

代数与矩阵基础

很多工程问题都可以通过数学建模转化成线性方程组，而矩阵是描述和求解线性方程组最基本和最常用的数学工具。本章将介绍矩阵的基本数学运算和重要性质。

1.1 代数与矩阵的基本概念

1.1.1 代数基本概念

群、环、域是代数的基本概念，但限于篇幅，不予赘述。这里主要介绍线性空间和线性映射（线性变换）。

在抽象代数中，域是一种可进行加、减、乘、除四则运算的代数结构。域的概念是数域及四则运算的推广。

首先，引入几个基本符号： P 代表数域， $\mathbb{R}$ 代表实数域， $\mathbb{C}$ 表示复数域， $\mathbb{Z}$ 为整数域。

一个 m 维列向量定义为

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \quad (1.1.1)$$

若其元素 x_i 取实数，即 $x_i \in \mathbb{R}$ ，则称其为 m 维实（数）向量，并记作 $\mathbf{x} \in \mathbb{R}^{m \times 1}$ ，或者简记为 $\mathbf{x} \in \mathbb{R}^m$ 。类似地，若 $x_i \in \mathbb{C}$ ，则称其为 m 维复向量，并记作 $\mathbf{x} \in \mathbb{C}^m$ 。

一个 m 维行向量定义为 $\mathbf{x} = [x_1, x_2, \dots, x_m]$ ，记作 $\mathbf{x} \in \mathbb{R}^{1 \times m}$ 或 $\mathbf{x} \in \mathbb{C}^{1 \times m}$ 。为了书写的方便，常将一个 m 维列向量写成 m 维行向量的转置形式，即 $\mathbf{x} = [x_1, x_2, \dots, x_m]^T$ 。

一个 $m \times n$ 矩阵定义为

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = [a_{ij}]_{i=1, j=1}^{m, n} \quad (1.1.2)$$

若其元素 $a_{ij} \in \mathbb{R}$ ，则称其为 $m \times n$ 实矩阵，用符号表示为 $\mathbf{A} \in \mathbb{R}^{m \times n}$ 。类似地， $\mathbf{A} \in \mathbb{C}^{m \times n}$ 表示 $\mathbf{A}$ 是一个 $m \times n$ 复矩阵。

$m \times n$ 矩阵可以利用其列向量 $\mathbf{a}_j = [a_{1j}, a_{2j}, \dots, a_{mj}]^T (j = 1, 2, \dots, n)$ 表示为 $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n]$ 。

定义 1.1.1 线性空间是指在一个集合 S 上定义了加法 (且对加法是交换群) 和数乘运算, 且数乘满足下列线性规则, 即 $\forall \alpha, \beta \in P, \forall x, y \in S$, 均有

$$\alpha(x + y) = \alpha x + \alpha y$$

$$(\alpha + \beta)x = \alpha x + \beta x$$

$$\alpha(\beta x) = (\alpha\beta)x$$

例 1.1.1 $\mathbb{R}^n$ 是线性空间, 其中“加法”运算定义为

$$\mathbf{a} + \mathbf{b} = [a_1 + b_1, a_2 + b_2, \dots, a_n + b_n]^T$$

数乘运算为

$$\lambda \mathbf{a} = [\lambda a_1, \lambda a_2, \dots, \lambda a_n]^T$$

其中 $\mathbf{a} = [a_1, a_2, \dots, a_n]^T$ 和 $\mathbf{b} = [b_1, b_2, \dots, b_n]^T$ 。

函数有定义域 (domain) 与值域 (range)。定义域是函数自变量所有可取值的集合; 值域则是由定义域中一切元素所能产生的所有函数值的集合。

一个 $m \times n$ 矩阵 $\mathbf{A} \in \mathbb{C}^{m \times n}$ 所对应的线性变换 $\mathbf{x} = \mathbf{A}\mathbf{y} (\mathbf{y} \in \mathbb{C}^n)$ 的值域定义为 $\text{Range}(\mathbf{A}) = \{\mathbf{x} | \mathbf{x} = \mathbf{A}\mathbf{y}, \mathbf{y} \in \mathbb{C}^n\}$, 而零空间 (null space) 则是矩阵方程 $\mathbf{A}\mathbf{v} = \mathbf{0}$ 的所有解向量 $\mathbf{v}$ 的集合, 也称为 $\mathbf{A}$ 的核或核空间, 常用数学符号表示为 $\text{Null}(\mathbf{A}) = \{\mathbf{v} \in V : \mathbf{A}\mathbf{v} = \mathbf{0}\}$ 。值域 $\text{Range}(\mathbf{A})$ 也写作 $\text{Span}(\mathbf{A})$, 或者简记为 $R(\mathbf{A})$ 。

例 1.1.2 给定 $\mathbf{A} \in \mathbb{C}^{m \times n}$ ($m \times n$ 复系数矩阵), 则该矩阵的值域 $R(\mathbf{A}) = \{\mathbf{x} | \mathbf{x} = \mathbf{A}\mathbf{y}, \mathbf{y} \in \mathbb{C}^n\}$ 和零空间 $N(\mathbf{A}) = \{\mathbf{x} | \mathbf{A}\mathbf{x} = \mathbf{0}\}$ 都是线性空间, 但 $R(\mathbf{A}) \subset \mathbb{C}^m$, 而 $N(\mathbf{A}) \subset \mathbb{C}^n$ 。

例 1.1.3 次数 $\leq n$ 的复系数多项式的全体 $C_n[\lambda] = \{a | a = a_n \lambda^n + \dots + a_1 \lambda + a_0, a_i \in \mathbb{C}\}$ 是线性空间; 但次数 $= n$ 的复系数多项式的集合不是线性空间。

思考: 最小的线性空间是什么?

S 和 Q 同为线性空间, 且 $Q \subset S$, 则称 Q 为 S 的线性子空间。

例 1.1.4 $\mathbf{A} \in \mathbb{C}^{m \times n}$, 则 $R(\mathbf{A})$ 是 $\mathbb{C}^m$ 的子空间, 而 $N(\mathbf{A})$ 是 $\mathbb{C}^n$ 的子空间。

S 是线性空间, 元素组 $\{x_1, x_2, \dots, x_m\} \subset S$ 是线性相关的, 系指存在不全为零的系数 $a_i \in P$, 使 $\sum_{i=1}^m a_i x_i = 0$ 。反之, 称 $\{x_1, x_2, \dots, x_m\}$ 是线性无关的。当 $\{x_1, x_2, \dots, x_m\}$ 线性无关时, $\sum_{i=1}^m a_i x_i = 0 \Rightarrow a_i = 0, i = 1, 2, \dots, m$ 。

若在线性空间 S 中存在 n 个向量线性无关, 而任何 $n+1$ 个向量均线性相关, 则称 S 的维数为 n , 记为 $\dim(S) = n$ 。

S 是域 P 上的线性空间, 元素组 $\{x_1, x_2, \dots, x_n\}$ 称为 S 的一组基, 是指:

① $\{x_1, x_2, \dots, x_n\}$ 线性无关;

② $\text{Span}[x_1, x_2, \dots, x_n] \stackrel{\text{def}}{=} \left\{ \sum_{i=1}^n a_i x_i, a_i \in P \right\} = S$ 。

注: 若 $\forall x \in S, x = \sum_{i=1}^n a_i x_i$, 则坐标向量 $\mathbf{a} = [a_1, a_2, \dots, a_n]^T$ 唯一。

因为线性空间也是集合, 因此可以定义交集和并集。线性空间之间还可以定义“和”运算以及“直(接)和”运算。线性空间的交、并、和、直和四种运算的定义如下:

$$T \cap V = \{x \mid x \in T \text{ 且 } x \in V\}$$

$$T \cup V = \{x \mid x \in T \text{ 或 } x \in V\}$$

$$T + V = \{x \mid x = y + z, y \in T, z \in V\}$$

$$T \oplus V \text{ 是指 } T + V, \text{ 当 } T \cap V = \{0\}$$

思考: 以上哪种集合还是线性空间?

映射 $\sigma: S \rightarrow T$ 称为线性映射 (线性算子), 是指它满足

$$\sigma(\mathbf{a} + \mathbf{b}) = \sigma(\mathbf{a}) + \sigma(\mathbf{b})$$

$$\sigma(\lambda \mathbf{a}) = \lambda \sigma(\mathbf{a})$$

1.1.2 矩阵与向量

科学和工程中的很多问题都可以通过数学建模转化成一个线性方程组

$$\left. \begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned} \right\} \quad (1.1.3)$$

它使用 m 个方程描述 n 个未知量之间的线性关系。

线性方程组 (1.1.3) 的简洁表示形式为

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad (1.1.4)$$

式中

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = [a_{ij}]_{i=1, j=1}^{m, n} \quad (1.1.5)$$

为 $m \times n$ 矩阵, 是一个按照长方阵列排列的实数或复数集合, 其中 a_{ij} 表示矩阵 $\mathbf{A}$ 的第 i 行、第 j 列元素, 简称第 (i, j) 个元素。

在工程问题的数学建模中, 矩阵 $\mathbf{A}$ 往往是某个物理系统的符号表示。其中, $\mathbf{A}$ 代表一线性系统, $\mathbf{x}$ 和 $\mathbf{b}$ 则分别表示该系统的输入激励 (不可观测) 和输出响应 (可观测)。于是, 矩阵方程 $\mathbf{A}\mathbf{x} = \mathbf{b}$ 的求解问题便可叙述为: 根据已知的线性系统参数和输出观测值, 求未知的输入激励。

当 $m = n$ 时, 矩阵 $\mathbf{A}$ 称为正方矩阵 (square matrix); 若 $m < n$, 则矩阵 $\mathbf{A}$ 形象地称为宽矩阵 (broad matrix); 而当 $m > n$ 时, 便称矩阵 $\mathbf{A}$ 为高矩阵 (tall matrix)。

科学和工程中遇到的向量可分为以下 3 种^[59]:

(1) 物理向量 泛指既有幅值, 又有方向的物理量, 如速度、加速度、位移等。

(2) 几何向量 是物理向量的可视化表示, 常用带方向的 (简称“有向”) 线段表示。这种有向线段称为几何向量。例如, $\boldsymbol{v} = \overrightarrow{AB}$ 表示一有向线段, 其起点为 A , 终点为 B 。

(3) 代数向量 几何向量的代数形式表示。例如, 若平面上的几何向量 $\boldsymbol{v} = \overrightarrow{AB}$ 的起点坐标为 $A = (a_1, a_2)$, 终点坐标为 $B = (b_1, b_2)$, 则该几何向量可以用代数形式表示为 $\boldsymbol{v} = \begin{bmatrix} b_1 - a_1 \\ b_2 - a_2 \end{bmatrix}$ 。这种用代数形式表示的几何向量称为代数向量。

根据元素取值种类的不同, 代数向量又可分为以下 3 种:

(1) 常数向量 向量的元素全部为实常数或者复常数, 例如 $\boldsymbol{a} = [1, 5, 4]^T$ 等。

(2) 函数向量 向量的元素包含了函数值, 例如 $\boldsymbol{x} = [1, x^2, \dots, x^n]^T$ 等。

(3) 随机向量 向量的元素为随机变量, 如 $\boldsymbol{x}(n) = [x_1(n), x_2(n), \dots, x_m(n)]^T$, 其中 $x_1(n), x_2(n), \dots, x_m(n)$ 是 m 个随机过程或随机信号。

图 1.1.1 归纳了向量的分类。

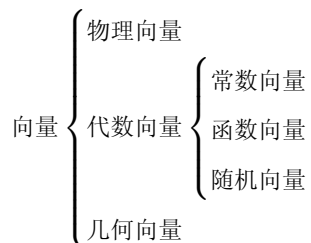


图 1.1.1 向量的分类

工程应用中遇到的往往是物理向量, 几何向量是物理向量的可视化表示, 而代数向量则可看作是物理向量的运算化工具。

一个 $n \times n$ 正方矩阵 $\boldsymbol{A}$ 的主对角线是指从左上角到右下角沿 $i = j, j = 1, 2, \dots, n$ 相连接的线段。与主对角线平行的对角线称为次对角线。位于主对角线上的元素称为 $\boldsymbol{A}$ 的对角元素, 它们是 $a_{ii}, i = 1, 2, \dots, n$ 。

主对角线以外元素全部为零的 $n \times n$ 矩阵称为对角矩阵, 记作

$$\boldsymbol{D} = \text{diag}(d_{11}, d_{22}, \dots, d_{nn}) \quad (1.1.6)$$

若对角矩阵主对角线元素全部等于 1, 则称其为单位矩阵, 用符号 $\boldsymbol{I}_{n \times n}$ 示之。所有元素为零的 $m \times n$ 矩阵称为零矩阵, 记为 $\boldsymbol{O}_{m \times n}$ 。

一个全部元素为零的向量称为零向量。为了书写的简洁, 单位矩阵、零矩阵和零向量分别表示为 $\boldsymbol{I}, \boldsymbol{O}$ 和 $\mathbf{0}$ 。

1.1.3 矩阵的基本运算

矩阵的基本运算包括矩阵的转置、共轭、共轭转置、加法、乘法和求逆。

1. 矩阵的转置

若 $\boldsymbol{A} = [a_{ij}]$ 是一个 $m \times n$ 矩阵, 则 $\boldsymbol{A}$ 的转置记作 $\boldsymbol{A}^T$, 它是一个 $n \times m$ 矩阵, 定义为

$$\mathbf{A}^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix} \quad (1.1.7)$$

矩阵 $\mathbf{A}$ 的复数共轭 $\mathbf{A}^*$ 仍然是一个 $m \times n$ 矩阵, 定义为

$$\mathbf{A}^* = \begin{bmatrix} a_{11}^* & a_{12}^* & \cdots & a_{1n}^* \\ a_{21}^* & a_{22}^* & \cdots & a_{2n}^* \\ \vdots & \vdots & & \vdots \\ a_{m1}^* & a_{m2}^* & \cdots & a_{mn}^* \end{bmatrix} \quad (1.1.8)$$

矩阵 $\mathbf{A}$ 的 (复) 共轭转置 $\mathbf{A}^H$ 是一个 $n \times m$ 矩阵, 定义为

$$\mathbf{A}^H = \begin{bmatrix} a_{11}^* & a_{12}^* & \cdots & a_{m1}^* \\ a_{21}^* & a_{22}^* & \cdots & a_{m2}^* \\ \vdots & \vdots & & \vdots \\ a_{n1}^* & a_{n2}^* & \cdots & a_{mn}^* \end{bmatrix} \quad (1.1.9)$$

共轭转置又叫 Hermitian 转置或 Hermitian 共轭。

满足 $\mathbf{A}^T = \mathbf{A}$ 的正方实矩阵和 $\mathbf{A}^H = \mathbf{A}$ 的正方复矩阵分别称为对称矩阵和 Hermitian 矩阵 (复共轭对称矩阵)。共轭转置与转置之间存在关系

$$\mathbf{A}^H = (\mathbf{A}^*)^T = (\mathbf{A}^T)^* \quad (1.1.10)$$

2. 矩阵的加法与乘法

两个 $m \times n$ 矩阵 $\mathbf{A} = [a_{ij}]$ 和 $\mathbf{B} = [b_{ij}]$ 之和记作 $\mathbf{A} + \mathbf{B}$, 定义为 $[\mathbf{A} + \mathbf{B}]_{ij} = a_{ij} + b_{ij}$ 。矩阵的加法服从下面的运算法则:

(1) 加法交换律 $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$

(2) 加法结合律 $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$

矩阵的乘法分为以下三种:

矩阵与标量的乘法 $m \times n$ 矩阵 $\mathbf{A} = [a_{ij}]$ 与标量 α 的乘积的元素定义为 $[\alpha \mathbf{A}]_{ij} = \alpha a_{ij}$ 。

矩阵与向量的乘法 $m \times n$ 矩阵 $\mathbf{A} = [a_{ij}]$ 与 $r \times 1$ 向量 $\mathbf{x} = [x_1, x_2, \dots, x_r]^T$ 的乘积 $\mathbf{Ax}$ 只有当 $n = r$ 时才存在, 它是一个 $m \times 1$ 向量, 定义为

$$[\mathbf{Ax}]_i = \sum_{j=1}^n a_{ij} x_j, \quad i = 1, 2, \dots, m$$

两个矩阵的乘法 $m \times n$ 矩阵 $\mathbf{A} = [a_{ij}]$ 与 $r \times s$ 矩阵 $\mathbf{B} = [b_{ij}]$ 的乘积 $\mathbf{AB}$ 只有当 $n = r$ 时才存在, 它是一个 $m \times s$ 矩阵, 定义为

$$[\mathbf{AB}]_{ij} = \sum_{k=1}^n a_{ik} b_{kj}, \quad i = 1, 2, \dots, m; \quad j = 1, 2, \dots, s$$

矩阵的乘积服从下面的运算法则:

(1) 乘法结合律 若 $A \in \mathbb{C}^{m \times n}$, $B \in \mathbb{C}^{n \times p}$, $C \in \mathbb{C}^{p \times q}$, 则 $A(BC) = (AB)C$ 。

(2) 乘法左分配律 若 A 和 B 是两个 $m \times n$ 矩阵, 且 C 是一个 $n \times p$ 矩阵, 则 $(A+B)C = AC + BC$ 。

(3) 乘法右分配律 若 A 是一个 $m \times n$ 矩阵, 并且 B 和 C 是两个 $n \times p$ 矩阵, 则 $A(B+C) = AB + AC$ 。

(4) 若 α 是一个标量, 并且 A 和 B 是两个 $m \times n$ 矩阵, 则 $\alpha(A+B) = \alpha A + \alpha B$ 。

注意, 矩阵的乘法一般不满足交换律, 即 $AB \neq BA$ 。

3. 矩阵的求逆

矩阵与向量的乘积 $Ax = y$ 可视为向量 x 的线性变换。此时, 矩阵 A 称为线性变换矩阵。若 A 为 $n \times n$ 矩阵且向量 y 到 x 的线性逆变换 A^{-1} 存在, 则

$$x = A^{-1}y \quad (1.1.11)$$

这相当于原线性变换 $Ax = y$ 两边左乘 A^{-1} 得到的结果 $A^{-1}Ax = A^{-1}y$ 。因此, 线性逆变换 A^{-1} 应该满足 $A^{-1}A = I$ 之关系。另一方面, $x = A^{-1}y$ 也应该是可逆的, 即两边左乘 A 后得到的 $Ax = AA^{-1}y$ 应该与原线性变换 $Ax = y$ 一致, 故 A^{-1} 还应该满足 $AA^{-1} = I$ 。

定义 1.1.2 令 A 是一个 $n \times n$ 矩阵。若 $n \times n$ 矩阵 A^{-1} 满足 $AA^{-1} = A^{-1}A = I$, 则称矩阵 A 可逆, 并称 A^{-1} 是矩阵 A 的逆矩阵。

下面是共轭、转置、共轭转置和逆矩阵的性质。

(1) 矩阵的共轭、转置和共轭转置满足分配律

$$(A+B)^* = A^* + B^*, \quad (A+B)^T = A^T + B^T, \quad (A+B)^H = A^H + B^H$$

(2) 矩阵乘积的转置、共轭转置和逆矩阵满足关系式

$$\begin{aligned} (AB)^T &= B^T A^T, \quad (AB)^H = B^H A^H \\ (AB)^{-1} &= B^{-1} A^{-1} \quad (A, B \text{ 为可逆的正方矩阵}) \end{aligned}$$

(3) 共轭、转置和共轭转置等符号均可与求逆符号交换, 即有

$$A^{-*} = (A^*)^{-1} = (A^{-1})^*, \quad A^{-T} = (A^T)^{-1} = (A^{-1})^T, \quad A^{-H} = (A^H)^{-1} = (A^{-1})^H$$

1.2 矩阵的初等变换

涉及矩阵行与行 (或者列与列) 之间的简单运算称为初等行变换或者初等列变换, 二者统称初等变换。在应用中, 矩阵的初等运算往往可以解决一些重要问题。例如, 只使用初等行运算, 就可以解决矩阵方程求解、矩阵求逆和矩阵基本空间的基向量构造等复杂问题。又如, 在矩阵的相似变换中, 也需要使用矩阵的初等变换。

1.2.1 初等行变换与阶梯型矩阵

定义 1.2.1 令矩阵 $A \in \mathbb{C}^{m \times n}$ 的 m 个行向量分别为 $r_1, r_2, \dots, r_m$ 。下列运算称为矩阵 A 的初等行运算 (elementary row operation) 或初等行变换 (elementary row transformation):

- (1) 互换矩阵的任意两行, 如 $r_p \leftrightarrow r_q$, 称为 I 型初等行变换。
- (2) 一行元素同乘一个非零常数 α , 如 $\alpha r_p \rightarrow r_p$, 称为 II 型初等行变换。
- (3) 将第 p 行元素同乘一个非零常数 β 后, 加给第 q 行, 即 $\beta r_p + r_q \rightarrow r_q$, 称为 III 型初等行变换。

假设矩阵 $A_{m \times n}$ 经过一系列初等行运算, 变换成为矩阵 $B_{m \times n}$, 则称矩阵 A 和 B 为行等价矩阵 (row equivalent matrix)。

一个非零行最左边的非零元素称为该行的首项元素 (leading entry)。如果首项元素等于 1, 便称为首一元素 (leading 1 entry)。

从线性方程组的求解以及基本空间的基向量构造等实际应用出发, 往往希望将一个矩阵经过初等行运算之后, 变换为阶梯型矩阵。

定义 1.2.2 一个 $m \times n$ 矩阵称为阶梯型 (echelon form) 矩阵, 若下列条件都满足:

- (1) 全部由零组成的所有行都位于矩阵的底部。
- (2) 每一个非零行的首项元素总是出现在上一个非零行的首项元素的右边。
- (3) 首项元素下面的同列元素全部为零。

例如, 下面是阶梯型矩阵的两个例子:

$$A = \begin{bmatrix} 2 & * & * \\ 0 & 5 & * \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & * & * \\ 0 & 3 & * \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

式中, $*$ 表示该元素可以为任意值。

定义 1.2.3 ^[59] 阶梯型矩阵 A 称为行简约阶梯型 (row reduced echelon form, RREF), 若 A 的每一非零行的首项元素等于 1 (即为首一元素), 并且每一个首一元素也是它所在列唯一的非零元素。

行简约阶梯型也称行阶梯标准型或 Hermite 标准型。

给定一个 $m \times n$ 矩阵 B , 下面的算法通过初等行变换将 B 化成行简约阶梯型矩阵。

算法 1.2.1 将 $m \times n$ 矩阵化成行简约阶梯型 ^[59]

- 步骤 1 将含有一个非零元素的列设定为最左边的第 1 列。
- 步骤 2 若需要, 将第 1 行与其他行互换, 使第 1 个非零列在第 1 行有一个非零元素。
- 步骤 3 如果第 1 行的首项元素为 a , 则将该行的所有元素乘以 $1/a$, 以使该行的首项元素等于 1, 成为首一元素。
- 步骤 4 通过初等行变换, 将其他行位于第 1 行首一元素下面的全部元素变成 0。
- 步骤 5 对第 $i = 2, 3, \dots, m$ 行依次重复以上步骤, 以使每一行的首一元素出现在上一行的首一元素的右边, 并使与第 i 行首一元素同列的其他各行元素都变为 0。

定理 1.2.1 任何一个矩阵 $\mathbf{A}_{m \times n}$ 都与一个并且唯一的一个行简约阶梯型矩阵是行等价的。

证明 参见文献 [63, Appendix A]。 ■

当矩阵的初等行变换产生一个行阶梯型矩阵时, 如果将行阶梯型矩阵进一步简化为行简约阶梯型, 则相应的初等行变换将不会改变行阶梯型矩阵各非零行首项元素的位置。也就是说, 任何一个矩阵的行阶梯型的首项元素与行简约阶梯型的首一元素总是处于相同的位置。由此可以引出下面的定义。

定义 1.2.4 [63, p.15] 矩阵 $\mathbf{A}_{m \times n}$ 的主元位置 (pivot position) 就是矩阵 $\mathbf{A}$ 中与其阶梯型的首项元素相对应的位置。矩阵 $\mathbf{A}$ 中包含主元位置的每一列都称为 $\mathbf{A}$ 的主元列 (pivot column)。

下面的例子说明如何通过初等行运算, 将一个矩阵变换为行阶梯型和行简约阶梯型, 以及如何判断原矩阵的主元列。

例 1.2.1 已知 3×5 矩阵

$$\mathbf{A} = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}$$

第 2 行乘 -2 , 加到第 3 行; 并且第 2 行乘 3, 加到第 1 行, 则

$$\mathbf{A} \sim \begin{bmatrix} 0 & 0 & 5 & 10 & -10 \\ 1 & -2 & 2 & 3 & -1 \\ 0 & 0 & 1 & 2 & -2 \end{bmatrix}$$

第 1 行乘 $-2/5$, 加到第 2 行; 同时第 1 行乘 $-1/5$, 加到第 3 行, 得

$$\mathbf{A} \sim \begin{bmatrix} 0 & 0 & 5 & 10 & -10 \\ 1 & -2 & 0 & -1 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

交换第 1 行和第 2 行, 又得到

$$\mathbf{A} \sim \begin{bmatrix} 1 & -2 & 0 & -1 & 3 \\ 0 & 0 & 5 & 10 & -10 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (\text{阶梯型})$$

下面画横杠的元素所在的位置称为主元位置。因此, 矩阵 $\mathbf{A}$ 的主元列为第 1 列和第 3 列, 即有

$$\begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix} \quad (\mathbf{A} \text{ 的主元列})$$

进一步地, 阶梯型矩阵的第 2 行乘以 $1/5$, 行阶梯型简化为

$$\mathbf{A} \sim \begin{bmatrix} 1 & -2 & 0 & -1 & 3 \\ 0 & 0 & 1 & 2 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (\text{行简约阶梯型})$$

1.2.2 初等行变换的两个应用

下面介绍初等行变换的两个重要应用：线性方程组求解和矩阵求逆。

1. 线性方程组求解

考察 $n \times n$ 线性方程组 $\mathbf{Ax} = \mathbf{b}$ 的求解，其中矩阵 $\mathbf{A}$ 存在逆矩阵 $\mathbf{A}^{-1}$ 。现在，希望通过初等行变换，得到方程的解 $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ 。

$n \times n$ 线性方程组 $\mathbf{Ax} = \mathbf{b}$ 中的 $\mathbf{A}$ 和 $\mathbf{b}$ 分别称作数据矩阵和数据向量，而 $\mathbf{x}$ 称为未知数 (或未知参数) 向量。为了方便讨论线性方程组的求解，常将数据矩阵和数据向量组合成一个 $n \times (n+1)$ 的新矩阵 $\mathbf{B} = [\mathbf{A}, \mathbf{b}]$ ，并称为线性方程组 $\mathbf{Ax} = \mathbf{b}$ 的增广矩阵。

注意到线性方程组的解 $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ 也可以写成线性方程组的形式 $\mathbf{Ix} = \mathbf{A}^{-1}\mathbf{b}$ ，其对应的增广矩阵为 $[\mathbf{I}, \mathbf{A}^{-1}\mathbf{b}]$ 。于是，我们可以将线性方程组的这一求解过程与它们对应的增广矩阵形式分别书写为

$$\begin{array}{lll} \text{方程求解} & \mathbf{Ax} = \mathbf{b} & \xrightarrow{\text{初等行变换}} \mathbf{x} = \mathbf{A}^{-1}\mathbf{b} \\ \text{增广矩阵} & [\mathbf{A}, \mathbf{b}] & \xrightarrow{\text{初等行变换}} [\mathbf{I}, \mathbf{A}^{-1}\mathbf{b}] \end{array}$$

这表明，若对增广矩阵 $[\mathbf{A}, \mathbf{b}]$ 使用初等行变换，使得左边变成一个 $n \times n$ 单位矩阵，则变换后的增广矩阵的第 $n+1$ 列给出原线性方程组的解 $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ 。这样一种求解线性方程组的初等行变换方法称为高斯消去 (Gauss elimination) 法或 Gauss-Jordan 消去法。

例 1.2.2 用高斯消去法求解线性方程组

$$\begin{cases} x_1 + x_2 + 2x_3 = 6 \\ 3x_1 + 4x_2 - x_3 = 5 \\ -x_1 + x_2 + x_3 = 2 \end{cases}$$

对其增广矩阵进行初等行变换

$$\begin{aligned} & \begin{bmatrix} 1 & 1 & 2 & 6 \\ 3 & 4 & -1 & 5 \\ -1 & 1 & 1 & 2 \end{bmatrix} \xrightarrow{\text{第 2 行减去第 1 行的 3 倍}} \begin{bmatrix} 1 & 1 & 2 & 6 \\ 0 & 1 & -7 & -13 \\ -1 & 1 & 1 & 2 \end{bmatrix} \xrightarrow{\text{第 1 行加到第 3 行}} \\ & \begin{bmatrix} 1 & 1 & 2 & 6 \\ 0 & 1 & -7 & -13 \\ 0 & 2 & 3 & 8 \end{bmatrix} \xrightarrow{\text{第 1 行减去第 2 行}} \begin{bmatrix} 1 & 0 & 9 & 19 \\ 0 & 1 & -7 & -13 \\ 0 & 2 & 3 & 8 \end{bmatrix} \xrightarrow{\text{第 3 行减去第 2 行的 2 倍}} \\ & \begin{bmatrix} 1 & 0 & 9 & 19 \\ 0 & 1 & -7 & -13 \\ 0 & 0 & 17 & 34 \end{bmatrix} \xrightarrow{\text{第 3 行乘以 } 1/17} \begin{bmatrix} 1 & 0 & 9 & 19 \\ 0 & 1 & -7 & -13 \\ 0 & 0 & 1 & 2 \end{bmatrix} \xrightarrow{\text{第 1 行减去第 3 行的 9 倍}} \\ & \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & -7 & -13 \\ 0 & 0 & 1 & 2 \end{bmatrix} \xrightarrow{\text{第 3 行乘以 7 后，再添加到第 2 行}} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix} \end{aligned}$$

即通过高斯消去法得到方程组的解为 $x_1 = 1$, $x_2 = 1$ 和 $x_3 = 2$ 。

初等行变换方法也适用于系数矩阵为 $m \times n$ 矩阵的线性方程组 $\mathbf{Ax} = \mathbf{b}$ 的求解。此时, 需要将增广矩阵化成行简约阶梯型矩阵。具体算法如下。

算法 1.2.2 系数矩阵为 $m \times n$ 矩阵的线性方程组 $\mathbf{Ax} = \mathbf{b}$ 的求解^[59]

步骤 1 构造增广矩阵 $\mathbf{B} = [\mathbf{A}, \mathbf{b}]$ 。

步骤 2 使用算法 1.2.1 将增广矩阵 $\mathbf{B}$ 变成行简约阶梯型, 它与原增广矩阵等价。

步骤 3 从简化的矩阵得到对应的线性方程组, 它与原线性方程组等价。

步骤 4 得到新的线性方程组的通解 (general solution)。

例 1.2.3 考察线性方程组及其增广矩阵

$$\begin{cases} 2x_1 + 2x_2 - x_3 = 1 \\ -2x_1 - 2x_2 + 4x_3 = 1 \\ 2x_1 + 2x_2 + 5x_3 = 5 \\ -2x_1 - 2x_2 - 2x_3 = -3 \end{cases} \quad \text{和} \quad \mathbf{B} = \begin{bmatrix} 2 & 2 & -1 & 1 \\ -2 & -2 & 4 & 1 \\ 2 & 2 & 5 & 5 \\ -2 & -2 & -2 & -3 \end{bmatrix}$$

第 1 行元素乘以 $1/2$, 使第 1 个元素为 1

$$\begin{bmatrix} 2 & 2 & -1 & 1 \\ -2 & -2 & 4 & 1 \\ 2 & 2 & 5 & 5 \\ -2 & -2 & -2 & -3 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & -\frac{1}{2} & \frac{1}{2} \\ -2 & -2 & 4 & 1 \\ 2 & 2 & 5 & 5 \\ -2 & -2 & -2 & -3 \end{bmatrix}$$

利用初等行变换, 使第 2~4 行的第 1 个元素都变成 0

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

第 2 行元素乘以 $1/3$, 使得其第 3 列元素等于 1, 即有

$$\begin{bmatrix} 1 & 1 & -\frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 3 & 2 \\ 0 & 0 & 6 & 4 \\ 0 & 0 & -3 & -2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & -\frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 & \frac{2}{3} \\ 0 & 0 & 6 & 4 \\ 0 & 0 & -3 & -2 \end{bmatrix}$$

利用初等行变换, 使第 2 行首项元素 1 的上边和下边的元素全部变为 0, 得到

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

对应的线性方程组为 $x_1 + x_2 = \frac{5}{6}$ 和 $x_3 = \frac{2}{3}$ 。该方程组有无穷多组解, 其通解为 $x_1 = \frac{5}{6} - x_2, x_3 = \frac{2}{3}$ 。若 $x_2 = 1$, 则得一特解 (particular solution) 为 $x_1 = -\frac{1}{6}, x_2 = 1$ 和 $x_3 = \frac{2}{3}$ 。