矩阵计算

李宇峰
liyf@nju.edu.cn
人工智能学院

5. 矩阵的特征值

- 矩阵的特征值是n阶矩阵的最重要的数值特征之一。而且也是一个有重要应用背景的数值特征。

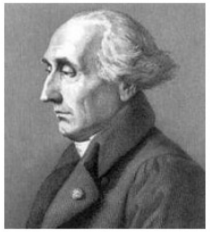
- 线性变换T的输出向量与输入向量只相差一个比例因子λ，也就是

$$Tu = \lambda u,\quad u \neq 0 \qquad (5.1)$$

即称标量λ和向量u分别是线性变换T的特征值(eigenvalue)和特征向量(eigenvector)。

5. 矩阵的特征值

- Arose in the study of *quadratic forms and differential equations* (18th century)
  - Euler
  - Lagrange
  - Cauchy


- Hilbert was the first to use the German word “*eigen*” to denote eigenvalues and eigenvectors in 1904

5. 矩阵的特征值

- 当线性变换T在给定基底下的矩阵表示是一个$n \times n$的矩阵时，确定数λ和非零向量u使得

$$Au = \lambda u \qquad (5.2)$$

这个问题就是矩阵A的特征值问题，数λ就是A的一个特征值，非零向量u就是A的对应于λ的一个特征向量。

注意：特征向量不是唯一的(αu也是特征向量)。

- 由特征值的定义，A的特征值问题可以表示为，确定非零向量$u \neq 0$，使得

$$(A - \lambda I)u = 0 \quad (5.3)$$

5. 矩阵的特征值

根据线性代数知识，(5.3)式存在有非零解$u \neq 0$的充要条件是矩阵$A - \lambda I$的行列式等于零，即

$$\det(A - \lambda I) = 0 \qquad (5.4)$$

显然，只要(5.4)式中有一个特征值为零，则就有$\det(A) = 0$.这表明矩阵只要有一个特征值为零，那么这个矩阵A一定是奇异矩阵。相反，非奇异矩阵一定没有特征值为零。

5. 矩阵的特征值

矩阵A的特征值常用符号$eig(A)$来表示。下面我们给出特征值的一些有用的基本性质：

(1) 若A, B都为$n\times n$矩阵，则矩阵乘积的特征值$eig(AB)=eig(BA)$；

(2) 若A为$m\times n$矩阵，B为$n\times m$矩阵，那么$eig(AB)$和$eig(BA)$有相同的非零特征值，所不同的是零特征值的重数不一样；

(3) 逆矩阵的特征值$eig(A^{-1})=1/eig(A)$；

(4) 设I_n为$n\times n$单位矩阵，c为标量，则

$$eig(I_n+cA)=1+c\cdot eig(A)$$

$$eig(A-cI_n)=eig(A)-c;$$

5. 矩阵的特征值

(5) A的相异特征值所对应的特征向量是线性无关的，就是说若向量$\lambda_i \neq \lambda_j$，向量u_i和u_j分别是对应于λ_i，λ_j的特征向量，那么u_i, u_j一定线性无关；

(6) 若A是$n \times n$实对称矩阵，则其所有特征值都是实数，其n个特征向量可以构成一个完备系，即一定存在n个互相正交的特征向量可以构成n维空间的基底。

5. 矩阵的特征值

我们这里考虑矩阵的正定性、半正定性等都是实对称矩阵，其实这些矩阵的特性都可以用其特征值来描述：

(1) 正定矩阵 $\Leftrightarrow$ 所有特征值都大于零；

(2) 半正定矩阵 $\Leftrightarrow$ 所有特征值为非负实数；

(3) 负定矩阵 $\Leftrightarrow$ 所有特征值为负的实数；

(4) 半正定矩阵 $\Leftrightarrow$ 每个特征值为非正实数；

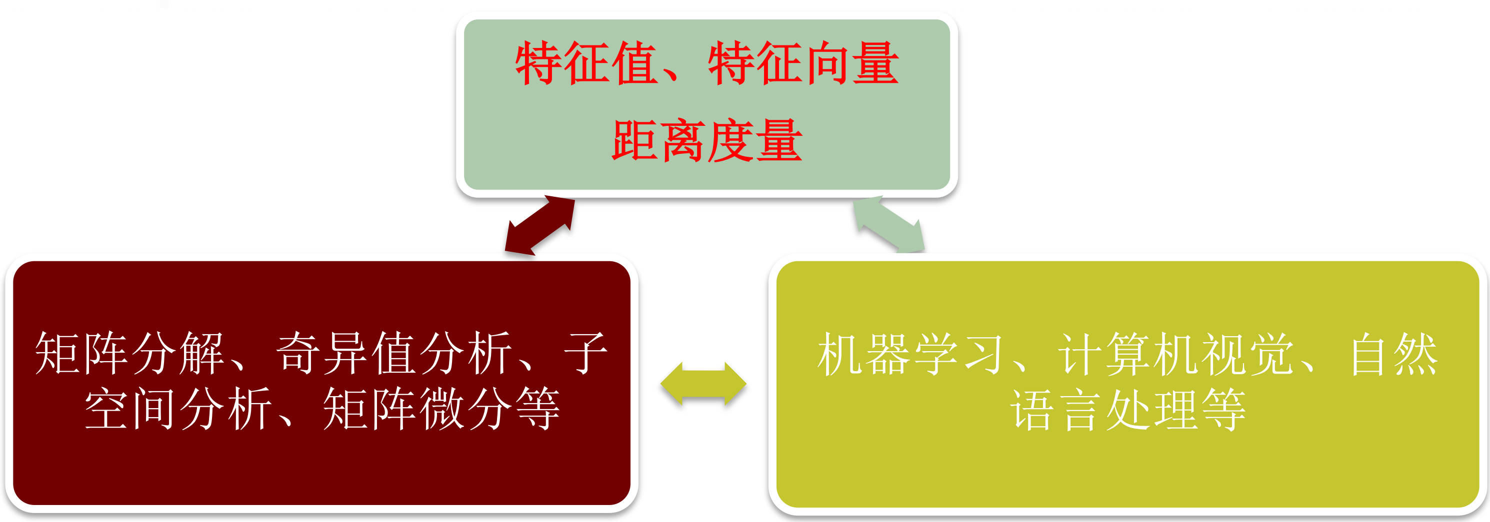
(5) 不定矩阵 $\Leftrightarrow$ 特征值有正有负的实数。

就是说，矩阵的特征值是刻画矩阵特性（奇异性、正定性等）的重要特征。

矩阵的性能指标

表 1.3.1　矩阵的性能指标

性能指标	描述的矩阵性能
二次型	矩阵的正定性与负定性
行列式	矩阵的奇异性
特征值	矩阵的奇异性、正定性和对角元素的结构
迹	矩阵对角元素之和、特征值之和
秩	行 (或列) 之间的线性无关性；线性方程组的适定性



矩阵的迹

矩阵的迹

定义 设A为一个$n\times n$矩阵，则称其对角线元素之和为A的迹(trace)，记作$tr(A)$. 即

$$tr(A) = a_{11} + a_{22} + \cdots + a_{nn}$$

矩阵A的迹对于A的特征值之和，即$tr(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$.

这说明，矩阵的迹反映矩阵所有特征值之和。

矩阵的迹

- 只有方阵才有迹，长方矩阵没有迹的概念。矩阵的迹有一些基本性质：

(1) 线性性：$tr(\alpha A+\beta B)=\alpha tr(A)+\beta tr(B)$； $tr(cA)=ctr(A)$；

$$tr(A\pm B)=tr(A)\pm tr(B);$$

(2) 设$A\in R^{m\times n}$，$B\in R^{n\times m}$，则$tr(AB)=tr(BA)$；

(3) 若A是一个$m\times n$矩阵，则$tr(A^H A)=0 \Leftrightarrow A=0_{m\times n}$(零矩阵)；

(4) $tr(xy^H)=y^H x$； $x^H Ax=tr(Axx^H)$；

矩阵的迹

- $\mathrm{trace}(\boldsymbol{A}+\boldsymbol{B})=$ \trace$\boldsymbol{A}+\mathrm{trace}\ \boldsymbol{B}$

  推导

  $\mathrm{trace}(\boldsymbol{A}+\boldsymbol{B})=\sum_{i=1}^{n}(a_{ii}+b_{ii})=\sum_{i=1}^{n}a_{ii}+\sum_{i=1}^{n}b_{ii}=\mathrm{trace}\ \boldsymbol{A}+\mathrm{trace}\ \boldsymbol{B}$

- $\mathrm{trace}(c\boldsymbol{A})=c(\mathrm{trace}\ \boldsymbol{A})$

  推导

  $\mathrm{trace}(c\boldsymbol{A})=\sum_{i=1}^{n}ca_{ii}=c\sum_{i=1}^{n}a_{ii}=c(\mathrm{trace}\ \boldsymbol{A})$

- $\mathrm{trace}(\boldsymbol{AB})=\mathrm{trace}(\boldsymbol{BA})$

  推导

  $\mathrm{trace}(\boldsymbol{AB})=\sum_{i=1}^{n}(\boldsymbol{AB})_{ii}=\sum_{i=1}^{m}\left(\sum_{j=1}^{n}a_{ij}b_{ji}\right)$

  $\mathrm{trace}(\boldsymbol{BA})=\sum_{j=1}^{n}(\boldsymbol{BA})_{jj}=\sum_{j=1}^{n}\left(\sum_{i=1}^{m}b_{ji}a_{ij}\right)$

矩阵的秩

- 前面我们讲到，向量系的线性相关和线性无关这些概念是非常重要的，它反映了向量系的一个本质特性。向量系线性相关意味着这个向量系中至少有一个向量可以由其他向量线性表示出来，一般地说这向量系中的数据有冗余。

- 一组m维n个向量组成的向量系$u_1, u_2, \ldots, u_n$线性无关，是指只有当n个系数$\alpha_1, \alpha_2, \ldots, \alpha_n$全为零时，其线性组合

$$\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n = 0$$

才成立。若向量系是线性无关的，意味着这个向量系中的任何一个向量不能由其他向量线性表示出来。

矩阵的秩

- 若由n个向量组成的向量系中，至少能确定由$r(r \le n)$个向量组成的向量系是线性无关的，而任何r+1个向量都是线性相关的，那就称这r个向量组成的向量系是极大线性无关组。对于给定的向量系，极大线性无关组并不唯一，但其所含向量的个数r必定是唯一的。

定义：设A是一个$m \times n$矩阵，如果该矩阵的列向量系的极大线性无关组所含向量的个数是r ($\le \min(m, n)$)，则称矩阵A的秩($rank$)是r，记作$rank(A)=r$.

矩阵的秩

$$A = \begin{bmatrix} 2 & 4 & 1 & 3 \\ -1 & -2 & 1 & 0 \\ 0 & 0 & 2 & 2 \\ 3 & 6 & 2 & 5 \end{bmatrix}$$

$$\boldsymbol{A} = \begin{bmatrix} -2 & -4 & 4 \\ 2 & -8 & 0 \\ 8 & 4 & -12 \end{bmatrix}$$

上述两个矩阵的秩分别为多少？

矩阵的秩

- 可以证明A的行向量系的极大线性无关组所含向量的个数也是r。所以对于给定的矩阵A它的秩是唯一的。

- 矩阵的秩是矩阵数值特征中的重要特征。设$A \in C^{m \times n}$，则A的秩$rank(A)$有一些基本性质：

(1) $rank(A)$是一个正整数；

(2) $rank(A) \leq \ min(m, n)$，就是说$rank(A)$不大于A的行数与列数；

(3) $rank(A^H) = rank(A^T) = rank(A)$；

矩阵的秩

(4) 若$c \neq 0$, 则$rank(cA) = rank(A)$;

(5) $rank(AB) \leq \min\{rank(A), rank(B)\}$ （思考题）

(6) 对A左乘或者右乘一个非奇异矩阵后，其秩保存不变。

即设$B \in C^{m\times m}, C \in C^{n\times n}$都为非奇异矩阵，则

$$rank(BA) = rank(AC) = rank(BAC) = rank(A);$$

(7) $rank(A^T A) = rank(AA^T) = rank(A)$;

$$rank(A^H A) = rank(AA^H) = rank(A).$$

矩阵的秩

(6) 对A左乘或者右乘一个非奇异矩阵后，其秩保存不变。

即设$B \in C^{m\times m}, C \in C^{n\times n}$都为非奇异矩阵，则

$$rank(BA) = rank(AC) = rank(BAC) = rank(A);$$

证明：根据性质5，$rank(AB) \leq \min\{rank(A), rank(B)\}$

可知 $rank(BA) \leq \min\{rank(A), rank(B)\} \leq rank(A)$

此外 $rank(A) = rank(B$^-1$BA) \leq rank(BA)$

从而 $rank(A) = rank(BA)$

矩阵的秩

- 根据矩阵秩的大小又可以将矩阵分成以下4类：

(1) 满秩(full rank)矩阵，秩为n的$n \times n$方阵，满秩矩阵就是非奇异矩阵；

(2) 秩亏缺(rank deficient)矩阵，$rank(A_{m\times n}) < \ min(m,n)$；

(3) 行满秩(full row rank)矩阵，$rank(A_{m\times n}) = m$, 行满秩矩阵一定是方阵或“宽”矩阵($m \leq n$)；

(4) 列满秩(full column rank)矩阵，$rank(A_{m\times n}) = n$, 列满秩矩阵一定是方阵或“高”矩阵($m \geq n$).

矩阵的秩

- 根据矩阵$A \in R^{m\times n}$的秩的大小，线性方程组可以分为下面三种类型：

(1) 适定线性方程组 若m=n且$rank(A) = n$, 即A为非奇异矩阵，则称$Ax = b$为适定方程组(well-determined systems)，适定方程组存在唯一解；

(2) 亚定线性方程组 若$m < n$，则称$Ax = b$为亚定线性方程组(under-determined systems), 这时如果$rank(A) = m$, 则方程组$Ax = b$一定有解，且有无穷多组解。

矩阵的秩

(3) 超定线性方程组 若$m>n$, 则称$Ax = b$为超定方程组(over-determined systems)。

超定方程组有可能有解、有可能无解。

- 如果线性方程组有解，则称方程组是相容线性方程组(consistent systems)，相容性条件是$rank(A) = rank(A\colon b)$, 也就是增广矩阵$(A\colon b)$的秩等于矩阵$A$的秩。

6. 内积与范数

在实际应用中，我们通常会对一个向量的长度以及两个向量之间的关系感兴趣，这些问题都涉及向量的内积和范数。

1. 向量的内积与范数

- 两个实n维向量$x=(x_1,x_2,\dots,x_n)^T$, $y=(y_1,y_2,\dots,y_n)^T$之间的内积（inner product）记为(x,y)。定义为

$$(x,y)=x_1y_1+x_2y_2+\cdots+x_ny_n=\sum_{i=1}^{n}x_iy_i=x^Ty$$

- 两个n维复向量$x=(x_1,x_2,\dots,x_n)^T$, $y=(y_1,y_2,\dots,y_n)^T$之间的内积也记为(x,y), 定义为

$$(x,y)=x_1^*y_1+x_2^*y_2+\cdots+x_n^*y_n=\sum_{i=1}^{n}x_i^*y_i=x^Hy$$

6. 内积与范数

- For a scalar function $\rho(x)$ to be a vector norm of vector x, the following three conditions must be satisfied:

  - $\rho(x) \geq 0$ and $\rho(x) = 0$ if and only if $x = 0$
  - $\rho(ax) = |a|\rho(x)$, a is any given scalar
  - For any vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, the following holds

$$\rho(x + y) \leq \rho(x) + \rho(y)$$

6. 内积与范数

- 若两个实向量之间定义了内积，我们就可以利用这个内积，来定义两个实向量夹角的余弦：

$$cos\theta = \frac{\langle x, y \rangle}{\sqrt{\langle x, x \rangle}\sqrt{\langle y, y \rangle}}$$

- 定义了内积的线性空间就是内积空间，实的内积空间称为欧几里德空间(Euclidean)。有了内积的定义以后，我们还要用到向量的长度(length) 、距离(distance)和邻域(neighborhood)等测度。

6. 内积与范数

- 向量$x=(x_1,x_2,\dots,x_n)^T\in C^n$的最常用的范数是*Euclidean*范数，或者称*L*2-范数(norm), 定义为

$$\|x\|_2=\sqrt{|x_1|^2+|x_2|^2+\cdots+|x_n|^2}=\sqrt{\langle x,x\rangle}$$

这就是常用的向量x长度的度量，用内积来表示长度，$|x_i|$表示复数x_i的模，若x_i是实数，则$|x_i|$表示实数x_i的绝对值。有这个长度的定义，前面定义实向量x,y的夹角就可以表示成

$$cos\theta=\frac{\langle x,y\rangle}{\|x\|_2\|y\|_2}$$

6. 内积与范数

- $L2$-范数还可以度量两个向量之间的距离

$$d(x,y)=\|x-y\|_2=\sqrt{|x_1-y_1|^2+|x_2-y_2|^2+\cdots+|x_n-y_n|^2}$$

以及一个向量的ε-邻域($\varepsilon>0$)

$$N_\varepsilon(x)=\{y\mid \|y-x\|_2\le\varepsilon\}$$

邻域的概念是最优化理论和方法中的一个重要概念，因为当我们讨论一种优化算法的性能时，局部最优（在某个邻域最优）比全局最优（在整个定义域最优）更方便进行比较和确定。

6. 内积与范数

- 对于向量$x=(x_1,x_2,\dots,x_n)^T$而言，综合起来有下列常用的范数：

(1) L_0范数(也称0- “范数”)

$$\|x\|_0 \triangleq \text{非零元素的个数}$$

(2) L_1范数(也称和范数或1-范数)

$$\|x\|_1 \triangleq \sum_{i=1}^{n}|x_i| = |x_1|+|x_2|+\cdots+|x_n|$$

(3) L_2范数(常称*Euclidean*范数或2-范数)

$$\|x\|_2 \triangleq \sqrt{\sum_{i=1}^{n}|x_i|^2} = \left(|x_1|^2+|x_2|^2+\cdots+|x_n|^2\right)^{\frac{1}{2}}$$

6. 内积与范数

(4) L_∞范数(也称无穷范数)

$$\|x\|_\infty \triangleq \max\{|x_1|, |x_2|, \dots, |x_n|\}$$

(5) L_p范数(也称*Holder*范数)

$$\|x\|_p \triangleq \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} \quad p \geq 1$$

6. 内积与范数

例6.2 设$x = (1, 2, -3)^T$, 试求$\|x\|_0$, $\|x\|_1$, $\|x\|_2$, $\|x\|_\infty$, $\|x\|_p$.

6. 内积与范数

例6.2 设$x=(1,2,-3)^T$, 试求$\|x\|_0, \|x\|_1, \|x\|_2, \|x\|_\infty, \|x\|_p$.

解：

$$\|x\|_0=3;$$

$$\|x\|_1 = |1| + |2| + |-3| = 6;$$

$$\|x\|_2 = (|1|^2 + |2|^2 + |-3|^2)^{\frac{1}{2}} = \sqrt{14};$$

$$\|x\|_\infty = \max\{|1|, |2|, |-3|\} = 3;$$

$$\|x\|_p = (|1|^p, |2|^p, |-3|^p)^{\frac{1}{p}}, \quad p > 1$$

6. 内积与范数

2.函数空间的内积与范数

- 若$x(t)$和$y(t)$都是闭区间$[a, b]$上的连续函数，则它们的内积：

$$\langle x(t), y(t)\rangle \triangleq \int_a^b x(t)y(t)dt$$

变量t可以是时间变量、频率变量或者空间变量。

函数$x(t)$的范数$\|x(t)\|$定义为：$\|x(t)\| \triangleq \sqrt{\int_a^b x(t)^2 dt}$

函数$x(t)$和$y(t)$的夹角的余弦是：

$$cos\theta \triangleq \frac{\langle x, y\rangle}{\sqrt{\langle x, x\rangle}\sqrt{\langle y, y\rangle}} = \frac{\int_a^b x(t)y(t)dt}{\|x(t)\| \cdot \|y(t)\|}$$

6. 内积与范数

有了内积的定义，我们就可以引入矩阵论中最重要的一个概念，就是正交性的概念。

- 如果两个非零向量x, y的内积等于零，也就是$\langle x, y\rangle = 0$，我们就称向量x与y是正交的。
- 正交性的概念可以在不同的层面上来解释：

(1) 数学定义，表明这两个向量的内积等于零；

(2) 几何解释，也称为是垂直的，表明两个向量的夹角是90°，并且一个向量到另一个向量的投影为零；

(3) 物理意义，两个向量正交意味着一个向量不含另一个向量的任何成分，不存在相互作用和干扰。

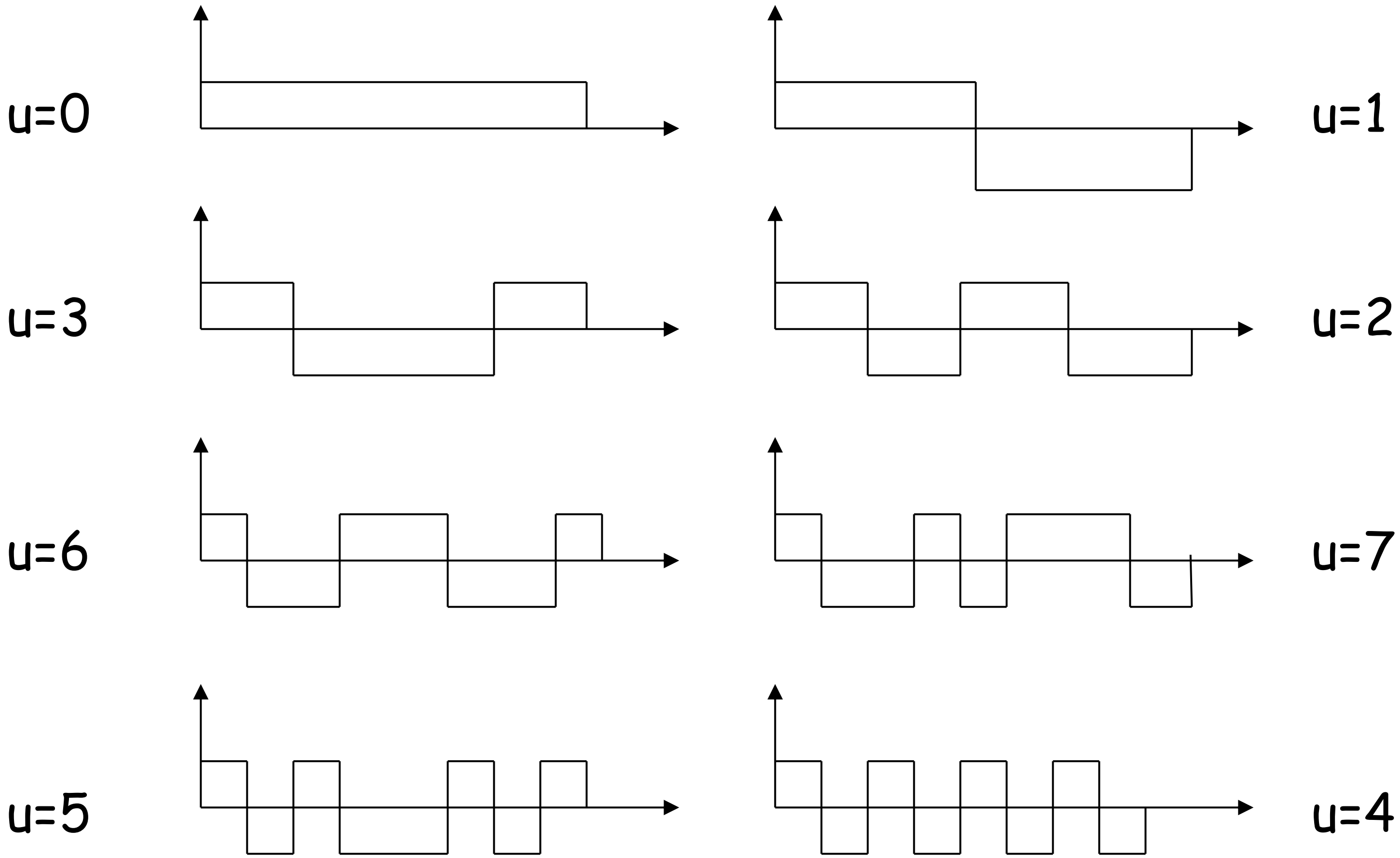
6. 内积与范数

给定一组n维向量系$x_1, x_2, \dots, x_m (m \le n)$, 如果两两正交满足：

$$\langle x_i, x_j \rangle = \begin{cases} \alpha_i > 0 & i = j \\ 0 & i \neq j \end{cases} \quad i, j = 1,2, \dots, m$$

则称向量$x_1, x_2, \dots, x_m (m \le n)$是一组正交向量系，若还有$\|x_i\| = 1,\ i = 1,2, \dots, m$, 则称该向量系是一组标准正交向量系。

正交基



矩阵范数

- A ***matrix norm*** is a function $||*||$ from the set of complex matrices into R that satisfies the following properties:

1. $||A|| \geq 0$ and $||A|| = 0 \iff A = 0$.
2. $||aA|| = |a| \cdot ||A||$ for all scalars a.
3. $||A + B|| \leq ||A|| + ||B||$ for matrices of the same size.
4. $||A \cdot B|| \leq ||A|| \cdot ||B||$.

矩阵范数

矩阵范数

将向量范数的概念推广到矩阵，作为矩阵大小的度量。

定义6.2 设$\|\cdot\|$是一种向量范数

$$\|\boldsymbol{A}\|_p=\left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^p\right)^{1/p}$$

$$\|A\|=\sup_{x\in R^n,x\neq 0}\frac{\|Ax\|}{\|x\|}=\sup_{x\in R^n,\|x\|=1}\|Ax\|$$

称之为由向量范数派生的矩阵算子范数 (induced norm)

矩阵范数

以下是三种常用的 p 范数：

(1) L_1 范数 (和范数) ($p=1$)

$$\|\boldsymbol{A}\|_1 \stackrel{\text{def}}{=} \sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|$$

(2) Frobenius 范数 ($p=2$)

$$\|\boldsymbol{A}\|_{\mathrm{F}} \stackrel{\text{def}}{=} \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2}$$

(3) 最大范数 (max norm) 即 $p=\infty$ 的 p 范数，定义为

$$\|\boldsymbol{A}\|_\infty = \max_{i=1,2,\cdots,m;\, j=1,2,\cdots,n}\{|a_{ij}|\}$$

矩阵范数

例6.4 求矩阵A的范数$\|A\|_1$， $\|A\|_\infty$， $\|A\|_F$

$$A=\begin{array}{cc} \begin{pmatrix}1&2&0\\-1&2&-1\\0&1&1\end{pmatrix} & \begin{array}{c}3\\4\\2\end{array} \\ \begin{array}{ccc}2&5&2\end{array} & \end{array}$$

矩阵范数

例6.4 求矩阵A的范数$\|A\|_1$， $\|A\|_\infty$， $\|A\|_F$

$$A=\begin{array}{cc} \begin{pmatrix} 1 & 2 & 0 \\ -1 & 2 & -1 \\ 0 & 1 & 1 \end{pmatrix} & \begin{array}{c} 3 \\ 4 \\ 2 \end{array} \\ \begin{array}{ccc} 2 & 5 & 2 \end{array} & \end{array}$$

解：

$$\|A\|_1=\sum_{j=1}^{n}\sum_{i=1}^{m}\left|a_{ij}\right| = 9;$$

$$\|A\|_\infty = \max_{1\le i\le m}\max_{1\le j\le n}\left|a_{ij}\right| = 2;$$

$$\|A\|_F=\sqrt{4+4+1+1+1+1+1}=\sqrt{13}$$

矩阵范数

以下是矩阵的内积与范数之间的关系[50]：

(1) Cauchy-Schwartz 不等式

$$|\langle \boldsymbol{A}, \boldsymbol{B}\rangle|^2 \leqslant \|\boldsymbol{A}\|^2\|\boldsymbol{B}\|^2$$

等号成立，当且仅当 $\boldsymbol{A} = c\boldsymbol{B}$，其中，$c$ 是某个复常数。

(2) Pathagoras 定理：$\langle \boldsymbol{A}, \boldsymbol{B}\rangle = 0 \ \Rightarrow \ \|\boldsymbol{A}+\boldsymbol{B}\|^2 = \|\boldsymbol{A}\|^2 + \|\boldsymbol{B}\|^2$。

对于 $\boldsymbol{A} \in \mathbb{C}^{m\times n}, \boldsymbol{x} \in \mathbb{C}^n$，有

$$\|\boldsymbol{A}\boldsymbol{x}\|_{\mathrm{F}} \leqslant \|\boldsymbol{A}\|_{\mathrm{F}}\|\boldsymbol{x}\|_{\mathrm{F}}, \text{ 即 } \|\boldsymbol{A}\boldsymbol{x}\|_2 \leqslant \|\boldsymbol{A}\|_2\|\boldsymbol{x}\|_2.$$

代数与矩阵基础

- 向量，矩阵
- 线性空间，线性子空间
  - 零空间，像空间
- 线性相关，线性无关
  - 空间的维数与基底
- 矩阵的基本运算
  - 转置，共轭转置
  - 相乘，求逆
- 矩阵的数值特征
  - 行列式，二次型，特征值，trace，rank
- 内积与范数
  - 向量范数
  - 矩阵范数