## 1.2线性代数

### 1.2.1线性方程组

先从线性方程组开始讲起，线性方程组的一般形式如下所示：



举个例子，现在有方程组如下：



每个方程都只有两个未知数，这样的方程就是二维空间中的一条直线。而求含有两个未知数的两个方程组成的方程组的解，等价于求两条直线的交点。很容易求出以上线性方程组的解为，图形结果如图1-17所示。

[](http://124.220.164.99:8090/upload/2022/11/image-1669479695647.png)

图1-17方程组存在一个解

此时，方程组存在一个唯一解。当然，两条直线并不一定交于一点，它们可能平行，也可能重合，重合的两条直线上的每个点都是交点。考虑下面两个方程组：



其中第一个方程组中的两条直线平行，没有交点，即方程组无解，如图1-18(a)所示；第二个方程组中的两条直线重合，有无数交点，即方程组有无穷多解，如图1-18(b)所示。

[](http://124.220.164.99:8090/upload/2022/11/image-1669480022061.png)

(a)无解 (b) 无穷解

图1-18方程组存在无穷解或者无解

通过上面的例子，可以总结一个重要的结论：线性方程组的解只有三种情况：一个解、无穷解和无解。

现在把方程扩展到三个未知数的线性方程组，这样每个方程将确定三维空间中的一个平面，如图1-19所示。

[](http://124.220.164.99:8090/upload/2022/11/image-1669480261827.png)

图1-19 三个未知数的线性方程组

现在想象一下三个这样的平面在三维空间中的分布会有几种情况？其实也是上述的三种情况：

（1）当三个平面相互平行时，无解。

（2）当三个平面相交于一条线时，无穷解。

（3）当三个平面相较于一点时，只有一个解。

为了更直观地理解三个平面交于一点的情境，可以将其比喻为桌子的一个角：桌面与两个侧面分别代表三个相交的平面，它们交于一点，即桌角。如果继续推广到高维空间其实也是这个结论，这里就不做演示了。

我们在来看线性方程组的一般公式：



其实写起来是非常麻烦的，我们可以通过矩阵系统进行简化，如图1-20所示。

[](http://124.220.164.99:8090/upload/2022/11/image-1669480581532.png)

图1-20 线性方程组的矩阵表示

（1）把系数从线性方程组中提取出来，写成的矩阵称为系数矩阵。

（2）把常数项从线性方程组中提取出来，写成的矩阵称为常数项矩阵。

（3）把系数矩阵和常数项矩阵左右拼接在一起，写出的矩阵称为增广矩阵。

特别地，若, 方程组变为:



称上面线性方程组为齐次线性方程组。齐次线性方程组与其系数矩阵一一对应。

### 1.2.2 线性方程组的矩阵求解法

线性方程组的矩阵求解法：



步骤一：即：



步骤二：即：



最后可以解得，再将代回式，解得。

小结一下：在上面的例子中，我们用到了行列式基本的三个初等行变换操作：

（1）对某个行乘以一个不为0的常数k （第一步中用到的操作）。

（2）某个行可以被它自己或另一个的和所替换 （第二步中用到的操作）。

（3）行与行之间可以交换顺序 （例子中没用到，但是可以想象两个函数上下交换位置不影响）。

下面介绍一下阶梯形矩阵，一个矩阵成为阶梯型矩阵，需满足两个条件：

（1）如果它既有零行，又有非零行，则零行在下，非零行在上。

（2）对于所有的非零行，其第一个非零元素所在的列号必须从上到下递增。

换句话说，如果所给矩阵为阶梯型矩阵则矩阵中每一行的第一个不为零的元素的左边及其所在列以下全为零。如下所示：



行最简形矩阵是阶梯形矩阵的特殊例子，在阶梯形矩阵中，若非零行的第一个非零元素全是1，且非零行的第一个元素1所在列的其余元素全为零，就称该矩阵为行最简形矩阵。



如果在行最简形矩阵中，非零行有且只有一个非零元素且为1，则称该矩阵为标准形矩阵。



基本概念介绍完了，那么阶梯形矩阵有什么用呢？我们再回过头看刚刚线性方程组的矩阵求解过程就会发现，这个求解过程实际上就是求线性方程组增广矩阵的阶梯形矩阵的过程，它有个高大上的名字，叫做 “高斯消元法“。

考虑方程组：



其对应的增广矩阵为：



步骤一：处理第一列，使第一列只有顶部的元素为1，其余为0。

首先从第二行中减去第一行的2倍：



再从第三行减去第一行的3倍：



步骤二：处理第二列，由于第二列除第一个元素外的所有元素都是0，可以直接跳过这一步并处理第三列。

步骤三：处理第三列，使第三列的第二行元素为1，其余为0。

为了将第二行第三列的-5变为1，将第二行除以-5：



接下来为了使第三行第三列的元素变为0，将第二行加到第三行上：



结论：通过上述的高斯消元过程，我们得到了一个上三角形的增广矩阵。根据这个矩阵，可以通过回代法得到方程组的解。但请注意，最后一个方程实际上是一个恒等式 ()，这意味着原方程组具有无数多的解。

最后，介绍几个矩阵的性质：

（1）任一矩阵可经过有限次初等行变换化成阶梯形矩阵。

（2）任一矩阵可经过有限次初等行变换化成行最简形矩阵。

（3）矩阵在经过初等行变换化为最简形矩阵后，再经过初等列变换，还可以化为最简形矩阵，因此，任一矩阵可经过有限次初等变换化成标准形矩阵。

（4）一个矩阵的行最简形矩阵是唯一确定的。

### 1.2.3 矩阵乘法

矩阵乘法（也称为matmul product）是两个矩形相乘的操作，其结果是另一个矩阵。定义如下：

设有两个矩阵和，令是一个的矩阵，而是一个的矩阵。那么矩阵和的乘积是一个的矩阵，每个元素由公式(1-53)给出：



其中是结果矩阵的第行第列的元素。

设有矩阵和如下：



计算的结果为：



哈达玛积（也称为element-wise product）表示两个矩阵对应元素相乘，二者维数必须相同，用表示。如：



克罗内克积（也称为Kronecker product）表示两个任意大小矩阵间的运算，矩阵的每个元素逐个与矩阵相乘，用表示。如：



### 1.2.4 向量的数乘

在上述讲解中，我们已经涉及了三个主要的数学系统：线性方程组、函数图形和矩阵。现在将介绍第四个系统：向量。线性代数的一个核心挑战是它涵盖了多个数学系统。要成功掌握线性代数，关键的学习策略是理解这些系统之间的联系和互动。通过将知识点在不同系统中的应用相互关联，可以更深入地理解概念，实现知识的融会贯通。

首先看一下向量的通俗理解： 向量是一个指令，不是一个坐标，可以存在于坐标系下的任何位置。怎么理解呢？如图1-21所示。

[](http://124.220.164.99:8090/upload/2022/11/image-1669483473105.png)

图1-21 向量

向量可以看作向右走两个单位，向上走一个单位。它可以存在于坐标系下的任何位置。并不代表其在坐标系中的轴和轴坐标。

向量的数乘（scalor）指用一个标量来乘向量，改变的是向量长短，不改变方向。如：



### 1.2.5 向量的加法

向量的加法（vector addition）计算采用平行四边形法则（首尾相连）：以同一起点的两个向量为邻边作平行四边形，则以公共起点为起点的对角线所对应向量就是和向量。如图1-22所示。

[](http://124.220.164.99:8090/upload/2022/11/image-1669483792718.png)

图1-22 向量的平行四边形法则

举个例子：



按照指令翻译的话：向右移动两个单位→向上移动一个单位→向左移动一个单位→向上移动一个单位。

### 1.2.6向量的线性组合

向量的线性组合（linear combination）：实际上就是向量数乘和向量加法的组合。可以用公式表示。其中​是常数，如：



向量的线性组合与线性方程组紧密相关：当寻求向量的线性组合的解时，实际上是在解决一个对应的线性方程组，如：



### 1.2.7 向量空间

向量空间指的是线性组合的集合，例如的向量空间是整个二维空间：



即：在二维空间中的任何一个向量，都可以通过向量和的线性组合进行表示。向量空间的严谨定义是：对向量加法和数乘（即线性组合）都封闭的非空集合，就是向量空间。

基本单位向量（standard basis vector）指向量中只有一个标量为1,其余标量均为0。如公式(1-63) 为二维向量空间的基本单位向量：



怎么确定一个向量是否在的向量空间中呢? 其实就是去求解向量是否可以在向量空间中被表示。如判断向量是否存在于向量空间中。可以把这个问题转化成线性方程组求解的问题。即求解一个向量是否在向量空间中，就是求向量对应的线性方程组是否有解。其转化成线性方程组的过程如下所示：









注意，这只是一种表达方式，并不能把的值解出来，要求解的话还是要转换成增广矩阵再进行求解。最后可解得：



### 1.2.8向量的线性相关和线性无关

如果，可以找到至少一个不为 0 ，即**不全为 0 ，则线性相关。

如果 ，只在的情况下成立，则线性无关。

关于线性相关性存在一个定理：个维向量必线性相关。例如三个3维向量可以线性无关，如图1-23(a)所示；但三个2维向量一定线性相关，如图1-23(b)所示。

[](http://124.220.164.99:8090/upload/2022/11/image-1669496955963.png)

图1-23(a) 线性相关 图1-23(b) 线性无关

图1-23线性相关性

如何判断一组向量是否线性相关呢？直接进行矩阵求解就好，如：

，判断是否线性相关？



解得，根据之前线性相关性的定义，线性无关。如果把的值进行修改，相关性会有什么变化呢？



解得：



此时有无穷解，说明线性相关。

小结一下向量线性相关性的判断思路：判断向量的线性相关性可以通过将其转化为线性组合问题，然后表示为线性方程组。接着使用增广矩阵来描述该方程组并将其转化为阶梯形矩阵。最终，如果该矩阵有唯一解，则向量线性无关；如果有无穷多个解（对应一个自由列），则向量线性相关。

结合下面这五种情况，你会发现，向量的数量和维度都不是决定向量空间（也称为张成空间）的决定性因素，而是需要结合向量的线性无关性来进行考量。

第一种情况：，显然，向量和是两个线性无关的二维向量，它构成了二维空间的一组基，因此它们的张成空间就是整个二维空间。

第二种情况：，。因此和是线性相关的共线向量，它们的张成空间是一条穿过原点的一维直线。

第三种情况：，和两个三维向量线性无关，但是由于向量的个数只有两个，因此它们的张成空间是三维空间中的一个穿过原点的平面。

第四种情况：，虽然向量的个数是3，但，因此它们是三个线性相关的共面向量，张成的空间仍然只是三维空间中的一个穿过原点的平面。

第五种情况：，三个向量线性无关，构成三维空间中的一组基，因此它们的张成空间是整个三维空间。

### 1.2.9 向量乘法

向量的点积和内积（Inner Product, dot product），用表示，两个向量的行列数必须相同，点积的结果是对应元素相乘后求和，结果是一个标量，如：



点积的几何意义可以用来计算两个向量的夹角：

至于向量长度的求解，*n*维向量的长度：，当时称为单位向量。向量的长度的求解满足齐次性，即；也满足三角不等式，即。

注意： Dot product 和 Inner product其实还是有区别的，目前暂时将二者视为同一个概念，后续再来细究。

向量的外积（Outer product），用表示，如：



向量的叉积（Outer product），用表示，如：



读者可能对叉积不太熟悉，其计算结果的几何意义是两个向量的法向量。举个例子：是*x*轴的单位向量，是*y*轴的单位向量，二者叉积的结果就是*z*轴的单位向量，即：



### 1.2.10 向量的正交

两两正交的非零向量组成的向量组称为正交向量组，若**是两两正交的非零向量，则**线性无关。例如：已知三维空间中的两个向量正交，试求一个非零向量，使得两两正交。

解题思路：内积等于0时，意味着两个向量正交。

显然，设，若，则：



解系数矩阵：



得

从而有基础解系，令。

规范正交基：*n* 维向量是向量空间**中的向量，满足：（1）是向量空间中的一个基；（2）两两正交；（3）都是单位向量。则称是的一个规范正交基。**是的一个规范正交基。

### 1.2.11 向量与矩阵

#### 1.初等矩阵

矩阵可以看作对向量的变换，单位矩阵是对角线全1的矩阵，相当于0变换。

矩阵的初等行变换等价于对单位矩阵做初等行变换再乘上要变换的矩阵。

可以发现初等矩阵是对角矩阵，如果一个对角矩阵的值不为1会对向量产生什么影响呢？如图1-24所示。

[](http://124.220.164.99:8090/upload/2022/11/image-1669504335193.png)

图1-24 对角矩阵对向量的影响

可以发现，此时的影响是在*x*轴或*y*轴上对向量进行伸缩。那么普通的矩阵又是什么效果呢？如图1-25所示。

[](http://124.220.164.99:8090/upload/2022/11/image-1669504448896.png)

图1-25 普通矩阵对向量的影响

可以发现，普通的矩阵还可以对向量产生旋转的变换效果。

#### 2.可逆矩阵

可逆矩阵把求解向量**的问题转换成了求可逆矩阵本身的问题，如下公式推导所示：



可逆矩阵的求解与矩阵和初等矩阵有关，具体推导如下：

假设，求可逆矩阵的过程如下：



通过矩阵的初等行变换将左边矩阵推导成右面的格式即完成了可逆矩阵的计算，此时可逆矩阵：。

#### 矩阵的行列式

这里就先谈行列式的几何意义，最后再谈行列式的计算方法的由来。思考一下，经过线性变换，空间发生了变化，相应的面积也会发生变化，如图1-26所示。

[](http://124.220.164.99:8090/upload/2022/11/image-1669501010318.png)

图1-26 矩阵的行列式

在*x*轴基向量扩大2倍，*y*轴基向量扩大3倍的情况下，面积扩大了6倍，而这个6正好就是线性变换后矩阵的行列式，由此可见，行列式的几何意义就是：线性变换改变面积（体积、超平面）的比例。下面引出正式的定义：



从几何上来推理一下公式(1-79)，随意假设*x*轴基向量与*y*轴基向量在任意一个线性变化的作用下变换，结果如图1-27所示。

[](http://124.220.164.99:8090/upload/2022/11/image-1669501136446.png)

图1-27 矩阵行列式的几何推导

由于求的是图1-27中平行四边形的面积，从图上构建出整个长方形，得到每一部分区域，做减法就可以得到结果了，整个减法如下：



在上面的基础上来考虑何为行列式为0。

行列式为0证明空间变换的比例为0，那么说明空间进行了收缩，也就是降维或者说数据冗余。如图1-28所示，两个向量共线，本来两个向量应该撑起一个平面的，但是现在只有一条线，所以就从二维变为了一维：

[](http://124.220.164.99:8090/upload/2022/11/image-1669501251512.png)

图1-28 矩阵行列式为零

这同样也就是我们俗称的线性相关，于是我们将整个逻辑串联起来就是：

（1）线性相关 = 有冗余 = 两个或多个向量在同一平面（点、空间、超平面） = 空间变化率为0 = 行列式为0。

（2）线性无关 = 没有冗余 = 任何一个向量都不会被其他向量表示 = 空间变化率不为0 = 行列式不为0

#### 4. 矩阵的秩

秩表示什么呢? 假设四个行向量组成的矩阵如下：



求其极大线性无关组假设有：**（因为是零向量，跟谁都有关，所以只假设前三个向量线性相关）。



解得：，即线性无关。

矩阵的秩表示当前矩阵中线性无关的向量组的个数，在当前例子中即为3。

在之前说过矩阵可以看作对向量做变换，例如可以对二维图形进行旋转，比如用旋转矩阵。此时的旋转矩阵秩为2，变换后的效果如图1-29所示。

[](http://124.220.164.99:8090/upload/2022/11/image-1669501713768.png)

图1-29 旋转矩阵

变换后的结果依然是二维的。如果用矩阵进行变换呢？此时矩阵的秩为1，变换效果如图1-30所示。

[](http://124.220.164.99:8090/upload/2022/11/image-1669501838884.png)

图1-30 旋转降维

变换后的结果变成了一维。这里就体现出矩阵的秩对向量的变换作用。即：如果矩阵的秩低于向量空间的维度，那么会对向量进行降维。

最后强调一下，矩阵的秩实际上代表了矩阵中不重复的主要特征个数。举个生活中的例子：家里的有三只小猫咪，给它们拍摄了100张照片，组成了十行十列的矩阵，该矩阵的秩等于3，就算拍一千张照片，组成的矩阵秩还是3。

### 1.2.12 特征值和特征向量

通俗的理解特征值和特征向量描述了什么，怎么获得成功的人生？在正确的道路上坚持努力下去。可以把千百种人生选择看作特征向量（它是有方向的！）；把在这个方向上的努力看作特征值（它是一个衡量大小的量）。

下面引出它的数学定义：对于给定矩阵，寻找一个常数和非零向量，使得向量被矩阵作用后，所得的向量与原向量平行，并且满足。其中，是特征向量，是特征值，特征值越大表示该特征向量越重要。举个例子来理解：

向量和向量都是向量的特征向量。

因为它们都可写成的形式：



特征向量有无数个，且此时特征向量对原始向量只有伸缩作用，没有旋转作用。小结一下：矩阵和向量作乘法，向量会变成另一个方向或长度的新向量，主要会发生旋转、伸缩变化，如果矩阵乘以某些向量后，向量不发生旋转变换，只产生伸缩变换，那么就说这些向量是矩阵的特征向量，伸缩的比例就是特征值。

最后，怎么求解特征向量呢？公式如下：



公式(1-84)把解特征向量变成了求解齐次方程的问题。例如求**的特征值。



很容易看出**。把**代入式**：



说明，*x*轴和*y*轴上所有的向量都是特征向量，且经过矩阵的作用，会在*x*轴上拉伸两倍，在*y*轴上拉伸3倍。

特征值分解：**，其中是矩阵的特征向量组成的矩阵；是特征值组成的对角矩阵，里面的特征值是由大到小排列的，这些特征值所对应的特征向量就是描述这个矩阵变化方向。

特征值分解的过程就是求解，即求解下面的线性方程组：



可以解出*n*个特征值：，再把*n*个特征值代入式子。可以求出*n*个对应的特征向量。

对于每一个特征值与特征向量满足：**。

因为。

**的等号左边等于：。

**的等号右边等于：。

可得：，如果矩阵可逆，则有。

至于特征分解的意义，通过一个具体的例子来讲：对于矩阵而言，由于是方阵，所以不会对向量进行维度的升降，所以矩阵代表的运动实际上只有两种：旋转和拉伸。其特征值分解后的结果如图1-31所示。

[](http://124.220.164.99:8090/upload/2022/11/image-1669504966575.png)

图1-31 特征值分解

通过图1-311来看特征值分解，实际上把旋转和拉伸运动给分解开了。为什么要分解呢？其实是为了筛选出整体中最具有代表性的特征。举个例子：图像也可以被视为矩阵，图像的每一个点都是由RGB值定义的，所以每个图像可以被表示为三个巨型矩阵（分别是矩阵）。

SVD分解可以被认为是特征值分解（Eigen Value Decomposition，EVD）的延伸。特征值分解将一个矩阵分解为两组正交的特征向量和一个特征值对角线矩阵。

而特征值矩阵又是从大到小排列的，特征值大小的下降速度很快，我们可以通过丢弃一些特征值来压缩数据。对于压缩图像来说，只要人眼不可察觉便可以认为是成功的压缩。

简单来说，就是通过把一块大的数据分解为很多项，通过给数据的每个项的重要程度排序，挑选出一部分最重要的保留，丢弃一部分最不重要的，来实现数据压缩。