高级工程数学
2021-2022 (1)

沈超敏
计算机科学与技术学院
cmshen@cs.ecnu.edu.cn

教书院 219

上次课内容回顾

- 内积和范数的一些公式和性质. 推广: 柯西-施瓦茨不等式（Cauchy-Schwarz）推出的两边之和大于第三边和勾股定理
- 关于欧氏范数，范数取 $p = 2$ 情况下的一些定义，范数也可以描述连续函数，还可以运用让 $\|x - x^*\|$ 取最小，即最接近解，进而得出最优解

根据线性方程组求最优解，都需要用到内积和范数，运用性质得出最优解。

其目的是让我们知道解方程组不是仅仅得出无解就可以了，而是让我们理解优化的过程，往最优的方向努力。

2.4 Inner Products and Norms

The Euclidean norm of a vector $\|\boldsymbol{x}\|$ has the following properties:

1. Positivity: $\|\boldsymbol{x}\| > 0$, $\|\boldsymbol{x}\| = \mathbf{0}$ iff $\boldsymbol{x} = \mathbf{0}$.
2. Homogeneity: $\|r\boldsymbol{x}\| = |r| \cdot \|\boldsymbol{x}\|$, $r \in \mathbb{R}$.
3. Triangle inequality: $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$.

2.4 Inner Products and Norms

The Euclidean norm of a vector $\|\boldsymbol{x}\|$ has the following properties:

1. Positivity: $\|\boldsymbol{x}\| > 0$, $\|\boldsymbol{x}\| = \mathbf{0}$ iff $\boldsymbol{x} = \mathbf{0}$.
2. Homogeneity: $\|r\boldsymbol{x}\| = |r| \cdot \|\boldsymbol{x}\|$, $r \in \mathbb{R}$.
3. Triangle inequality: $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$.

三角不等式可用 Cauchy-Schwarz 不等式证明:

$$\begin{aligned} \|\boldsymbol{x} + \boldsymbol{y}\|^2 & \xlongequal{def} \|\boldsymbol{x}\|^2 + 2\langle \boldsymbol{x}, \boldsymbol{y} \rangle + \|\boldsymbol{y}\|^2 \\ & \overset{C-S}{\leq} \|\boldsymbol{x}\|^2 + 2\|\boldsymbol{x}\|\|\boldsymbol{y}\| + \|\boldsymbol{y}\|^2 \\ & = (\|\boldsymbol{x}\| + \|\boldsymbol{y}\|)^2 \end{aligned}$$

2.4 Inner Products and Norms

The Euclidean norm of a vector $\|\boldsymbol{x}\|$ has the following properties:

1. Positivity: $\|\boldsymbol{x}\| > 0$, $\|\boldsymbol{x}\| = \mathbf{0}$ iff $\boldsymbol{x} = \mathbf{0}$.

2. Homogeneity: $\|r\boldsymbol{x}\| = |r| \cdot \|\boldsymbol{x}\|$, $r \in \mathbb{R}$.

3. Triangle inequality: $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$.

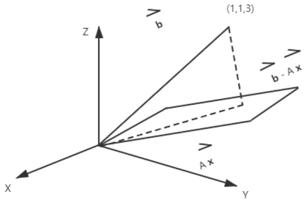
三角不等式可用 Cauchy-Schwarz 不等式证明:

$$\begin{aligned} \|\boldsymbol{x}+\boldsymbol{y}\|^2 & \xlongequal{def} \|\boldsymbol{x}\|^2 + 2\langle \boldsymbol{x}, \boldsymbol{y}\rangle + \|\boldsymbol{y}\|^2 \\ & \overset{C-S}{\leq} \|\boldsymbol{x}\|^2 + 2\|\boldsymbol{x}\|\|\boldsymbol{y}\| + \|\boldsymbol{y}\|^2 \\ & = (\|\boldsymbol{x}\| + \|\boldsymbol{y}\|)^2 \end{aligned}$$

If $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = 0$, then $\|\boldsymbol{x}+\boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2$.

投影的例子

$$\begin{cases} x+y=1 \\ x-y=1 \\ x+2y=3 \end{cases} \quad \text{归结为} \quad \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix}$$



$$A\boldsymbol{x} = \boldsymbol{b}$$
$$A^\top A\boldsymbol{x} = A^\top \boldsymbol{b}$$
$$\boldsymbol{x} = \left(A^\top A\right)^{-1} A^\top \boldsymbol{b}$$

投影的例子

给了一个向量 $\boldsymbol{b}$，和一个平面（由 $\boldsymbol{a}_1$ 和 $\boldsymbol{a}_2$ 生成），要求 P，使 $(P\boldsymbol{b})$ 在平面内，i.e., $\boldsymbol{b} = P\boldsymbol{b} + (\boldsymbol{b} - P\boldsymbol{b})$，且 $(P\boldsymbol{b})^\top \cdot (\boldsymbol{b} - P\boldsymbol{b}) = 0$.

如何求：从上次课，$\boldsymbol{b}$ 就是上次课的 $\boldsymbol{b}$，上次课的 $A = (\boldsymbol{a}_1, \boldsymbol{a}_2)$，且：$A\boldsymbol{x} = \boldsymbol{b}$ 的最优解（误差最小解）$x^* = \left(A^\top A\right)^{-1} A^\top \boldsymbol{b}$, 而 $A\boldsymbol{x}^*$ 就是 $P\boldsymbol{b}$. $\therefore A\boldsymbol{x}^* = A\left(A^\top A\right)^{-1} A^\top \boldsymbol{b}$. 由 $A\boldsymbol{x}^* = P\boldsymbol{b}$ 得 $P = A\left(A^\top A\right)^{-1} A^\top$

下面验证：

$$(P\boldsymbol{b})^\top (\boldsymbol{b} - P\boldsymbol{b}) = 0$$

验证：

$$\begin{aligned}
&(P\boldsymbol{b})^\top (\boldsymbol{b} - P\boldsymbol{b}) \\
&= (A\boldsymbol{x}^*)^\top (\boldsymbol{b} - A\boldsymbol{x}^*) \\
&= \boldsymbol{x}^{*\top} A^\top \boldsymbol{b} - \boldsymbol{x}^{*\top} A^\top A\boldsymbol{x}^* \\
&\xlongequal{x^* = \left(A^\top A\right)^{-1} A^\top b} \boldsymbol{x}^{*\top} A^\top \boldsymbol{b} - \boldsymbol{x}^{*\top} A^\top A\left(A^\top A\right)^{-1} A^\top \boldsymbol{b}
\end{aligned}$$

提问

$$\begin{bmatrix}1 & 2 & 3\\4 & 5 & 6\\7 & 8 & 9\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} = \begin{bmatrix}6\\15\\24\end{bmatrix} \text{的解集}\boldsymbol{x}=\underline{\qquad\qquad}$$

问: $[1, 1, 1]'$ 是不是某种意义下的最优解? (如要证明是 2-范数下的最优解, 需要证明任何一个解 2- 范数均 $>=$ (1 1 1) 的 2-范数)

[0 3 0] 是 0-范数下的最优解

这次课提问

习题 3.5 对 $A = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 2 & 5 & 2 & 1 \\ -1 & 1 & 0 & 3 \end{bmatrix}$ 给出的 linear transform, 找一组基, 使在这组基下 A 对应的矩阵是对角阵.

解: 该题意思是:$\boldsymbol{A}$ 代表一个变换, 将 $\boldsymbol{x}$ 变成 $\boldsymbol{y}$, 即 $Ax = \boldsymbol{y}$(这里 A 就是变换对应的矩阵, $\boldsymbol{A}$,$\boldsymbol{x}$,$\boldsymbol{y}$, 均在自然基下表示).

该题问, 如果换了一组基, 在新基下 $\boldsymbol{A}$ 对应的矩阵如记成 $\boldsymbol{B}$, 要使 $\boldsymbol{B}$ 是对角阵, 基应如何取法.

claim: 新基就是 $\boldsymbol{A}$ 的特征向量 $\left\{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}, \dots \boldsymbol{v_n}\right\}$(如果他们构成基),$\boldsymbol{B}$ 就是 $\boldsymbol{A}$ 的特征值构成的 $\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$.

理由:$\boldsymbol{x}$ 在 $\{\boldsymbol{e}_1, \cdots \boldsymbol{e}_n\}$ 下的坐标是 $\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, 在 $\{\boldsymbol{v}_1, \cdots \boldsymbol{v}_n\}$ 下坐标 $\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$.

$\boldsymbol{y}$ 在 $\{\boldsymbol{e}_1, \cdots \boldsymbol{e}_n\}$ 下的坐标是 $\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$, 在 $\{\boldsymbol{v}_1, \cdots \boldsymbol{v}_n\}$ 下坐标 $\begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$.

则 $$\{\boldsymbol{e_1}, \cdots \boldsymbol{e_n}\} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \{\boldsymbol{v}_1, \cdots \boldsymbol{v}_n\} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \Longrightarrow \boldsymbol{V} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}. \tag{1}$$

$$\{\boldsymbol{e}_1, \cdots \boldsymbol{e}_n\} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \boldsymbol{V} \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \tag{2}$$

对 (1) 两端左乘 $\boldsymbol{A}$, 得

$$\boldsymbol{A} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \boldsymbol{AV} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \boldsymbol{V} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

而 $\boldsymbol{A}\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \boldsymbol{y} = \boldsymbol{V}\begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$(由（2）式可得).

所以，$\boldsymbol{V}\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \boldsymbol{V}\begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$

$\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$

所以在 $\boldsymbol{V}$ 下,$\boldsymbol{A}$ 对应的 $\boldsymbol{B}$ 是对角阵.

Transformations

3.1 Linear Transformations

A function $\mathcal{L} : \mathbb{R}^n \to \mathbb{R}^m$ is called a linear transformation if:

1. $\mathcal{L}(a\boldsymbol{x}) = a\mathcal{L}(\boldsymbol{x})$ for every $\boldsymbol{x} \in \mathbb{R}^n$ and $a \in \mathbb{R}$.
2. $\mathcal{L}(\boldsymbol{x}_1 + \boldsymbol{x}_2) = \mathcal{L}(\boldsymbol{x}_1) + \mathcal{L}(\boldsymbol{x}_2)$ for $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^n$.

3.1 Linear Transformations

A function $\mathcal{L} : \mathbb{R}^n \to \mathbb{R}^m$ is called a linear transformation if:

1. $\mathcal{L}(a\boldsymbol{x}) = a\mathcal{L}(\boldsymbol{x})$ for every $\boldsymbol{x} \in \mathbb{R}^n$ and $a \in \mathbb{R}$.
2. $\mathcal{L}(\boldsymbol{x}_1 + \boldsymbol{x}_2) = \mathcal{L}(\boldsymbol{x}_1) + \mathcal{L}(\boldsymbol{x}_2)$ for $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^n$.

If we fix the bases for $\mathbb{R}^n$ and $\mathbb{R}^m$, then the linear transformation $\mathcal{L}$ can be represented by a matrix.

Suppose that $\boldsymbol{x} \in \mathbb{R}^n$ is a given vector, and $\boldsymbol{x}'$ is the representation of $\boldsymbol{x}$ with respect to the given basis for $\mathbb{R}^n$. If $\boldsymbol{y} = \mathcal{L}(\boldsymbol{x})$, and $\boldsymbol{y}'$ is the representation of $\boldsymbol{y}$ with respect to the given basis for $\mathbb{R}^m$. If there exists $A \in \mathbb{R}^{m \times n}$ such that the following representation holds, then

$$\boldsymbol{y}' = A\boldsymbol{x}'.$$

We call A the matrix representation of $\mathcal{L}$ with respect to the given bases for $\mathbb{R}^n$ and $\mathbb{R}^m$. In the special case where we assume the natural bases for $\mathbb{R}^n$ and $\mathbb{R}^m$, the matrix representation A satisfies

$$\mathcal{L}(\boldsymbol{x}) = A\boldsymbol{x}.$$

3.1 Linear Transformations

We call A the matrix representation of $\mathcal{L}$ with respect to the given bases for $\mathbb{R}^n$ and $\mathbb{R}^m$.

In the special case where we assume the natural bases for $\mathbb{R}^n$ and $\mathbb{R}^m$, the matrix representation A satisfies

$$\mathcal{L}(\boldsymbol{x}) = A\boldsymbol{x}.$$

3.1 Linear Transformation

这本书上用以下记号 (同华师大大一线性代数教科书不同)

Let $\{\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_n\}$ and $\{\boldsymbol{e}'_1, \boldsymbol{e}'_2, ..., \boldsymbol{e}'_n\}$ be two bases for $\mathbb{R}^n$. Define the matrix

$$T = [\boldsymbol{e}'_1, \boldsymbol{e}'_2, ..., \boldsymbol{e}'_n]^{-1}[\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_n].$$

We call T the transformation matrix from $\{\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_n\}$ to $\{\boldsymbol{e}'_1, \boldsymbol{e}'_2, ..., \boldsymbol{e}'_n\}$. It is clear that

$$[\boldsymbol{e}'_1, \boldsymbol{e}'_2, ..., \boldsymbol{e}'_n]T = [\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_n];$$

i.e., T 的第 i 列是 $\boldsymbol{e}_i$ 在基 $\{\boldsymbol{e}'_1, \cdots, \boldsymbol{e}'_n\}$ 下的坐标.

3.1 Linear Transformation

这本书上用以下记号 (同华师大大一线性代数教科书不同)

Let $\{\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_n\}$ and $\{\boldsymbol{e}'_1, \boldsymbol{e}'_2, ..., \boldsymbol{e}'_n\}$ be two bases for $\mathbb{R}^n$. Define the matrix

$$T = [\boldsymbol{e}'_1, \boldsymbol{e}'_2, ..., \boldsymbol{e}'_n]^{-1}[\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_n].$$

We call T the transformation matrix from $\{\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_n\}$ to $\{\boldsymbol{e}'_1, \boldsymbol{e}'_2, ..., \boldsymbol{e}'_n\}$. It is clear that

$$[\boldsymbol{e}'_1, \boldsymbol{e}'_2, ..., \boldsymbol{e}'_n]T = [\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_n];$$

i.e., T 的第 i 列是 $\boldsymbol{e}_i$ 在基 $\{\boldsymbol{e}'_1, \cdots, \boldsymbol{e}'_n\}$ 下的坐标.

我们大一教科书

$$(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \cdots, \boldsymbol{\eta}_n) = (\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \cdots, \boldsymbol{\varepsilon}_n)C,$$

我们称矩阵 C 为由基 $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \cdots, \boldsymbol{\varepsilon}_n$ 到基 $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \cdots, \boldsymbol{\eta}_n$ 的过渡矩阵.

Video 10 结束

3.1 Linear Transformations

对 $\mathbb{R}^n$ 中的一个向量 $\boldsymbol{v}$, 如 $\boldsymbol{x}$ 为其在 $\{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n\}$ 下的坐标,
$\boldsymbol{x}'$ 为其在 $\{\boldsymbol{e}'_1, \cdots, \boldsymbol{e}'_n\}$ 下的坐标,

定理 7.4 设 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n$ 和 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n$ 是 n 维线性空间 V 的两组基，V 中的线性变换 $\mathscr{A}$ 在这两组基下的矩阵分别为 A 和 B，且从 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n$ 到 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n$ 的过渡矩阵为 C，那么

$$B=C^{-1}AC. \tag{7.7}$$

证 据定理的条件可知

$$\mathscr{A}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)A,$$
$$\mathscr{A}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)=(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)B,$$
$$(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)C,$$

于是由 $\mathscr{A}$ 的线性性质

$$\begin{aligned}\mathscr{A}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)&=\mathscr{A}[(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)C]=[\mathscr{A}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)]C\\&=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)AC=(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)C^{-1}AC.\end{aligned}$$

因 $\mathscr{A}$ 在基 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n$ 下的矩阵是唯一的，故有

$$B=C^{-1}AC.$$

3.2 Eigenvalues and Eigenvectors

Let A be an $n \times n$ real square matrix. A scalar λ (possibly complex) and a nonzero vector $\boldsymbol{v}$ satisfying the equation $A\boldsymbol{v} = \lambda\boldsymbol{v}$ are said to be, respectively, an eigenvalue and an eigenvector of A.

3.2 Eigenvalues and Eigenvectors

Let A be an $n \times n$ real square matrix. A scalar λ (possibly complex) and a nonzero vector $\boldsymbol{v}$ satisfying the equation $A\boldsymbol{v} = \lambda\boldsymbol{v}$ are said to be, respectively, an eigenvalue and an eigenvector of A.

For λ to be an eigenvalue it is necessary and sufficient for the matrix $\lambda I - A$ to be singular; that is, $\det[\lambda I - A] = 0$, where I is the $n \times n$ identity matrix. This leads to an nth-order polynomial equation

$$\det[\lambda I - A] = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0 = 0$$

$\det[\lambda I - A]$ 称为 A 的特征多项式; 方程称为特征方程

方程有 n 个复数根 (可能有相同的). 若有 n 个不同的特征根, 则有 n 个独立的特征向量.

3.2 Eigenvalues and Eigenvectors

Theorem 3.1 Suppose that the characteristic equation $\det[\lambda I - A] = 0$ has n distinct roots $\lambda_1, \lambda_2, ..., \lambda_n$. Then, there exist n linearly independent vectors $\boldsymbol{v}_1, ..., \boldsymbol{v}_n$ such that

$$A\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i, \qquad i = 1, 2, ..., n$$

Proof. We now prove the linear independence of $\{\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n\}$. To do this, let $c_1, ..., c_n$ be scalars such that $\sum_{i=1}^{n} c_i \boldsymbol{v}_i = 0$. We show that $c_i = 0$, $i = 1, ..., n$.

令

$$Z = (\lambda_2 I - A)(\lambda_3 I - A) \cdots (\lambda_n I - A).$$

1) c_1=0. 理由

$$\begin{aligned} Z\boldsymbol{v}_n &= (\lambda_2 I - A)(\lambda_3 I - A) \cdots (\lambda_n I - A)\boldsymbol{v}_n \\ &= (\lambda_2 I - A)(\lambda_3 I - A) \cdots (\lambda_n \boldsymbol{v}_n - A\boldsymbol{v}_n) \\ &= \mathbf{0} \end{aligned}$$

同理: $Z\boldsymbol{v}_k = 0$ for $k = 2, 3, \cdots, n$.

但

$$
\begin{aligned}
Z\boldsymbol{v}_1 &= (\lambda_2 I - A)(\lambda_3 I - A)\cdots(\lambda_n I - A)\boldsymbol{v}_1 \\
&= (\lambda_2 I - A)(\lambda_3 I - A)\cdots(\lambda_{n-1}\boldsymbol{v}_1 - A\boldsymbol{v}_1)(\lambda_n - \lambda_1) \\
&\vdots \\
&= (\lambda_2 - \lambda_1)(\lambda_3 - \lambda_1)\cdots(\lambda_{n-1} - \lambda_1)(\lambda_n - \lambda_1)\boldsymbol{v}_1
\end{aligned}
$$

所以

$$
\begin{aligned}
Z(\sum_{i=1}^{n} c_i \boldsymbol{v}_i) &= \sum_{i=1}^{N} c_i Z\boldsymbol{v}_i \\
&= c_1 Z\boldsymbol{v}_1 \\
&= c_1(\lambda_2 - \lambda_1)(\lambda_3 - \lambda_1)\cdots(\lambda_n - \lambda_1)\boldsymbol{v}_1 = \mathbf{0}
\end{aligned}
$$

所以，$c_1 = 0$.

2) 同理 $c_i = 0$ for $i = 2, \cdots, n$. 所以，$\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\}$ 线性独立

Video 11 结束

若 A 的 n 个线性无关的特征向量是 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n$, 则

$$(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)^{-1} A (v_1, \cdots, \boldsymbol{v}_n) = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

Theorem 3.2 实对称阵 $(A = A^\top)$ 的所有特征值是实的.

证因为 $A\boldsymbol{x} = \lambda\boldsymbol{x}$ for $\boldsymbol{x} \neq \mathbf{0}$, 则

$$\langle A\boldsymbol{x}, \boldsymbol{x}\rangle = \langle \lambda\boldsymbol{x}, \boldsymbol{x}\rangle = \lambda\langle \boldsymbol{x}, \boldsymbol{x}\rangle.$$

另一方面,

$$\langle A\boldsymbol{x}, \boldsymbol{x}\rangle = \langle \boldsymbol{x}, A^\top\boldsymbol{x}\rangle = \langle \boldsymbol{x}, A\boldsymbol{x}\rangle = \langle \boldsymbol{x}, \lambda\boldsymbol{x}\rangle = \bar{\lambda}\langle \boldsymbol{x}, \boldsymbol{x}\rangle.$$

因为 $\langle \boldsymbol{x}, \boldsymbol{x}\rangle$ 是实的, 所以 > 0. 因此 $\lambda = \bar{\lambda}$.

3.2 Eigenvalues and Eigenvectors

Theorem 3.3 Any real symmetric $n \times n$ matrix has a set of n eigenvectors that are mutually orthogonal.

证这里只证 n 个特征值不同的情形.

Suppose that $A\boldsymbol{v}_1 = \lambda_1 \boldsymbol{v}_1$, $A\boldsymbol{v}_2 = \lambda_2 \boldsymbol{v}_2$, where $\lambda_1 \neq \lambda_2$. Then

$$\langle A\boldsymbol{v}_1, \boldsymbol{v}_2 \rangle = \langle \lambda_1 \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle = \lambda_1 \langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle.$$

因为 $A = A^\top$,

$$\langle A\boldsymbol{v}_1, \boldsymbol{v}_2 \rangle = \langle \boldsymbol{v}_1, A^\top \boldsymbol{v}_2 \rangle = \langle \boldsymbol{v}_1, A\boldsymbol{v}_2 \rangle = \lambda_2 \langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle.$$

所以

$$\lambda_1 \langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle = \lambda_2 \langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle.$$

因为 $\lambda_1 \neq \lambda_2$, **所以** $\langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle = 0$. □

正交阵

Video 12 结束

3.3 Orthogonal Projections

$\mathcal{V}$ is a subspace of $\mathbb{R}^n$ if $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{V} \Rightarrow \alpha\boldsymbol{x}_1 + \beta\boldsymbol{x}_2 \in \mathcal{V}$ for all $\alpha, \beta \in \mathbb{R}$.

Furthermore, the dimension of a subspace $\mathcal{V}$ is equal to the maximum number of linearly independent vectors in $\mathcal{V}$. If $\mathcal{V}$ is a subspace of $\mathbb{R}^n$, then the orthogonal complement of $\mathcal{V}$, denoted $\mathcal{V}^\perp$, consists of all vectors that are orthogonal to every vector in $\mathcal{V}$. Thus,

$$\mathcal{V}^\perp = \{\boldsymbol{x} : \boldsymbol{v}^\top \boldsymbol{x} = 0 \text{ for all } \boldsymbol{v} \in \mathcal{V}\}.$$

$\mathcal{V}$ and $\mathcal{V}^\perp$ span **成** $\mathbb{R}^n$, i.e., $\forall \boldsymbol{x} \in \mathbb{R}^n$ **可唯一表为**

$$\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2,$$

where $\boldsymbol{x}_1 \in \mathcal{V}$ and $\boldsymbol{x}_2 \in \mathcal{V}^\perp$. **正交分解**. $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are orthogonal projections of $\boldsymbol{x}$ onto the subspaces $\mathcal{V}$ and $\mathcal{V}^\perp$, respectively. $\mathbb{R}^n = \mathcal{V} \oplus \mathcal{V}^\perp$.

A linear transformation P is an *orthogonal projector* onto $\mathcal{V}$ if for all $\boldsymbol{x} \in \mathbb{R}^n$, we have $P\boldsymbol{x} \in \mathcal{V}$ and $\boldsymbol{x} - P\boldsymbol{x} \in \mathcal{V}^\perp$.

3.3 Orthogonal Projections

Let $A \in \mathbb{R}^{m \times n}$. Let the range, or image, of A be denoted

$$\mathcal{R}(A) = \{A\boldsymbol{x} : \boldsymbol{x} \in \mathbb{R}^n\},$$

and the nullspace, or kernel, of A be denoted

$$\mathcal{N}(A) = \{\boldsymbol{x} \in \mathbb{R}^n : A\boldsymbol{x} = \mathbf{0}\}.$$

Note that $\mathcal{R}(A)$ and $\mathcal{N}(A)$ are subspaces.

Theorem 3.4 Let A be a given matrix. Then, 1) $\mathcal{R}(A)^{\perp} = \mathcal{N}(A^{\top})$, and 2) $\mathcal{N}(A)^{\perp} = \mathcal{R}(A^{\top})$.

Proof. 1) Suppose that $\boldsymbol{x} \in \mathcal{R}(A)^{\perp}$. Then, $\boldsymbol{y}^{\top}(A^{\top}\boldsymbol{x}) = (A\boldsymbol{y})^{\top}\boldsymbol{x} = 0$ for all $\boldsymbol{y}$, so that $A^{\top}\boldsymbol{x} = \mathbf{0}$. Hence, $\boldsymbol{x} \in \mathcal{N}(A^{\top})$. This implies that $\mathcal{R}(A)^{\top} \subset \mathcal{N}(A^{\top})$.

If now $\boldsymbol{x} \in \mathcal{N}(A^{\top})$, then $(A\boldsymbol{y})^{\top}\boldsymbol{x} = \boldsymbol{y}^{\top}(A^{\top}\boldsymbol{x}) = 0$ for all $\boldsymbol{y}$, so that $\boldsymbol{x} \in \mathcal{R}(A)^{\perp}$, and consequently, $\mathcal{N}(A^{\top}) \subset \mathcal{R}(A)^{\perp}$. Thus, $\mathcal{R}(A)^{\perp} = \mathcal{N}(A^{\top})$.

2) **利用** 1), and $(\mathcal{V}^{\perp})^{\perp} = \mathcal{V}$.

Video 13 结束

3.3 Orthogonal Projections

Theorem 3.5 A matrix P is an orthogonal projector [onto the subspace $\mathcal{V} = \mathcal{R}(P)$] $\Longleftrightarrow$ $P^2 = P = P^\top$.

证: 利用 1) $\boldsymbol{x} = P\boldsymbol{x} + (\boldsymbol{x} - P\boldsymbol{x})$
2) $\mathcal{R}(P)^\perp = \mathcal{N}(P^\top)$

$\Rightarrow$ 由 1) $\mathcal{R}(I-P) \subset \mathcal{R}(P)^\perp$, 由 2) $\mathcal{R}(P)^\perp = \mathcal{N}(P^\top)$, 所以 $\mathcal{R}(I-p) \subset N(P^\top)$. 所以 $P^\top(I-P)\boldsymbol{y} = 0$ for all $\boldsymbol{y}$, 所以 $P^\top(I-P) = \boldsymbol{O}$. 所以 $P^\top = P^\top P$, i.e., $P = P^\top = P^2$.

$\Leftarrow$ 若 $P = P^\top = P^2$, 对任意 $\boldsymbol{x}$, 我们有 $(P\boldsymbol{y})^\top(I-P)\boldsymbol{x} = \boldsymbol{y}^\top P^\top(I-P)\boldsymbol{x} = \boldsymbol{y}^T P(I-P)\boldsymbol{x} = 0$ for any $\boldsymbol{y}$. Thus, $(I-P)\boldsymbol{x} \in \mathcal{R}(P)^\perp$, 这即意味着 P 是 orthogonal projector. $\square$

Video 14 结束

Transformations

3.4 Quadratic Forms

A quadratic form $f : \mathbb{R}^n \to \mathbb{R}$ is a function

$$f(\boldsymbol{x}) = \boldsymbol{x}^\top Q \boldsymbol{x},$$

where Q is an $n \times n$ real matrix.

假设 Q **是对称的**.

称 quadratic form $\boldsymbol{x}^\top Q \boldsymbol{x}$ 是 positive definite, if $\boldsymbol{x}^\top Q \boldsymbol{x} > 0$ for all nonzero $\boldsymbol{x}$. Semidefinite if $\boldsymbol{x}^\top Q \boldsymbol{x} \geq 0$ for all nonzero $\boldsymbol{x}$.

Negative definite, negative semidefinite.

Video 15 结束