^{\mathrm{T}}) = \Lambda(\Sigma\Sigma^{\mathrm{T}})$.

Let's compute the matrices $\Sigma^{\mathrm{T}}\Sigma$ and $\Sigma\Sigma^{\mathrm{T}}$:

- If $m \geq n$: Have $\Sigma = \mathrm{diag}_{m\times n}(\sigma_1, \ldots, \sigma_n) = \begin{pmatrix} \mathrm{diag}_{n\times n}(\sigma_1, \ldots, \sigma_n) \\ \hline 0_{(m-n)\times n} \end{pmatrix}$.

  Hence,

$$\begin{aligned} \Sigma^{\mathrm{T}}\Sigma &= \left(\mathrm{diag}_{n\times n}(\sigma_1, \ldots, \sigma_n) \,\middle|\, 0_{n\times(m-n)}\right) \begin{pmatrix} \mathrm{diag}_{n\times n}(\sigma_1, \ldots, \sigma_n) \\ \hline 0_{(m-n)\times n} \end{pmatrix} \\ &= \mathrm{diag}_{n\times n}(\sigma_1^2, \ldots, \sigma_n^2), \end{aligned}$$

  and

$$\begin{aligned} \Sigma\Sigma^{\mathrm{T}} &= \begin{pmatrix} \mathrm{diag}_{n\times n}(\sigma_1, \ldots, \sigma_n) \\ \hline 0_{(m-n)\times n} \end{pmatrix} \left(\mathrm{diag}_{n\times n}(\sigma_1, \ldots, \sigma_n) \,\middle|\, 0_{n\times(m-n)}\right) \\ &= \left(\begin{array}{c|c} \mathrm{diag}_{n\times n}(\sigma_1^2, \ldots, \sigma_n^2) & 0_{n\times(m-n)} \\ \hline 0_{(m-n)\times n} & 0_{(m-n)\times(m-n)} \end{array}\right) \\ &= \mathrm{diag}_{m\times m}(\sigma_1^2, \ldots, \sigma_n^2, 0, \ldots, 0). \end{aligned}$$

  $\Longrightarrow \{\sigma_1^2, \ldots, \sigma_n^2\} = \Lambda(\Sigma^{\mathrm{T}}\Sigma) = \Lambda(A^{\mathrm{T}}A)$.

- If $m < n$: Have
  $\Sigma = \mathrm{diag}_{m\times n}(\sigma_1, \ldots, \sigma_n) = \left(\mathrm{diag}_{m\times m}(\sigma_1, \ldots, \sigma_m) \,\middle|\, 0_{m\times(n-m)}\right)$.
  Hence,

$$
\begin{aligned}
\Sigma^{\mathrm{T}}\Sigma &= \begin{pmatrix} \mathrm{diag}_{m\times m}(\sigma_1, \ldots, \sigma_m) \\ \hline 0_{(n-m)\times m} \end{pmatrix} \left(\mathrm{diag}_{m\times m}(\sigma_1, \ldots, \sigma_m) \,\middle|\, 0_{m\times(n-m)}\right) \\
&= \left(\begin{array}{c|c} \mathrm{diag}_{m\times m}(\sigma_1^2, \ldots, \sigma_m^2) & 0_{m\times(n-m)} \\ \hline 0_{(n-m)\times m} & 0_{(n-m)\times(n-m)} \end{array}\right) \\
&= \mathrm{diag}_{n\times n}(\sigma_1^2, \ldots, \sigma_m^2, 0, \ldots, 0).
\end{aligned}
$$

  and

$$
\begin{aligned}
\Sigma\Sigma^{\mathrm{T}} &= \left(\mathrm{diag}_{m\times m}(\sigma_1, \ldots, \sigma_m) \,\middle|\, 0_{m\times(n-m)}\right) \begin{pmatrix} \mathrm{diag}_{m\times m}(\sigma_1, \ldots, \sigma_m) \\ \hline 0_{(n-m)\times m} \end{pmatrix} \\
&= \mathrm{diag}_{m\times m}(\sigma_1^2, \ldots, \sigma_m^2).
\end{aligned}
$$

$\Longrightarrow \{\sigma_1^2, \ldots, \sigma_m^2\} = \Lambda(\Sigma\Sigma^{\mathrm{T}}) = \Lambda(AA^{\mathrm{T}})$.

$\Longrightarrow$ Altogether, $\sigma_1^2, \ldots, \sigma_p^2$ uniquely determined. As singular values non-negative and non-increasing, find $\sigma_1, \ldots, \sigma_p$ are uniquely determined.

2) Remains to show: If $A \in \mathbb{R}^{n\times n}$ and singular values are positive and distinct, then the left and right singular vectors are unique up to signs.

(What do we mean by “unique up to signs”? Recall SVD is equivalent to

$$\begin{aligned} &\text{if } m \geq n: \quad Av_i = \sigma_i u_i \quad \forall 1 \leq i \leq n, \\ &\text{if } m < n: \quad Av_i = \sigma_i u_i \quad \forall 1 \leq i \leq m, \quad Av_j = 0 \quad \forall m < j \leq n. \end{aligned}$$

So, one can always find another SVD by replacing a chosen v_i by $-v_i$ when also replacing u_i by $-u_i$. We claim that this is the only way of obtaining another SVD.)

We do not give a rigorous proof, but argue geometrically:

If the lengths of the semiaxes of a hyperellipse (i.e., the singular values σ_i) are distinct, then the semiaxes (i.e., the vectors $\sigma_i u_i$) are determined uniquely up to signs from the geometry of the hyperellipse. (Note that if Σ and U is uniquely determined, then also V must be uniquely determined from $A = U\Sigma V^{\mathrm{T}}$ as U and Σ are invertible (recall assumption $\sigma_i > 0$).) □

SVD as a tool for transformation into diagonal form

A brief summary of what we have achieved:

- Every matrix has a SVD. In geometric terms: the image of the 2-norm unit sphere in $\mathbb{R}^n$ under any $m \times n$ matrix is a hyperellipse.
- The singular values are uniquely determined. For square matrices with positive distinct singular values, the left & right singular vectors are uniquely determined up to signs.

$\Longrightarrow$ Can transform any matrix into a diagonal matrix via change of bases:

For any $x \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$:

$$Ax = b \quad \Longleftrightarrow \quad U\Sigma V^{\mathrm{T}} x = b \quad \Longleftrightarrow \quad \Sigma V^{\mathrm{T}} x = U^{\mathrm{T}} b \quad \Longleftrightarrow \quad \Sigma x' = b',$$

where $x' = V^{\mathrm{T}} x$ ($\Leftrightarrow x = V x'$) coordinate vector for expansion of x in basis of right singular vectors and $b' = U^{\mathrm{T}} b$ ($\Leftrightarrow b = U b'$) coordinate vector for expansion of b in basis of left singular vectors.

Comparison with eigenvalue decomposition

If $A \in \mathbb{R}^{n\times n}$ is diagonalizable with eigenvalue decomposition

$$A = XDX^{-1}$$

for some invertible $X \in \mathbb{R}^{n\times n}$ and a diagonal matrix $D \in \mathbb{C}^{n\times n}$ containing the eigenvalues of A, then for any $x, b \in \mathbb{C}^n$:

$$Ax = b \iff XDX^{-1}x = b \iff DX^{-1}x = X^{-1}b \iff Dx' = b',$$

where $x' = X^{-1}x, b' = X^{-1}b$ coordinate vectors for expansions of x, b in the basis of columns of X (eigenvectors).

SVD vs. eigenvalue decomposition: ($A = U\Sigma V^{\mathrm{T}}$ vs. $A = XDX^{-1}$)

- SVD uses two orthonormal bases (left and right singular vectors). Eigenvalue decomposition uses only one – not necessarily orthogonal – basis (eigenvectors).
- **The huge advantage of the SVD: Every matrix has one.** An eigenvalue decomposition only exists for certain square matrices (geometric multiplicity = algebraic multiplicity for all eigenvalues).

2.3 Computation

Observation: can restrict to tall matrices

For any $A \in \mathbb{R}^{m \times n}$:

$$\begin{aligned} & A = U[\operatorname{diag}_{m\times n}(\sigma_1, \dots, \sigma_p)]V^{\mathrm{T}} \text{ is a SVD for } A \\ \iff \quad & A^{\mathrm{T}} = V[\operatorname{diag}_{n\times m}(\sigma_1, \dots, \sigma_p)]U^{\mathrm{T}} \text{ is a SVD for } A^{\mathrm{T}}. \end{aligned}$$

$\Longrightarrow$ We can restrict to the case $m \geq n$ when thinking about how to compute SVDs.

If we know how to compute a SVD of an arbitrary given "tall" matrix, we know how to compute a SVD of any given matrix.

Idea for computing a SVD

Let $A = (a_1|\cdots|a_n) \in \mathbb{R}^{m\times n}$ with $m \geq n$.

Idea: If $A = U\Sigma V^{\mathrm{T}}$ is a SVD of A, then

$$A^{\mathrm{T}}A = V\Sigma^{\mathrm{T}}U^{\mathrm{T}}U\Sigma V^{\mathrm{T}} = V\Sigma^{\mathrm{T}}\Sigma V^{\mathrm{T}}.$$

So, (firstly ignoring the question if the following is all possible)

- first, find decomposition $A^{\mathrm{T}}A = VDV^{\mathrm{T}}$ with $V \in \mathbb{R}^{n\times n}$ orthogonal and $D \in \mathbb{R}^{n\times n}$ diagonal.
- then, find $\Sigma \in \mathbb{R}^{m\times n}$ such that $\Sigma = \mathrm{diag}_{m\times n}(\sigma_1, \ldots, \sigma_n)$ with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$, and $\Sigma^{\mathrm{T}}\Sigma = D$.
- finally, find $U \in \mathbb{R}^{m\times m}$ orthogonal from $AV = U\Sigma$. Then, $A = U\Sigma V^{\mathrm{T}}$ is SVD.

Let's discuss this in detail:

Algorithm for computing SVD

Let $A = (a_1 | \cdots | a_n) \in \mathbb{R}^{m \times n}$ with $m \geq n$.

1) Compute the so-called **Gram matrix of $a_1, \ldots, a_n \in \mathbb{R}^m$ for the Euclidean inner product on $\mathbb{R}^m$**

$$A^{\mathrm{T}} A = \begin{pmatrix} \langle a_1, a_1 \rangle & \langle a_1, a_2 \rangle & \cdots & \langle a_1, a_n \rangle \\ \langle a_2, a_1 \rangle & \langle a_2, a_2 \rangle & \cdots & \langle a_2, a_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle a_n, a_1 \rangle & \langle a_n, a_2 \rangle & \cdots & \langle a_n, a_n \rangle \end{pmatrix} \in \mathbb{R}^{n \times n}.$$

Observe:

- $A^{\mathrm{T}} A$ is symmetric $\Longrightarrow$ Each eigenvalue of $A^{\mathrm{T}} A$ is real, and there are n orthonormal real eigenvectors ($\Rightarrow$ $A^{\mathrm{T}} A$ is orthogonally diagonalizable).
- All eigenvalues of $A^{\mathrm{T}} A$ are non-negative: If $\lambda \in \Lambda(A^{\mathrm{T}} A)$ and $x \in \mathbb{R}^n \backslash \{0\}$ is eigenvector corresponding to this eigenvalue, have

$$A^{\mathrm{T}} A x = \lambda x \implies x^{\mathrm{T}} A^{\mathrm{T}} A x = \lambda x^{\mathrm{T}} x \implies \|Ax\|_2^2 = \lambda \|x\|_2^2 \implies \lambda \geq 0.$$

2) Compute an eigenvalue decomposition

$$A^{\mathrm{T}}A = VDV^{\mathrm{T}}$$

with

$$V = (v_1|\cdots|v_n) \in \mathbb{R}^{n\times n} \text{ orthogonal},$$
$$D = \mathrm{diag}_{n\times n}(\lambda_1,\ldots,\lambda_n) \in \mathbb{R}^{n\times n},$$
$$\lambda_1,\ldots,\lambda_n \in \Lambda(A^{\mathrm{T}}A), \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0.$$

3) Set $\sigma_i := \sqrt{\lambda_i}$ for $i \in \{1,\ldots,n\}$, and set

$$\Sigma := \mathrm{diag}_{m\times n}(\sigma_1,\ldots,\sigma_n) \in \mathbb{R}^{m\times n}.$$

Then, $\Sigma^{\mathrm{T}}\Sigma = D$. Indeed:

$$\Sigma^{\mathrm{T}}\Sigma = \left(\mathrm{diag}_{n\times n}(\sigma_1,\ldots,\sigma_n)\,\middle|\,0_{n\times(m-n)}\right)\begin{pmatrix}\mathrm{diag}_{n\times n}(\sigma_1,\ldots,\sigma_n)\\ \hline 0_{(m-n)\times n}\end{pmatrix} =$$
$$\mathrm{diag}_{n\times n}(\sigma_1^2,\ldots,\sigma_n^2) = D.$$

4) Find an orthogonal matrix $U = (u_1 | \cdots | u_m) \in \mathbb{R}^{m \times m}$ such that

$$U\Sigma = AV, \quad \text{i.e.,} \quad \sigma_i u_i = Av_i \quad \forall 1 \leq i \leq n.$$

Then, we have that

$$A = U\Sigma V^{\mathrm{T}}$$

is a SVD of A.

An example will make things more clear:

Example for computing a SVD

We compute a SVD of

$$M := \begin{pmatrix} 1 & 0 & -1 & 1 \\ -1 & 1 & 0 & 1 \end{pmatrix} \in \mathbb{R}^{2\times 4}.$$

0) Set $A := M^{\mathrm{T}} \in \mathbb{R}^{4\times 2}$. Now start with our Algorithm applied to A.

1) Compute $A^{\mathrm{T}}A = MM^{\mathrm{T}} = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}$.

2) Eigenvalues of $A^{\mathrm{T}}A$: $\lambda_1 := 3$, $\lambda_2 := 3$.

   Corresponding normalized eigenvectors: $v_1 := \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $v_2 := \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

   Set $V := (v_1|v_2) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $D := \mathrm{diag}_{2\times 2}(\lambda_1, \lambda_2) = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}$.

   Then, V is orthogonal, D is diagonal with $\lambda_1 \geq \lambda_2 \geq 0$, and

$$A^{\mathrm{T}}A = VDV^{\mathrm{T}}.$$

3) Set $\Sigma := \operatorname{diag}_{4\times 2}(\sigma_1, \sigma_2) := \operatorname{diag}_{4\times 2}(\sqrt{\lambda_1}, \sqrt{\lambda_2}) = \begin{pmatrix} \sqrt{3} & 0 \\ 0 & \sqrt{3} \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$.

4) Find $U = (u_1|u_2|u_3|u_4) \in \mathbb{R}^{4\times 4}$ orthogonal with

$$Av_1 = \sigma_1 u_1, \qquad Av_2 = \sigma_2 u_2.$$

So, must have

$$u_1 = \frac{Av_1}{\sigma_1} = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 0 \\ -1 \\ 1 \end{pmatrix}, \qquad u_2 = \frac{Av_2}{\sigma_2} = \frac{1}{\sqrt{3}} \begin{pmatrix} -1 \\ 1 \\ 0 \\ 1 \end{pmatrix}.$$

Extend to an orthonormal basis u_1, u_2, u_3, u_4 of $\mathbb{R}^4$. Can take

$$U := \begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & 0 \\ -\frac{1}{\sqrt{3}} & 0 & 0 & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{pmatrix}.$$

$\Longrightarrow$ We have obtained the following SVD of $A = M^{\mathrm{T}} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \\ -1 & 0 \\ 1 & 1 \end{pmatrix}$:

$$A = \begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & 0 \\ -\frac{1}{\sqrt{3}} & 0 & 0 & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & \sqrt{3} \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{\mathrm{T}}.$$

Now, transpose the equation to obtain a SVD of $M = \begin{pmatrix} 1 & 0 & -1 & 1 \\ -1 & 1 & 0 & 1 \end{pmatrix}$:

$$M = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 & 0 & 0 \\ 0 & \sqrt{3} & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & 0 \\ -\frac{1}{\sqrt{3}} & 0 & 0 & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{pmatrix}^{\mathrm{T}}.$$

Further examples/exercises will be on problem sheets.

Alternative computation of SVDs for square matrices $A \in \mathbb{R}^{n\times n}$:

based on an eigenvalue decomposition of the symmetric matrix

$$\left(\begin{array}{c|c} 0_{n\times n} & A^{\mathrm{T}} \\ \hline A & 0_{n\times n} \end{array}\right) \in \mathbb{R}^{2n\times 2n}.$$

Question to SVD example: Given an orthonormal set $\{u_1, u_2\} \subseteq \mathbb{R}^4$, how to find $u_3, u_4 \in \mathbb{R}^4$ such that $U := (u_1|u_2|u_3|u_4) \in \mathbb{R}^{4\times 4}$ is orthogonal?

Observe that for $x \in \mathbb{R}^4$:

$$\{x\} \perp \{u_1, u_2\} \iff (u_1|u_2)^{\mathrm{T}} x = \begin{pmatrix} u_1^{\mathrm{T}} x \\ u_2^{\mathrm{T}} x \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \iff x \in \mathcal{N}((u_1|u_2)^{\mathrm{T}}).$$

So, need to find an orthonormal basis $\{u_3, u_4\}$ of $\mathcal{N}((u_1|u_2)^{\mathrm{T}})$.

$\Longrightarrow$ Compute a basis $\{w_3, w_4\}$ of $\mathcal{N}((u_1|u_2)^{\mathrm{T}})$ and transform this into an orthonormal basis $\{u_3, u_4\}$ of $\mathcal{N}((u_1|u_2)^{\mathrm{T}})$ using the Gram–Schmidt algorithm (Chapter 3).

A way without using things we didn't cover yet: Let's consider our explicit example, i.e., $u_1 := \frac{1}{\sqrt{3}}(1,0,-1,1)^{\mathrm{T}}$, $u_2 := \frac{1}{\sqrt{3}}(-1,1,0,1)^{\mathrm{T}}$.
Compute

$$\mathcal{N}((u_1|u_2)^{\mathrm{T}}) = \mathcal{N}(\frac{1}{\sqrt{3}}\begin{pmatrix} 1 & 0 & -1 & 1 \\ -1 & 1 & 0 & 1 \end{pmatrix}) = \mathcal{N}(\begin{pmatrix} 1 & 0 & -1 & 1 \\ -1 & 1 & 0 & 1 \end{pmatrix})$$

$$= \mathcal{N}(\begin{pmatrix} 1 & 0 & -1 & 1 \\ 0 & 1 & -1 & 2 \end{pmatrix}) = \mathrm{span}(\begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ -2 \\ 0 \\ 1 \end{pmatrix})$$

Take u_3 to be a unit vector in $\mathcal{N}((u_1|u_2)^{\mathrm{T}})$, say $u_3 := \frac{1}{\sqrt{6}}(-1,-2,0,1)^{\mathrm{T}}$.

Now we have a set of three orthonormal vectors $\{u_1, u_2, u_3\}$. Remains to find an orthonormal basis $\{u_4\}$ of $\mathcal{N}((u_1|u_2|u_3)^{\mathrm{T}})$. Then, $\{u_1, u_2, u_3, u_4\}$ is an orthonormal basis of $\mathbb{R}^4$.

Compute

$$\mathcal{N}((u_1|u_2|u_3)^{\mathrm{T}}) = \mathcal{N}(\begin{pmatrix} \frac{1}{\sqrt{3}} & 0 & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{6}} \end{pmatrix}) = \mathcal{N}(\begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -2 \end{pmatrix})$$

$$= \mathrm{span}(\begin{pmatrix} 1 \\ 0 \\ 2 \\ 1 \end{pmatrix}).$$

Take u_4 to be a unit vector in $\mathcal{N}((u_1|u_2|u_3)^{\mathrm{T}})$, say $u_4 := \frac{1}{\sqrt{6}}(1,0,2,1)^{\mathrm{T}}$.

Done: $U := (u_1|u_2|u_3|u_4) = \begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & 0 \\ -\frac{1}{\sqrt{3}} & 0 & 0 & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{pmatrix}$ is orthogonal.

2.4 Matrix properties

Connection of the SVD to matrix properties

Theorem (Matrix properties via SVD)

Let $A \in \mathbb{R}^{m \times n}$*, set* $p := \min(m, n)$*, and let*

$$A = U\Sigma V^{\mathrm{T}} = (u_1|\cdots|u_m)[\mathrm{diag}_{m\times n}(\sigma_1, \ldots, \sigma_p)](v_1|\cdots|v_n)^{\mathrm{T}}$$

be a SVD for A*. Let* $0 \le r \le p$ *denote the number of non-zero singular values of* A*, so that* $\sigma_1, \ldots, \sigma_r > 0$ *and* $\sigma_i = 0 \ \forall r < i \le p$*. Then,*

(i) $\mathrm{rk}(A) = r$.

(ii) $\mathscr{R}(A) = \mathrm{span}(u_1, \ldots, u_r)$ *and* $\mathcal{N}(A) = \mathrm{span}(v_{r+1}, \ldots, v_n)$.

(iii) $\|A\|_2 = \sigma_1$ *and* $\|A\|_F = \sqrt{\sum_{i=1}^r \sigma_i^2}$.

(iv) $\{\sigma_1, \ldots, \sigma_r\} = \{\sqrt{\lambda} \,|\, \lambda \in \Lambda(A^{\mathrm{T}}A)\}\backslash\{0\} = \{\sqrt{\lambda} \,|\, \lambda \in \Lambda(AA^{\mathrm{T}})\}\backslash\{0\}$.

(v) *If* $m = n$*, then* $|\det(A)| = \prod_{i=1}^n \sigma_i$.

(vi) *If* $m = n$ *and* $A = A^{\mathrm{T}}$*, then* $\{\sigma_1, \ldots, \sigma_n\} = \{|\lambda| \,|\, \lambda \in \Lambda(A)\}$.

(Define $\{x_i, \ldots, x_j\} := \emptyset$ for $i, j \in \mathbb{N}_0$ with $i > j$. Define $\mathrm{span}(\emptyset) := \{0\}$.)

This theorem lays the foundation for many practical algorithms.

In particular, from a computational point of view,

- the standard way to compute the rank of a matrix is via (i) ($\mathrm{rk}(A) = r$, in practice: count the number of singular values greater than some very small tolerance),
- the most accurate method for computing orthonormal bases of the range and the nullspace of a matrix is via (ii) ($\mathscr{R}(A) = \mathrm{span}(u_1, \ldots, u_r)$ and $\mathscr{N}(A) = \mathrm{span}(v_{r+1}, \ldots, v_n)$),
- the standard way to compute the spectral norm of a matrix A is via (iii) ($\|A\|_2 = \sigma_1$).

Now, let's prove the theorem:

Proof of (i)

Recall set-up: $A \in \mathbb{R}^{m\times n}$, $p := \min(m,n)$,

$$A = U\Sigma V^{\mathrm{T}} = (u_1|\cdots|u_m)[\mathrm{diag}_{m\times n}(\sigma_1,\ldots,\sigma_p)](v_1|\cdots|v_n)^{\mathrm{T}}$$

a SVD for A. Let $0 \le r \le p$ denote the number of non-zero singular values of A, so that $\sigma_1,\ldots,\sigma_r > 0$ and $\sigma_{r+1},\ldots,\sigma_p = 0$.

Claim: $\mathrm{rk}(A) = r$.

Proof: First, two observations:

- For any invertible matrices $M_m \in \mathbb{R}^{m\times m}$ and $M_n \in \mathbb{R}^{n\times n}$:

$$\mathrm{rk}(M_m A) = \mathrm{rk}(A) = \mathrm{rk}(AM_n).$$

- $\mathrm{rk}(\Sigma) = r$.

Hence,

$$\mathrm{rk}(A) = \mathrm{rk}(U\Sigma V^{\mathrm{T}}) = \mathrm{rk}(\Sigma) = r.$$

□

Proof of (ii)

Recall set-up: $A \in \mathbb{R}^{m\times n}$, $p := \min(m,n)$,

$$A = U\Sigma V^{\mathrm{T}} = (u_1|\cdots|u_m)[\mathrm{diag}_{m\times n}(\sigma_1,\ldots,\sigma_p)](v_1|\cdots|v_n)^{\mathrm{T}}$$

a SVD for A. Let $0 \le r \le p$ denote the number of non-zero singular values of A, so that $\sigma_1,\ldots,\sigma_r > 0$ and $\sigma_{r+1},\ldots,\sigma_p = 0$.

Claim: $\mathcal{R}(A) = \mathrm{span}(u_1,\ldots,u_r)$, $\quad \mathcal{N}(A) = \mathrm{span}(v_{r+1},\ldots,v_n)$.

Proof: Note that $\mathcal{R}(A) = \mathcal{R}(U\Sigma V^{\mathrm{T}}) =$

$$\{U\Sigma V^{\mathrm{T}}x \,|\, x \in \mathbb{R}^n\} = \{U\Sigma y \,|\, y \in \mathbb{R}^n\} = \{Uz \,|\, z \in \mathcal{R}(\Sigma)\},$$

and $\mathcal{N}(A) = \mathcal{N}(U\Sigma V^{\mathrm{T}}) = \{x \in \mathbb{R}^n \,|\, U\Sigma V^{\mathrm{T}}x = 0\} =$

$$\{x \in \mathbb{R}^n \,|\, \Sigma V^{\mathrm{T}}x = 0\} = \{x \in \mathbb{R}^n \,|\, V^{\mathrm{T}}x \in \mathcal{N}(\Sigma)\} = \{Vy \,|\, y \in \mathcal{N}(\Sigma)\}.$$

Observing

$$\mathcal{R}(\Sigma) = \mathrm{span}(e_1,\ldots,e_r) \subseteq \mathbb{R}^m, \qquad \mathcal{N}(\Sigma) = \mathrm{span}(e_{r+1},\ldots,e_n) \subseteq \mathbb{R}^n,$$

it follows $\mathcal{R}(A) = \mathrm{span}(u_1,\ldots,u_r)$ and $\mathcal{N}(A) = \mathrm{span}(v_{r+1},\ldots,v_n)$. $\square$

Proof of (iii)

Recall set-up: $A \in \mathbb{R}^{m \times n}$, $p := \min(m, n)$,

$$A = U\Sigma V^{\mathrm{T}} = (u_1|\cdots|u_m)[\mathrm{diag}_{m\times n}(\sigma_1, \ldots, \sigma_p)](v_1|\cdots|v_n)^{\mathrm{T}}$$

a SVD for A. Let $0 \leq r \leq p$ denote the number of non-zero singular values of A, so that $\sigma_1, \ldots, \sigma_r > 0$ and $\sigma_{r+1}, \ldots, \sigma_p = 0$.

Claim: $\|A\|_2 = \sigma_1$ and $\|A\|_F = \sqrt{\sum_{i=1}^{r} \sigma_i^2}$.

Proof: Have already shown that $\|A\|_2 = \sigma_1$.

For second equality, recall that the Frobenius norm is invariant under multiplication by orthogonal matrices. Hence,

$$\|A\|_F = \|U\Sigma V^{\mathrm{T}}\|_F = \|\Sigma\|_F = \sqrt{\sum_{i=1}^{r} \sigma_i^2}.$$

□

Proof of (iv)

Recall set-up: $A \in \mathbb{R}^{m \times n}$, $p := \min(m, n)$,

$$A = U\Sigma V^{\mathrm{T}} = (u_1|\cdots|u_m)[\mathrm{diag}_{m\times n}(\sigma_1, \ldots, \sigma_p)](v_1|\cdots|v_n)^{\mathrm{T}}$$

a SVD for A. Let $0 \le r \le p$ denote the number of non-zero singular values of A, so that $\sigma_1, \ldots, \sigma_r > 0$ and $\sigma_{r+1}, \ldots, \sigma_p = 0$.

Claim:
$\{\sigma_1, \ldots, \sigma_r\} = \{\sqrt{\lambda} \,|\, \lambda \in \Lambda(A^{\mathrm{T}}A)\}\backslash\{0\} = \{\sqrt{\lambda} \,|\, \lambda \in \Lambda(AA^{\mathrm{T}})\}\backslash\{0\}$.

Proof: Recall from proof of uniqueness of singular values that

$$\Lambda(A^{\mathrm{T}}A) = \Lambda(\Sigma^{\mathrm{T}}\Sigma), \quad \Lambda(AA^{\mathrm{T}}) = \Lambda(\Sigma\Sigma^{\mathrm{T}})$$

and

$$\Sigma^{\mathrm{T}}\Sigma = \mathrm{diag}_{n\times n}(\sigma_1^2, \ldots, \sigma_n^2), \quad \Sigma\Sigma^{\mathrm{T}} = \mathrm{diag}_{m\times m}(\sigma_1^2, \ldots, \sigma_n^2, 0, \ldots, 0) \quad \text{if } m \ge n,$$
$$\Sigma^{\mathrm{T}}\Sigma = \mathrm{diag}_{n\times n}(\sigma_1^2, \ldots, \sigma_m^2, 0, \ldots, 0), \quad \Sigma\Sigma^{\mathrm{T}} = \mathrm{diag}_{m\times m}(\sigma_1^2, \ldots, \sigma_m^2) \quad \text{if } m < n.$$

Done. (note $\{\sigma_1, \ldots, \sigma_r\} = \{\sigma_1, \ldots, \sigma_p\}\backslash\{0\}$). □

Proof of (v)

Recall set-up: $A \in \mathbb{R}^{m\times n}$, $p := \min(m,n)$,

$$A = U\Sigma V^{\mathrm{T}} = (u_1|\cdots|u_m)[\mathrm{diag}_{m\times n}(\sigma_1,\ldots,\sigma_p)](v_1|\cdots|v_n)^{\mathrm{T}}$$

a SVD for A. Let $0 \le r \le p$ denote the number of non-zero singular values of A, so that $\sigma_1,\ldots,\sigma_r > 0$ and $\sigma_{r+1},\ldots,\sigma_p = 0$.

Claim: If $m = n$, then $|\det(A)| = \prod_{i=1}^{n} \sigma_i$.

Proof: Compute

$$\det(A) = \det(U\Sigma V^{\mathrm{T}}) = \det(U)\det(\Sigma)\det(V^{\mathrm{T}}) = \det(U)\det(\Sigma)\det(V).$$

Recall $|\det(Q)| = 1$ for any orthogonal matrix $Q \in \mathbb{R}^{n\times n}$. Thus,

$$|\det(A)| = |\det(\Sigma)| = \prod_{i=1}^{n} \sigma_i.$$

□

Proof of (vi)

Recall set-up: $A \in \mathbb{R}^{m \times n}$, $p := \min(m, n)$,

$$A = U\Sigma V^{\mathrm{T}} = (u_1|\cdots|u_m)[\mathrm{diag}_{m\times n}(\sigma_1, \ldots, \sigma_p)](v_1|\cdots|v_n)^{\mathrm{T}}$$

a SVD for A. Let $0 \leq r \leq p$ denote the number of non-zero singular values of A, so that $\sigma_1, \ldots, \sigma_r > 0$ and $\sigma_{r+1}, \ldots, \sigma_p = 0$.

Claim: If $m = n$ and $A = A^{\mathrm{T}}$, then $\{\sigma_1, \ldots, \sigma_n\} = \{|\lambda| \,|\, \lambda \in \Lambda(A)\}$.

Proof: Since A is symmetric, $\exists$ orthogonal matrix $Q \in \mathbb{R}^{n\times n}$, a diagonal matrix $D = \mathrm{diag}_{n\times n}(\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^{n\times n}$ with $\{\lambda_1, \ldots, \lambda_n\} = \Lambda(A)$ s.t.

$$A = QDQ^{\mathrm{T}}.$$

Assume that entries of D are ordered s.t. $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$. Define

$$\tilde{\Sigma} := \mathrm{diag}_{n\times n}(|\lambda_1|, \ldots, |\lambda_n|), \quad S := \mathrm{diag}_{n\times n}(\mathrm{sign}(\lambda_1), \ldots, \mathrm{sign}(\lambda_n)),$$

and note $D = \tilde{\Sigma} S^{\mathrm{T}}$. Setting $\tilde{U} := Q$ and $\tilde{V} := QS$,

$$A = Q\tilde{\Sigma}S^{\mathrm{T}}Q^{\mathrm{T}} = \tilde{U}\tilde{\Sigma}\tilde{V}^{\mathrm{T}}.$$

This is a SVD of A! Singular values unique $\Longrightarrow$ $\{|\lambda_i|\}$ singular values of A.

Short alternative proof of (vi)

Recall singular values $\{\sigma_i\}$ of any $A \in \mathbb{R}^{m\times n}$ are unique and

$$\{\sigma_1^2, \ldots, \sigma_p^2\} = \begin{cases} \Lambda(A^{\mathrm{T}}A) & \text{, if } m \geq n, \\ \Lambda(AA^{\mathrm{T}}) & \text{, if } m < n. \end{cases}$$

Claim: If $A \in \mathbb{R}^{n\times n}$ is symmetric, then $\{\sigma_1, \ldots, \sigma_n\} = \{|\lambda| \mid \lambda \in \Lambda(A)\}$.

Proof:

$$\{\sigma_1^2, \ldots, \sigma_n^2\} = \Lambda(A^{\mathrm{T}}A) = \Lambda(A^2) = \{\lambda^2 \mid \lambda \in \Lambda(A)\}.$$

□

Proof of $\Lambda(A^2) = \{\lambda^2 \mid \lambda \in \Lambda(A)\}$ for A symmetric:

$A = QDQ^{\mathrm{T}}$ for some $Q \in \mathbb{R}^{n\times n}$ orthogonal, $D = \mathrm{diag}_{n\times n}(\lambda_1, \ldots, \lambda_n)$ with $\{\lambda_1, \ldots, \lambda_n\} = \Lambda(A) \subseteq \mathbb{R}$.

$\implies A^2 = QDQ^{\mathrm{T}}QDQ^{\mathrm{T}} = QD^2Q^{\mathrm{T}} \implies \Lambda(A^2) = \Lambda(D^2) = \{\lambda_1^2, \ldots, \lambda_n^2\}$.

2.5 Low-rank approximation

The problem of low-rank approximation

Given: some matrix $A \in \mathbb{R}^{m \times n} \backslash \{0\}$ and some $\nu \in \mathbb{N}_0$ with $0 \leq \nu < \mathrm{rk}(A)$.

Goal: find the best approximation to A from $\{B \in \mathbb{R}^{m \times n} \,|\, \mathrm{rk}(B) \leq \nu\}$.
$\Longrightarrow$ question: what do we mean by "best" approximation?

Low-rank approximation problem:

$$\begin{cases} \text{minimize} & \|A - B\|, \\ \text{subject to} & B \in \mathbb{R}^{m \times n},\ \mathrm{rk}(B) \leq \nu, \end{cases}$$

for some given matrix norm $\|\cdot\| : \mathbb{R}^{m \times n} \to [0, \infty)$.

Practical application: image compression.

This section: solve this problem for spectral and Frobenius norms.

Key tool: SVD.

A crucial observation

Using SVD, **any matrix A can be written as the sum of $r := \mathrm{rk}(A)$ rank-one matrices**.

Let $A \in \mathbb{R}^{m \times n}$, set $p := \min(m,n)$, and let

$$A = U\Sigma V^{\mathrm{T}} = (u_1|\cdots|u_m)[\mathrm{diag}_{m\times n}(\sigma_1,\ldots,\sigma_p)](v_1|\cdots|v_n)^{\mathrm{T}}$$

be a SVD of A. Setting $r := \mathrm{rk}(A)$, we have

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^{\mathrm{T}}.$$

This follows from the fact that we can write Σ as the sum of the r matrices $\mathrm{diag}_{m\times n}(\sigma_1, 0, \ldots, 0),\ \mathrm{diag}_{m\times n}(0, \sigma_2, 0, \ldots, 0),\ \ldots,\ \mathrm{diag}_{m\times n}(0, \ldots, 0, \sigma_r)$.

Exercise: find other, more simple, ways to express $A \in \mathbb{R}^{m\times n}$ as a sum of rank-one matrices.

So, why is this SVD-based rank-one decomposition more useful?

$\Longrightarrow$ its ν-th partial sum captures largest possible amount of "energy" of A.

Solution of the low-rank approximation problem

Theorem (Eckart–Young–Mirsky theorem)

Let $A \in \mathbb{R}^{m\times n}\backslash\{0\}$, *set* $p := \min(m,n)$, *and let*

$$A = U\Sigma V^{\mathrm{T}} = (u_1|\cdots|u_m)[\mathrm{diag}_{m\times n}(\sigma_1,\ldots,\sigma_p)](v_1|\cdots|v_n)^{\mathrm{T}}$$

be a SVD for A. *Further, let* $\nu \in \mathbb{N}_0$ *with* $0 \leq \nu < \mathrm{rk}(A)$, *and set*

$$A_\nu = \sum_{i=1}^{\nu} \sigma_i u_i v_i^{\mathrm{T}}.$$

Then,

$$\inf_{\substack{B\in\mathbb{R}^{m\times n}\\ \mathrm{rk}(B)\leq\nu}} \|A-B\|_2 = \|A-A_\nu\|_2 = \sigma_{\nu+1},$$

$$\inf_{\substack{B\in\mathbb{R}^{m\times n}\\ \mathrm{rk}(B)\leq\nu}} \|A-B\|_F = \|A-A_\nu\|_F = \sqrt{\sum_{i=\nu+1}^{r} \sigma_i^2}.$$

Proof of Eckart–Young–Mirsky theorem for spectral norm

Let $A \in \mathbb{R}^{m\times n}\backslash\{0\}$, set $p := \min(m,n)$, $r := \mathrm{rk}(A)$, and let

$$A = U\Sigma V^{\mathrm{T}} = (u_1|\cdots|u_m)[\mathrm{diag}_{m\times n}(\sigma_1,\ldots,\sigma_p)](v_1|\cdots|v_n)^{\mathrm{T}}$$

be a SVD for A. Further, let $\nu \in \mathbb{N}_0$ with $0 \le \nu < \mathrm{rk}(A)$, and set $A_\nu = \sum_{i=1}^{\nu} \sigma_i u_i v_i^{\mathrm{T}}$.

Claim: $\inf_{\substack{B\in\mathbb{R}^{m\times n}\\ \mathrm{rk}(B)\le\nu}} \|A-B\|_2 = \|A-A_\nu\|_2 = \sigma_{\nu+1}$.

Step 1: Show $\|A-A_\nu\|_2 = \sigma_{\nu+1}$.

Compute

$$\|A-A_\nu\|_2 = \left\|\sum_{i=1}^{r}\sigma_i u_i v_i^{\mathrm{T}} - \sum_{i=1}^{\nu}\sigma_i u_i v_i^{\mathrm{T}}\right\|_2 = \left\|\sum_{i=\nu+1}^{r}\sigma_i u_i v_i^{\mathrm{T}}\right\|_2 = \sigma_{\nu+1},$$

where we have used in the last step that the the largest singular value of the matrix $\sum_{i=\nu+1}^{r}\sigma_i u_i v_i^{\mathrm{T}}$ is given by $\sigma_{\nu+1}$.

As $\mathrm{rk}(A_\nu) \le \nu$:

$$\inf_{\substack{B\in\mathbb{R}^{m\times n}\\ \mathrm{rk}(B)\le\nu}} \|A-B\|_2 \le \|A-A_\nu\|_2 = \sigma_{\nu+1}.$$

Proof of Eckart–Young–Mirsky theorem for spectral norm

Let $A \in \mathbb{R}^{m\times n}\backslash\{0\}$, set $p := \min(m,n)$, $r := \mathrm{rk}(A)$, and let

$$A = U\Sigma V^{\mathrm{T}} = (u_1|\cdots|u_m)[\mathrm{diag}_{m\times n}(\sigma_1,\ldots,\sigma_p)](v_1|\cdots|v_n)^{\mathrm{T}}$$

be a SVD for A. Further, let $\nu \in \mathbb{N}_0$ with $0 \le \nu < \mathrm{rk}(A)$, and set $A_\nu = \sum_{i=1}^{\nu} \sigma_i u_i v_i^{\mathrm{T}}$.

Claim: $\inf_{\substack{B\in\mathbb{R}^{m\times n}\\ \mathrm{rk}(B)\le\nu}} \|A-B\|_2 = \|A-A_\nu\|_2 = \sigma_{\nu+1}$.

Step 2: Remains to prove $\inf_{\substack{B\in\mathbb{R}^{m\times n}\\ \mathrm{rk}(B)\le\nu}} \|A-B\|_2 \ge \sigma_{\nu+1}$.

Suppose $\exists\ B \in \mathbb{R}^{m\times n}$ with $\mathrm{rk}(B) \le \nu$ and $\|A-B\|_2 < \sigma_{\nu+1}$. Then,

- $\dim(\mathcal{N}(B)) = n - \mathrm{rk}(B) \ge n-\nu$.
- $\forall x \in \mathcal{N}(B)\backslash\{0\}$: $\|Ax\|_2 = \|(A-B)x\|_2 \le \|A-B\|_2\|x\|_2 < \sigma_{\nu+1}\|x\|_2$.
- $\forall v = \sum_{i=1}^{\nu+1} \alpha_i v_i \in \mathrm{span}(v_1,\ldots,v_{\nu+1})$:

(use $Av_i = \sigma_i u_i$ for all $1 \le i \le \nu+1$ (note $\nu+1 \le r \le p$), and Pythagorean theorem: $\forall a,b \in \mathbb{R}^n : a \perp b \Rightarrow \|a+b\|_2^2 = \|a\|_2^2 + \|b\|_2^2$)

$$\|Av\|_2^2 = \left\|\sum_{i=1}^{\nu+1} \alpha_i\sigma_i u_i\right\|_2^2 = \sum_{i=1}^{\nu+1} \alpha_i^2\sigma_i^2 \ge \sigma_{\nu+1}^2 \sum_{i=1}^{\nu+1} \alpha_i^2 = \sigma_{\nu+1}^2 \left\|\sum_{i=1}^{\nu+1} \alpha_i v_i\right\|_2^2 = \sigma_{\nu+1}^2\|v\|_2^2.$$

Proof of Eckart–Young–Mirsky theorem for spectral norm

Let $A \in \mathbb{R}^{m\times n}\setminus\{0\}$, set $p := \min(m,n)$, $r := \mathrm{rk}(A)$, and let

$$A = U\Sigma V^{\mathrm{T}} = (u_1|\cdots|u_m)[\mathrm{diag}_{m\times n}(\sigma_1,\ldots,\sigma_p)](v_1|\cdots|v_n)^{\mathrm{T}}$$

be a SVD for A. Further, let $\nu \in \mathbb{N}_0$ with $0 \le \nu < \mathrm{rk}(A)$, and set $A_\nu = \sum_{i=1}^{\nu} \sigma_i u_i v_i^{\mathrm{T}}$.

Claim: $\inf_{\substack{B\in\mathbb{R}^{m\times n}\\ \mathrm{rk}(B)\le\nu}} \|A-B\|_2 = \|A-A_\nu\|_2 = \sigma_{\nu+1}$.

Write $N := \mathcal{N}(B)$, $S := \mathrm{span}(v_1,\ldots,v_{\nu+1})$. Have shown:

- $\dim(N) \ge n-\nu$.
- $\forall x \in N\setminus\{0\}$: $\|Ax\|_2 < \sigma_{\nu+1}\|x\|_2$.
- $\forall x \in S$: $\|Ax\|_2 \ge \sigma_{\nu+1}\|x\|_2$.

We have

$$\dim(N\cap S) \ge \underbrace{\dim(N)}_{\ge n-\nu} + \underbrace{\dim(S)}_{\ge \nu+1} - \underbrace{\dim(N+S)}_{\le n} \ge 1.$$

$\Longrightarrow$ $\exists$ non-zero vector which is contained in $N\cap S$. Contradiction. $\square$

Application: image compression

Application: image compression

low-rank approximation: $\nu = 50$

Application: image compression

low-rank approximation: $\nu = 100$

$\Longrightarrow$ Try yourself in MATLAB. To transform an image into a matrix, do
$\text{image} = \text{imread}([\text{'filename.jpg'}]);\ \text{A} = \text{im2double}(\text{rgb2gray}(\text{image}));$

End of “Chapter 2: Singular Value Decomposition”.