

矩阵基础知识

🕒 *Posted on May 15, 2019*

☰ Algorithm (<https://paul.pub/category/#Algorithm>) 🔍 Matrix (<https://paul.pub/tags/#Matrix>)
🔍 Singular (<https://paul.pub/tags/#Singular>) 🔍 Eigenvalue (<https://paul.pub/tags/#Eigenvalue>)
🔍 Transpose (<https://paul.pub/tags/#Transpose>)

矩阵是高等代数学中的常见工具，也常见于统计分析等应用数学学科中。在人工智能的项目中，无论是使用机器学习，还是做数值优化，都会用到矩阵的知识。因此借助这篇文章，让我们来一起了解一下有关矩阵的一些基础概念。

- 介绍
- 词源
- 发展
- 标量，向量，矩阵和张量
- 矩阵与线性方程组
 - 增广矩阵
- 矩阵算术
 - 加减法
 - 标量乘法
 - 转置
 - 矩阵乘法
- 范数
- 特殊类型的矩阵
 - 对角矩阵
 - 对称矩阵
 - 正交矩阵
- 方阵
 - 单位矩阵
 - 矩阵逆元
 - 行列式
- 特征值与特征向量
 - 正定性
- 奇异值分解
- 向量空间
 - 子空间
 - 线性无关
 - 基和维数
 - 行空间和列空间

- 矩阵的秩
- 线性变换
 - 推移
 - 镜射
 - 挤压
 - 伸缩
 - 旋转
- Hessian矩阵
- 雅克比矩阵
- 附录：术语中英文对照
- 参考资料与推荐读物

Morpheus: This is your last chance. After this, there is no turning back.

You take the blue pill - the story ends, you wake up in your bed and believe whatever you want to believe.

You take the red pill - you stay in Wonderland and I show you how deep the rabbit-hole goes.

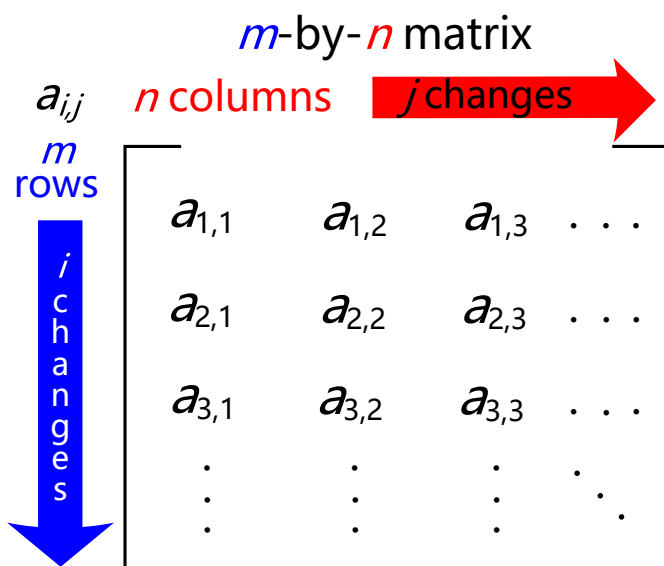
– The Matrix

介绍

将一些元素排列成若干行，每行放上相同数量的元素，就是一个矩阵（Matrix）。

数学上，一个 $m \times n$ 的矩阵是一个由 m 行和 n 列元素排列成的矩形阵列。

矩阵里的元素可以是数字、符号或数学式。



我们通常用大写字母来表示矩阵，例如：A，B，C。用带下标的小写字母描述矩阵中的元素，例如： A_{ij} 。

词源

在古罗马，Matrix (<https://www.merriam-webster.com/dictionary/matrix#note-1>)是为繁衍而饲养的雌性动物，或是其种子用于生产其他植物的植物（有时称为“亲本植物”或“母本植物”）。

中文中矩阵的概念最早见于1922年。北京师范大学附属中学数学老师程廷熙在一篇介绍文章中将矩阵译为“纵横阵”。

1925年，科学名词审查会算学名词审查组在《科学》第十卷第四期刊登的审定名词表中，矩阵被翻译为“矩阵式”，方块矩阵翻译为“方阵式”，而各类矩阵如“正交矩阵”、“伴随矩阵”中的“矩阵”则被翻译为“方阵”。

1935年，中国数学会审查后，中华民国教育部审定的《数学名词》（并“通令全国各院校一律遵用，以昭划一”）中，“矩阵”作为译名首次出现。

1938年，曹惠群在接受科学名词审查会委托就数学名词加以校订的《算学名词汇编》中，认为应当的译名是“长方阵”。中华人民共和国成立后编订的《数学名词》中，则将译名定为“（矩）阵”。

1993年，中国自然科学名词审定委员会公布的《数学名词》中，“矩阵”被定为正式译名，并沿用至今。

发展

日本数学家关孝和（1683年）与微积分的发现者之一戈特弗里德·威廉·莱布尼茨（1693年）近乎同时地独立建立了行列式论。其后行列式作为解线性方程组的工具逐步发展。1750年，加布里尔·克拉默发现了克莱姆法则。

进入十九世纪后，行列式的研究进一步发展，矩阵的概念也应运而生。奥古斯丁·路易·柯西是最早将行列式排成方阵并将其元素用双重下标表示的数学家。他还在1829年就在行列式的框架中证明了实对称矩阵特征根为实数的结论。

其后，詹姆斯·约瑟夫·西尔维斯特注意到，在作为行列式的计算形式以外，将数以行和列的形式作出的矩形排列本身也是值得研究的。在他希望引用数的矩形阵列而又不能用行列式来形容的时候，就用“**matrix**”一词来形容。而在此之前，数学家已经开始将增广矩阵作为独立的对象引用了。

西尔维斯特使用“**matrix**”一词是因为他希望讨论行列式的子式，即将矩阵的某几行和某几列的共同元素取出来排成的矩阵的行列式，所以实际上“**matrix**”被他看做是生成各种子式的“母体”：我在先前的文章中将矩形排布的序列称为“**Matrix**”，盖因从中可以产生出各种不同的行列式，就如由同一个母体的子宫中孕育出来一样。

阿瑟·凯莱被公认为矩阵论的奠基人。他开始将矩阵作为独立的数学对象研究时，许多与矩阵有关的性质已经在行列式的研究中被发现了，这也使得凯莱认为矩阵的引进是十分自然的。他说：“我决然不是通过四元数而获得矩阵概念的；它或是直接从行列式的概念而来，

或是作为一个表达线性方程组的方便方法而来的。”他从1858年开始，发表了《矩阵论的研究报告》等一系列关于矩阵的专门论文，研究了矩阵的运算律、矩阵的逆以及转置和特征多项式方程。

标量，向量，矩阵和张量

- 标量（scalar）：一个标量就是一个单独的数。
- 向量（vector）：一个向量是一列数。这些数是有序排列的。通过次序中的索引，我们可以确定每个单独的数。通常我们赋予向量小写变量名称，比如 x 。向量中的元素可以通过脚标表示。向量 x 的第一个元素是 x_1 ，第二个元素是 x_2 ，等等。可以把向量看作空间中的点，每个元素是不同坐标轴上的坐标。
- 矩阵（matrix）：矩阵是一个二维数组，其中的每一个元素被两个索引所确定。如果一个实数矩阵高度为 m ，宽度为 n ，那么我们说 $A \in R^{m \times n}$ 。
- 张量（tensor）：在某些情况下，我们会讨论坐标超过两维的数组。一般地，一个数组中的元素分布在若干维坐标的规则网格中，我们称之为张量。

矩阵与线性方程组

形如 $a_1x_1 + a_2x_2 + \dots + a_nx_n = b$ 的方程称为含有 n 个未知量的线性方程。

含有 m 个方程和 n 个未知量的线性方程组定义为：

$$\begin{aligned}a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2 \\&\dots \\a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &= b_m\end{aligned}\tag{1}$$

若有序 n 元组 (x_1, x_2, \dots, x_n) 满足方程组中的所有方程，则称其为 $m \times n$ 方程组的解。

线性方程组的所有解的集合称之为方程组的解集。

如果不存在实数能同时满足所有方程，则称方程组无解，或者说其解集为空。

根据解的情况，线性方程组可以分为三类：

- 有唯一解的恰定方程组
- 没有解的超定方程组
- 有无穷多解的欠定方程组

方程组系数所组成的矩阵称之为系数矩阵（coefficient matrix），如下所示：

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & & & \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \quad (2)$$

增广矩阵

增广矩阵（**augmented matrix**），是在线性代数中系数矩阵的右边添上线性方程组等号右边的常数列得到的矩阵。

方程组 $AX = B$ 的系数矩阵为 A ，它的增广矩阵为 $[A|B]$ 。如下所示：

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \dots & & & & \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_n \end{array} \right] \quad (3)$$

如果线性方程组的右端全为零，则称其为齐次（**homogeneous**）线性方程组。

齐次线性方程组总是有解：只有令所有未知量为0即可，这个解称之为平凡解。

另外，若 $n > m$ ，则 $m \times n$ 的齐次线性方程组有非凡解。

矩阵算术

加减法

矩阵的加减法只在两个矩阵具有相同大小时才有意义。

两个矩阵的加（减）法通过对应元素相加（减）得到。

即： $C = A \pm B$ ，其中：

$$C_{ij} = A_{ij} \pm B_{ij} \quad (4)$$

标量乘法

设 A 为一矩阵， b 为一标量，则 bA 为将 A 中的每一个元素乘以 b 而构成的一个矩阵。即：

$$bA = b \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & & & \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} = \begin{bmatrix} ba_{11} & ba_{12} & \dots & ba_{1n} \\ ba_{21} & ba_{22} & \dots & ba_{2n} \\ \dots & & & \\ ba_{m1} & ba_{m2} & \dots & ba_{mn} \end{bmatrix} \quad (5)$$

转置

转置（transpose）是矩阵的重要操作之一。矩阵的转置是以对角线为轴的镜像。

这条从左上角到右下角的对角线被称之为主对角线（main diagonal）。

我们将矩阵A的转置表示为 A^T ，定义如下：

$$(A^T)_{ij} = A_{ji} \quad (6)$$

假设矩阵A如下：

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \ddots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \quad (7)$$

则其转置为：

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \dots & \dots & \ddots & \dots \\ a_{1m} & a_{2m} & \dots & a_{mn} \end{bmatrix} \quad (8)$$

很显然，一个 $m \times n$ 的矩阵转置之后是一个 $n \times m$ 的矩阵。

并且矩阵的转置和数乘运算对加法满足分配律：

$$\begin{aligned} (A + B)^T &= A^T + B^T \\ c(A + B) &= cA + cB \end{aligned} \quad (9)$$

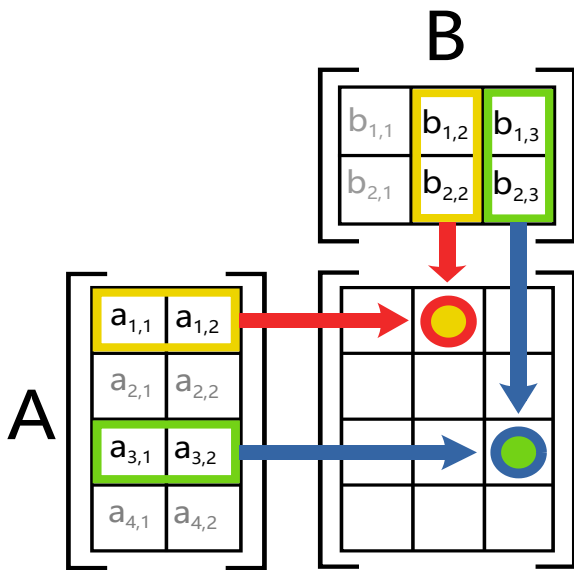
矩阵乘法

如果矩阵A的列数等于矩阵B的行数，则矩阵A可以和矩阵B相乘。

其定义为：若A为一个 $m \times n$ 的矩阵，且B为一个 $n \times r$ 的矩阵，则 $AB = C$ 为一个 $m \times r$ 的矩阵，它的元素定义为：

$$c_{ij} = a_i b_j = \sum_{k=1}^n a_{ik} b_{kj} \quad (10)$$

矩阵乘法的运算方法图示如下：



矩阵的乘法满足结合律和分配律，但不满足交换律。

两个矩阵乘积不是指两个矩阵中对应元素的乘积。不过，这样的操作也存在，称之为元素对应乘积（element-wise product）或者Hadamard乘积，记为 $A \odot B$ 。

范数

我们使用被称为范数（norm）的函数衡量向量的大小。形式上， L^p 范数定义如下：

$$\|x\|_p = \left(\sum_i |x_i|^p \right)^{\frac{1}{p}} \quad (11)$$

其中， $p \geq 1$ 。

范数是将向量映射到非负值的函数。直观上来说，向量x的范数衡量从原点到点x的距离。

当 $p = 2$ 时， L^2 范数被称之为欧几里得范数（Euclidean norm）。

特殊类型的矩阵

对角矩阵

对角矩阵（**diagonal matrix**）只在主对角线上含有非零元素，其他位置都是零。即：当且仅当对于所有的 $i \neq j$, $D_{i,j} = 0$ 。

我们用 $\text{diag}(v)$ 表示一个对角元素由向量 v 中元素给定的对角方阵。

对称矩阵

对称（**symmetric**）矩阵是转置和自己相等的矩阵，即： $A = A^T$ 。

满足 $A = -A^T$ 的矩阵称为反对称矩阵。

正交矩阵

单位向量（**unit vector**）是具有单位范数（**unit norm**）的向量：

$$\|x\|_2 = 1 \quad (12)$$

如果 $x^T y = 0$ ，那么向量 x 和向量 y 互相正交（**orthogonal**）。如果两个向量都有非零范数，则这两个向量之间的夹角是90度。在 \mathbb{R}^n 中，至多有 n 个范数非零向量互相正交。如果这些向量不仅互相正交，并且范数都为1，那么我们称它们是标准正交（**orthonormal**）。

正交矩阵（**orthogonal matrix**）是指行向量和列向量是分别标准正交的方阵，即：

$$A^T A = A A^T = I。$$

这意味着： $A^{-1} = A^T$ 。

方阵

行数与列数相同的矩阵称为方块矩阵，简称方阵。

对角矩阵不一定是方阵，长方形的矩阵也可能是对角矩阵。

但是所有的对称矩阵和正交矩阵一定是方阵。

单位矩阵

如实数乘法中的单位元一样，也存在一个特殊矩阵 I 是矩阵乘法中的单位元，即 $IA = AI = A$ ，对于任意方阵 A 都成立。

更为正式地，我们有如下定义：单位矩阵 I 的中：

$$a_{ij} = \begin{cases} 1, & \text{当 } i = j \\ 0, & \text{当 } i \neq j \end{cases} \quad (13)$$

可见，单位矩阵的结构很简单：所有沿主对角线的元素都是1，所有其他元素都是0。

很显然，单位矩阵是对称矩阵。

矩阵逆元

如存在一个矩阵B使得 $AB = BA = I$ ，则称方阵A为非奇异的（nonsingular）或者可逆的（invertible），矩阵B称为矩阵A的乘法逆元（multiplicative inverse）。

若方阵不存在乘法逆元，则称为奇异的（singular）。

只有方阵有乘法逆元。对于非方阵，不应使用术语奇异或非奇异。

行列式

方阵A的行列式（determinant）是一个将方阵映射到标量的函数，记作 $\det(A)$ 。

行列式反映了矩阵自身的一定特性。一个方阵的行列式等于0当且仅当该方阵不可逆。

方阵A的行列式定义如下：

$$\det(A) = \sum_{\sigma \in S_n} \text{sgn}(\sigma) \prod_{i=1}^n a_{i,\sigma(i)} \quad (14)$$

其中：

- S_n 是集合 $\{1, 2, \dots, n\}$ 到自身上的双射的全体。
- $\sum_{\sigma \in S_n}$ 表示对 S_n 全部元素的求和,即对于每个 $\sigma \in S_n$ 来说, $\text{sgn}(\sigma) \prod_{i=1}^n a_{i,\sigma(i)}$ 在加法中出现一次；对每一个满足 $1 \leq i, j \leq n$ 的数对 (i, j) , $a_{i,j}$ 是矩阵A的第i行第j列的元素。
- $\text{sgn}(\sigma)$ 表示置换 $\sigma \in S_n$ 的符号差，具体地说，满足 $1 \leq i < j \leq n$ 但 $\sigma(i) > \sigma(j)$ 的有序数对 (i, j) 称为 σ 的一个逆序。
- 如果 σ 的逆序共有偶数个，则 $\text{sgn}\sigma = 1$ 。如果共有奇数个，则 $\text{sgn}\sigma = -1$ 。

2阶矩阵的行列式计算如下：
$$\begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} = a_{1,1}a_{2,2} - a_{1,2}a_{2,1}。$$

特征值与特征向量

许多数学对象可以通过将它们分解成多个组成部分或者找到它们的一些属性而更好地理解。

正如我们可以通过分解质因数来发现整数的一些内在性质，我们也可以通过分解矩阵来发现矩阵表示成数组元素时不明显的函数性质。

特征分解（**eigendecomposition**）是使用最广的矩阵分解之一，即我们将矩阵分解成一组特征向量和特征值。

方阵 A 的特征向量（**eigenvector**）是指与 A 相乘后相当于对该向量进行缩放的非零向量 v ，即：

$$Av = \lambda v \quad (15)$$

标量 λ 被称为这个特征向量对应的特征值（**eigenvalue**）。

假设矩阵 A 有 n 个线性无关（见下文）的特征向量 $\{v^{(1)}, \dots, v^{(n)}\}$ ，对应的特征值是 $\{\lambda_1, \dots, \lambda_n\}$ 。

我们将特征向量连接成一个矩阵，使得每一列是一个特征向量： $V = [v^{(1)}, \dots, v^{(n)}]$ 。类似地，我们也可以将特征值连接成一个向量 $\lambda = [\lambda_1, \dots, \lambda_n]^T$ 。

由此， A 的特征分解（**eigendecomposition**）可以记作：

$$A = V \text{diag}(\lambda) V^{-1} \quad (16)$$

不是每一个矩阵都可以分解成特征值和特征向量。但每个实对称矩阵都可以分解成实特征向量和实特征值。

正定性

如果对称方阵 A 满足对所有非零向量 x ，对应的二次型： $Q(x) = x^T A x$ 函数值都是正数，就称 A 为正定矩阵（**positive definite matrix**）。对称矩阵的正定性与其特征值密切相关。矩阵是正定的当且仅当其特征值都是正数。

另外：

- 所有特征值都是非负数的矩阵被称为半正定（**positive semidefinite**）矩阵。
- 所有特征值都是负数的矩阵被称为负定（**negative definite**）矩阵。
- 所有特征值都是非正数的矩阵被称为半负定（**negative semidefinite**）矩阵。

奇异值分解

除了将矩阵分解成特征向量和特征值，还有另一种分解矩阵的方法称为奇异值分解（**singular value decomposition**，简称 **SVD**）。这种分解方法将矩阵分为奇异向量（**singular vector**）和奇异值（**singular value**）。

非方阵的矩阵没有特征分解。但每个实数矩阵都有一个奇异值分解。

奇异值分解是将矩阵A分解成三个矩阵的乘积：

$$A = UDV^T \quad (17)$$

假设A是一个 $m \times n$ 的矩阵，那么U是一个 $m \times m$ 的矩阵，D是一个 $m \times n$ 的矩阵，V是一个 $n \times n$ 的矩阵。

这些矩阵中的每一个经定义后都拥有特殊的结构。矩阵U和V都定义为正交矩阵，而矩阵D定义为对角矩阵（注意，矩阵D不一定是方阵）。

对角矩阵D对角线上的元素被称为矩阵A的奇异值（singular value）。矩阵U的列向量被称为左奇异向量（left singular vector），矩阵V的列向量被称为右奇异向量（right singular vector）。

SVD最有用的一个性质可能是拓展矩阵求逆到非方矩阵上。

向量空间

向量空间是现代数学中的一个基本概念。是线性代数研究的基本对象。

定义：令V为一定义了加法和标量乘法运算的集合。这意味着，对V中的每一对元素x和y，可唯一对应于V中的一个元素x+y，且对每一个V中的元素x和每一个标量a，可唯一对应于V中的元素ax。如果集合V连同其上的加法和标量乘法运算满足下面的公理，则称之为向量空间（vector space）。

- **【交换律】** 对V中的任何x和y， $x + y = y + x$
- **【结合律】** 对V中的任何x，y和z， $(x + y) + z = x + (y + z)$
- **【单位元】** V中存在一个元素0，满足对任意的 $x \in V$ ，有 $x + 0 = x$
- **【逆元素】** 对每一个 $x \in V$ ，存在V中的一个元素x，满足 $x + (-x) = 0$
- **【分配律】** 对任意标量a和V中的元素x和y，有 $a(x + y) = ax + ay$
- **【分配律】** 对任意变量a和b即 $x \in V$ ，有 $(a + b)x = ax + bx$
- 对任意标量a和b即 $x \in V$ ，有 $(ab)x = a(bx)$
- **【单位元】** 对所有 $x \in V$ ，有 $1 \cdot x = x$

子空间

若S为向量空间V的非空子集，且S满足如下条件：

- 对任意标量a，若 $x \in S$ ，则 $ax \in S$ 。
- 若 $x \in S$ 且 $y \in S$ ，则 $x + y \in S$ 。

则S称为V的子空间。

上面两个条件的意义是：S在标量乘法和加法意义下都是封闭的。也就是说，以S中的元素进行这两种运算后，得到的结果仍然是S中的元素。

重新回到线性方程组的概念。假设我们通过 $Ax = b$ 来描述线性方程组。

其中：

- A 表示参数矩阵
- x 表示待求解的向量
- b 表示等式右边值组成的向量

为了分析方程有多少个解，我们可以将 A 的列向量看做从原点出发的不同方向，确定有多少种方法可以达到向量 b 。在这个观点下，向量 x 中的每个元素表示我们应该沿着这些方向走多远。

即 x_i 表示我们需要沿着第 i 个向量的方向走多远：

$$Ax = \sum_i x_i A_{:,i} \quad (18)$$

一般而言，这种操作被称之为线性组合（linear combination）。形式上，一组向量的线性组合，是指每个向量乘以对应标量系数之后的和，即： $\sum_i c_i v^{(i)}$ 。

一组向量的生成子空间（span）是原始向量线性组合后所能抵达的点的集合。

确定 $Ax=b$ 是否有解相当于确定向量 b 是否在 A 列向量的生成子空间中。

线性无关

为了使方程 $Ax=b$ 对于任意向量 b 都存在解，我们要求 A 的列空间构成整个 R^m 。如果 R^m 中的某个点不在 A 的列空间中，那么该点对应的 b 会使得该方程没有解。

矩阵 A 的列空间是整个 R^m 的要求，意味着 A 至少有 m 列，即 $n \geq m$ 。否则， A 列空间的维数会小于 m 。

不等式 $n \geq m$ 仅是方程对每一点都有解的必要条件。这不是一个充分条件，因为有些列向量可能是冗余的。假设有一个 $R^{2 \times 2}$ 中的矩阵，它的两个列向量是相同的。那么它的列空间和它的一个列向量作为矩阵的列空间是一样的。换言之，虽然该矩阵有2列，但是它的列空间仍然只是一条线，不能涵盖整个 R^2 空间。

正式地说，这种冗余被称为线性相关（linear dependence）。

如果一组向量中的任意一个向量都不能表示成其他向量的线性组合，那么称这组向量为线性无关（linearly independent）。如果某个向量是一组向量中某些向量的线性组合，那么我们将这个向量加入这组向量后不会增加这组向量的生成子空间。

这意味着，如果一个矩阵的列空间涵盖整个 R^m ，那么该矩阵必须包含至少一组 m 个线性无关的向量。这是对于每一个向量 b 的取值都有解的充分必要条件。值得注意的是，这个条件是说该向量集恰好有 m 个线性无关的列向量，而不是至少 m 个。不存在一个 m 维向量的集合具有多于 m 个彼此线性不相关的列向量，但是一个有多于 m 个列向量的矩阵有可能拥有不止一个大小为 m 的线性无关向量集。

基和维数

如果向量空间 V 中的向量 v_1, v_2, \dots, v_n 满足：

- v_1, v_2, \dots, v_n 是线性无关的。
- v_1, v_2, \dots, v_n 构成 V 。

则，称它们是向量空间 V 的基（**basis**）。

向量空间的基是它的一个特殊的子集，基的元素称为基向量。向量空间中任意一个元素，都可以唯一地表示成基向量的线性组合。如果基中元素个数有限，就称向量空间为有限维向量空间，将元素的个数称作向量空间的维数。

一个向量空间可以由多个基，但是所有的基都有相同的维数（**dimension**）。

不是所有空间都拥有由有限个元素构成的基底。这样的空间称为无限维空间。

如果一组基中每一个基向量的范数都是1，则称之为标准基（**standard basis**）。

若对于两个向量， $x^T y = 0$ ，则称这两个向量正交。

如果标准基两两之间互相正交，则称为标准正交基或”规范正交基”（**Orthonormal basis**）。

行空间和列空间

对于 $m \times n$ 矩阵， m 个行的向量称之为行向量（**row vector**）。类似地， n 个列的向量称之为列向量。

行向量构成的子空间称为行空间（**row space**）。类似地，列向量构成的子空间称之为列空间。

矩阵的秩

矩阵 A 线性无关的列的极大数目，称之为矩阵 A 的列秩。类似地，矩阵 A 线性无关的行的极大数目，称之为矩阵 A 的行秩。

矩阵的行秩和列秩总是相等的。因此可以简称为矩阵的秩（**rank**），通常记为 $r(A)$ 或者 $rank(A)$ 。

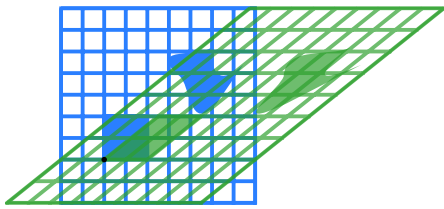
线性变换

矩阵是线性变换的便利表达法。矩阵乘法的本质在联系到线性变换的时候最能体现，因为矩阵乘法和线性变换的合成有以下的联系：以 R^n 表示所有长度为 n 的行矢量的集合。每个 $m \times n$ 的矩阵 A 都代表了一个从 R^n 射到 R^m 的线性变换。

以下是一些典型的2维实平面上的线性变换对平面矢量（图形）造成的效果，以及它们对应的2维矩阵。其中每个线性变换将蓝色图形映射成绿色图形；平面的原点(0,0)用黑点表示。

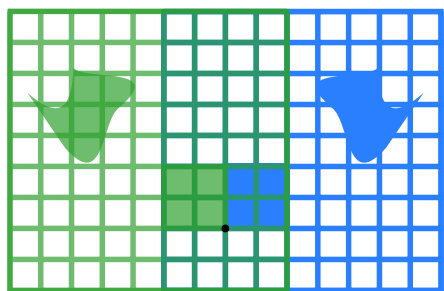
推移

矩阵: $\begin{bmatrix} 1 & 1.25 \\ 0 & 1 \end{bmatrix}$



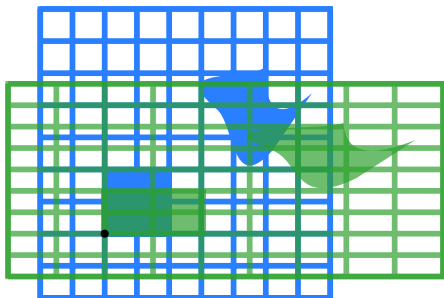
镜射

矩阵: $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$



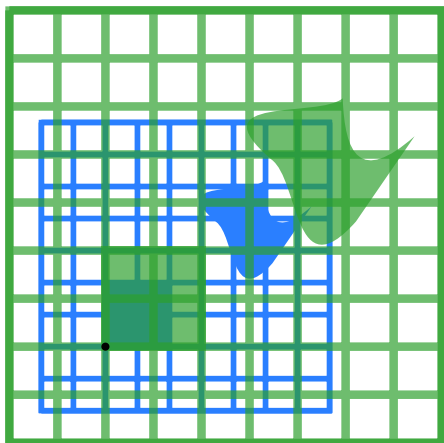
挤压

矩阵: $\begin{bmatrix} \frac{3}{2} & 0 \\ 0 & \frac{2}{3} \end{bmatrix}$



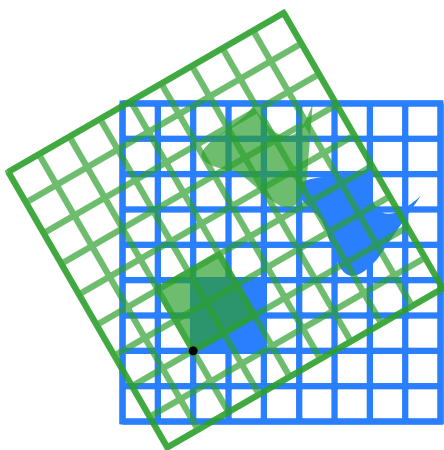
伸缩

矩阵: $\begin{bmatrix} \frac{3}{2} & 0 \\ 0 & \frac{3}{2} \end{bmatrix}$



旋转

矩阵:
$$\begin{bmatrix} \cos(\frac{\pi}{6}) & -\sin(\frac{\pi}{6}) \\ \sin(\frac{\pi}{6}) & \cos(\frac{\pi}{6}) \end{bmatrix}$$



Hessian矩阵

黑塞矩阵（德语：Hesse-Matrix；英语：Hessian matrix 或 Hessian），又译作海森矩阵、海塞矩阵或海瑟矩阵，是一个以德国数学家奥托·黑塞命名的多变量实值函数的二阶偏导数组成的方块矩阵，假设有一实数函数 $f(x_1, x_2, \dots, x_n)$ ，如果f所有的二阶偏导数都存在，都么f的Hessian矩阵的第ij项为：

$$H(f)_{ij}(x) = D_i D_j f(x) \quad (19)$$

其中 $x = (x_1, x_2, \dots, x_n)$ ，即：

$$H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \quad (20)$$

Hessian矩阵被应用于牛顿法解决大规模优化问题。

如果f函数在区域D内的每个二阶导数都是连续函数，那么f的黑塞矩阵在D区域内为对称矩阵。

雅克比矩阵

雅克比矩阵是函数的一阶偏导数以一定方式排列成的矩阵，其行列式称为雅克比行列式。它们全部都以普鲁士数学家卡尔·雅可比命名。

假设某函数从 R^n 映到 R^m ，其雅可比矩阵是从 R^n 到 R^m 的线性映射，其重要意义在于它表现了一个多变数向量函数的最佳线性逼近。因此，雅可比矩阵类似于单变数函数的导数。假设 $F: R^n \rightarrow R^m$ 是一个从 n 维欧氏空间映射到 m 维欧氏空间的函数。这个函数由 m 个实函数组成: $y_1(x_1, \cdots, x_n), \cdots, y_m(x_1, \cdots, x_n)$ 。这些函数的偏导数如果存在可以组成一个 m 行 n 列的矩阵，这个矩阵就是所谓的雅可比矩阵：

$$\begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix} \quad (21)$$

附录：术语中英文对照

- 主对角线：main diagonal
- 乘法逆元：multiplicative inverse
- 半正定：positive semidefinite
- 单位向量：unit vector
- 单位矩阵：identity matrix
- 可逆的：invertible
- 向量：vector
- 增广矩阵：augmented matrix
- 奇异的：singular

- 对称矩阵: symmetric matrix
- 对角矩阵: diagonal matrix
- 张量: tensor
- 方阵: square
- 标准正交: orthonormal
- 标量: scalar
- 正交矩阵: orthogonal matrix
- 正交: orthogonal
- 正定: positive definite
- 点积: dot product
- 特征值: eigenvalue
- 特征分解: eigendecomposition
- 特征向量: eigenvector
- 生成子空间: span
- 矩阵乘积: matrix product
- 矩阵逆: matrix inversion
- 矩阵: matrix
- 系数矩阵: coefficient matrix
- 线性相关: linear dependence
- 线性组合: linear combination
- 范数: norm
- 行列式: determinant
- 负定: negative definite
- 转置: transpose
- 非奇异的: nonsingular
- 齐次: homogeneous
- 向量空间: vector space

参考资料与推荐读物

- 《线性代数》 (<https://www.amazon.cn/dp/B015H5F99Y/>)
- 维基百科: 矩阵 (<https://zh.wikipedia.org/wiki/矩阵>)
- 《深度学习》- 第二章 线性代数 (<https://www.amazon.cn/dp/B073LWHBBY/>)
- 《运筹学》- 第五章 线性规划的单纯形法 (<https://www.amazon.cn/dp/B07FHLYBG5/>)
- 知乎话题: 矩阵 (<https://www.zhihu.com/topic/19650614/hot>)

原文地址: 《矩阵基础知识》 (<https://paul.pub/the-matrix/>) by 保罗的酒吧 (<https://paul.pub>)



说点什么

① 支持 Markdown 语法 (<https://guides.github.com/features/mastering-markdown/>)

预览

使用 GitHub 登录

来做第一个留言的人吧！

≡ Contents

- 介绍
- 词源
- 发展
- 标量，向量，矩阵和张量
- 矩阵与线性方程组
 - 增广矩阵
- 矩阵算术
 - 加减法
 - 标量乘法
 - 转置
 - 矩阵乘法
- 范数
- 特殊类型的矩阵
 - 对角矩阵
 - 对称矩阵
 - 正交矩阵
- 方阵
 - 单位矩阵
 - 矩阵逆元
 - 行列式
- 特征值与特征向量
 - 正定性
- 奇异值分解
- 向量空间
 - 子空间
 - 线性无关
 - 基和维数
 - 行空间和列空间
 - 矩阵的秩

- 线性变换
 - 推移
 - 镜射
 - 挤压
 - 伸缩
 - 旋转
- Hessian矩阵
- 雅克比矩阵
- 附录：术语中英文对照
- 参考资料与推荐读物

© 保罗的酒吧 (<https://paul.pub>)