0.1 矩阵的定义

矩阵的来源 矩阵起源于线性方程组的求解，像简单的二元一次方程组 $\begin{cases} a_{11}x + a_{12}y &= b_1 \\ a_{21}x + a_{22}y &= b_2 \end{cases}$ 我们在求解的过程仅需要对其系数进行操作，于是乎，变量符号 x, y 对于我们来说，在计算过程中，并不显得十分必要；因此，我们可以略掉其变量名而将系数写在一对方括号之中，以便于进行计算。从该二元方程组我们可以得到如下形式的系数“方块”

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

类似的，我们可以从 n 元方程组（不一定有 n 个方程式）

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

得到系数“方块”

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

于是，为求得方程组的解集，一般来说我们需要进行两种操作：一，将某一行乘以一个非零常数并不改变方程组的解集；二，将某一行的非零常数倍加到另一行，也不改变方程组的解集。这样，对应到我们新得到的“方块中”，就是将某一行乘以非零常数倍，或者将某一行的常数倍加到另一行之后，得到的新“方块”在方程组有同样的解集的意义下等价。

至此，我们已经抽象出了矩阵的基本概念，接下来我们正式定义矩阵。

矩阵及其运算的定义 在数学上，一个 $m \times n$ 矩阵，是一个由 m 行，每行 n 个数字的列表组成。行列数相等，即 $m = n$ 的矩阵被称为方阵，方阵对应很多独特的性质，今后将会谈到。在矩阵上我们定义定义加法和乘法：

1. 加法：

   并非任意种类的矩阵都可以进行加法运算，必须是同种类型的矩阵才可以，即，假设我们用记号 $M_{m\times n}$ 表示所有的 $m\times n$ 矩阵，那么对任意的矩阵 A, B，当且仅当存在某个 $M_{m\times n}$ 包含他们二者时，$A+B$ 才有意义；用数学符号表示为：$\exists M_{m\times n}, A\in M_{m\times n}\&B\in M_{m\times n}\Leftrightarrow \exists A+B.$ 直观的看来就是，两个矩阵相同位置的元素相加，所得到的新矩阵就是矩阵的和。

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

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

   很明显，能够进行加法的矩阵 A 和 B，$A+B=B+A$，因此矩阵的加法满足交换律。

2. 乘法：

   类似于加法，矩阵的乘法对矩阵的类型也有要求。我们假定 $A\in M_{m\times n}$ 及 $B\in M_{l\times k}$，当且仅当 $n=l$ 时，$A\cdot B$ 才有意义，并且得到一个 n 行 k 列的新矩阵 $C\in M_{m\times k}$。后面我会详细讨论这样定义矩阵乘法的实际用途。

$$\begin{bmatrix} a_{11} & a_{12} & \cdots a_{1n} \\ a_{21} & a_{22} & \cdots a_{2n} \\ \vdots & & \\ a_{m1} & a_{m2} & \cdots a_{mn} \end{bmatrix} \cdot \begin{bmatrix} b_{11} & b_{12} & \cdots b_{1k} \\ b_{21} & b_{22} & \cdots b_{2k} \\ \vdots & & \\ b_{n1} & b_{n2} & \cdots b_{nk} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & \cdots c_{1k} \\ c_{21} & c_{22} & \cdots c_{2k} \\ \vdots & & \\ c_{m1} & c_{m2} & \cdots c_{mk} \end{bmatrix}$$

   其中元素 c_{ij} 的值为

$$c_{ij}=\sum_{r=1}^{n} a_{ir}b_{rk}$$

如果我们把一个矩阵的每一行看做一个有限维向量（称为行向量），也把每一列看做一个有限维向量（称为列向量）。再考虑向量的数量积的定义，如果

$$\vec{a} = (a_1, a_2, \cdots, a_n)$$

$$\vec{b} = (b_1, b_2, \cdots, b_n)$$

那么向量 $\vec{a}$ 与 $\vec{b}$ 的数量积为：

$$\vec{a} \cdot \vec{b} = a_1b_1 + a_2b_2 + \cdots + a_nb_n = \sum_{i=1}^{n} a_ib_i$$

现在我们可以给予矩阵乘法的一种比较直观的解释，积矩阵的元素 c_{ij} 是来自第一个因子矩阵的第 i 个行向量和第二个因子矩阵的第 j 个列向量的数量积；现在我们明白了为什么要求第一个矩阵的列数要与第二个矩阵的行数相等了，因为只有这样，才能保证矩阵 A 的行向量的维数与矩阵 B 的列向量的维数相同，而只有维数相同的向量才能进行数量积。

0.2 矩阵的现实及其运算意义

矩阵的代数意义 前面已经提到，矩阵是从线性方程组中抽象出来的数学对象，它最重要的运用就是简化线性代数方程组的求解。对于每一个矩阵 $A \in M_{m\times n}$，A 代表某个线性方程组的系数矩阵，看看前述“矩阵的来源”一段中引入的“方块”，左侧称为“系数矩阵”，如果把右侧的 $n \times 1$ 的矩阵（实际上它是一个 m 维列向量）“粘贴到” A 的最右侧，即

$$\left[\begin{array}{ccccc} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right]$$

我们称新得到的矩阵为系数矩阵 A 的“增广矩阵”，每一个增广矩阵完全表征了一个线性方程组。这种形式很方便于利用高斯-若尔当消元法求解。

高斯-若尔当 (Gauss-Jordan) 消元法是判定线性方程组解集的性态，求解线性方程组，求矩阵的秩，求方阵的行列式、逆矩阵的极有效的方法。后

面我们会详细解释它，此时我们只是先了解它的存在。

进一步，在这种代数意义下我们来研究矩阵的乘法的比较明显的意义的例子：我们考虑系数矩阵 A, 同时定义一个 $n \times 1$ 的矩阵（一个 n 维列向量）

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

称为未知量矩阵（n 维向量），我们研究 $A \cdot X$，显然我们可以得到一个 m 维的列向量，其每一个分量都是一个方程式的等号的左边部分，如果我们将每一个方程式等号的右边部分也写成一个 m 维列向量 B

$$B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

那么

$$A \cdot X = B$$

就完全是代数方程组的简写。

矩阵的几何意义 在线性代数中，一个矩阵总是会和线性空间、线性变换、空间的基等概念联系起来。此阶段我仅仅介绍而不详细解释，它涉及的基本概念太多。

我们对笛卡尔坐标很熟悉，对其中的向量的概念也不陌生；我们知道，在一个 n 维欧几里得空间（可以理解为有 n 个坐标轴的笛卡尔坐标系），每一个向量都可以表示成一个 n 维数组，这样如果我们记和每个坐标轴 x_i 平行的单位向量（长度为 1 的向量）为 $\vec{e_i}$，那么 $\vec{e_i}$ 可以表示成这样的数组:

$$\vec{e_i} = (\underbrace{0, 0, \cdots, 0}_{i-1}, 1, 0, \cdots, 0)$$

回想向量的加法以及向量与实数的乘法，我们就可以通过 $\vec{e_1}, \vec{e_2}, \cdots, \vec{e_n}$ 的线性组合表示一切 n 维空间中的向量了。我们知道任意一个 n 维空间中的

向量可以写成 $\vec{v} = (v_1, v_2, \cdots, v_n)$，于是：

$$\begin{aligned}\vec{v} &= (v_1, v_2, \cdots, v_n) \\ &= (v_1, 0, 0, \cdots, 0) + (0, v_2, \cdots, 0) + \cdots + (0, \cdots, 0, v_n) \\ &= v_1(1, 0, \cdots, 0) + v_2(0, 1, 0, \cdots, 0) + \cdots + v_n(0, \cdots, 0, 1) \\ &= v_1\vec{e_1} + v_2\vec{e_2} + \cdots + v_n\vec{e_n}\end{aligned}$$

于是，任何一个向量都可以通过这 n 个向量的线性组合来表示。我们把这样的 n 个向量的集合称为 n 维空间的一组基。

当然空间的基的形式不止一种，我们可以把其中一个 $\vec{e_i}$ 换成 $\vec{e_i'} = -\vec{e_i} = (0, 0, \cdots, -1, 0, \cdots, 0)$，甚至换成另外它和另外任意一个向量 $\vec{v}$ 的和，我们依然可以保证空间中的任意一个向量都能通过他们组合出来。所以，空间的基的形式并不唯一，但是可以确定的是，基的个数一定是 n，否则一定存在不能用他们表示的向量，比如 $\vec{e_n}$ 就不能用前面的 $\vec{e_1}, \vec{e_2}, \cdots, \vec{e_{n-1}}$ 表示出来，因为这些向量的最后一个坐标值总是 0，而 $\vec{e_n}$ 的最后一个坐标值是 1。

值得强调的是，并不是任意 n 个向量都能称为一组基，我们考虑如下 n 个向量，$\vec{e_1}, 2\vec{e_1}, \cdots, n\vec{e_1}$，这样的 n 个向量的组合，用 $\vec{v_i}$ 表示 $i\vec{e_1}$

$$\sum_{i=1}^{n} a_i\vec{v_i} = (\sum_{i=1}^{n} ia_i)\vec{e_1}$$

这样的向量总是在坐标轴上，除此之外的点不可能由它们的组合表示出来。

而矩阵就相当于是这样的向量组的集合，方阵（行列数相同的矩阵）则有可能成为一个空间的基，矩阵的乘法就相当于对这些向量的线性组合，而成为一些新的向量集合。对于方阵，如果它包含的向量能够成为一个空间的基，那么他也可以作为一个坐标变换的方法存在（因为从一个坐标系到另一个坐标系的变换，总是对应于 n 次空间基的线性组合而成为 n 个新的基）。这里要再多指出一点，空间坐标变换指的是同一个向量在该空间的不同的基下的表示形式，向量本身并不改变，变换的只是表示形式；比如，对于同一个图形，如果我们把坐标轴的每个刻度的值增加到原来的 2 倍，但是图形本身在纸上并不改变，那么，得到的在新的意义下的坐标表示的每个分量都将会是原来的 2 倍；于是我们得到了同图形在新坐标下的表示方

式。详细的过程我们会在今后进行。

小结 因此，一个数学对象的实际意义会根据所处情况的不同采取不同的解释，这种行为称为模型化，对于纯粹的数学概念，他们本身是没有任何实际上的意义的，只有当我们对其进行某种解释后，他们的意义就能显现出来；采用不同的解释，数学理论里面的每一个纯概念就会得到不同的实际意义。关于我们如何提取到这些纯粹的数学概念的，又是如何能对他们加以解释的原因，涉及很深层次的逻辑学和哲学问题，讨论极其复杂，你有兴趣的话，我再给你深入解释。

0.3 高斯-若尔当消元法 (Gauss-Jordan)

在求解方程组时的运用 对于一个具有 n 个变量的 $x_1, x_2, \cdots, x_n$ 的线性方程组，对任意的变量 x_i 我们总能找到至少一个方程组中的方程式，其中 x_i 的系数不为 0；否则，如果对于任意的方程式 x_i 的系数都为 0，那么这个变量就是多余的，不应当被写出。我们先写出这个方程组和它的增广矩阵（请回想增广矩阵的定义）。

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

$$\left[\begin{array}{ccccc} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right]$$

在求解线性方程组时，我们知道，将两个等式相加不改变方程组的解；在任意一组等式的等号两边同时乘以一个不为 0 的常数，也不改变方程组的解；当然，交换两个等式的位置，也不会改变方程组的解。对于前两点，我们很容易就能延伸到对增广矩阵的操作：将某一行加到另一行，矩阵对应的方

程组的解保持不变，称为在解的意义下，变换后的矩阵和变换前的矩阵等价；将某一行乘以一个非零常数也能得到等价矩阵，结合这两点，我们可以断言，将矩阵的某一行的任意倍加到另一行，将得到等价矩阵；以及交换两行的位置，同样可以得到等价矩阵。

这三种操作，对应了矩阵的基础操作，今后会时常用到。现在，我们详细述说如何实现矩阵的高斯消元法，我们前面已经提到，高斯-若尔当消元法不仅仅可以用求解线性方程组，还可以进行很多其他的应用。我们先详细地介绍如何在理论层面中进行这种方法，然后在以 Python 语言为原型，详细讲解如何实现一个矩阵类，以及实现这些运算和操作。

首先我们对第一个变量 x_1 进行考量，如前述的约定，我们总可以找到一个包含 x_1 非零系数的方程组，不是一般性，我们可以假定第一行已经是这样的方程式，如若不然，我们可以交换两行，让第一行的 x_1 系数总不为零，即我们已经得到断言 $a_11 \neq 0$。接下来，我们可以进行对第 2 至 n 行进行这样的操作，即让他们加上第一行的某个倍数得到一个新方程式，从而使分别他们的 x_1 系数为 0；即：

$$\begin{aligned} a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{1n}x_n &= b_i \\ &+ \\ -\frac{a_{i1}}{a_{11}}(a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}) &= -\frac{a_{i1}}{a_{11}}b_1 \end{aligned}$$

我们可以得到新的等式：

$$0x_1 + a^*_{i2}x_2 + \cdots + a^*_{in}x_n = b^*_i$$

其中 $a^*_{ij} = a_{ij} - \frac{a_{i1}}{a_{11}}a_{1j}, b^*_i = b_i - \frac{a_{i1}}{a_{11}}b_1$ 以同样的方式进行增广矩阵的操作，我们可以得到等价矩阵：

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ 0 & a^*_{22} & \cdots & a^*_{2n} & b^*_2 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & a^*_{m2} & \cdots & a^*_{mn} & b^*_m \end{bmatrix}$$

至此我们已经把第一列的第一行之下全部变为 0，同时保证了矩阵的等价性；现在，我们保持第一行不动，对第二行以及之后的所有行执行同样的操

作，意即，首先判断第二行的第二个系数是否为 0，如果是，那么判断在剩下的行中是否存在第二个系数不为零的行，如果有，那么就交换他们的行，然后进行上述操作；此时我们可以得到等价矩阵的形式如下：

如果所有的 $a_{22}^*, a_{32}^*, \cdots, a_{m2}^*$ 都是零，那么我们就得到行