随机信号分析

线性代数导论

主讲: 赵国亮

内蒙古大学电子信息工程学院

August 29, 2020

目录

目录

在本小节中，我们首先引入矩阵的概念并介绍其基本运算方法。

定义 .1 数域

设它是复数集合 $\mathbb{C}$ 的一个非空子集，若 F 中的任意两个数 a, b 的和 $a+b$ 从差 $a-b$, 积 ab, 商 $a/b(b$ 不等于 0) 仍在 $\mathbb{F}$ 中，则称 $\mathbb{F}$ 为一个数域，我们常用 $\mathbb{R}$ 表示实数域，$\mathbb{C}$ 表示复数域。

定义 .2 数域上的矩阵

设 A 是复数域 $\mathbb{C}$ 中 $m \times n$ 个数 $a_{ij}(i = 1, 2, \cdots, m; j = 12, \cdots, n)$ 所组成的数集，若规定了其中各个数之间的相对位置并排成 m 行 n 列的数表

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

则称数集 A 为复数域 $\mathbb{C}$ 上的 m 行 n 列矩阵, 简称为 $m \times n$ 矩阵, 记为 $A = [a_{ij}] = (a_{ij})_{m \times n}$, 其中数 $a_{ij}(i = 1, 2, \cdots, m; j = 1, 2, \cdots, n)$ 称为矩阵 A 的第 i 行第 j 列元素。数域 $\mathbb{C}$ 上的全体 $m \times n$ 矩阵的集合记为 $\mathbb{C}^{m \times n}$(或 $\mathbb{R}^{m \times n}$) .

定义 .3 矩阵相等

若 $A, B \in C^{m\times n}$, 并且它们的对应元素相等，即

$$a_{ij} = b_{ij} \quad (i = 1, 2, \cdots, m; j = 1, 2, \cdots, n)$$

则称矩阵 A 与矩阵 B 相等，记作 $A = B$.

定义 .4 矩阵转置

设 $A \in C^{m\times n}$, 矩阵 A 的第 $i(i = 1, 2, \cdots, m)$ 行作为新矩阵 B 的第 i 列，则称 B 时 A 的转置矩阵, 记作 $B = A^T$, 显然有 $\mathbf{A}^\top \in \mathbb{C}^{n\times m}$, $(\mathbf{A}^\top)^\top = A$.

定义 .5 矩阵共轭

设 $A \in C^{m\times n}$, 矩阵 A 的第 i 个元素换成其共轭，则称 A^* 是 A 的共轭矩阵, 记作 $B = A^* = A^H$, 显然有 $A^H \in C^{n\times m}$, $(A^H)^H = A$.

定义 .6 矩阵和

设 $A, B \in C^{m\times n}$, A, 和矩阵 $C = [a_{ij} \pm b_{ij}]$, 记作 $C = A \pm B$, 显然, $A \pm B \in C^{m\times n}$.

定义 .7 矩阵积

设 $A \in \mathbb{C}^{m\times s}, B \in \mathbb{C}^{s\times n}$, A, B 中的元素按 $\sum_{k=1}^{s} a_k b_{kj}$ 方式计算, 得到新矩阵 $C = [c_{ij}]$ 成为是 A 与 B 的乘积, 记为 $C = AB$.

矩阵乘法具有如下运算性质:
设 A, B, C 是矩阵，K 是数，并且运算都是可行的，则
(1) 结合律

$$\begin{aligned} (AB)C &= A(BC) \\ (kB)C &= k(BC) \end{aligned} \tag{1}$$

(2) 分配律

$$\begin{aligned} A(B+C) &= AB+AC \\ (A+B)C &= AC+BC \end{aligned} \tag{2}$$

(3) 转置律

$$(AB)^T = B^T A^T \tag{3}$$

矩阵的初等变换也是我们经常用到的矩阵操作，它主要包括：位置变换，即互换矩阵的第 I 行 (列) 与第 j 行 (列)；数乘变换，即用非零数 k 乘矩阵的第 i 行 (列)。

定义 .8 迹

$A = [a_{ij}] \in C^{n\times n}$ 的所有对角元素之和称为 A 的迹，记为 $\mathrm{tr}(\mathbf{A}) = a_{11} + a_{22} + \cdots + a_{nn}$

定义 .9 行列式的阶

$$\det\{\mathbf{A}\} = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \tag{4}$$

表示关于矩阵 A 的 n^2 个元素的一种特定代数运算，称为矩阵 A 的行列式，其中 n 称为行列式的阶.

定义 .10 迹

$A \in \mathbb{C}^{n\times n}$, 把 A 中元素 a_{ij} 所在的第 i 行和第 j 列划去后，留下来的 $n-1$ 阶方阵称为元素 a_{ij} 的余子矩阵，记为 $A(i,j)$, 并且称 $\text{cof}\,(A_{ij}) = \det\{A(i,j)\}$ 为元素 a_{ij} 的余子式，$(-1)^{i+j}\det\langle A(i,j)\rangle$ 称为元素 a_{ij} 的代数余子式，记为 $\mathbf{A}_{ij}$.

行列式具体的运算法则是
当 $n=1$ 时, $\det\{\mathbf{A}\} = a_{11}$
当 $n>1$ 时,
$\det\{\mathbf{A}\} = a_{11}\mathbf{A}_{11} + a_{21}\mathbf{A}_{21} + \cdots + a_{n1}\mathbf{A}_{n1} = \sum_{k=1}^{n} a_{k1}\mathbf{A}_{k1}$, 其中 $\mathbf{A}_{k1}$ 是 a_{ij} 的代数余子式.
特别地当 $\det\langle\mathbf{A}\rangle = 0$ 时，A 称为奇异 (退化) 方阵，否则称为非奇异 (非退化) 方阵，也称正则矩阵。

n 阶行列式的性质：

(1) 转置变换不改变行列式的值，即 $\det\left\{\mathbf{A}^{T}\right\} = \det\{\mathbf{A}\}$;

(2) 位置变换仅改变行列式的符号，即两行 (列) 互换，行列式的值变号。

定理 .1

数乘变换改变行列式的值，并且仿方阵 $A = [a_1, \cdots, a_i, \cdots, a_n], f_j|, B = [a_1, \cdots, ka_i, \cdots, a_n]$，其中列矢量 $a_i \in \mathbb{C}^{n\times 1}(i = 1, 2, \cdots, n)$, 则

$$\det\langle\mathbf{B}\} = k\det\langle\mathbf{A}\rangle \tag{5}$$

$$\det\{k\mathbf{A}\} = k^{n}\det\{\mathbf{A}\} \tag{6}$$

$$\det\langle AB\} = \det\langle A\rangle\det\langle B\}, A, B \in \mathbb{R}^{n\times n} \tag{7}$$

引入矩阵子式的概念，对于矩阵 $\mathbf{A} \in \mathbb{C}^{m\times n}$, 其 k 行、k 列相交的元素所组成的矩阵的行列式称为 A 的 K 阶子式。

定义 .11 满秩

设矩阵 $\mathbf{A} \in \mathbb{C}^{m\times n}$, 有一个 k 阶子式不等于零，而所有 $k+1$ 阶子式 (若存在) 都等于零，则称正整数 k 为矩阵 A 的秩，记为 $\mathrm{rank}A = k$. 若 $m = n = k$, 则称 A 为满秩矩阵。

定义 .12 迹

矩阵 $\mathbf{A} \in \mathbb{C}^{m\times n}$, 如果对于任意的非零列矢量 $x \in C^n$, 有

(1) 当 $\mathrm{Re}\left\{x^H A x\right\} > 0$ 时，称矩阵 A 为正定矩阵；

(2) 当 $\mathrm{Re}\left\{x^H A x\right\} \geq 0$ 时，称矩阵 A 为半正定 (非负定) 矩阵；

(3) 当 $\mathrm{Re}\left\{x^H A x\right\} < 0$ 时，称矩阵 A 为负定矩阵

定义 .13

设 A 和 B 都是 n 阶方阵，则

(1) 当方阵 $C = A - B$ 为正定矩阵时，称矩阵 $A > B$.

(2) 当方阵 $C = A - B$ 为半正定矩阵时，称矩阵 $A \geq B$;

(3) 当方阵 $C = A - B$ 为负定矩阵时，称矩阵 $A < B$.

定义 .14 范数

矩阵 $A \in \mathrm{C}^{m\times n}$ 的某个行列式 $\|\mathbf{A}\|$ 满足如下性质
(1) A 为非零矩阵时，$\|A\| > 0$; A 为零矩阵时，$\|A\| = 0$;
(2) 对于任意复数 c 有 $\|c\mathbf{A}\| = |c|\|\mathbf{A}\|$;
(3) $\|\mathbf{A} + \mathbf{B}\| \leqslant \|\mathbf{A}\| + \|\mathbf{B}\|$;
(4) $\|AB\| \leqslant \|A\|\|B\|$.
则称 $\|A\|$ 是 $\mathrm{C}^{m\times n}$ 上的一个范数.

常见的范数有 F 范数、行和范数和列和范数：

$$\begin{aligned} \|A\|_F &= \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2} \\ \|A\|_{row} &= \max_{1\le i\le m}\left\{\sum_{j=1}^{n}|a_{ij}|\right\} \\ \|A\|_{ol} &= \max_{1\le i\le n}\left\{\sum_{i=1}^{m}|a_{ij}|\right\} \end{aligned} \tag{8}$$

定义 .15 矩阵的内积

对于同型矩阵 $\mathsf{A}, \mathsf{B} \in \mathbb{C}^{m \times n}$, 可以定义

$$\langle \mathbf{A}, \mathbf{B} \rangle = \mathrm{tr}\left\{ \mathbf{A}\mathbf{B}^{\mathrm{H}} + \mathbf{B}\mathbf{A}^{\mathrm{H}} \right\} / 2 \tag{9}$$

为矩阵的内积.

定理 .2 Laurent-Moïse Schwartz 不等式

设 $A \in C^{m\times n}, B \in \mathbb{C}^{n\times l}, AA^T$ 可逆，则

$$B^TB \geqslant (AB)^T\left(AA^T\right)^{-1}(AB) \tag{10}$$

等号成立的条件为存在 $C \in \mathbb{C}^{m\times l}$, $B = \mathbf{A}^T\mathbf{C}$.

目录

定义 .16 矩阵

设矩阵 $\mathbf{A} \in \mathbb{C}^{m \times n}$, 如果 $\mathbf{A}$ 在主对角线上的元素全为 1 而其他元素全为 0, 则称 $\mathbf{A}$ 为 n 阶单位矩阵，通常用 $\mathbf{I}$ 表示；如果 $\mathbf{A}$ 在交叉对角线上的元素全为 1 而其他元素全为 0，则称 $\mathbf{A}$ 为 n 阶反向单位矩阵，通常用 $\mathbf{J}$ 表示。

定义 .17 范德蒙矩阵

设矩阵 $\mathbf{A} \in \mathbb{C}^{m \times n}$, 如果 $\mathbf{A} = \mathbf{A}^{T}$, 则称 $\mathbf{A}$ 为对称矩阵，通常用 $\mathbf{I}$ 表示；如果 $\mathbf{A} = -\mathbf{A}^{T}$，则称 $\mathbf{A}$ 为反对称矩阵；如果 $\mathbf{A} = \mathbf{A}^{H}$，则称 $\mathbf{A}$ 为共轭对称矩阵；如果 $\mathbf{A} = -\mathbf{A}^{H}$，则称 $\mathbf{A}$ 为反共轭对称矩阵，也称反厄米特矩阵。

定义 .18 范德蒙矩阵

设矩阵 $\mathbf{A} \in \mathbb{C}^{m\times n}$,

$$A = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{m-1} & x_2^{m-1} & \cdots & x_n^{m-1} \end{bmatrix} \tag{11}$$

称为范德蒙矩阵，记为 $\mathbf{V}$, 其中第 j 列是比值为 x_j 的等比数列, $j = 1, 2, \cdots, n$.

定理 .3

范德蒙矩阵 $\mathbf{V} \in \mathbb{C}^{n \times n}$, 其行列式

$$\det\{\mathbf{V}\} = \prod_{\Omega^i} (x_j - x_i) \equiv \prod_{i=1}^{n} \prod_{j=i+1}^{n} (x_j - x_i) \tag{12}$$

定理 .4 汉克矩阵

设矩阵 $\mathbf{A} \in \mathbb{C}^{n+1\times n+1}$,

$$\mathbf{A} = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_n \\ a_1 & a_2 & a_3 & \cdots & a_{n+1} \\ a_2 & a_3 & a_4 & \cdots & a_{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_n & a_{n+1} & a_{n+2} & \cdots & a_{2n} \end{bmatrix} \quad (13)$$

称为汉克矩阵 (Hankel matrix)，其中 $a_{ij} = a_{i+j-2}(i,j = 1,2,\cdots,n+1)$. 汉克矩阵是对称矩阵，完全由其第 1 行和第 $n+l$ 行的 $2n+l$ 个元素确定。

定理 .5 汉克矩阵

非奇异矩阵 $H \in \mathbb{C}^{n\times n}$ 为汉克矩阵的重要统计是存在 $n \times n$ 阶矩阵

$$\mathbf{C} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \\ p_0 & p_1 & \cdots & p_{n-2} & p_{n-1} \end{bmatrix} \tag{14}$$

使得

$$CH = HC^T \tag{15}$$

定义 .19 托普利兹矩阵

$$\mathbf{T} = \begin{bmatrix} t_0 & t_1 & t_2 & \cdots & t_{n-1} \\ t_{-1} & t_0 & t_1 & \cdots & t_{n-2} \\ t_{-2} & t_{-1} & t_0 & \cdots & t_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_{-n+1} & t_{-n+2} & t_{-n+3} & \cdots & t_0 \end{bmatrix} \tag{16}$$

则称 T 为托普利兹矩阵 (Toeplitz matrix) , 其中 T 完全由其第 1 行和第 1 列的 $2n-1$ 个元素确定。托普利兹矩阵沿平行主对角线的每一对角线上的元素是相等的，是关于交叉对角线对称的. 显然

$$\mathbf{J}\mathbf{T}^{\mathrm{T}}\mathbf{J} = \mathbf{T} \tag{17}$$

其中 **J** 为反向单位矩阵

简单的托普利兹矩阵有

$$\mathbf{F}=\begin{bmatrix} 0 & 1 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 0 & \cdots & 0 & 0 \end{bmatrix}\in\mathbb{R}^{n\times n},\quad \mathbf{B}=\begin{bmatrix} 0 & 0 & \cdots & 0 \\ 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 1 & 0 \end{bmatrix}\in\mathbb{R}^{n\times n} \tag{18}$$

因它们作用于标准基 $\{\mathbf{e}_1,\cdots,\mathbf{e}_n\}$(见定义 (60) 时所产生的直观影响，故分别称为前向移位矩阵 (forward shift matrix) 和后向移位矩阵 (backward shift matrix) .

定义 .20 托普利兹矩阵

$\mathbf{T} \in \mathbb{C}^{n\times n}$ 为托普利兹矩阵的充分必要条件是 T 可以表示为

$$\mathbf{T} = \sum_{k=1}^{n-1} t_{-k}\mathbf{B}^k + \sum_{k=0}^{n-1} t_k\mathbf{F}^k \tag{19}$$

其中 **F** 是 **B** 由 1.3.22 给定.

定理 .6

非奇异矩阵 $\mathbf{T} \in \mathbb{C}^{n\times n}$ 为托普利兹矩阵的充分必要条件是存在 $n \times n$ 阶矩阵

$$\mathbf{C}^{(1)} = \begin{bmatrix} p_{n-1} & p_{n-2} & \cdots & p_1 & p_0 \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix} \tag{20}$$

$$\mathbf{C}^{(2)} = \begin{bmatrix} 0 & \cdots & \cdots & 0 & p_0 \\ 1 & \ddots & & \vdots & p_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & p_{n-2} \\ 0 & \cdots & 0 & 1 & p_{n-1} \end{bmatrix} \tag{21}$$

定理 .7

若矩阵 $\mathbf{H} \in \mathbb{C}^{n\times n}$ 是汉克矩阵，则 $\mathbf{JH}$ 和 $\mathbf{HJ}$ 是托普利兹矩阵. 如 $\mathbf{T} \in \mathbb{C}^{n\times n}$ 是托普利兹矩阵，则 $\mathbf{JT}$ 和 $\mathbf{TJ}$ 是汉克矩阵, 其中 $\mathbf{J}$ 是反向单位矩阵.

定义 .21 托普利兹矩阵

如果满足条件 $A^HA = AA^H$, 则称 $\mathbf{A}$ 为正规矩阵; 如 $A^HA = AA^H = \mathbf{I}$, 则称 $\mathbf{A}$ 为酉矩阵；如果满足条件 $A^TA = AA^T = \mathbf{I}$, 称 $\mathbf{A}$ 为正交矩阵.

目录

对矩阵 $\mathbf{A} \in \mathbb{C}^{n\times n}$, 若 $\det\{\mathbf{A}\} \neq 0$, 则必存在唯一的同阶方阵 $\mathbf{B}$ 使得

$$\mathrm{AB} = \mathrm{BA} = \mathrm{I} \tag{23}$$

称 $\mathbf{B}$ 是 $\mathbf{A}$ 的逆矩阵，记为 $\mathbf{A}^{-1}$. 但当 $\mathbf{A}$ 不是方阵或者 $\mathbf{A}$ 是方阵但 $\det\mathbf{A} = 0$ 时，则上述的逆矩阵就不存在。

逆矩阵的性质：

(1) $\left(\mathrm{A}^{-1}\right)^{-1} = \mathrm{A}$;

(2) 若 $\mathrm{A} = k\mathrm{I}(k \neq 0)$, 则 $\mathrm{A}^{-1} = k^{-1}\mathrm{I}$;

(3) 若 $\mathbf{A}, \mathbf{B}$ 都是 n 阶可逆方阵，则 $\mathbf{AB}$ 也是可逆矩阵，且 $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$ 主

定义 .22

(Sherman-Morrison 公式) 设 $\mathbf{A}$ 是一个 $n \times n$ 的可逆矩阵，并且 x 和 y 是两个 $n \times l$ 矢量。若 $(A + xy^H)$ 可逆，则

$$\left(\mathbf{A} + xy^H\right)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}xy^H\mathbf{A}^{-1}}{1 + y^H\mathbf{A}^{-1}x} \tag{24}$$

上式可推广为矩阵之和的求逆公式：

$$\begin{aligned}(\mathbf{A} + \mathbf{UBV})^{-1} &= \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{UB}\left(\mathbf{B} + \mathbf{BVA}^{-1}\mathbf{UB}\right)^{-1}\mathbf{BVA}^{-1} \\ &= \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}\left(\mathbf{I} + \mathbf{BVA}^{-1}\mathbf{U}\right)^{-1}\mathbf{BVA}^{-1} \\ &= \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{UBVA}^{-1}\left(\mathbf{I} + \mathbf{UBVA}^{-1}\right)^{-1}\end{aligned} \tag{25}$$

或者

$$(\mathsf{A}-\mathsf{U}\mathsf{V})^{-1}=\mathsf{A}^{-1}+\mathsf{A}^{-1}\mathsf{U}\left(\mathsf{I}-\mathsf{V}\mathsf{A}^{-1}\mathsf{U}\right)^{-1}\mathsf{V}\mathsf{A}^{-1} \tag{26}$$

分块矩阵的几种求逆公式:

(1) 矩阵 **A** 可逆时，为

$$\begin{bmatrix} \mathbf{A} & \mathbf{U} \\ \mathbf{V} & \mathbf{D} \end{bmatrix}^{-1} = \Bigg[\begin{matrix} \mathbf{A}^{-1}+\mathbf{A}^{-1}\mathbf{U}\left(\mathbf{D}-\mathbf{V}\mathbf{A}^{-1}\mathbf{U}\right)^{-1}\mathbf{V}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{U}\left(\mathbf{D}-\mathbf{V}\mathbf{A}^{-} \right. \\ -\left(\mathbf{D}-\mathbf{V}\mathbf{A}^{-1}\mathbf{U}\right)^{-1}\mathbf{V}\mathbf{A}^{-1} & \left(\mathbf{D}-\mathbf{V}\mathbf{A}^{-1}\mathbf{U}\right) \end{matrix} \tag{27}$$

2) 矩阵 **A** 和 **D** 可逆时，为

$$\begin{bmatrix} A & U \\ V & D \end{bmatrix}^{-1} = \begin{bmatrix} \left(A - UD^{-1}V\right)^{-1} & -A^{-1}U\left(D - VA^{-1}U\right)^{-1} \\ -D^{-1}V\left(A - UD^{-1}V\right)^{-1} & \left(D - VA^{-1}U\right)^{-1} \end{bmatrix} \tag{28}$$

(3) 矩阵 **A** 和 **D** 可逆时，为

$$\begin{bmatrix} A & U \\ V & D \end{bmatrix}^{-1} = \begin{bmatrix} \left(A - UD^{-1}V\right)^{-1} & -\left(A - UD^{-1}V\right)^{-1}UD^{-1} \\ -\left(D - VA^{-1}U\right)^{-1}VA^{-1} & \left(D - VA^{-1}U\right)^{-1} \end{bmatrix} \tag{29}$$

或

$$\begin{bmatrix} \mathbf{A} & \mathbf{U} \\ \mathbf{V} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \left(\mathbf{A}-\mathbf{U}\mathbf{D}^{-1}\mathbf{V}\right)^{-1} & -\left(\mathbf{V}-\mathbf{D}\mathbf{U}^{-1}\mathbf{A}\right)^{-1} \\ \left(\mathbf{U}-\mathbf{A}\mathbf{V}^{-1}\mathbf{D}\right)^{-1} & \left(\mathbf{D}-\mathbf{V}\mathbf{A}^{-1}\mathbf{U}\right)^{-1} \end{bmatrix} \tag{30}$$

$A \in \mathbb{C}^{m\times n}, G \in \mathbb{C}^{n\times m}$ 满足 Penrose - Moore 方程

$$AGA = A; \quad GAG = G; \quad (AG)^H = AG; \quad (GA)^H = GA \tag{31}$$

的全部或一部分，则称 G 为 **A** 的广义逆矩阵。其中满足第一个条件的广义逆矩阵记作 $\mathbf{A}^{-1}$ 满足全部四个条件的广义逆矩阵记作 $\mathbf{A}^{\dagger}$.

对于任意 $A \in \mathbb{C}^{m\times n}$, 满足 (31) 式条件的广义逆矩阵 $A^{\dagger}$ 是唯一存在的，其性质如下：

$$
\begin{aligned}
&(\mathbf{A}')' = \mathbf{A}\\
&\left(\mathbf{A}^{H}\right)' = (\mathbf{A}')^{H}\\
&\left(\mathbf{A}^{T}\right)' = (\mathbf{A}')^{H}\\
&\mathbf{A}' = \left(\mathbf{A}^{H}\mathbf{A}\right)'\mathbf{A}^{H} = \mathbf{A}^{H}\left(\mathbf{A}\mathbf{A}^{H}\right)'
\end{aligned}
\tag{32}
$$

$\mathbf{A}^{\dagger}$ 的计算一般需要用到矩阵的分解。设 $A \in \mathbb{C}^{m\times n}, \mathrm{rank}\{\mathbf{A}\} = r, \mathbf{A}$, $\mathbf{A}$ 的最大秩分解为 $\mathbf{A} = \mathbf{BC}$, 其中 $B \in \mathbb{C}^{m\times r}, C \in \mathbb{C}^{r\times n}, \mathrm{rank}\{B\} = \mathrm{rank}\{C\} = \mathrm{rank}\{A\} = r$, 则有

$$\mathbf{A}^{\dagger} = \mathbf{C}^{H}\left(\mathbf{CC}^{H}\right)^{-1}\left(\mathbf{BB}^{H}\right)^{-1}\mathbf{B}^{H} \tag{33}$$

特别是

当 $\mathbf{A} \in C^{m\times n}, \mathrm{rank}\{\mathbf{A}\} = m$ 时, 则 $A^{\dagger} = \mathbf{A}^{H}\left(\mathbf{AA}^{H}\right)^{-1}$.

当 $\mathbf{A} \in C^{m\times n}, \mathrm{rank}\{\mathbf{A}\} = n$ 时, 则 $A^{\dagger} = \left(\mathbf{AA}^{H}\right)^{-1}\mathbf{A}^{H}$.

目录

所谓矩阵分解，就是通过线性变换，将某个给定或已知的矩阵分解为两个或三个矩阵标准型的乘积。常用的矩阵分解方法主要有特征值分解 (EVD) 奇异值分解 (SVD)、Cholesky 分解、LU 分解、QR 分解等。

1. 特征值分解

定义 .23 特征值

设 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 如果存在数 λ 和 n 维非零列矢量 $\mathbf{x}$, 使得 $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$, 则称 λ 是矩阵 $\mathbf{A}$ 的特征值，$\mathbf{x}$ 是 $\mathbf{A}$ 的属于特征值 λ 的特征矢量。

特征值与特征矢量的性质：

(1) 若 **A** 是实对称矩阵或厄米特矩阵，则其所有特征值都是实数。

(2) 对一个 n $\times$ n 矩阵 **A**:

a 若 λ 是 **A** 的特征值，则 λ 也是 $\mathbf{A}^{T}$ 的特征值。

b 若 λ 是 **A** 的特征值，则 λ^* 是 $\mathbf{A}^{H}$ 的特征值

c. 若入是 **A** 的特征值，则 $\lambda + \sigma^2$ 是 $\mathbf{A} + \sigma^2\mathbf{I}$ 的特征值。

d. 若入是 **A** 的特征值，则 $1/\lambda$ 是逆矩阵 $\mathbf{A}^{-1}$ 的特征值。

(3) 若 $\mathbf{A}$ 是实正交矩阵，则其所有特征值为 1 或者- 1.
(4) 不同特征值入 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 对应的非零特征矢量 $u_1, u_2, \cdots, u_n$ 线性无关。
(5) 特征值与秩的关系：
a. 若 $n \times n$ 矩阵 $\mathbf{A}$ 有 k 个非零特征值，则 $\mathrm{rank}\mathbf{A} = k$.
b. 若 $n \times n$ 矩阵 $\mathbf{A}$ 只有一个零特征值，则 $\mathrm{rank}\mathbf{A} = n - l$.
c. 若 $\mathrm{rank}\mathbf{A} - \lambda\mathbf{I} \le n - 1$, 则入是矩阵 $\mathbf{A}$ 的特征值。

(6) 对于一个厄米特矩阵 **A**, 当且仅当它的所有特征值都为正 (或者非负) 值时，它是正定 (或半正定) 的。
(7) 若 **A** 的特征值不相同，则一定存在一个矩阵 **S**, 使 $\mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \mathbf{D}$(对角矩阵)，**D** 的对角元素是矩阵 **A** 的特征值。
(8) 一个 $n \times n$ 矩阵 $\mathbf{A} = [a_{ij}]$ 的最大特征值以该矩阵的列元素之和的最大值为界，即

$$\lambda_{max} \leqslant \max_{i} \left\{ \sum_{j=1}^{n} a_{ij} \right\} \tag{34}$$

(9) 若 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 是 $n \times n$ 矩阵 $\mathbf{A}$ 的特征值，$u_1, u_2, \cdots, u_n$ 分别是与特征值对应的特征矢量，则

$$A = E\Lambda E^{-1} \tag{35}$$

其中 $E = [u_1, u_2, \cdots, u_n], \Lambda = \text{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$

定理 .8

设 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 是 $n \times n$ 矩阵 $\mathbf{A}$ 的 n 个特征值，则

$$\begin{aligned} \textstyle\sum_{i=1}^n \lambda_i = \sum_{i=1}^n a_{ii} = \text{tr}\{\mathbf{A}\} \\ \textstyle\prod_{i=1}^n \lambda_i = \lambda_1\lambda_2\cdots\lambda_n = \det\{\mathbf{A}\} \end{aligned} \tag{36}$$

n 阶方阵 $\mathbf{A}$ 可逆等价于 $\mathbf{A}$ 的 n 个特征值全不为零。

定理 .9

设 λ 是 $\mathbf{A}$ 的特征值,t 是一个多项式,则 $f(\lambda)=a_0+a_1\lambda+\cdots+a_m\lambda^m$ 是方阵 $f(\mathbf{A})=a_0\mathbf{I}+a_1\mathbf{A}+\cdots+a_m\mathbf{A}^m$ 的一个特征值.

2. 奇异值分解

定义 .24 奇异值

设 $A\in\mathbb{C}^{m\times n}$, λ_i 为 $\mathbf{A}^H\mathbf{A}$ 的特征值, 则称 $\sigma_i=\lambda_i^{1/2}(i=1,2,\cdots,n)$ 为 $\mathbf{A}$ 的奇异值.

定理 .10

设 $\mathsf{A} \in \mathbb{C}^{m\times n}$, $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$ 为 **A** 的奇异值，则存在酉矩阵 $\mathbf{U} \in \mathbb{C}^{m\times m}, \mathbf{V} \in \mathbb{C}^{n\times n}$, 使得

$$\mathsf{A} = \mathsf{U}\Sigma\mathsf{V}^H \tag{37}$$

式中 $\Sigma = \begin{bmatrix} \boldsymbol{\Sigma}_1 & 0 \\ 0 & 0 \end{bmatrix}, \boldsymbol{\Sigma}_1 = \mathrm{diag}\{\delta_1, \delta_2, \cdots, \delta,\}$, 其对角元素 δ_i 为复数，且 $|\delta_i| = \sigma_i, r = \mathrm{rank}\{A\}$.

令矩阵 $\mathbf{A}$ 为 $m \times n$ 矩阵，并且 $r = \mathrm{rank}\mathbf{A}$, $p = \min\{m, n\}$. 设矩阵 $\mathbf{A}$ 的奇异值排列为

$$\sigma_{\max} = \sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_{p-1} \geqslant \sigma_p = \sigma_{\min} \geqslant 0 \tag{38}$$

矩阵的各种变形与奇异值的变化有以下关系：
(1) 矩阵 $\mathbf{A}$ 的共轭转置 $\mathbf{A}^H$ 的奇异值分解为

$$A^H = V\Sigma^H U^H \tag{39}$$

即矩阵 $\mathbf{A}$ 和 $\mathbf{A}^H$ 的具有完全相同的奇异值。

(2) **P** 和 **Q** 分别为 $m \times m$ 和 $n \times n$ 酉矩阵时，$\mathbf{PAQ}^H$ 的奇异值分解由

$$\mathbf{PAQ}^H = \tilde{U}\Sigma\tilde{V}^H \tag{40}$$

给出。其中，$\tilde{\mathbf{U}} = \mathbf{PU}, \tilde{\mathbf{V}} = \mathbf{QV}$. 也就是说，矩阵 $\mathbf{PAQ}^H$ 与 **A** 具有相同的奇异值，即奇异值具有酉不变性，但奇异矢量不同. $\mathbf{A}^H\mathbf{A}, \mathbf{AA}^H$ 的奇异值分解分别为

$$A^HA = V\Sigma^H\Sigma V^H, \quad AA^H = U\Sigma\Sigma^H U^H \tag{41}$$

其中

$$\Sigma^{\mathrm{H}}\Sigma = \mathrm{diag}\left\{\sigma_1^2, \sigma_2^2, \cdots, \sigma_{\mathrm{r}}^2, \overbrace{0}, \cdots, 0\right\}$$
$$\Sigma\Sigma^{\mathrm{H}} = \mathrm{diag}\left\{\sigma_1^2, \sigma_2^2, \cdots, \sigma_{\mathrm{r}}^2, \overbrace{0}, \cdots, 0\right\} \tag{42}$$

注：$\mathrm{A}^{\mathrm{H}}\mathrm{A}, \mathrm{A}\mathrm{A}^{\mathrm{H}}$ 均为厄米特矩阵，厄米特矩阵的奇异值分解与特征值分解是一致的。

4) $m \times n$ 矩阵 **A** 的奇异值分解与 $n \times m$ 维 Moore - Penrose 广义逆矩阵 $\mathbf{A}^{\dagger}$

$$A^{\dagger} = V\Sigma^{\dagger}U^{H} \tag{43}$$

其中 $\boldsymbol{\Sigma}^{+} = \begin{bmatrix} \boldsymbol{\Sigma}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$。

(5) 矩阵 **A** 的谱范数为 **A** 的最大奇异值，即

$$\|A\|_{spec} = \sigma_{max} \tag{44}$$

(6) 若 $m \geq n$, 则对于矩阵 $\mathbf{A}$, 有

$$\begin{aligned}\sigma_{\min}(\mathbf{A}) &= \min_{\mathbf{x}\neq 0}\left\{\left(\frac{\mathbf{x}^H\mathbf{A}^H\mathbf{A}\mathbf{x}}{\mathbf{x}^H\mathbf{x}}\right)^{1/2}\right\} \\ &= \min_{\mathbf{x}^H\mathbf{x}=1}\ \min_{\mathbf{x}\in\mathbb{C}^{n\times 1}}\left\{\left(\mathbf{x}^H\mathbf{A}^H\mathbf{A}\mathbf{x}\right)^{1/2}\right\}\end{aligned} \tag{45}$$

(7) $m \geq n$, 则对于矩阵 $\mathbf{A}$, 有

$$\begin{aligned}\sigma_{\max}(\mathbf{A}) &= \max_{\mathbf{x}\neq 0}\left\{\left(\frac{\mathbf{x}^H\mathbf{A}^H\mathbf{A}\mathbf{x}}{\mathbf{x}^H\mathbf{x}}\right)^{1/2}\right\} \\ &= \max_{\mathbf{x}^H\mathbf{x}=1}\ \max_{\mathbf{x}\in\mathbb{C}^{n\times 1}}\left\{\left(\mathbf{x}^H\mathbf{A}^H\mathbf{A}\mathbf{x}\right)^{1/2}\right\}\end{aligned} \tag{46}$$

(8) 若 $\mathbf{A}$ 和 $\mathbf{B}$ 均为 $m \geq n$ 矩阵，则对千 $1 \leqslant i, j \leqslant p, i+j \leqslant p+1$，有

$$\sigma_{i+j-1}(A+B) \leqslant \sigma_i(A) + \sigma_j(B) \tag{47}$$

其中 $\sigma_i(\mathbf{A})$ 表示矩阵 $\mathbf{A}$ 的第 i 个大奇异值。特别地，当 $j = l$ 时，$\sigma_i(\mathbf{A}+\mathbf{B}) \leqslant \sigma_i(\mathbf{A}) + \sigma_1(B)(i = 1, 2, \cdots, p)$ 成立.
(9) 对于 $m \geq n$ 矩阵 $\mathbf{A}$ 和 $\mathbf{B}$，有

$$\sigma_{max}(\mathbf{A}+\mathbf{B}) \leqslant \sigma_{max}(\mathbf{A}) + \sigma_{max}(\mathbf{B}) \tag{48}$$

3. Cholesky 分解

设 $\mathbf{A} \in \mathbb{C}^{n\times n}$ 是正定矩阵，$\mathbf{A} = \mathbf{GAG}^H$ 称为矩阵 $\mathbf{A}$ 的 Cholesky 分解，其中 $\mathbf{G}$

$$\mathbf{G} = \begin{bmatrix} g_{11} & 0 & \cdots & 0 \\ g_{21} & g_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ g_{n1} & g_{n2} & \cdots & g_{mn} \end{bmatrix} \tag{49}$$

定理 .11

若 $\mathbf{A} \in \mathbb{C}^{n\times n}$ 是正定矩阵，则 Cholesky 分解 $\mathbf{A} = \mathbf{AGAG}^{-1}$ 是唯一的.

4. LU 分解

设 $\mathbf{A} \in \mathbb{C}^{m\times n}$, 则 $\mathbf{A} = \mathbf{LU}$ 称为矩阵的 **LU** 分解。其中，**L** 为 $m \times m$ 单位下三角矩阵 (对角线为 1 的下三角矩阵)，**U** 是 **A** 的为 $m \times n$ 上阶梯型矩阵。

定理 .12

如果 $\mathbf{A} \in \mathbb{C}^{n\times n}$ 非奇异，并且其 **LU** 分解存在，则 **A** 的 **LU** 分解是唯一的，且 $\det\{\mathbf{A}\} = u_{11}u_{22}\cdots u_{nn}$ ▪

5. QR 分解

定理 .13

若 $\mathbf{A} \in \mathbb{C}^{m\times n}$, 且 $m \geq n$, 则存在列正交的矩阵 $\mathbf{Q} \in \mathbb{C}^{m\times n}$ 和上三角矩阵 $\mathbf{R} \in \mathbb{C}^{n\times n}$ 使得 $\mathbf{A} = \mathbf{QR}$. 当 $m = n$ 时，$\mathbf{Q}$ 是正交矩阵。若 $\mathbf{A}$ 是非奇异的 $n \times n$ 矩阵，则 R 的所有对角线元素均为正，并且 $\mathbf{Q}$ 和 $\mathbf{R}$ 二者是唯一的。

目录

所有 n 维复矢量的集合构成 n 维复矢量空间 $\mathbb{C}^n$, $m(\leq n)$ 个 n 维复矢量的子集合构成 $\mathbb{C}^n$ 内的一个矢量子空间。
若 $S = \{u_1, u_2, \cdots, u_m\}$ 是矢量空间 **V** 的矢量子集合，则 $u_1, u_2, \cdots, u_m$ 的所有线性组合的集合 **W** 称为由 $u_1, u_2, \cdots, u_m$ 张成的子空间，为

$$\begin{aligned} W &= \mathrm{Span}\{u_1, u_2, \cdots, u_m\} \\ &= \{u; u = a_1u_1 + a_2u_2 + \cdots + a_mu_m\} \end{aligned} \tag{50}$$

V 本身是 **V** 的一个子空间；只由零矢量构成的矢量集合也是 **V** 的一个子空间，称为零子空间 **V** 和零子空间称为 **V** 的平凡子空间；不是平凡的子空间称为 **V** 的真子空间.

定理 .14

令 **S** $=\{u_1,u_2,\cdots,u_m\}$ 是矢量空间 **V** 的一个集合，且 **W** $=$ Span $\{u_1,u_2,\cdots,u_m\}$ 是由 **S** 的 m 个列矢量张成的子空间。
(1) 在 **S** 内，若有某个矢量 (例如 U_k) 是其他矢量的线性组合，则从 **S** 中删去矢量 U_k 后，其他矢量张成的子空间仍为 **W**.
(2) 若 **W** 为非平凡子空间，则在 **S** 内一定存在某个由线性无关的矢扯组成的集合，它张成子空间 **W**.

定义 .25 基

若 $\mathbf{W}$ 是由矢量集合 $\mathbf{S}=\{u_1,u_2,\cdots,u_m\}$ 张成的子空间，则 S 内的最大线性无关矢量最组称为 $\mathbf{W}$ 的一组基.

定义 .26 子空间

子空间 $\mathbf{W}$ 的任何一组基的矢量个数称为 $\mathbf{W}$ 的维数，用符号 $\dim(\mathbf{W})$ 表示若 W 的任何一组基都不是由有限个线性无关的矢量组成时，则称 $\mathbf{W}$ 是无限维矢量子空间.

定义 .27 标准基

$\{e_1, e_2, \cdots, e_n\}$ 是 $\mathbb{R}^n$ 上的一组基矢量，其中

$$e_1 = [1, 0, \cdots, 0]^T, \quad e_2 = [0, 1, \cdots, 0]^\top, \quad \cdots, \quad e_n = [0, 0, \cdots, 1]^T \tag{51}$$

任一矢量 $x = [x_1, x_2, \cdots, x_n]^T \in \mathbb{R}^n$ 可以用矢量 e_i 表示为 $x = \sum_{i=1}^{n} x_i e_i$，我们称 $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\}\mathbb{R}^n$ 上的标准基.

定理 .15

$B=\{\mathbf{b}_1,\mathbf{b}_2,\cdots,\mathbf{b}_n\}$ 是 n 维矢量子空间 **W** 的一组基，则对于 **W** 中的任何一个矢量 **x**, 都存在一组唯一的标量 $c_1,c_2,\cdots,c_n$ 使得 **x** 可以表示为

$$\mathbf{x}=c_1b_1+c_2b_2+\cdots+c_nb_n \tag{52}$$

一矢量与子空间 **W** 的所有矢量都正交，则称该矢最正交子空间 **W**, 若 $\forall a_i\in W_i, a_j\in W_j(i\neq j)$，恒有 $a_i\perp a_j$, 则子空间 W_i, W_j 正交子空间，记作 $W_i\perp W_j(i\neq j)$.

特别地，与子空间 **W** 正交的所有矢量的集合组成的子空间，称为 **W** 的正交补空间，记作 $\mathbf{W}^{\perp}$, 即

$$W^{\perp}=\left\{x|x^Ty=0,\forall y\in W\right\} \tag{53}$$

1. 实矢量的梯度算子

$n \times l$ 实矢量 $\mathbf{x}$ 的梯度算子定义为

$$\begin{aligned} \nabla_x &= \left[\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \cdots, \frac{\partial}{\partial x_n}\right]^T = \frac{\partial}{\partial x} \\ \nabla_{x^T} &= \left[\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \cdots, \frac{\partial}{\partial x_n}\right] = \frac{\partial}{\partial x^T} \end{aligned} \tag{54}$$

对于 $n \times l$ 实矢量 $\mathbf{x}$, 有一实标量函数 $f(\mathbf{x})$, 则 $f(\mathbf{x})$ 相对于 $\mathbf{x}$ 和 $\mathbf{x}^T$ 的梯度分别为

$$\begin{aligned} \nabla_x f(x) &= \frac{\partial f(x)}{\partial x} = \left[\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \cdots, \frac{\partial f(x)}{\partial x_n}\right]^T \\ \nabla_x^T f(x) &= \frac{\partial f(x)}{\partial x^T} = \left[\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \cdots, \frac{\partial f(x)}{\partial x_n}\right] \end{aligned} \tag{55}$$

从梯度的定义式可以看出：
(1) 一个以矢量为变元的标量函数的梯度为一矢量；
(2) 梯度的每个分量给出了标量函数在该分量方向上的变化率。梯度矢量反映了当变元增大时函数 t 的最大增大率。
对于 $n \times l$ 实矢量 $\mathbf{x}$，有一实函数矢量 $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \cdots, f_m(\mathbf{x})]$，则 $\mathbf{f}(\mathbf{x})$ 相对于 $\mathbf{x}$ 的梯度定义为

$$\nabla_x f(x) = \frac{\partial f(x)}{\partial x} = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \frac{\partial f_2(x)}{\partial x_1} & \cdots & \frac{\partial f_m(x)}{\partial x_1} \\ \frac{\partial f_1(x)}{\partial x_2} & \frac{\partial f_2(x)}{\partial x_2} & \cdots & \frac{\partial f_m(x)}{\partial x_2} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_1(x)}{\partial x_n} & \frac{\partial f_2(x)}{\partial x_n} & \cdots & \frac{\partial f_m(x)}{\partial x_n} \end{bmatrix} \quad (56)$$

2. 实矩阵的梯度算子

对于 $m \times n$ 实矩阵 $\mathbf{A}$, 有一实值函数 $f(\mathbf{A})$, 则 $f(\mathbf{A})$ 相对于 $\mathbf{A}$ 的梯度 (简称梯度矩阵) 为

$$\nabla_A f(A) = \frac{\partial f(A)}{\partial A} = \begin{bmatrix} \frac{\partial f(A)}{\partial a_{11}} & \frac{\partial f(A)}{\partial a_{12}} & \cdots & \frac{\partial f(A)}{\partial x_{1n}} \\ \frac{\partial f(A)}{\partial a_{21}} & \frac{\partial f(A)}{\partial a_{22}} & \cdots & \frac{\partial f(A)}{\partial a_{2n}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f(A)}{\partial a_{m1}} & \frac{\partial f(A)}{\partial a_{m2}} & \cdots & \frac{\partial f(A)}{\partial a_{mn}} \end{bmatrix} \tag{57}$$

式中, a_{ij} 是矩阵 $\mathbf{A}$ 的第 i 行、第 j 列元素。

3. 迹函数的梯度矩阵

(1) 单个矩阵的迹的梯度矩阵对于 $m \times n$ 矩阵 $\mathbf{W}$, 有

$$\frac{\partial \operatorname{tr}\{W\}}{\partial W} = I_m \tag{58}$$

对于 $m \times n$ 可逆矩阵 $\mathbf{W}$, 有

$$\frac{\partial \operatorname{tr}\left\{\mathbf{W}^{-1}\right\}}{\partial \mathbf{W}} = -\left(\mathbf{W}^{-2}\right)^T \tag{59}$$

对于两个列矢量的外积，有

$$\frac{\partial \operatorname{tr}\left\{xy^T\right\}}{\partial x} = \frac{\partial \operatorname{tr}\left\{yx^T\right\}}{\partial x} = y \tag{60}$$

(2) 两个矩阵乘积的迹的梯度

对于 $\mathbf{W} \in \mathbb{R}^{m \times n}, A \in \mathbb{R}^{n \times m}$, 有

$$\frac{\partial \operatorname{tr}\{WA\}}{\partial W} = \frac{\partial \operatorname{tr}\{AW\}}{\partial W} = A^T \tag{61}$$

当 $\mathbf{W}$ 是对称矩阵时, 有

$$\frac{\partial \operatorname{tr}\{WA\}}{\partial W} = \frac{\partial \operatorname{tr}\{AW\}}{\partial W} = A + A^T - \operatorname{diag}\{A\} \tag{62}$$

对于 $\mathbf{W} \in \mathbb{R}^{m \times n}, A \in \mathbb{R}^{n \times m}$, 有

$$\frac{\partial \operatorname{tr}\left\{\mathbf{W}^T\mathbf{A}\right\}}{\partial \mathbf{W}} = \frac{\partial \operatorname{tr}\left\{\mathbf{A}\mathbf{W}^T\right\}}{\partial \mathbf{W}} = \mathbf{A} \tag{63}$$

对于 $\mathbf{W} \in \mathbb{R}^{m\times n}$, 有

$$\frac{\partial\, \mathrm{tr}\left\{\mathbf{W}\mathbf{W}^{\mathrm{T}}\right\}}{\partial \mathbf{W}} = \frac{\partial\, \mathrm{tr}\left\{\mathbf{W}^{\mathrm{T}}\mathbf{W}\right\}}{\partial \mathbf{W}} = 2\mathbf{W} \tag{64}$$

对于 $\mathbf{W} \in \mathbb{R}^{m\times m}$, 有

$$\frac{\partial\, \mathrm{tr}\left\{\mathbf{W}^{2}\right\}}{\partial \mathbf{W}} = \frac{\partial\, \mathrm{tr}\{\mathbf{W}\mathbf{W}\}}{\partial \mathbf{W}} = 2\mathbf{W}^{\mathrm{T}} \tag{65}$$

对于 $\mathbf{W} \in \mathbb{R}^{m\times n}, \mathrm{A} \in \mathbb{R}^{n\times m}$ 且 $\mathbf{W}$ 非奇异, 有

$$\frac{\partial\, \mathrm{tr}\left\{\mathbf{A}\mathbf{W}^{-1}\right\}}{\partial \mathbf{W}} = -\left(\mathbf{W}^{-1}\mathbf{A}\mathbf{W}^{-1}\right)^{\mathrm{T}} \tag{66}$$

(3) 三个矩阵乘积的迹的梯度

对于 $\mathbf{W} \in \mathbb{R}^{m\times n}, A \in \mathbb{R}^{m\times m}$, 有

$$\frac{\partial \operatorname{tr}\left\{\mathbf{W}^{\mathsf{T}}\mathbf{A}\mathbf{W}\right\}}{\partial \mathbf{W}} = \left(\mathbf{A} + \mathbf{A}^{\mathsf{T}}\right)\mathbf{W} \tag{67}$$

当 $\mathbf{W}$ 是对称矩阵时, 有

$$\frac{\partial \operatorname{tr}\left\{W^{\top}AW\right\}}{\partial W} = 2AW \tag{68}$$

对于 $\mathbf{W} \in \mathbb{R}^{m\times n}, A \in \mathbb{R}^{n\times n}$, 有

$$\frac{\partial \operatorname{tr}\left\{WAW^{\mathsf{T}}\right\}}{\partial W} = W\left(A + A^{\mathsf{T}}\right) \tag{69}$$

当 $\mathbf{A}$ 是对称矩阵时, 有

$$\frac{\partial \operatorname{tr}\left\{\mathsf{W}\mathsf{A}\mathsf{W}^{\mathsf{T}}\right\}}{\partial \mathsf{W}} = 2\mathsf{W}\mathsf{A} \tag{70}$$

对于 $\mathbf{W} \in \mathbb{R}^{m \times n}, A \in \mathbb{R}^{n \times m}$ 且 $\mathbf{W}$ 非奇异, 有

$$\frac{\partial \operatorname{tr}\left\{\mathbf{A}\mathbf{W}^{-1}\mathbf{B}\right\}}{\partial \mathbf{W}} = -\left(\mathbf{W}^{-1}\mathbf{B}\mathbf{A}\mathbf{W}^{-1}\right)^{\mathsf{T}} \tag{71}$$

(4) 四个矩阵乘积的迹的梯度

对于 $\mathbf{W} \in \mathbb{R}^{m\times n}, A \in \mathbb{R}^{p\times m}$, 有

$$\frac{\partial \operatorname{tr}\left\{AWW^{T}A^{T}\right\}}{\partial W} = 2A^{T}AW \tag{72}$$

对于 $\mathbf{W} \in \mathbb{R}^{m\times n}, A \in \mathbb{R}^{p\times n}$, 有

$$\frac{\partial \operatorname{tr}\left\{\mathbf{A}\mathbf{W}^{T}\mathbf{W}\mathbf{A}^{T}\right\}}{\partial \mathbf{W}} = 2\mathbf{W}\mathbf{A}^{T}\mathbf{A} \tag{73}$$

对于 $\mathbf{W} \in \mathbb{R}^{m\times n}, A \in \mathbb{R}^{p\times m}, B \in \mathbb{R}^{m\times p}$, 有

$$\frac{\partial \operatorname{tr}\left\{AWW^{T}\mathbf{B}\right\}}{\partial \mathbf{W}} = \left(\mathbf{B}\mathbf{A} + \mathbf{A}^{T}\mathbf{B}^{T}\right)\mathbf{W} \tag{74}$$

对于 $\mathbf{W} \in \mathbb{R}^{m\times n}, A \in \mathbb{R}^{p\times n}, B \in \mathbb{R}^{n\times p}$, 有

$$\frac{\partial \operatorname{tr}\left\{\mathbf{A}\mathbf{W}^{T}\mathbf{W}\mathbf{B}\right\}}{\partial \mathbf{W}} = \mathbf{W}\left(\mathbf{B}\mathbf{A} + \mathbf{A}^{T}\mathbf{B}^{T}\right) \tag{75}$$

对于 $\mathbf{W} \in \mathbb{R}^{m\times n}, A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{m\times m}$, 有

$$\frac{\partial \operatorname{tr}\left\{WA^{\top}W^{T}B\right\}}{\partial W} = BWA + B^{T}WA^{T} \tag{76}$$

对于 $\mathbf{W} \in \mathbb{R}^{m\times n}, A, B \in \mathbb{R}^{n\times m}$, 有

$$\frac{\partial \operatorname{tr}\{WAWB\}}{\partial W} = B^{\top}W^{\top}A^{\top} + A^{\top}W^{\top}B^{\top} \tag{77}$$

4. 行列式的梯度矩阵

若矩阵 $\mathbf{W}$ 的元素 W_{ij} 相互独立，

$$\frac{\partial W}{\partial w_{ij}} = E_{ij} = \mathbf{e}_i e_j^T \tag{78}$$

其中，$\mathbf{E}_{ij}$ 是一个除 (i,j) 元素为 1 外其余元素均为 0 的矩阵；$\mathbf{e}_i$ 为(51)式给出的标准基，其第 I 个元素为 1, 其余元素全部为 0.

若矩阵 $\mathbf{W}$ 为对称矩阵，有

$$\frac{\partial W}{\partial w_{ij}} = E_{ij} + E_{ji} - \delta_{ij} E_{ij} \tag{79}$$

其中 $\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$

记 $n \times n$ 矩阵 $\mathbf{W}$ 中元素 W_{ij} 的余子式为 $\mathrm{cof}(\mathbf{W};;)$, 则有 $\det(\mathbf{W}) = \sum_{j=1}^{n} w_{ij}\,\mathrm{cof}\left(W_{ij}\right)\ (i = 1, 2, \cdots, n)$, 有

$$\frac{\partial \det\{W\}}{\partial w} = \mathrm{cof}\left(W_{ij}\right),$$

$\mathbf{W}$ 的元素相互独立.

$$\begin{aligned}\frac{\partial \det\{W\}}{\partial w_{ij}} &= \mathrm{cof}\left(W_{ij}\right) + \mathrm{cof}\left(W_{ji}\right) - \delta_{ij}\,\mathrm{cof}\left(W_{ij}\right) \\ &= \left(1 - \delta_{ij}\right)\mathrm{cof}\left(W_{ij}\right) + \mathrm{cof}\left(W_{ji}\right),\end{aligned} \tag{80}$$

$\mathbf{W}$ 为对称阵.

令 $\mathsf{i},\mathsf{j} = 1,2,\cdots,\mathsf{n}$ 写成矩阵形式，有

$$\frac{\partial \det\{\mathbf{W}\}}{\partial \mathbf{W}} = \left(\mathbf{W}^{\sharp}\right)^{\mathsf{T}} = \det\{\mathbf{W}\}\left(\mathbf{W}^{-1}\right)^{\mathsf{T}}, \tag{81}$$

W 的元素相互独立.

$$\begin{aligned}\frac{\partial \det\{\mathbf{W}\}}{\partial \mathbf{W}} &= 2\mathbf{W}^{\sharp} - \mathrm{diag}\left\{\mathbf{W}^{\sharp}\right\} \\ &= \det\{\mathbf{W}\}\left(2\mathbf{W}^{-1} - \mathrm{diag}\left\{\mathbf{W}^{-1}\right\}\right)\end{aligned} \tag{82}$$

W 为对称阵, $\mathbf{W}^{\#}$ 是 **W** 伴随矩阵，$\mathbf{W}^{-1} = \mathbf{W}^{\#}/\det\{\mathbf{W}\}$.

若是 **W** 为非奇异矩阵，有

$$\frac{\partial}{\partial \mathsf{W}}\log(\det\{\mathsf{W}\}) = \frac{1}{\det\{\mathsf{W}\}}\frac{\partial \det\{\mathsf{W}\}}{\partial \mathsf{W}} \tag{83}$$

将 (81) 式和 (82) 式的结果代入 (83) 式，可得

$$\frac{\partial}{\partial \mathsf{W}}\log(\det\{\mathsf{W}\}) = (\mathsf{W}^{-1})^{\mathsf{T}} \tag{84}$$

W 的元素相互独立.

$$\frac{\partial}{\partial \mathsf{W}} \log(\det\{\mathsf{W}\}) = 2\mathsf{W}^{-1} - \mathrm{diag}\left\{\mathsf{W}^{-1}\right\} \qquad (85)$$

W 为对称阵, **W**$^{\#}$ 是 **W** 伴随矩阵，$\mathsf{W}^{-1} = \mathsf{W}^{\#}/\det\{\mathsf{W}\}$.
行列式梯度其他性质还有：
(1) 满秩矩阵的行列式的梯度

$$\frac{\partial \det\left\{\mathbf{W}^{-1}\right\}}{\partial \mathbf{W}} = -\det{}^{-1}\{\mathbf{W}\}\left(\mathbf{W}^{-1}\right)^{\mathsf{T}} \qquad (86)$$

(2) 两个矩阵乘积的行列式的梯度

$$\begin{aligned}
\frac{\partial \det\{\mathsf{W}^{\mathsf{T}}\}}{\partial \mathsf{W}} &= 2\det\left\{\mathsf{W}\mathsf{W}^{\mathsf{T}}\right\}\left(\mathsf{W}\mathsf{W}^{\mathsf{T}}\right)^{-1}\mathsf{W}, \quad \mathrm{rank}\left\{\mathsf{W}_{\max n}\right\} = m\\
\frac{\partial \det\{\mathsf{W}^{\mathsf{T}}\mathsf{W}\}}{\partial \mathsf{W}} &= 2\det\left\{\mathsf{W}^{\mathsf{T}}\mathsf{W}\right\}\mathsf{W}\left(\mathsf{W}^{\mathsf{T}}\mathsf{W}\right)^{-1}, \quad \mathrm{rank}\left\{\mathsf{W}_{m\times n}\right\} = n\\
\frac{\partial \det\{\mathsf{W}^{2}\}}{\partial \mathsf{W}} &= 2\det{}^{2}\{\mathsf{W}\}\left(\mathsf{W}^{-1}\right)^{\mathsf{T}}, \quad \mathrm{rank}\left\{\mathsf{W}_{m\times m}\right\} = m
\end{aligned} \qquad (87)$$

(3) 三个矩阵乘积的行列式的梯度

$$\frac{\partial \det\{\mathbf{AWB}\}}{\partial \mathbf{W}} = \det\{\mathbf{AWB}\}\mathbf{A}^T\left(\mathbf{B}^T\mathbf{W}^T\mathbf{A}^T\right)^{-1}\mathbf{B}^T$$

$$\frac{\partial \det\left\{\mathbf{W}^T\mathbf{AW}\right\}}{\partial \mathbf{W}} = 2\mathbf{AW}\left(\mathbf{W}^T\mathbf{AW}\right)^{-1}, \quad \det\left\{\mathbf{W}^T\mathbf{AW}\right\} > 0$$

$$\begin{aligned}\frac{\partial \det\left\{\mathbf{WAW}^T\right\}}{\partial \mathbf{W}} &= \left[\left(\mathbf{WAW}^T\right)^{-1}\right]^T\mathbf{W}\left(\mathbf{A}^T+\mathbf{A}\right)\\ &= 2\left(\mathbf{WAW}^T\right)^{-1}\mathbf{WA}, \quad \mathbf{A}=\mathbf{A}^T\end{aligned} \tag{88}$$

5. 复矢量函数的梯度
首先引入复数求导的定义，若 $f(w)$ 是复数 w 的函数，其中 $w=w_r+jw_i, w^*=w_r-jw_i$，则 $f(w)$ 对 w 求导定义如下：

$$\begin{aligned}\frac{df(w)}{dw}&=\frac{1}{2}\left[\frac{\partial f(w)}{\partial w_r}-j\frac{\partial f(w)}{\partial w_i}\right]\\ \frac{df(w)}{dw^*}&=\frac{1}{2}\left[\frac{\partial f(w)}{\partial w_r}+j\frac{\partial f(w)}{\partial w_i}\right]\end{aligned}\tag{89}$$

那么，目标函数 $f(\mathbf{w})$ 相对于复矢量 $\mathbf{w}$ 的梯度定义为

$$\frac{\partial f(\mathbf{w})}{\partial \mathbf{w}}=\nabla_{\mathbf{w}}f(\mathbf{w})=\left[\frac{\partial f(\mathbf{w})}{\partial w_1},\frac{\partial f(\mathbf{w})}{\partial w_2},\cdots,\frac{\partial f(\mathbf{w})}{\partial w_n}\right]^T\tag{90}$$

而目标函数 $f(\mathbf{w})$ 相对于复数共扼矢量 $\mathbf{w}^*$ 的梯度常简称为共枙梯度 (矢量)，定义为

$$\frac{\partial f(w)}{\partial w^*} = \nabla_w \cdot f(w) = \left[\frac{\partial f(w)}{\partial w_i^*}, \frac{\partial f(w)}{\partial w_2^*}, \cdots, \frac{\partial f(w)}{\partial w_n^*}\right]^T \tag{91}$$

若 $\mathbf{f}(\mathbf{w}) = [f_1(\mathbf{w}), f_2(\mathbf{w}), \cdots, \mathbf{f}_m(\mathbf{w})]$ 则 $f(\mathbf{w})$ 相对于复列矢量 $\mathbf{w}$ 的梯度为一 $n \times m$ 矩阵，并定义为

$$\frac{\partial \mathbf{f}(\mathbf{w})}{\partial \mathbf{w}} = \begin{bmatrix} \frac{\partial f_1(\mathbf{w})}{\partial \mathbf{w}_1} & \frac{\partial f_2(\mathbf{w})}{\partial w_1} & \cdots & \frac{\partial f_m(\mathbf{w})}{\partial w_1} \\ \frac{\partial f_1(\mathbf{w})}{\partial w_2} & \frac{\partial f_2(\mathbf{w})}{\partial w_2} & \cdots & \frac{\partial f_m(\mathbf{w})}{\partial w_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1(w)}{\partial w_n} & \frac{\partial f_2(w)}{\partial w_n} & \cdots & \frac{\partial f_m(\mathbf{w})}{\partial w_n} \end{bmatrix} \tag{92}$$

类似地，行矢量函数 $f(\mathbf{w})$ 相对于复共枙列矢量 $\mathbf{w}^*$ 的梯度称为共枙梯度矩阵，定义为

$$\frac{\partial f(w)}{\partial w^*}=\begin{bmatrix} \frac{\partial f_1(w)}{\partial w_1^*} & \frac{\partial f_2(w)}{\partial w_i^*} & \cdots & \frac{\partial f_m(w)}{\partial w_i^*} \\ \frac{\partial f_1(w)}{\partial w_i^*} & \frac{\partial f_2(w)}{\partial w_i^*} & \cdots & \frac{\partial f_m(w)}{\partial w_i^*} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1(w)}{\partial w_n^*} & \frac{\partial f_2(w)}{\partial w_n^*} & \cdots & \frac{\partial f_m(w)}{\partial w_n^*} \end{bmatrix} \tag{93}$$

可得

$$\frac{\partial \mathbf{w}^{T}}{\partial \mathbf{w}} = \frac{\partial w^{H}}{\partial \mathbf{w}^{*}} = \mathbf{I} \tag{94}$$

$$\frac{\partial w^{T}}{\partial w^{*}} = \frac{\partial w^{H}}{\partial w} = 0 \tag{95}$$

其中，1 和 0 分别为单位矩阵和零矩阵。
由以上两式可得：任何一个复矢量 $\mathbf{w}$ 和它的共轭矢量 $\mathbf{w}^*$ 都可以当作两个独立的复变元处理，即在求梯度的过程中，复矢量 $\mathbf{w}$ 相对于其共轭矢量 $\mathbf{w}^*$ 可视为一常数；反之，$\mathbf{w}^*$ 相对于 $\mathbf{w}$ 也可视为一常数。
令 $f(\mathbf{w})$ 是复矢量 $\mathbf{w}$ 的实值函数。通过将 $\mathbf{w}$ 和 $\mathbf{w}^*$ 视为独立的变元，实目标函数 $f(\mathbf{w})$ 的曲率方向由共轭梯度矢量 $\nabla_{w} \cdot f(\mathbf{w})$ 给出 (曲率方向就是函数最大变化率方向)。

常用标量函数的共扼梯度公式：
(1) 若 $f(\mathbf{x}) = c$ 为常数，则共轭梯度

$$\frac{\partial c}{\partial x^*} = 0 \tag{96}$$

(2) 若 $n \times l$ 矢量 $\mathbf{a}$ 为与 $\mathbf{x}$ 无关的常数矢量，则

$$\frac{\partial \mathbf{a}^H \mathbf{x}}{\partial \mathbf{x}^*} = \mathbf{0}, \quad \frac{\partial \mathbf{x}^H \mathbf{a}}{\partial \mathbf{x}^*} = \mathbf{a} \tag{97}$$

(3) 若 $\mathbf{A}$ 是一个与矢量 x, y 无关的矩阵，则

$$\begin{aligned} \frac{\partial x^H A x}{\partial x} &= A^T x^*, & \frac{\partial x^H A x}{\partial x^*} &= Ax \\ \frac{\partial x^H A y}{\partial A} &= x^* y^T, & \frac{\partial x^H A x}{\partial A} &= x^* x^T \end{aligned} \tag{98}$$

(4) 迹函数的共扼梯度公式

$$\frac{\partial\,\mathrm{tr}\left\{x^H y\right\}}{\partial x^*} = \frac{\partial\,\mathrm{tr}\left\{y x^H\right\}}{\partial x^*} = y$$
$$\frac{\partial\,\mathrm{tr}\left\{BA^H\right\}}{\partial A^*} = \frac{\partial\,\mathrm{tr}\left\{A^H B\right\}}{\partial A^*} = B$$
$$\frac{\partial\,\mathrm{tr}\left\{A^H\right\}}{\partial A^*} = I, \quad \frac{\partial\,\mathrm{tr}\{A\}}{\partial A^*} = 0$$
$$\frac{\partial\,\mathrm{tr}\left\{A^H W A\right\}}{\partial A^*} = WA, \quad \frac{\partial\,\mathrm{tr}\left\{AWA^H\right\}}{\partial A^*} = AW \tag{99}$$

小结

随机信号和统计信号处理是机器学习和信号处理的数学基础，有了随机信号和统计信号处理构建出的思维，会容易很难理解目前机器学习，深度学习和神经网络的公共数学基础及其推导过程。在学习随机信号和统计信号处理前，需要具备两方面的知识：一是理解随机变量，二是涉概率论。随机信号和统计信号处理中的估计理论还需要一些数字信号处理中的 Z 变换、傅立叶变换和滤波器等知识。