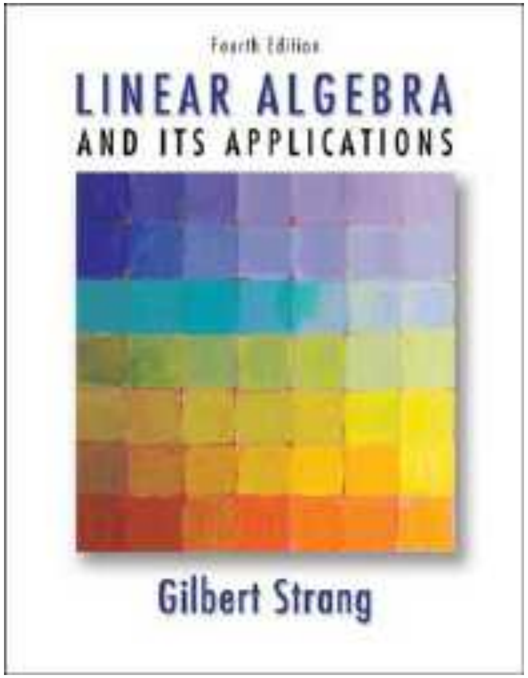
Linear Algebra



Instructor: Jing YAO

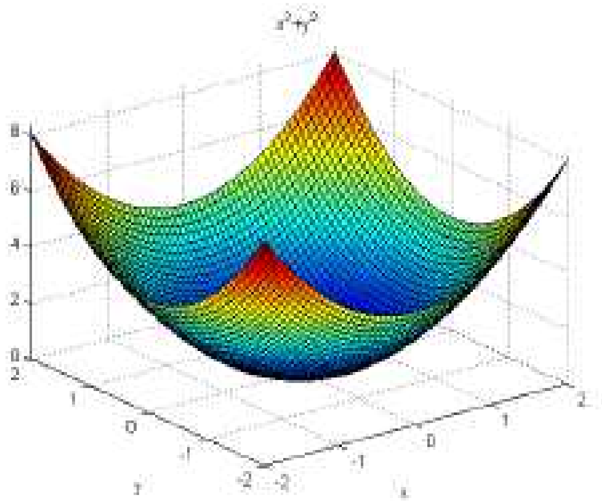
6 Positive Definite Matrices (正定矩阵)

6.3

SINGULAR VALUE DECOMPOSITION

(奇异值分解)

- $A^{\mathrm{T}}A$ and AA^{T}
- SVD Theorem
- Applications



In 1964, together with *William Kahan* and *Christian Reinsch*, he created an algorithm to compute the ***Singular Value Decomposition***, or ***SVD***, which will forever be an essential computational tool.

I. Facts about $A^{\mathrm{T}}A$ and AA^{T} ($A \in \mathrm{R}^{m\times n}$)

- $\operatorname{rank}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}) = \operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A}^{\mathrm{T}}) = r.$
- $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ and $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ are real symmetric (degree n and m respectively), and positive semidefinite. ($\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ and $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$的特征值为非负实数)
- The eigenvalues of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ and $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$:
  - $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ has n eigenvalues $\lambda_1, \dots, \lambda_n$, then
  $$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_n.$$
  - $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ has m eigenvalues $\mu_1, \dots, \mu_m$, then
  $$\mu_1 \geq \mu_2 \geq \cdots \geq \mu_r > 0 = \mu_{r+1} = \cdots = \mu_m.$$
  - We have the following conclusion: $\lambda_i = \mu_i > 0,\ i = 1, \dots, r.$ ($\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ and $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$的非零特征值集合相同)
  - **Definition**: $\sigma_i = \sqrt{\lambda_i} = \sqrt{\mu_i} > 0\ (i = 1, \dots, r)$ are called the **singular values (奇异值)** of $\boldsymbol{A}$.

Example 1 Find the singular values of the following matrices:

(1) $\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$;   (2) $\boldsymbol{A} = \begin{bmatrix} 0 & -1 & 1 \\ 2 & 0 & 0 \end{bmatrix}$.

Solution

(1) $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, so the eigenvalues of $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ are 5, 0 and 0,

and the singular value of $\boldsymbol{A}$ is $\sqrt{5}$.

(We can also check the eigenvalues of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$)

(2) $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$, so the eigenvalues of $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ are 2 and 4,

and the singular values of $\boldsymbol{A}$ are $\sqrt{2}$ and 2.

II. Singular Value Decomposition (奇异值分解)

We have seen that any real *symmetric* matrix $\boldsymbol{A}$ can be factored into

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathrm{T}} \quad (\textit{eigenvalue-eigenvector factorization}),$$

where $\boldsymbol{Q}$ is orthogonal, and $\boldsymbol{\Lambda}$ is diagonal.

There is an extension of this result.

Theorem 1 Any $n \times n$ invertible real matrix can be factored into

$$\boldsymbol{A} = \boldsymbol{Q}_1 \boldsymbol{S} \boldsymbol{Q}_2^{\mathrm{T}},$$

(invertible) = (orthogonal) (positive definite diagonal) (orthogonal)

where $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ are *orthogonal* matrices of degree n, and $\boldsymbol{S}$ is a positive definite *diagonal* matrix of degree n.

$$\boldsymbol{A} = \boldsymbol{Q}_1 \boldsymbol{S} \boldsymbol{Q}_2^{\mathrm{T}}$$

Proof Since $\boldsymbol{A}$ is invertible, we have

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = (\boldsymbol{A}\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}) = \|\boldsymbol{A}\boldsymbol{x}\|^2 > 0, \text{ for any } \boldsymbol{x} \neq \boldsymbol{0},$$

so $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is positive definite.

Thus there exists an orthogonal matrix $\boldsymbol{Q}_2$, such that

$$\boldsymbol{Q}_2^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q}_2 = \boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n).$$

Let $s_i = \sqrt{\lambda_i}$, then $s_i > 0$, i.e., $\boldsymbol{S} = \mathrm{diag}(s_1, s_2, \ldots, s_n)$ is positive definite, and $\boldsymbol{Q}_2^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q}_2 = \boldsymbol{S}^2$. Therefore,

$$\boldsymbol{S}^{-1}\boldsymbol{Q}_2^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q}_2\boldsymbol{S}^{-1} = \boldsymbol{I}, \qquad (*)$$

which shows that $\boldsymbol{A}\boldsymbol{Q}_2\boldsymbol{S}^{-1}$ is an orthogonal matrix, denoted by $\boldsymbol{Q}_1$.

Rewrite (*) and we can get $\boldsymbol{Q}_1^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q}_2\boldsymbol{S}^{-1} = \boldsymbol{I}$, so $\boldsymbol{Q}_1^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q}_2 = \boldsymbol{S}$,

or equivalently, $\boldsymbol{A} = \boldsymbol{Q}_1\boldsymbol{S}\boldsymbol{Q}_2^{\mathrm{T}}$.

Theorem 2 (*Singular Value Decomposition -- "SVD"*)

Any $m \times n$ real matrix with rank r can be factored into

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

(real matrix) $=$ (orthogonal) (rectangular diagonal) (orthogonal)

where $\boldsymbol{U}$ is orthogonal of degree m, $\boldsymbol{V}$ is orthogonal of degree n, and $\boldsymbol{\Sigma}$ is diagonal (but rectangular: $m \times n$).

Further, the columns of $\boldsymbol{U}$ are eigenvectors of $\boldsymbol{AA}^{\mathrm{T}}$, the columns of $\boldsymbol{V}$ are eigenvectors of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$, and the r positive entries $\sigma_1, \ldots, \sigma_r$ (called 'singular values') on the diagonal of $\boldsymbol{\Sigma}$ are the square roots of the nonzero eigenvalues of both $\boldsymbol{AA}^{\mathrm{T}}$ and $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$.

The factorization is called a **singular value decomposition (奇异值分解)**, or **SVD** for short.

$$\boldsymbol{A}_{m\times n} = \boldsymbol{U}_{m\times m} \left[\begin{array}{cccc|c} \sigma_1 & & & & \\ & \sigma_2 & & & \mathbf{0}_{r\times(n-r)} \\ & & \ddots & & \\ & & & \sigma_r & \\ \hline & \mathbf{0}_{(m-r)\times r} & & & \mathbf{0}_{(m-r)\times(n-r)} \end{array}\right]_{m\times n} (\boldsymbol{V}_{n\times n})^T$$

(orthogonal) (rectangular diagonal) (orthogonal)

where $\sigma_1, \sigma_2, \cdots, \sigma_r$ are the square roots of the nonzero eigenvalues of both $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ and $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$.

- $\boldsymbol{U}$: the columns are orthonormal eigenvectors for $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$.
- $\boldsymbol{V}$: the columns are orthonormal eigenvectors for $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$.

Remark 1 (other versions of SVD)

- For positive definite matrices $\boldsymbol{A}$: $\boldsymbol{\Sigma}$ is $\boldsymbol{\Lambda}$ and $\boldsymbol{U\Sigma V}^{\mathrm{T}}$ is identical to $\boldsymbol{Q\Lambda Q}^{\mathrm{T}}$.
- For other symmetric matrices $\boldsymbol{A}$, any negative eigenvalues in $\boldsymbol{\Lambda}$ become positive in $\boldsymbol{\Sigma}$.
- For complex matrices $\boldsymbol{A}$, $\boldsymbol{\Sigma}$ remains real but $\boldsymbol{U}$ and $\boldsymbol{V}$ become *unitary* (the complex version of *orthogonal*). We take $\boldsymbol{A} = \boldsymbol{U\Sigma V}^{\mathrm{H}}$.

Remark 2 $\boldsymbol{A} = \boldsymbol{U\Sigma V}^{\mathrm{T}}$, so $\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{V\Sigma}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}$, and hence

$$\boldsymbol{AA}^{\mathrm{T}} = (\boldsymbol{U\Sigma V}^{\mathrm{T}})(\boldsymbol{V\Sigma}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}) = \boldsymbol{U\Sigma\Sigma}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}},$$
$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = (\boldsymbol{V\Sigma}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}})(\boldsymbol{U\Sigma V}^{\mathrm{T}}) = \boldsymbol{V\Sigma}^{\mathrm{T}}\boldsymbol{\Sigma V}^{\mathrm{T}}.$$

$\boldsymbol{U}$ *must be the eigenvector matrix for* $\boldsymbol{AA}^{\mathrm{T}}$. The eigenvalue matrix is $\boldsymbol{\Sigma\Sigma}^{\mathrm{T}}$ (m by m) with $\sigma_1^2, \ldots, \sigma_r^2$ on the diagonal.

$\boldsymbol{V}$ *must be the eigenvector matrix for* $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$. The eigenvalue matrix is $\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{\Sigma}$ (n by n) with $\sigma_1^2, \ldots, \sigma_r^2$ on the diagonal.

$$\mathbf{\Sigma} = \left[\begin{array}{cccc|c} \sigma_1 & & & & \\ & \sigma_2 & & & \mathbf{0}_{r\times(n-r)} \\ & & \ddots & & \\ & & & \sigma_r & \\ \hline & \mathbf{0}_{(m-r)\times r} & & & \mathbf{0}_{(m-r)\times(n-r)} \end{array}\right]_{m\times n}$$

$$\mathbf{\Sigma}\mathbf{\Sigma}^{\mathrm{T}} = \begin{bmatrix} \sigma_1^2 & & & & \\ & \ddots & & & \\ & & \sigma_r^2 & & \\ & & & 0 & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}_{m\times m} \qquad \mathbf{\Sigma}^{\mathrm{T}}\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & & & & \\ & \ddots & & & \\ & & \sigma_r^2 & & \\ & & & 0 & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}_{n\times n}$$

Diagonal entries are eigenvalues for $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$

Diagonal entries are eigenvalues for $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$

Remark 3 $\boldsymbol{AV} = \boldsymbol{A}[\boldsymbol{V}_r \ \vdots \ \boldsymbol{V}_{n-r}] = [\boldsymbol{U}_r \ \vdots \ \boldsymbol{U}_{m-r}] \begin{bmatrix} \boldsymbol{\Sigma}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \boldsymbol{U\Sigma}.$

where $\boldsymbol{A}: m \times n$; $\boldsymbol{V}: n \times n$; $\boldsymbol{V}_r: n \times r$; $\boldsymbol{V}_{n-r}: n \times (n-r)$; $\boldsymbol{U}: m \times m$; $\boldsymbol{U}_r: m \times r$; $\boldsymbol{U}_{m-r}: m \times (m-r)$; $\boldsymbol{\Sigma}_r: r \times r$; $\boldsymbol{\Sigma}: m \times n$.

So $\boldsymbol{AV}_r = \boldsymbol{U}_r\boldsymbol{\Sigma}_r$ *and* $\underline{\boldsymbol{AV}_{n-r} = \mathbf{0}.}$

Similarly, $\boldsymbol{A} = \boldsymbol{U\Sigma V}^{\mathrm{T}}$, so $\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{V\Sigma}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}$, and $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{U} = \boldsymbol{V\Sigma}^{\mathrm{T}}$, therefore

$$\boldsymbol{A}^{\mathrm{T}}[\boldsymbol{U}_r \ \vdots \ \boldsymbol{U}_{m-r}] = [\boldsymbol{V}_r \ \vdots \ \boldsymbol{V}_{n-r}] \begin{bmatrix} \boldsymbol{\Sigma}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

So $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{U}_r = \boldsymbol{V}_r\boldsymbol{\Sigma}_r$ *and* $\underline{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{U}_{m-r} = \mathbf{0}.}$

$\boldsymbol{U}$ **and** $\boldsymbol{V}$ **give orthonormal bases for *all four fundamental subspaces*:**	
first r columns of $\boldsymbol{U}$	**column space** of $\boldsymbol{A}$
last $m-r$ columns of $\boldsymbol{U}$	**left nullspace** of $\boldsymbol{A}$
first r columns of $\boldsymbol{V}$	**row space** of $\boldsymbol{A}$
last $n-r$ columns of $\boldsymbol{V}$	**nullspace** of $\boldsymbol{A}$

Example 2 (SVD) Let $\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$.

Then $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$.

The nonzero eigenvalue of $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ (and also $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$) is 5, so the singular value of $\boldsymbol{A}$ is $\sqrt{5}$, and $\boldsymbol{\Sigma} = \begin{bmatrix} \sqrt{5} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$.

Three unit eigenvectors of $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ are $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. Thus $\boldsymbol{U} = \boldsymbol{I}$.

Finding two eigenvectors of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ gives rise to $\boldsymbol{V} = \begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}$.

Therefore, the matrix $\boldsymbol{A}$ can be decomposed into

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

It is easy to check that

$$\begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{5} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}.$$

Any problem here? (see next slide)

Attention! The SVD chooses those bases in an extremely special way. They are more than just orthonormal. Actually,

$$\boldsymbol{A}\boldsymbol{v}_j = \sigma_j \boldsymbol{u}_j, \quad j = 1, \ldots, r.$$

Reason: Let $\boldsymbol{v}_j$ be a unit eigenvector of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ corresponding to the eigenvalue σ_j^2, that is, $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_j = \sigma_j^2 \boldsymbol{v}_j$. Thus

$$(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})(\boldsymbol{A}\boldsymbol{v}_j) = \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_j = \boldsymbol{A}\sigma_j^2 \boldsymbol{v}_j = \sigma_j^2(\boldsymbol{A}\boldsymbol{v}_j),$$

namely, $\boldsymbol{A}\boldsymbol{v}_j$ is an eigenvector of $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ corresponding to the eigenvalue σ_j^2. The length of the vector $\boldsymbol{A}\boldsymbol{v}_j$ is σ_j because $\boldsymbol{v}_j^{\mathrm{T}}\boldsymbol{v}_j = 1$:

$$\|\boldsymbol{A}\boldsymbol{v}_j\|^2 = (\boldsymbol{A}\boldsymbol{v}_j)^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{v}_j) = \boldsymbol{v}_j^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_j = \boldsymbol{v}_j^{\mathrm{T}}\sigma_j^2\boldsymbol{v}_j = \sigma_j^2\boldsymbol{v}_j^{\mathrm{T}}\boldsymbol{v}_j = \sigma_j^2.$$

So the unit eigenvector $\boldsymbol{A}\boldsymbol{v}_j/\sigma_j = \boldsymbol{u}_j$, and $\boldsymbol{A}\boldsymbol{v}_j = \sigma_j\boldsymbol{u}_j$.

In other words, $\boldsymbol{A}\boldsymbol{V} = \boldsymbol{U}\boldsymbol{\Sigma}$.

How to construct the matrix V

$\boldsymbol{V}$: the columns are orthonormal eigenvectors for $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$.

Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$ are columns of $\boldsymbol{V}$, and let

$$\boldsymbol{V} = [\boldsymbol{v}_1 \quad \cdots \quad \boldsymbol{v}_r \quad \boldsymbol{v}_{r+1} \quad \cdots \quad \boldsymbol{v}_n] = [\boldsymbol{V}_r \vdots \boldsymbol{V}_{n-r}]$$

Then $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{V}(\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{\Sigma})\boldsymbol{V}^{\mathrm{T}}$ becomes

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}[\boldsymbol{v}_1 \quad \cdots \quad \boldsymbol{v}_r \quad \boldsymbol{v}_{r+1} \quad \cdots \quad \boldsymbol{v}_n]$$

$$= [\boldsymbol{v}_1 \quad \cdots \quad \boldsymbol{v}_r \quad \boldsymbol{v}_{r+1} \quad \cdots \quad \boldsymbol{v}_n]\begin{bmatrix} \sigma_1^2 & & & & & \\ & \ddots & & & & \\ & & \sigma_r^2 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}$$

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_1 = \sigma_1^2\boldsymbol{v}_1, \cdots, \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_r = \sigma_r^2\boldsymbol{v}_r, \qquad \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_{r+1} = \boldsymbol{0}, \cdots, \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_n = \boldsymbol{0}$$

$\boldsymbol{v}_1, \cdots, \boldsymbol{v}_r$ are eigenvectors of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ belonging to nonzero eigenvalues $\sigma_1^2, \cdots, \sigma_r^2$ respectively.

$\boldsymbol{v}_{r+1}, \cdots, \boldsymbol{v}_n$ are eigenvectors of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ belonging to $\lambda = 0$.

How to construct the matrix U

$$\boldsymbol{A}_{m\times n} = \boldsymbol{U}_{m\times m}\left[\begin{array}{cccc|c} \sigma_1 & & & & \\ & \sigma_2 & & & \mathbf{0}_{r\times(n-r)} \\ & & \ddots & & \\ & & & \sigma_r & \\ \hline & \mathbf{0}_{(m-r)\times r} & & & \mathbf{0}_{(m-r)\times(n-r)} \end{array}\right]_{m\times n} (\boldsymbol{V}_{n\times n})^{\mathrm{T}}$$

$$\Longrightarrow \boldsymbol{A}_{m\times n}\boldsymbol{V}_{n\times n} = \boldsymbol{U}_{m\times m}\,\boldsymbol{\Sigma}_{m\times n}$$

$$\boldsymbol{A}[\boldsymbol{v}_1 \quad \cdots \quad \boldsymbol{v}_r \quad \boldsymbol{v}_{r+1} \quad \cdots \quad \boldsymbol{v}_n] = [\boldsymbol{u}_1 \quad \cdots \quad \boldsymbol{u}_r \quad \boldsymbol{u}_{r+1} \quad \cdots \quad \boldsymbol{u}_m]\begin{bmatrix} \boldsymbol{\Sigma}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

Comparing the first r columns of each side, we see that

$$\boldsymbol{A}\boldsymbol{v}_j = \sigma_j \boldsymbol{u}_j, j = 1, \cdots, r \Longrightarrow \boldsymbol{u}_j = \frac{1}{\sigma_j}\boldsymbol{A}\boldsymbol{v}_j, j = 1, \cdots, r$$

It follows from that each $\boldsymbol{u}_j, j = 1, \cdots, r$, is in the column space of $\boldsymbol{A}$.

The dimension of the column space is r, so $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_r$ form an orthonormal basis for $C(\boldsymbol{A})$. The vector space $\mathrm{C}(\boldsymbol{A})^{\perp} = \mathrm{N}(\boldsymbol{A}^{\mathrm{T}})$ has dimension $m - r$. $\boldsymbol{u}_{r+1}, \cdots, \boldsymbol{u}_m$ is an orthonormal basis for $\mathrm{N}(\boldsymbol{A}^{\mathrm{T}})$.

Example 2' (SVD) Let $\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$.

Then $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$.

The nonzero eigenvalue of $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ (and also $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$) is 5, so the singular value of $\boldsymbol{A}$ is $\sigma_1 = \sqrt{5}$, and $\boldsymbol{\Sigma} = \begin{bmatrix} \sqrt{5} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$.

Two unit eigenvectors of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ are $\boldsymbol{v}_1 = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix}$, $\boldsymbol{v}_2 = \begin{bmatrix} -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix}$, corresponding to the eigenvalues 5 and 0 respectively, so $\boldsymbol{V} = \begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}$.

By $\boldsymbol{A}\boldsymbol{v}_1/\sigma_1 = \boldsymbol{u}_1$, we get $\boldsymbol{u}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, and we choose $\boldsymbol{u}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, $\boldsymbol{u}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ to make $\boldsymbol{U}$ an orthogonal matrix.

Therefore, the matrix $\boldsymbol{A}$ can be decomposed into

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

$$\text{i.e., } \begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{5} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}.$$

Example 3 (SVD) Let $\boldsymbol{A} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$.

Then $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$, $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$.

The eigenvalues of $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ are 3 and 1, which are also the non-zero eigenvalues of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$. So $\sigma_1 = \sqrt{3}$, $\sigma_2 = 1$.

Finding three eigenvectors of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ gives rise to

$$\boldsymbol{V} = \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ -\frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{bmatrix} = [\boldsymbol{v}_1 \quad \boldsymbol{v}_2 \quad \boldsymbol{v}_3].$$

Two unit eigenvectors of $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ are

$$\boldsymbol{u}_1 = \frac{\boldsymbol{A}\boldsymbol{v}_1}{\sigma_1} = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \quad \boldsymbol{u}_2 = \frac{\boldsymbol{A}\boldsymbol{v}_2}{\sigma_2} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Thus $\boldsymbol{U} = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$.

Now $\boldsymbol{\Sigma}$ is a (2×3)-matrix $\boldsymbol{\Sigma} = \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, it is easy to get that

$$\boldsymbol{A} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}} = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}.$$

Example 4 Find the singular value decomposition(SVD) for

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix}.$$

Solution.

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$$

has eigenvalues 4 and 0, so the singular value of $\boldsymbol{A}$ is $\sqrt{4} = 2$, and

$$\boldsymbol{\Sigma} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Finding two eigenvectors corresponding to 4,0 of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ gives rise to

$$\boldsymbol{V} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}.$$

Solution(continued).

$$\boldsymbol{u}_1 = \frac{1}{\sigma_1}\boldsymbol{A}\boldsymbol{v}_1 = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}$$

The remaining column vectors of $\boldsymbol{U}$ must form an orthonormal basis for $\mathrm{N}(\boldsymbol{A}^{\mathrm{T}})$, we can compute a basis for $\mathrm{N}(\boldsymbol{A}^{\mathrm{T}})$ in the usual way,

$$\boldsymbol{x}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \qquad \boldsymbol{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Since these vectors are already orthogonal, it is not necessary to use the Gram–Schmidt process to obtain an orthonormal basis. We need only set

$$\boldsymbol{u}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \qquad \boldsymbol{u}_3 = \boldsymbol{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Solution(continued).

It then follows that

$$\boldsymbol{A}=\begin{bmatrix}1 & 1\\ 1 & 1\\ 0 & 0\end{bmatrix}=\boldsymbol{U\Sigma V}^{\mathrm{T}}$$

$$=\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}2 & 0\\ 0 & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{bmatrix}.$$

Type	Form	The Factors	Notes
***LU* factorization** *(Gaussian elimination)*	$\boldsymbol{PA} = \boldsymbol{LU}$ ($\boldsymbol{A}$ is any $m \times n$ matrix)	$\boldsymbol{P}$: permutation matrix $\boldsymbol{L}$: lower triangular matrix with unit diagonal $\boldsymbol{U}$: $m \times n$ echelon matrix (When $m = n$, $\boldsymbol{U}$ is upper triangular.)	The permutation matrix $\boldsymbol{P}$ is needed when there are row exchanges during the row reduction. (Otherwise, $\boldsymbol{A} = \boldsymbol{LU}$) $\boldsymbol{PA} = \boldsymbol{LDU}$ if $\boldsymbol{U}$ is upper triangular with unit diagonal.
***QR* factorization** *(Gram-Schmidt orthogonalization)*	$\boldsymbol{A} = \boldsymbol{QR}$ ($\boldsymbol{A}$ is any $m \times n$ matrix with independent columns)	$\boldsymbol{Q}$: matrix with orthonormal columns (When $m = n$, $\boldsymbol{Q}$ becomes an orthogonal matrix.) $\boldsymbol{R}$: upper triangular and invertible	When $m = n$, any invertible matrix can be factorized as a product of an orthogonal matrix and an upper triangular matrix.
Singular Value Decomposition *(SVD)*	$\boldsymbol{A} = \boldsymbol{U\Sigma V}^{\mathrm{T}}$ ($\boldsymbol{A}$ is any $m \times n$ matrix with rank r)	$\boldsymbol{U}$: $m \times m$ orthogonal matrix $\boldsymbol{V}$: $n \times n$ orthogonal matrix $\boldsymbol{\Sigma}$ is diagonal (but rectangular: $m \times n$).	The columns of $\boldsymbol{U}$ are eigenvectors of $\boldsymbol{AA}^{\mathrm{T}}$, the columns of $\boldsymbol{V}$ are eigenvectors of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$, and the r positive entries on the diagonal of $\boldsymbol{\Sigma}$ are the square roots of the nonzero eigenvalues of both $\boldsymbol{AA}^{\mathrm{T}}$ and $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$.

III. Applications of SVD

There are lots of applications of SVD in various areas, including in communication of information and in mathematics.

***Application I* —** polar decomposition / factorization (极分解)

Theorem 3 (*Polar factorization*) Every real square matrix can be factored into $\boldsymbol{A} = \boldsymbol{QS}$, where $\boldsymbol{Q}$ is *orthogonal* and $\boldsymbol{S}$ is *symmetric positive semidefinite*. If $\boldsymbol{A}$ is invertible then $\boldsymbol{S}$ is positive definite.

Proof. Let $\boldsymbol{A} = \boldsymbol{U\Sigma V}^{\mathrm{T}}$ as above. Then

$$\boldsymbol{A} = \boldsymbol{U\Sigma V}^{\mathrm{T}} = (\boldsymbol{UV}^{\mathrm{T}})(\boldsymbol{V\Sigma V}^{\mathrm{T}})$$

Let $\boldsymbol{Q} = \boldsymbol{UV}^{\mathrm{T}}$ and $\boldsymbol{S} = \boldsymbol{V\Sigma V}^{\mathrm{T}}$,

then $\boldsymbol{S}$ is symmetric and positive semidefinite (because $\boldsymbol{\Sigma}$ is), and $\boldsymbol{Q}$ is an orthogonal matrix.

Example 5 Polar decomposition $\boldsymbol{A} = \boldsymbol{QS}$, where $\boldsymbol{A} = \begin{bmatrix} 1 & -2 \\ 3 & -1 \end{bmatrix}$.

Solution $\boldsymbol{A} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix}$.

Remark 1 **(Analogy)** Every nonzero complex number z can be written as $z = r\mathbf{e}^{i\theta}$ (polar coordinates: 极坐标)

If we think of z as a 1 by 1 matrix, r corresponds to a *positive definite matrix* and $\mathbf{e}^{i\theta}$ corresponds to an *orthogonal matrix*.

More exactly, since $\mathbf{e}^{i\theta}$ is complex and satisfies $\mathbf{e}^{-i\theta}\mathbf{e}^{i\theta} = 1$, it forms a 1 by 1 *unitary matrix*: $\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U} = \boldsymbol{I}$.

The SVD extends this “polar factorization” to matrices of any size.

Remark 2 A major use of the polar decomposition is in continuum mechanics (连续介质力学) and recently in robotics.

In any deformation, it is important to separate stretching from rotation, and that is exactly what $\boldsymbol{QS}$ achieves. ($\boldsymbol{Q}$: a rotation, and possibly a reflection. The material feels no strain; $\boldsymbol{S}$: has eigenvalues $\sigma_1, \ldots, \sigma_r$, which are the stretching factors (or compression factors).

Application II — the effective rank (有效秩)

The rank of a matrix: the number of independent rows (columns).

That can be hard to decide in computations!

- In exact arithmetic, counting the pivots is correct.
- Real arithmetic can be misleading—but discarding small pivots is not the answer.

For example, consider the following (**$\boldsymbol{\epsilon}$ is very small**):

$$\boldsymbol{A}=\begin{bmatrix}\epsilon & 2\epsilon \\ 1 & 2\end{bmatrix} \text{ and } \boldsymbol{B}=\begin{bmatrix}\epsilon & 1 \\ 0 & 0\end{bmatrix} \text{ and } \boldsymbol{C}=\begin{bmatrix}\epsilon & 1 \\ \epsilon & 1+\epsilon\end{bmatrix}$$

(1) $\boldsymbol{A}$ has rank 1, although roundoff error will probably produce a second pivot. Both pivots will be small; how many do we ignore?

(2) $\boldsymbol{B}$ has one small pivot, but we cannot pretend that its row is insignificant.

(3) $\boldsymbol{C}$ has two pivots and its rank is 2, but its "effective rank" ought to be 1.

$$\boldsymbol{A} = \begin{bmatrix} \epsilon & 2\epsilon \\ 1 & 2 \end{bmatrix} \text{ and } \boldsymbol{B} = \begin{bmatrix} \epsilon & 1 \\ 0 & 0 \end{bmatrix} \text{ and } \boldsymbol{C} = \begin{bmatrix} \epsilon & 1 \\ \epsilon & 1+\epsilon \end{bmatrix}$$

Solution (a more stable measure of rank):

Step 1 use $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ or $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$, which are symmetric but share the same rank as $\boldsymbol{A}$.

Their eigenvalues—the singular values squared—are not misleading.

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \begin{bmatrix} \epsilon^2+1 & 2\epsilon^2+2 \\ 2\epsilon^2+2 & 4\epsilon^2+4 \end{bmatrix}$$

Step 2 Based on the accuracy of the data, we decide on a tolerance (like 10^{-6}) and count the singular values above it—that is the effective rank.

The examples above have effective rank 1 (when ϵ is very small).

Application III — image processing (图像处理)

$$\boldsymbol{A} = \boldsymbol{U\Sigma V}^{\mathrm{T}} = [\boldsymbol{U}_r \quad \boldsymbol{U}_{m-r}]\begin{bmatrix}\boldsymbol{\Sigma}_r & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{V}_r^{\mathrm{T}}\\ \boldsymbol{V}_{n-r}^{\mathrm{T}}\end{bmatrix} = \boldsymbol{U}_r\boldsymbol{\Sigma}_r\boldsymbol{V}_r^{\mathrm{T}}.$$

where $\boldsymbol{A}: m\times n$; $\boldsymbol{V}: n\times n$; $\boldsymbol{V}_r: n\times r$; $\boldsymbol{V}_{n-r}: n\times(n-r)$; $\boldsymbol{U}: m\times m$; $\boldsymbol{U}_r: m\times r$; $\boldsymbol{U}_{m-r}: m\times(m-r)$; $\boldsymbol{\Sigma}_r: r\times r$; $\boldsymbol{\Sigma}: m\times n$.

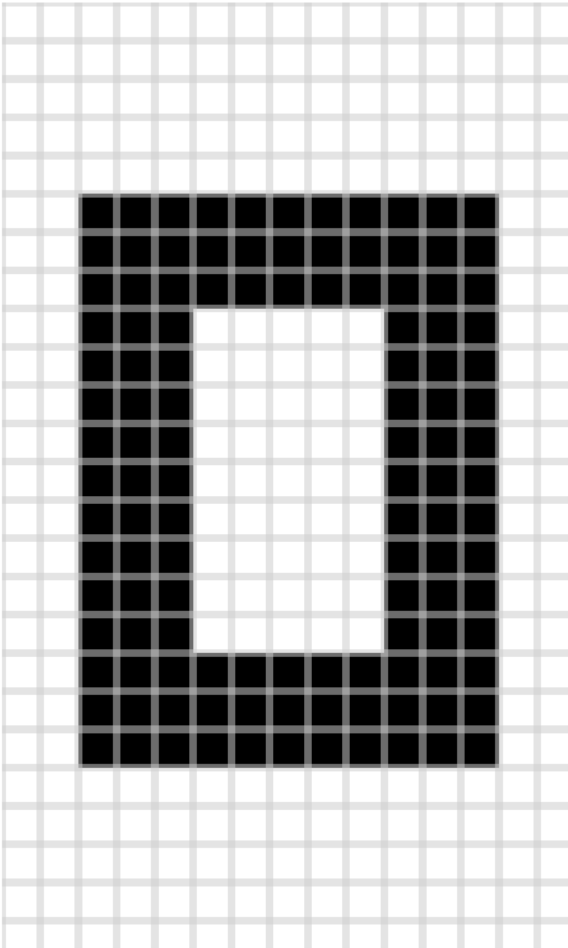
- **Image compression (图像压缩)**

例如一张$m\times n$的图像, 需要$m\times n$的矩阵$\boldsymbol{A}$来存储它.

而利用奇异值分解, 则只需存储矩阵的奇异值($\boldsymbol{\Sigma}_r$的对角线), 奇异向量$\boldsymbol{U}_r$和$\boldsymbol{V}_r^{\mathrm{T}}$, 数目为$r\times(m+n+1)$, 而不再是$m\times n$.

通常$r \ll m,\ n$, 所以存储该图像所需的存储量减小了.

比值称$\frac{m\times n}{r\times(m+n+1)}$为<u>图像的压缩比</u>, 其倒数称为<u>数据压缩率</u>.



15 × 25 black or white pixels

$$M = \begin{bmatrix}
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&0&0&0&0&0&0&0&0&0&0&0&1&1\\
1&1&0&0&0&0&0&0&0&0&0&0&0&1&1\\
1&1&0&0&0&0&0&0&0&0&0&0&0&1&1\\
1&1&0&0&0&1&1&1&1&1&0&0&0&1&1\\
1&1&0&0&0&1&1&1&1&1&0&0&0&1&1\\
1&1&0&0&0&1&1&1&1&1&0&0&0&1&1\\
1&1&0&0&0&1&1&1&1&1&0&0&0&1&1\\
1&1&0&0&0&1&1&1&1&1&0&0&0&1&1\\
1&1&0&0&0&1&1&1&1&1&0&0&0&1&1\\
1&1&0&0&0&1&1&1&1&1&0&0&0&1&1\\
1&1&0&0&0&1&1&1&1&1&0&0&0&1&1\\
1&1&0&0&0&1&1&1&1&1&0&0&0&1&1\\
1&1&0&0&0&0&0&0&0&0&0&0&0&1&1\\
1&1&0&0&0&0&0&0&0&0&0&0&0&1&1\\
1&1&0&0&0&0&0&0&0&0&0&0&0&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1
\end{bmatrix}$$

0: black ; 1: white pixel

The matrix ***M*** has 375 entries.

If we perform a singular value decomposition on $\boldsymbol{M}$, we find there are only three non-zero singular values.

$$\sigma_1 = 14.72\ ,\ \sigma_2 = 5.22\ ,\ \sigma_3 = 3.31$$

$$\boldsymbol{M} = \boldsymbol{u}_1 \sigma_1 \boldsymbol{v}_1{}^{\mathrm{T}} + \boldsymbol{u}_2 \sigma_2 \boldsymbol{v}_2{}^{\mathrm{T}} + \boldsymbol{u}_3 \sigma_3 \boldsymbol{v}_3{}^{\mathrm{T}}$$

This implies that we may represent the matrix using only 123 numbers rather than the 375 that appear in the matrix.

In this way, the singular value decomposition discovers the redundancy in the matrix and provides a format for eliminating it.

The organizing committee for the 1964 Gatlinburg/Householder meeting on ***Numerical Algebra***. All six members of the committee – J. H. Wilkinson, Wallace Givens, George Forsythe, Alston Householder, Peter Henrici, and F. L. Bauer – have influenced MATLAB.

The following figure shows an image corresponding to a 176×260 matrix $\boldsymbol{A}$ and three images corresponding to lower rank approximations of $\boldsymbol{A}$. The gentlemen in the picture are (from left to right) James H. Wilkinson, Wallace Givens, and George Forsythe (three pioneers in the field of numerical linear algebra).

Original 176 by 260 Image

Rank 5 Approximation to Image

Rank 15 Approximation to Image

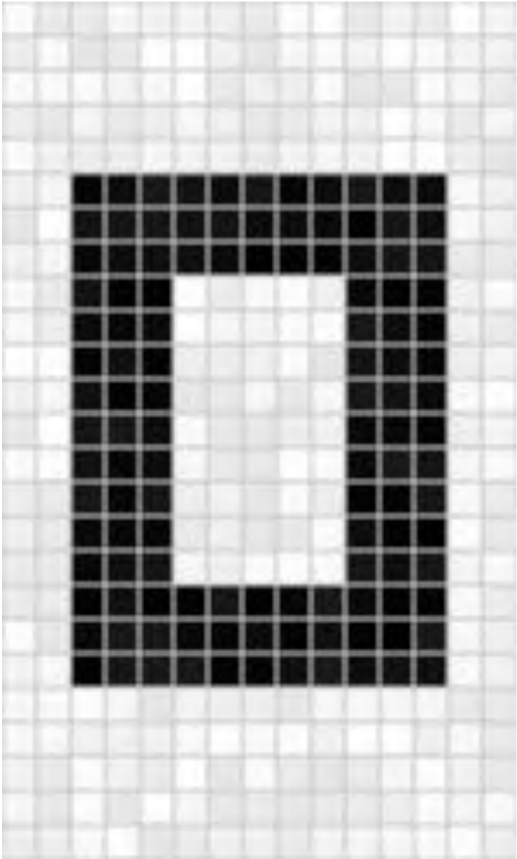
Rank 30 Approximation to Image

Courtesy *Oakridge National Laboratory*

The pictures are really striking, as more and more singular values are included.
At first you see nothing, and suddenly you recognize everything.
The cost is in computing the SVD–this had become much more efficient, but it is expensive for a big matrix.

- **Noise reduction (降噪)**

如果矩阵的奇异值从一个数开始值远小于前面的奇异值, 则可以删去, 这样在保证图像不失真的前提下, 进一步减小了存储量.



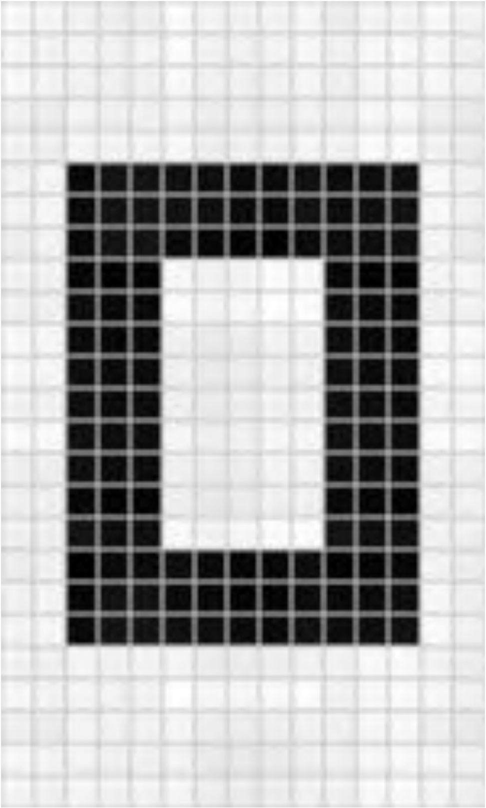
$\sigma_1 = 14.15$
$\sigma_2 = 4.67$
$\sigma_3 = 3.00$
$\sigma_4 = 0.21$
$\sigma_5 = 0.19$
...
$\sigma_{15} = 0.05$

Clearly, the *first three singular values* are the most important, so we will assume that the others are due to the noise in the image and make the approximation:

$$\boldsymbol{M} \approx \boldsymbol{u}_1\sigma_1\boldsymbol{v}_1^{\mathrm{T}} + \boldsymbol{u}_2\sigma_2\boldsymbol{v}_2^{\mathrm{T}} + \boldsymbol{u}_3\sigma_3\boldsymbol{v}_3^{\mathrm{T}}$$

Noisy image

Improved image



Application IV — least squares (最小二乘)

Recall that:

Theorem. *If a system* $\boldsymbol{Ax} = \boldsymbol{b}$ *is inconsistent* (*has no solution*), *its least-squares solution minimizes* $||\boldsymbol{Ax} - \boldsymbol{b}||^2$:

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\widehat{\boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}. \quad \textbf{(Normal equations)}$$

Moreover, if $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ *is invertible, then*

$$\widehat{\boldsymbol{x}} = \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}. \quad \textbf{(Best estimate)}$$

The projection of $\boldsymbol{b}$ *onto the column space is the nearest point* $\boldsymbol{A}\widehat{\boldsymbol{x}}$:

$$\boldsymbol{p} = \boldsymbol{A}\widehat{\boldsymbol{x}} = \boldsymbol{A}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}. \quad \textbf{(Projection)}$$

There is a simple way to decide *whether* $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ *is invertible*.

Theorem. *The matrices* $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ *and* $\boldsymbol{A}$ *have the same nullspace.*

In particular, if $\boldsymbol{A}$ *has full column rank, then* $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ *is invertible.*

Problem: ***If A has dependent columns then*** $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ ***is not invertible and*** $\widehat{\boldsymbol{x}}$ ***is not determined (not unique)***. Any vector in the nullspace could be added to $\widehat{\boldsymbol{x}}$.

We choose the solution: *The optimal solution of $\boldsymbol{Ax}=\boldsymbol{b}$ is the minimum length solution of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\widehat{\boldsymbol{x}}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$.*

That minimum length solution will be called $\boldsymbol{x}^{+}$. It is our preferred choice as the best solution to $\boldsymbol{Ax}=\boldsymbol{b}$ (which had no solution), and also to $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\widehat{\boldsymbol{x}}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ (which had too many).

We start with a *diagonal example*.

Example 6 $\boldsymbol{A}$ is diagonal, with dependent rows and dependent columns:

$$\boldsymbol{A}\widehat{\boldsymbol{x}}=\boldsymbol{p} \quad \text{is} \quad \begin{bmatrix}\sigma_1 & 0 & 0 & 0\\ 0 & \sigma_2 & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}\begin{bmatrix}\hat{x}_1\\ \hat{x}_2\\ \hat{x}_3\\ \hat{x}_4\end{bmatrix}=\begin{bmatrix}b_1\\ b_2\\ 0\end{bmatrix}.$$

The columns all end with zero. In the column space, the closest vector to $\boldsymbol{b}=(b_1,b_2,b_3)^{\mathrm{T}}$ is $\boldsymbol{p}=(b_1,b_2,0)^{\mathrm{T}}$.

The best we can do with $\boldsymbol{Ax}=\boldsymbol{b}$ is to solve the first two equations, then $\hat{x}_1=b_1/\sigma_1$ *and* $\hat{x}_2=b_2/\sigma_2$.

Example 6 (Continued) To make $\hat{\boldsymbol{x}}$ as short as possible, we choose the totally arbitrary $\hat{x}_3 = \hat{x}_4 = 0$.

The minimum length solution is $\boldsymbol{x}^+$:

$$\boldsymbol{x}^+ = \begin{bmatrix} b_1/\sigma_1 \\ b_2/\sigma_2 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/\sigma_1 & 0 & 0 \\ 0 & 1/\sigma_2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

($\boldsymbol{A}^+$ is **pseudoinverse** of $\boldsymbol{A}$, and $\boldsymbol{x}^+ = \boldsymbol{A}^+\boldsymbol{b}$ is the shortest solution.)

***Remark*.** Based on this example, we know $\boldsymbol{\Sigma}^+$ and $\boldsymbol{x}^+$ for any diagonal matrix $\boldsymbol{\Sigma}$:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \end{bmatrix}_{m\times n}, \quad \boldsymbol{\Sigma}^+ = \begin{bmatrix} 1/\sigma_1 & & & \\ & \ddots & & \\ & & 1/\sigma_r & \\ & & & \end{bmatrix}_{n\times m},$$

$$\boldsymbol{x}^+ = \boldsymbol{\Sigma}^+\boldsymbol{b} = \begin{bmatrix} b_1/\sigma_1 \\ \vdots \\ b_r/\sigma_r \\ \\ \end{bmatrix}_{n\times 1}.$$ *and obviously* $(\boldsymbol{\Sigma}^+)^+ = \boldsymbol{\Sigma}$.

Now we find $\boldsymbol{x}^+$ **in the general case**.

We claim that: ***The shortest solution x^+ is always in the row space of A***.

Remember that any vector $\widehat{\boldsymbol{x}}$ can be split into a row space component $\boldsymbol{x}_r$ and a nullspace component: $\widehat{\boldsymbol{x}} = \boldsymbol{x}_r + \boldsymbol{x}_n$.

There are three important points about that splitting:

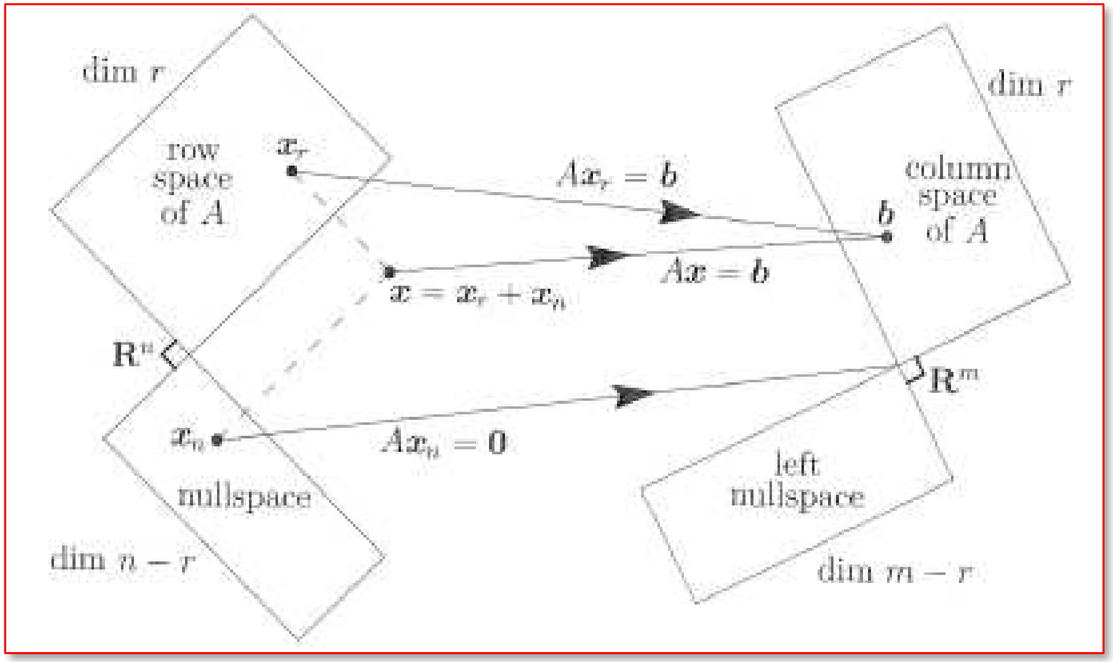
1. The row space component also solves $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}_r = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$, because $\boldsymbol{A}\boldsymbol{x}_n = \boldsymbol{0}$.

2. The components are orthogonal, and they obey Pythagoras's law:

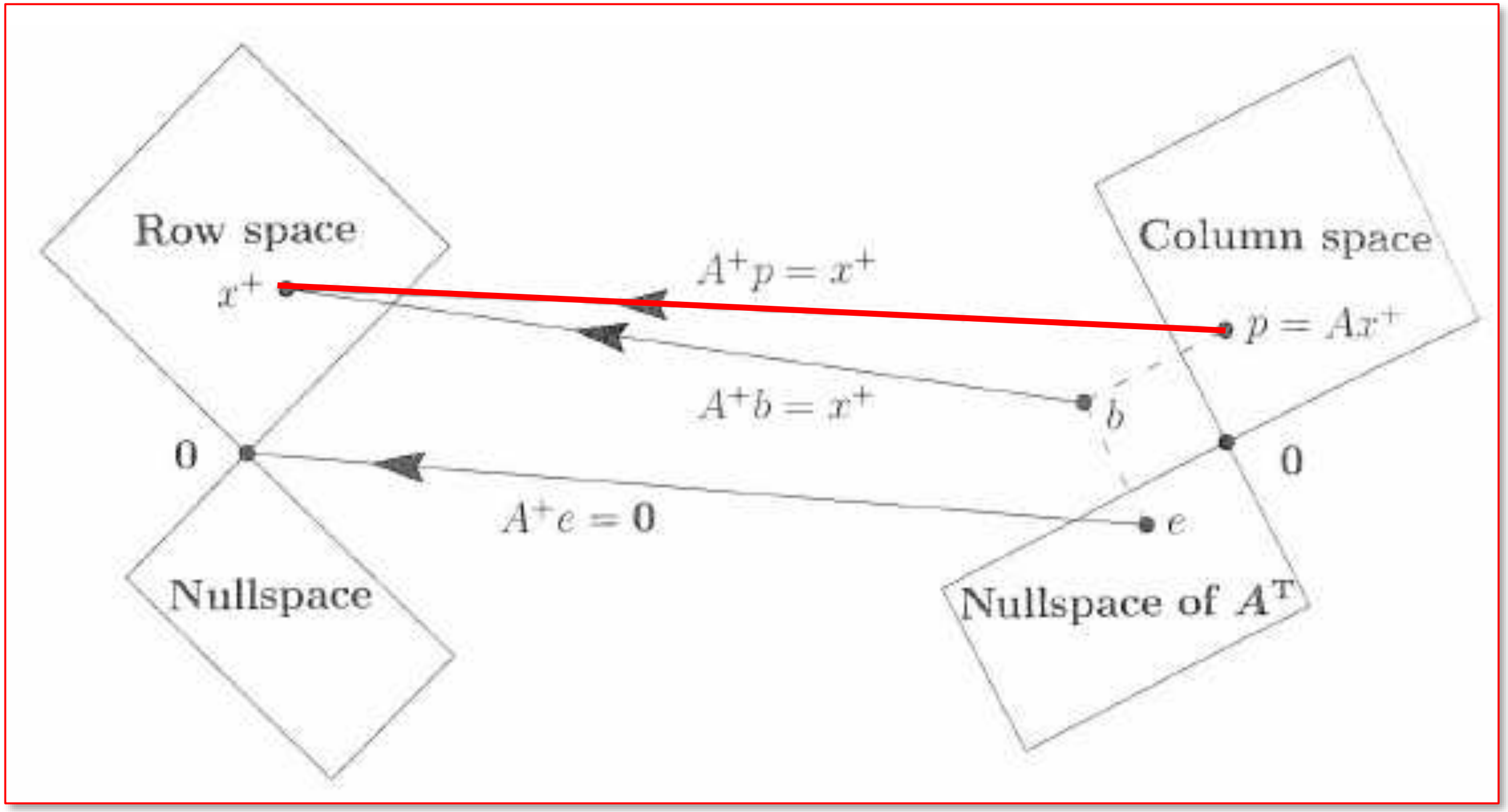
$$\|\widehat{\boldsymbol{x}}\|^2 = \|\boldsymbol{x}_r\|^2 + \|\boldsymbol{x}_n\|^2,$$

so $\widehat{\boldsymbol{x}}$ is shortest when $\boldsymbol{x}_n = \boldsymbol{0}$.

3. All solutions of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\widehat{\boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ have the same $\boldsymbol{x}_r$. ***That vector is*** $\boldsymbol{x}^+$.



All we are doing is to choose that vector, $x^+ = x_r$, as the best solution to $Ax = b$.



The pseudoinverse $\boldsymbol{A}^+$ in the figure above starts with $\boldsymbol{b}$ and comes back to $\boldsymbol{x}^+$. *It inverts $\boldsymbol{A}$ where $\boldsymbol{A}$ is invertible*—between row space and column space. The pseudoinverse knocks out the left nullspace by sending it to zero, and it knocks out the nullspace by choosing $\boldsymbol{x}_r$ as $\boldsymbol{x}^+$.

Example 7 $\boldsymbol{Ax} = \boldsymbol{b}$ is $-x_1 + 2x_2 + 2x_3 = 18$, with a whole plane of solutions.

According to our theory, the shortest solution should be in the row space of $\boldsymbol{A} = [-1 \;\; 2 \;\; 2]$.

The multiple of that row that satisfies the equation is $\boldsymbol{x}^+ = (-2,4,4)^{\mathrm{T}}$.

The matrix that produces $\boldsymbol{x}^+$ from $\boldsymbol{b} = [18]$ is the pseudoinverse $\boldsymbol{A}^+$.

Whereas $\boldsymbol{A}$ was 1 by 3, this $\boldsymbol{A}^+$ is 3 by 1:

$$\boldsymbol{A}^+ = [-1 \quad 2 \quad 2]^+ = \begin{bmatrix} -\frac{1}{9} \\ \frac{2}{9} \\ \frac{2}{9} \end{bmatrix} \quad \text{and} \quad \boldsymbol{A}^+\boldsymbol{b} = \begin{bmatrix} -\frac{1}{9} \\ \frac{2}{9} \\ \frac{2}{9} \end{bmatrix} [18] = \begin{bmatrix} -2 \\ 4 \\ 4 \end{bmatrix}.$$

The row space of $\boldsymbol{A}$ is the column space of $\boldsymbol{A}^+$.

Remark There are longer solutions like $(-2,5,3)^{\mathrm{T}}$, $(-2,7,1)^{\mathrm{T}}$, or $(-6,3,3)^{\mathrm{T}}$, but they all have nonzero components from the nullspace.

Here is a formula for finding $\boldsymbol{A}^+$ (next slide).

Theorem 4 *If* $\boldsymbol{A} = \boldsymbol{U\Sigma V}^T$ *(the SVD), then its* ***pseudoinverse*** *is*

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

Remark

We can check this proposition in Example 7, where $\boldsymbol{A} = [-1\ \ 2\ \ 2]$, and the singular value of $\boldsymbol{A}$ is $\sigma = 3$ — the square root of the eigenvalue of $\boldsymbol{AA}^{\mathrm{T}} = [9]$.

$$\boldsymbol{A} = [-1\ \ 2\ \ 2] = \boldsymbol{U\Sigma V}^{\mathrm{T}} = [1][3 \quad 0 \quad 0]\begin{bmatrix} -\frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \end{bmatrix},$$

and

$$\boldsymbol{V\Sigma}^+\boldsymbol{U}^{\mathrm{T}} = \begin{bmatrix} -\frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \end{bmatrix}\begin{bmatrix} \frac{1}{3} \\ 0 \\ 0 \end{bmatrix}[1] = \begin{bmatrix} -\frac{1}{9} \\ \frac{2}{9} \\ \frac{2}{9} \end{bmatrix} = \boldsymbol{A}^+.$$

Theorem 4 *If* $\boldsymbol{A} = \boldsymbol{U\Sigma V}^{\mathrm{T}}$ (*the SVD*), *then its* ***pseudoinverse*** *is*

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

Proof Multiplication by the orthogonal matrix $\boldsymbol{U}^{\mathrm{T}}$ leaves lengths unchanged:

$$\|\boldsymbol{Ax} - \boldsymbol{b}\| = \left\|\boldsymbol{U\Sigma V}^{\mathrm{T}}\boldsymbol{x} - \boldsymbol{b}\right\| = \left\|\boldsymbol{\Sigma V}^{\mathrm{T}}\boldsymbol{x} - \boldsymbol{U}^{\mathrm{T}}\boldsymbol{b}\right\|.$$

Introduce the new unknown $\boldsymbol{y} = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{V}^{-1}\boldsymbol{x}$, which has the same length as $\boldsymbol{x}$.

Then, minimizing $\|\boldsymbol{Ax} - \boldsymbol{b}\|$ is the same as minimizing $\left\|\boldsymbol{\Sigma y} - \boldsymbol{U}^{\mathrm{T}}\boldsymbol{b}\right\|$.

Now $\boldsymbol{\Sigma}$ is diagonal and we know the best $\boldsymbol{y}^{+}$.

It is $\boldsymbol{y}^{+} = \boldsymbol{\Sigma}^{+}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{b}$, so the best $\boldsymbol{x}^{+}$ is $\boldsymbol{V}\boldsymbol{y}^{+}$:

$$\textbf{Shortest solution} \qquad \boldsymbol{x}^{+} = \boldsymbol{V}\boldsymbol{y}^{+} = \boldsymbol{V\Sigma}^{+}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{b} = \boldsymbol{A}^{+}\boldsymbol{b}.$$

$\boldsymbol{V}\boldsymbol{y}^{+}$ is in the row space, and $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}^{+} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ from the **SVD**.

Key words:

Singular values

Singular Value Decomposition

Applications (I, II, III, IV)

Homework

See Blackboard