矩阵论

2024年秋学期

第五讲

2024年9月23日

稀疏表示和压缩感知
特殊矩阵
矩阵微分

稀疏表示和压缩感知

一个含有大多数零元素的向量或矩阵称为稀疏向量（sparse vector）或者稀疏矩阵（sparse matrix）

信号向量 $\boldsymbol{y} \in \mathbb{R}^m$ 最多可分解为 m 个正交基 (向量) $\boldsymbol{g}_k \in \mathbb{R}^m, k = 1, \cdots, m$，这些正交基的集合称为完备正交基 (complete orthogonal basis)。此时，信号分解

$$\boldsymbol{y} = \boldsymbol{G}\boldsymbol{c} = \sum_{i=1}^{m} c_i \boldsymbol{g}_i$$

c 非稀疏

字典或库

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} = \sum_{i=1}^{n} x_i \boldsymbol{a}_i \quad (n > m)$$

x稀疏

则 $n(> m)$ 个向量 $\boldsymbol{a}_i \in \mathbb{R}^m, i = 1, \cdots, n$ 不可能是正交基的集合。

$\boldsymbol{A}$ 的列称为原子（atom）或框架，组成的集合为过完备集合。

稀疏表示和压缩感知

欠定方程

$$y = Ax = \sum_{i=1}^{n} x_i a_i \quad (n > m)$$

1. 经典方法：求最小L_2范数解

$$\min \|x\|_2 \quad \text{subject to } Ax = y$$

最小能量解，具有唯一性，解不符合稀疏表示要求

2. 现代方法：求最小L_0范数解

$$\min \|x\|_0 \quad \text{subject to } Ax = y$$

非零元素个数

解符合稀疏表示要求，求解较复杂

存在观测误差/噪声（随机）情况下：求最小L_0范数解

$$\min \|x\|_0 \quad \text{subject to } \|Ax - y\|_2 \leqslant \varepsilon$$

估计 信号

稀疏表示和压缩感知

稀疏信号：大多数采样时刻取值等于零或者近似等于零（频域）

压缩感知：低维的采样数据向量恢复或重构高维数据向量

压缩+低速率采样

$$y = Ax$$

$$m \ll n$$

$m \times 1$　　$m \times n$　　$n \times 1$

低维感知波形　感知矩阵　待重构：高维信号向量

变换域　$x = \Phi\alpha$

***x*的K-稀疏表示**

1. **估计稀疏向量α，仅含K个非零元素**
2. **重构*x***

Hermitian矩阵

$$
\boldsymbol{R}=\begin{bmatrix} r_{11} & r_{21}^* & r_{31}^* & r_{41}^* \\ r_{21} & r_{22} & r_{32}^* & r_{31}^* \\ r_{31} & r_{32} & r_{22} & r_{21}^* \\ r_{41} & r_{31} & r_{21} & r_{11} \end{bmatrix}
$$

复共轭对称矩阵 $\mathbf{R}=\mathbf{R}^H$

反Hermitian矩阵? $\mathbf{R}=-\mathbf{R}^H$

$$\alpha \boldsymbol{A}+\beta \boldsymbol{B}$$

置换矩阵

定义 2.2.1 一个正方矩阵称为置换矩阵 (permutation matrix)，若它的每一行和每一列有一个且仅有一个非零元素 1。

置换矩阵 $\boldsymbol{P}$ 有下列性质[61]：

(1) $(\boldsymbol{P}_{m\times n})^{\mathrm{T}}=\boldsymbol{P}_{n\times m}$。

(2) $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}=\boldsymbol{P}\boldsymbol{P}^{\mathrm{T}}=\boldsymbol{I}$，这说明置换矩阵是正交矩阵。

(3) $\boldsymbol{P}^{\mathrm{T}}=\boldsymbol{P}^{-1}$。

$$\boldsymbol{A}=\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \\ a_{51} & a_{52} & a_{53} & a_{54} \end{bmatrix}$$

$$\boldsymbol{P}_4=\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix},\ \boldsymbol{P}_5=\begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\boldsymbol{P}_5\boldsymbol{A}=\begin{bmatrix} a_{51} & a_{52} & a_{53} & a_{54} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{41} & a_{42} & a_{43} & a_{44} \\ a_{11} & a_{12} & a_{13} & a_{14} \end{bmatrix},$$

对行进行重新排列

$$\boldsymbol{A}\boldsymbol{P}_4=\begin{bmatrix} a_{13} & a_{12} & a_{14} & a_{11} \\ a_{23} & a_{22} & a_{24} & a_{21} \\ a_{33} & a_{32} & a_{34} & a_{31} \\ a_{43} & a_{42} & a_{44} & a_{41} \\ a_{53} & a_{52} & a_{54} & a_{51} \end{bmatrix}$$

对列进行重新排列

互换矩阵

$$J = \begin{bmatrix} 0 & & & 1 \\ & & 1 & \\ & \cdot^{\cdot^{\cdot}} & & \\ 1 & & & 0 \end{bmatrix}$$

左乘：矩阵的行顺序反转

$$J_m A = \begin{bmatrix} a_{m1} & a_{m2} & \cdots & a_{mn} \\ \vdots & \vdots & \vdots & \vdots \\ a_{21} & a_{22} & \cdots & a_{2n} \\ a_{11} & a_{12} & \cdots & a_{1n} \end{bmatrix}$$

flipud

右乘：矩阵的列顺序反转

$$AJ_n = \begin{bmatrix} a_{1n} & \cdots & a_{12} & a_{11} \\ a_{2n} & \cdots & a_{22} & a_{21} \\ \vdots & \vdots & \vdots & \vdots \\ a_{mn} & \cdots & a_{m2} & a_{m1} \end{bmatrix}$$

fliplr

广义置换矩阵

定义 2.2.2 一个正方矩阵称为广义置换矩阵 (generalized permutation matrix)，简称 g 矩阵，若其每行和每列有一个并且仅有一个非零元素。

$$\boldsymbol{G}=\begin{bmatrix}0&0&0&0&\alpha\\0&0&\beta&0&0\\0&\gamma&0&0&0\\0&0&0&\lambda&0\\\rho&0&0&0&0\end{bmatrix}=\begin{bmatrix}0&0&0&0&1\\0&0&1&0&0\\0&1&0&0&0\\0&0&0&1&0\\1&0&0&0&0\end{bmatrix}\begin{bmatrix}\rho&&&&0\\&\gamma&&&\\&&\beta&&\\&&&\lambda&\\0&&&&\alpha\end{bmatrix}$$

g矩阵　　　　置换矩阵　　非奇异对角阵

观测数据模型 $\boldsymbol{x}(t)=\boldsymbol{A}\boldsymbol{s}(t)=\sum_{i=1}^{n}\boldsymbol{a}_i s_i(t)$

对信号进行恢复 $\boldsymbol{s}(t)=\boldsymbol{A}^{\dagger}\boldsymbol{x}(t)$　　$\boldsymbol{A}^{\dagger}=(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}$ **广义逆矩阵**

两种不确定性：1）累加导致信号顺序不确定

2）信号幅度不确定 $\boldsymbol{x}(t)=\sum_{i=1}^{n}\frac{\boldsymbol{a}_i}{\alpha_i}\alpha_i s_i(t)$

这两种不确定性可以通过广义置换矩阵进行描述　　*x(t)=GAs(t)*

正交矩阵与酉矩阵

定理 2.3.2[238] 若 $\boldsymbol{U} \in \mathbb{C}^{n\times n}$，则下列叙述等价：

(1) $\boldsymbol{U}$ 是酉矩阵；

(2) $\boldsymbol{U}$ 是非奇异的，并且 $\boldsymbol{U}^{\mathrm{H}} = \boldsymbol{U}^{-1}$；

(3) $\boldsymbol{U}\boldsymbol{U}^{\mathrm{H}} = \boldsymbol{U}^{\mathrm{H}}\boldsymbol{U} = \boldsymbol{I}$；

(4) $\boldsymbol{U}^{\mathrm{H}}$ 是酉矩阵；

(5) $\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n]$ 的列组成标准正交组，即

$$\boldsymbol{u}_i^{\mathrm{H}}\boldsymbol{u}_j = \delta(i-j) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

(6) $\boldsymbol{U}$ 的行组成标准正交组；

(7) 对所有 $\boldsymbol{x} \in \mathbb{C}^n$ 而言，$\boldsymbol{y} = \boldsymbol{U}\boldsymbol{x}$ 的 Euclidean 长度与 $\boldsymbol{x}$ 的 Euclidean 长度相同，即 $\boldsymbol{y}^{\mathrm{H}}\boldsymbol{y} = \boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}$。

正交矩阵与酉矩阵

酉变换

(1) 向量内积在酉变换下是不变的，即

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{A}\boldsymbol{y} \rangle$$

这是因为 $\langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{A}\boldsymbol{y} \rangle = (\boldsymbol{A}\boldsymbol{x})^{\mathrm{H}}\boldsymbol{A}\boldsymbol{y} = \boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{y} = \boldsymbol{x}^{\mathrm{H}}\boldsymbol{y} = \langle \boldsymbol{x}, \boldsymbol{y} \rangle$。

(2) 向量范数在酉变换下是不变的，即

$$\|\boldsymbol{A}\boldsymbol{x}\|^2 = \|\boldsymbol{x}\|^2$$

因为 $\|\boldsymbol{A}\boldsymbol{x}\|^2 = \langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{A}\boldsymbol{x} \rangle = \langle \boldsymbol{x}, \boldsymbol{x} \rangle = \|\boldsymbol{x}\|^2$。

(3) 两个向量的夹角在酉变换下也是不变的，即

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

实向量/矩阵与复向量/矩阵的性质比较

表 2.3.1　实向量、实矩阵与复向量、复矩阵的性质比较

实向量、实矩阵	复向量、复矩阵
范数 $\|\boldsymbol{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$	范数 $\|\boldsymbol{x}\| = \sqrt{\vert x_1\vert^2 + \vert x_2\vert^2 + \cdots + \vert x_n\vert^2}$
转置 $\boldsymbol{A}^{\mathrm{T}} = [a_{ji}]$,　$(\boldsymbol{AB})^{\mathrm{T}} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}$	共轭转置 $\boldsymbol{A}^{\mathrm{H}} = [a_{ji}^*]$,　$(\boldsymbol{AB})^{\mathrm{H}} = \boldsymbol{B}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}$
内积 $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$	内积 $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^{\mathrm{H}}\boldsymbol{y}$
正交性 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = 0$	正交性 $\boldsymbol{x}^{\mathrm{H}}\boldsymbol{y} = 0$
对称矩阵 $\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}$	Hermitian 矩阵 $\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{A}$
正交矩阵 $\boldsymbol{Q}^{\mathrm{T}} = \boldsymbol{Q}^{-1}$	酉矩阵 $\boldsymbol{U}^{\mathrm{H}} = \boldsymbol{U}^{-1}$
特征值分解 $\boldsymbol{A} = \boldsymbol{Q\Sigma Q}^{\mathrm{T}} = \boldsymbol{Q\Sigma Q}^{-1}$	特征值分解 $\boldsymbol{A} = \boldsymbol{U\Sigma U}^{\mathrm{H}} = \boldsymbol{U\Sigma U}^{-1}$
范数的正交不变性 $\|\boldsymbol{Qx}\| = \|\boldsymbol{x}\|$	范数的酉不变性 $\|\boldsymbol{Ux}\| = \|\boldsymbol{x}\|$
内积的正交不变性 $\langle \boldsymbol{Qx}, \boldsymbol{Qy} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle$	内积的酉不变性 $\langle \boldsymbol{Ux}, \boldsymbol{Uy} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle$

三角矩阵

$$\boldsymbol{U} = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u_{nn} \end{bmatrix} \qquad \boldsymbol{L} = \begin{bmatrix} l_{11} & & & 0 \\ l_{12} & l_{22} & & \\ \vdots & \vdots & \ddots & \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{bmatrix}$$

(1) 下三角矩阵，若 $a_{ij} = 0\,(i < j)$;

(2) 严格下三角矩阵，若 $a_{ij} = 0\ (i \leqslant j)$;

(3) 单位下三角矩阵，若 $a_{ij} = 0\ (i < j),\ a_{ii} = 1\ (\forall\ i)$;

(4) 上三角矩阵，若 $a_{ij} = 0\ (i > j)$;

(5) 严格上三角矩阵，若 $a_{ij} = 0\ (i \geqslant j)$;

(6) 单位上三角矩阵，若 $a_{ij} = 0\ (i > j),\ a_{ii} = 1\ (\forall\ i)$。

相似矩阵

定义 2.6.1 (相似矩阵与相似变换) 矩阵 $\boldsymbol{B} \in \mathbb{C}^{n\times n}$ 称为矩阵 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ 的相似矩阵，若存在一非奇异矩阵 $\boldsymbol{S} \in \mathbb{C}^{n\times n}$ 使得 $\boldsymbol{B} = \boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}$。此时，线性变换 $\boldsymbol{A} \mapsto \boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}$ 称为矩阵 $\boldsymbol{A}$ 的相似变换。关系"$\boldsymbol{B}$ 相似于 $\boldsymbol{A}$"常简写作 $\boldsymbol{B} \sim \boldsymbol{A}$。

特征值相同
特征向量存在线性变换关系

$$\det(\boldsymbol{B}) = \det(\boldsymbol{A}) \qquad \mathrm{tr}(\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A})$$

Vandermonde矩阵

矩阵每行（或列）的元素组成一个等比序列

$$A = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & \vdots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{bmatrix} \qquad A = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{bmatrix}$$

性质：第二行/第二列元素各不相同时，矩阵非奇异

Fourier矩阵

Fourier矩阵是一种特殊结构的Vandermonde

DFT

$$X_k = \sum_{n=0}^{N-1} x_n \mathrm{e}^{-\mathrm{j}\,2\pi nk/N} = \sum_{n=0}^{N-1} x_n w^{nk}, \quad k = 0, 1, \cdots, N-1$$

$$\begin{bmatrix} X_0 \\ X_1 \\ \vdots \\ X_{N-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & w & \cdots & w^{N-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & w^{N-1} & \cdots & w^{(N-1)(N-1)} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{bmatrix}$$

$$\hat{\boldsymbol{x}} = \boldsymbol{F}\boldsymbol{x}$$

$$\boldsymbol{F} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & w & \cdots & w^{N-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & w^{N-1} & \cdots & w^{(N-1)(N-1)} \end{bmatrix}, \quad w = \mathrm{e}^{-\mathrm{j}\,2\pi/N}$$ **称为Fourier矩阵**

Fourier矩阵

$$\boldsymbol{F}^{\mathrm{H}}\boldsymbol{F} = \boldsymbol{F}\boldsymbol{F}^{\mathrm{H}} = N\boldsymbol{I}$$

$$\boldsymbol{F}^{-1} = \frac{1}{N}\boldsymbol{F}^{\mathrm{H}} = \frac{1}{N}\boldsymbol{F}^{*}$$

$$\boldsymbol{x} = \boldsymbol{F}^{-1}\hat{\boldsymbol{x}} = \frac{1}{N}\boldsymbol{F}^{*}\hat{\boldsymbol{x}}$$

$$\begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{bmatrix} = \frac{1}{N}\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & w^{*} & \cdots & (w^{N-1})^{*} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (w^{N-1})^{*} & \cdots & (w^{(N-1)(N-1)})^{*} \end{bmatrix}\begin{bmatrix} X_0 \\ X_1 \\ \vdots \\ X_{N-1} \end{bmatrix}$$

DFT 反变换

$$x_n = \frac{1}{N}\sum_{k=0}^{N-1} X_k \mathrm{e}^{\mathrm{j}2\pi nk/N}, \quad n = 0, 1, \cdots, N-1$$

Hadamard矩阵

定义 2.9.1 $\boldsymbol{H}_n \in \mathbb{R}^{n\times n}$ 称为 Hadamard 矩阵，若它的所有元素取 +1 或者 −1，且

$$\boldsymbol{H}_n\boldsymbol{H}_n^{\mathrm{T}} = \boldsymbol{H}_n^{\mathrm{T}}\boldsymbol{H}_n = n\boldsymbol{I}_n \tag{2.9.1}$$

第 1 列和第 1 行的所有元素为 +1 的 Hadamard 矩阵 称为规范化 Hadamard 矩阵。

只有当 $n = 2$ 或者 n 是 4 的整数倍时，Hadamard 矩阵才存在。

容易验证 $\frac{1}{\sqrt{n}}\boldsymbol{H}_n$ 为标准正交矩阵。

$n\times n$ Hadamard 矩阵 $\boldsymbol{H}_n$ 的行列式 $\det(\boldsymbol{H}_n) = n^{n/2}$。

定理 2.9.1 令 $n = 2^k$，$k = 1, 2, \cdots$，则规范化的标准正交 Hadamard 矩阵具有通用构造公式

$$\bar{\boldsymbol{H}}_n = \frac{1}{\sqrt{2}}\begin{bmatrix} \bar{\boldsymbol{H}}_{n/2} & \bar{\boldsymbol{H}}_{n/2} \\ \bar{\boldsymbol{H}}_{n/2} & -\bar{\boldsymbol{H}}_{n/2} \end{bmatrix} \tag{2.9.2}$$

其中

$$\bar{\boldsymbol{H}}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{2.9.3}$$

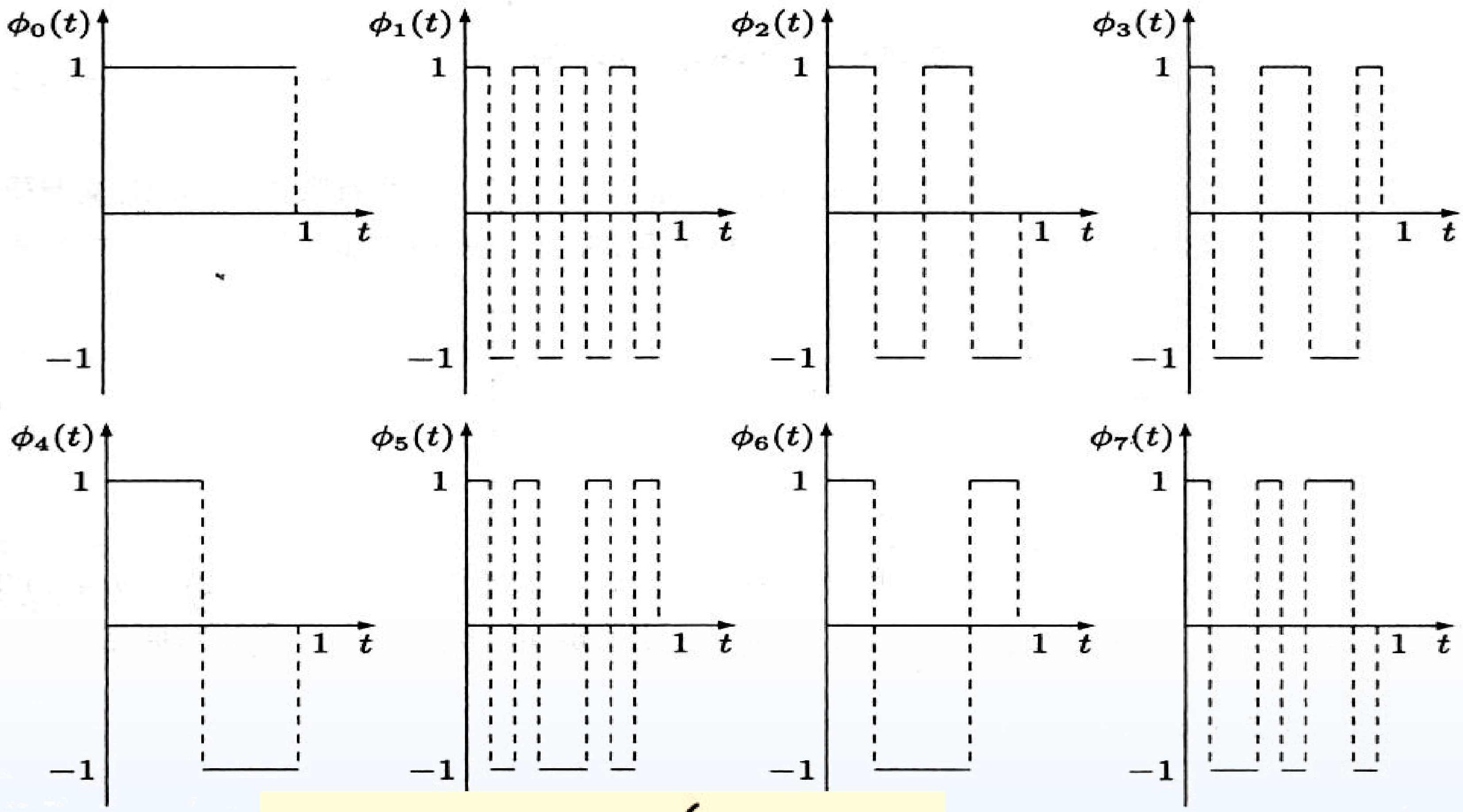
Hadamard矩阵

例 2.9.1 当 $n=2^3=8$ 时，Hadamard 矩阵

$$\boldsymbol{H}_8=\begin{bmatrix}1&1\\1&-1\end{bmatrix}\otimes\begin{bmatrix}1&1\\1&-1\end{bmatrix}\otimes\begin{bmatrix}1&1\\1&-1\end{bmatrix}$$

$$=\begin{bmatrix}1&1&1&1&1&1&1&1\\1&-1&1&-1&1&-1&1&-1\\1&1&-1&-1&1&1&-1&-1\\1&-1&-1&1&1&-1&-1&1\\1&1&1&1&-1&-1&-1&-1\\1&-1&1&-1&-1&1&-1&1\\1&1&-1&-1&-1&-1&1&1\\1&-1&-1&1&-1&1&1&-1\end{bmatrix}$$

Hadamard矩阵



$$\int_0^1 \phi_i(t)\phi_j(t)\mathrm{d}t = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

Toeplitz矩阵

任何一条对角线的元素取相同值

$$\boldsymbol{A}=\begin{bmatrix} a_0 & a_{-1} & a_{-2} & \cdots & a_{-n} \\ a_1 & a_0 & a_{-1} & \cdots & a_{-n+1} \\ a_2 & a_1 & a_0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & a_{-1} \\ a_n & a_{n-1} & \cdots & a_1 & a_0 \end{bmatrix} = [a_{i-j}]_{i,j=0}^{n}$$

对称 Toeplitz 矩阵 $\boldsymbol{A}=[a_{|i-j|}]_{i,j=0}^{n}$

若一个复 Toeplitz 矩阵的元素满足复共轭对称关系 $a_{-i}=a_i^*$

称为 Hermitian Toeplitz 矩阵

$$\boldsymbol{A}=\begin{bmatrix} a_0 & a_1^* & a_2^* & \cdots & a_n^* \\ a_1 & a_0 & a_1^* & \cdots & a_{n-1}^* \\ a_2 & a_1 & a_0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & a_1^* \\ a_n & a_{n-1} & \cdots & a_1 & a_0 \end{bmatrix}$$

Hankel矩阵

正方矩阵 $\boldsymbol{A} \in \mathbb{C}^{(n+1)\times(n+1)}$ 称为 Hankel 矩阵，若

$$\boldsymbol{A} = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_n \\ a_1 & a_2 & a_3 & \cdots & a_{n+1} \\ a_2 & a_3 & a_4 & \cdots & a_{n+2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_n & a_{n+1} & a_{n+2} & \cdots & a_{2n} \end{bmatrix}$$

矩阵微分

矩阵微分是矩阵分析和多变量微积分中的一个重要工具，它在理论研究和实际应用中都有广泛的用途。通过矩阵微分，可以高效地处理和解决涉及多维数据的问题。

矩阵微分的主要用途：

1. 优化问题：在机器学习、统计学和工程优化等领域，经常需要最小化或最大化某个多变量函数（通常表达为矩阵形式）。矩阵微分提供了一种寻找这些函数最优点（极值点）的方法。例如，通过求解梯度等于零的点，可以找到函数的局部最小值或最大值。

2. 机器学习：在机器学习中，矩阵微分用于计算损失函数的梯度，这是许多优化算法（如梯度下降法）的核心步骤。此外，在神经网络的反向传播算法中，矩阵微分被用来高效地计算权重的更新。

函数对变量（元素、向量或者矩阵）的导数/微分

变元、函数和映射

$\boldsymbol{x}=\left[x_1,\cdots,x_m\right]^{\mathrm{T}}\in\mathbb{R}^m$ 为实向量变元

$\boldsymbol{X}=\left[\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n\right]\in\mathbb{R}^{m\times n}$ 为实矩阵变元

$f(\boldsymbol{x})\in\mathbb{R}$ 为实值标量函数，其变元为 $m\times 1$ 实值向量 $\boldsymbol{x}$， $f:\mathbb{R}^m\to\mathbb{R}$ ；

$f(\boldsymbol{X})\in\mathbb{R}$ 为实值标量函数，其变元为 $m\times n$ 实值矩阵 $\boldsymbol{X}$， $f:\mathbb{R}^{m\times n}\to\mathbb{R}$；

$\boldsymbol{f}(\boldsymbol{X})\in\mathbb{R}^p$ 为p维实列向量函数，其变元为 $m\times 1$ 实值向量 $\boldsymbol{x}$，$\boldsymbol{f}:\mathbb{R}^m\to\mathbb{R}^p$;

$\boldsymbol{f}(\boldsymbol{X})\in\mathbb{R}^p$ 为p维实列向量函数，其变元为 $m\times n$ 实值矩阵 $\boldsymbol{X}$，$\boldsymbol{f}:\mathbb{R}^{m\times n}\to\mathbb{R}^p$;

$\boldsymbol{F}(\boldsymbol{X})\in\mathbb{R}^{p\times q}$ 为$p\times q$实矩阵函数，其变元为 $m\times 1$ 实值向量 $\boldsymbol{x}$，$F:\mathbb{R}^m\to\mathbb{R}^{p\times q}$;

$\boldsymbol{F}(\boldsymbol{X})\in\mathbb{R}^{p\times q}$ 为$p\times q$实矩阵函数，其变元为 $m\times n$实值向量 $\boldsymbol{X}$, $F:\mathbb{R}^{m\times n}\to\mathbb{R}^{p\times q}$;

变元、函数和映射

表 3.1.1 实值函数的分类

函数类型	向量变元 $\boldsymbol{x} \in \mathbb{R}^m$	矩阵变元 $\boldsymbol{X} \in \mathbb{R}^{m\times n}$
标量函数 $f \in \mathbb{R}$	$f(\boldsymbol{x})$ $f:\ \mathbb{R}^m \to \mathbb{R}$	$f(\boldsymbol{X})$ $f:\ \mathbb{R}^{m\times n} \to \mathbb{R}$
向量函数 $\boldsymbol{f} \in \mathbb{R}^p$	$\boldsymbol{f}(\boldsymbol{x})$ $\boldsymbol{f}:\ \mathbb{R}^m \to \mathbb{R}^p$	$\boldsymbol{f}(\boldsymbol{X})$ $\boldsymbol{f}:\ \mathbb{R}^{m\times n} \to \mathbb{R}^p$
矩阵函数 $\boldsymbol{F} \in \mathbb{R}^{p\times q}$	$\boldsymbol{F}(\boldsymbol{x})$ $\boldsymbol{F}:\ \mathbb{R}^m \to \mathbb{R}^{p\times q}$	$\boldsymbol{F}(\boldsymbol{X})$ $\boldsymbol{F}:\ \mathbb{R}^{m\times n} \to \mathbb{R}^{p\times q}$

Jacobian矩阵

行向量偏导算子

$$\mathrm{D}_x \overset{\text{def}}{=} \frac{\partial}{\partial \boldsymbol{x}^{\mathrm{T}}} = \left[\frac{\partial}{\partial x_1}, \cdots, \frac{\partial}{\partial x_m}\right]$$

$$\mathrm{D}_x f(\boldsymbol{x}) = \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathrm{T}}} = \left[\frac{\partial f(\boldsymbol{x})}{\partial x_1}, \cdots, \frac{\partial f(\boldsymbol{x})}{\partial x_m}\right]$$

当实值标量函数的变元为实值矩阵时，存在两种定义

1) $f(X)$ **关于矩阵变元** X **的Jacobian矩阵**

$$\boldsymbol{X} \in \mathbb{R}^{m\times n} \quad \mathrm{D}_X f(\boldsymbol{X}) = \frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}^{\mathrm{T}}} = \begin{bmatrix} \frac{\partial f(\boldsymbol{X})}{\partial x_{11}} & \cdots & \frac{\partial f(\boldsymbol{X})}{\partial x_{m1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\boldsymbol{X})}{\partial x_{1n}} & \cdots & \frac{\partial f(\boldsymbol{X})}{\partial x_{mn}} \end{bmatrix} \in \mathbb{R}^{n\times m}$$

雅可比矩阵的每一行包含了映射到函数中某个分量的导数

Jacobian矩阵

2）行偏导向量

$$\mathrm{D}_{\mathrm{vec}\boldsymbol{X}}\, f(\boldsymbol{X}) = \frac{\partial f(\boldsymbol{X})}{\partial\, \mathrm{vec}^{\mathrm{T}}(\boldsymbol{X})} = \left[\frac{\partial f(\boldsymbol{X})}{\partial x_{11}}, \cdots, \frac{\partial f(\boldsymbol{X})}{\partial x_{m1}}, \cdots, \frac{\partial f(\boldsymbol{X})}{\partial x_{1n}}, \cdots, \frac{\partial f(\boldsymbol{X})}{\partial x_{mn}}\right]$$

$$\mathrm{D}_{\mathrm{vec}\boldsymbol{X}}\, f(\boldsymbol{X}) = \mathrm{rvec}\left(\mathrm{D}_{\boldsymbol{X}}\, f(\boldsymbol{X})\right) = \left(\mathrm{vec}\left(\mathrm{D}_{\boldsymbol{X}}^{\mathrm{T}}\, f(\boldsymbol{X})\right)\right)^{\mathrm{T}}$$

Jacobian矩阵

函数为矩阵，变元为矩阵

Jacobian 矩阵

$$\mathrm{D}_{\boldsymbol{X}}\boldsymbol{F}(\boldsymbol{X}) \overset{\text{def}}{=} \frac{\partial \mathrm{vec}(\boldsymbol{F}(\boldsymbol{X}))}{\partial (\mathrm{vec}\boldsymbol{X})^{\mathrm{T}}} \in \mathbb{R}^{pq \times mn}$$

其具体表达式为

$$\mathrm{D}_{\boldsymbol{X}}\boldsymbol{F}(\boldsymbol{X}) = \begin{bmatrix} \frac{\partial f_{11}}{\partial (\mathrm{vec}\boldsymbol{X})^{\mathrm{T}}} \\ \vdots \\ \frac{\partial f_{p1}}{\partial (\mathrm{vec}\boldsymbol{X})^{\mathrm{T}}} \\ \vdots \\ \frac{\partial f_{1q}}{\partial (\mathrm{vec}\boldsymbol{X})^{\mathrm{T}}} \\ \vdots \\ \frac{\partial f_{pq}}{\partial (\mathrm{vec}\boldsymbol{X})^{\mathrm{T}}} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_{11}}{\partial x_{11}} & \cdots & \frac{\partial f_{11}}{\partial x_{m1}} & \cdots & \frac{\partial f_{11}}{\partial x_{1n}} & \cdots & \frac{\partial f_{11}}{\partial x_{mn}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_{p1}}{\partial x_{11}} & \cdots & \frac{\partial f_{p1}}{\partial x_{m1}} & \cdots & \frac{\partial f_{p1}}{\partial x_{1n}} & \cdots & \frac{\partial f_{p1}}{\partial x_{mn}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_{1q}}{\partial x_{11}} & \cdots & \frac{\partial f_{1q}}{\partial x_{m1}} & \cdots & \frac{\partial f_{1q}}{\partial x_{1n}} & \cdots & \frac{\partial f_{1q}}{\partial x_{mn}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_{pq}}{\partial x_{11}} & \cdots & \frac{\partial f_{pq}}{\partial x_{m1}} & \cdots & \frac{\partial f_{pq}}{\partial x_{1n}} & \cdots & \frac{\partial f_{pq}}{\partial x_{mn}} \end{bmatrix}$$