胡说八道——线性代数

引言

线性代数是现代工程应用和科学研究的基本工具，其地位就好比智能手机在人们日常生活中的地位。大学二年级就会学线性代数，但鉴于中国教材的写法，许多人学完线性代数，却并不知道如何用线性代数。理论和实践之间有一条鸿沟，这导致很多研究生（尽管考研也要考线性代数）一看到线性代数就头大。可偏偏实际应用中涉及大量的线性代数知识。我所讲的实际应用，是指那些真正可以赚钱的应用——诸如发射卫星、控制汽车、金融分析，而不是吃饭、睡觉、洗脚、泡妹。

代数

代数是算术概念的进一步抽象，在代数发明之前，数学上只有算术式：

$$
\begin{aligned}
1+1&=2\\
3*3&=9\\
12/3&=4\\
24=2*3&*4
\end{aligned}
\tag{1}
$$

假设用字母a代替数字2，用字母b代替数字3，算术式（1）就可以写成代数式：

$$
\begin{aligned}
1+1 &= a \\
b*b &= 9 \\
12/b &= 4 \\
24 &= a*b*4
\end{aligned} \tag{2}
$$

代数的产生，使得人们可以研究数字间更普遍的关系，而不再局限于某个具体的数字。例如乘法交换律$a*b=b*a$，乘法分配率$(a+b)*c=a*c+b*c$。这些性质具有普遍意义，不仅在实数范围内有效，而且在复数以及其他集合上均有效。

上面的代数称为初等代数（线性代数则属于高等代数），初等代数只涉及到代数的一般性质，这些性质比较直接，很容易在具体的数学运算中被观察到。而高等代数涉及到更为抽象的代数性质，这些性质不那么容易观察到，因此理解起来也有一定难度，但这些性质却是非常有用的。下面我们以方程为例，讲解一下高等代数中一些性质的应用。

方程

考虑如下问题：

> 养禽场共养鸡鸭3000只,鸡的只数是鸭的三倍,鸡鸭各有多少只?

用方程很容易解答这个问题。设鸡的数量为x，鸭的数量为y，列出如下方程：

$$
\begin{aligned}
x+y &= 3000 \\
3*y &= x
\end{aligned} \tag{3}
$$

解方程可知，鸭子有750只，鸡有2250只。方程是代数的一个重大应用。上面这个问题，如果不列方程，那么解答的难度就会高很多，所以方程的是解应用题的“大杀器”。但现实中的应用题不会那么简单（上面这个题目仅仅只能作为学生的练习），众多的工程应用（例如建筑、制造、物理、天文、火箭、导弹之类）并不只有两个未知数，通常未知数有几百、几千，甚至几万，这样的方程如何求解呢?

一般线性代数教材均从高斯消元法求方程开始，可见线性代数的一个主要用途就是解多元方程。方程组（3）可以写成如下的矩阵运算形式：

$$
\begin{bmatrix} 1 & 1 \\ 1 & -3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3000 \\ 0 \end{bmatrix} \tag{4}
$$

矩阵运算（4）可以简写为：

$$
AX=b \tag{5}
$$

其中$A=\begin{bmatrix}1 & 1\\1 & -3\end{bmatrix}$，$X=\begin{bmatrix}x\\y\end{bmatrix}$，$b=\begin{bmatrix}3000\\0\end{bmatrix}$。
我们发现矩阵形式的方程组(5)相对于(4)有很多优点：

1. 表达非常简洁。无论方程的数量有多少，只用3个符号（$AX=b$）就可以表示。
2. 方程的解很容易计算。显然$X=\frac{b}{A}$（准确的应该是$X=A^{-1}b$），并且如果$|A|=0$方程无解。

无论是求A^{-1}，还是计算$|A|$都用到了代数性质，这些性质属于高等代数（线性代数），不是初等代数。可以想见，如果没有线性代数的这些性质，高元方程组（比如100元一次方程组）很难表示，也很难求解。

线性方程

线性方程的数学定义比较复杂，之所以复杂是为了数学逻辑的完整。如果仅仅为了理解概念，可以认为$AX=b$形式的就是线性方程。那么什么是非线性方程呢？含有x^2，$sinx$，$\int_0^1 xdx$等项目的都可以称为非线性的方程。线性方程其实是最容易求解的那种方程。显然还有一个数学分支是专门研究非线性方程的，并且非线性才是最常见的情况。懂了吧？大学教科书为什么主要讲线性方程，那是因为按照大多数本科生的智力水平，能搞懂线性方程就不错啦。至于非线性方程，数学家也没有怎么搞明白，除非是专业研究数学，其他行当只要能够用上非线性一定是很高级的。没错，**线性是基本应该搞懂的，最低层次的要求**。

$AX=b$的方程有什么好研究的？这可不是简单的问题，线性代数可是洋洋洒洒讲了大半的篇幅。从高斯消元开始，讲到行列式，再讲到矩阵，讲到矩阵运算，讲到基，讲到秩，讲到其次线性方程有解的条件，讲到非其次线性方程有解的条件等等（**显然只有将这些概念全部搞清楚，才能搞定解方程，否则你能看懂A^{-1}吗**）。事实上，如果数学书能够更结合一下实际，线性代数上的诸多内容，还远远不够——哪怕仅仅是解最简单的线性方程。

从实际角度来看解方程这个问题。$AX=b$如果没有解，也就是$|A|=0$难道就不解了吗？显然不是。因为从实际情况来看，无解并非真的无解，有可能是误差导致的。实际应用中误差无处不在。以天体物理为例，星球间的距离差个几千米是非常正常的，完全在误差范围之内。可是这样的误差，完全足够让一个方程组无解。怎么办呢？可以求最优解。最小二乘法就是做这个事情的。最小二乘法超出了本文的范畴，因为本文希望在道理上写得透彻些，在细节上写得粗糙些。就目前而言，如果用计算机程序来解方程，几乎所有的高级语言都有最小二乘法的函数库。所以只要知道最小二乘法是干这个事情的，然后知道怎么调用函数即可。例如python可以按照如下方式求解方程：

```
from numpy import *
from numpy.linalg import *
a=array([[1, 3, 4],[2, -1, 2],[1, 3, 4]])
```

```
b=array([[1],[2],[3]])
print lstsq(a,b)
```

上述代码求解了如下方程组：

$$\begin{bmatrix} 1 & 3 & 4 \\ 2 & -1 & 2 \\ 1 & 3 & 4 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \tag{6}$$

方程组(6)是无解的，但可以用最小二乘法求得一个最优解。
某些方程组的解还可能有特殊约束，例如X必须大于0或者X必须是整数。一旦有了这些约束，就不能直接按照标准算法求解，具体问题需要具体分析，但通常都是求满足某种条件的最优解。牛顿迭代法是最常用的方法。

线性方程的另外一种常见情况是：原问题可能不是线性方程，但通过一些变换，可以变换为线性方程；在某些情况下，甚至是将非齐次线性方程转换为齐次线性方程。这类情况比较复杂，也是具体问题解法不同。计算机视觉中有大量的算法可以归为这一类型。以后讲到具体算法时，大家可以体会一下。

矩阵分解

矩阵分解是一个很难懂的问题，如果只看线性代数的教材，根本搞不明白为什么要做分解。这玩意是干吗的？很多人都有这个疑问吧。简单来说，矩阵分解有些类似于因式分解，或者因数分解。例如：

$$\begin{gathered} a^2 + 2ab + b^2 = (a+b)^2 \\ 12 = 2 * 2 * 3 \end{gathered} \tag{7}$$

同理，假设对矩阵A做奇异值分解，即：

$$AX = U\Sigma V^T X = b \tag{8}$$

其中$A = U\Sigma V^T$与因式分解、因数分解有着类似的目的：

1. 分解后$U\Sigma V^T$能够表达A某方面的性质，如同$12 = 2 * 2 * 3$，通过这个分解我们知道12可以被3和2整除。
2. 分解后$U\Sigma V^T$的性质能够被利用。不同的性质用法不同，所以这个问题也是需要具体分析的。

总而言之，矩阵分解绝对是有用的，至于怎么用要看具体的情况。所以会计算矩阵分解几乎是毫无意义的，搞懂（或者发现）某个矩阵分解的用途才是真正有意义的。具体到奇异值分

解，它的应用**非常非常非常……**广泛。

其他应用非常广泛的矩阵分解还包括：特征值分解、RQ分解、QR分解、cholesky分解等等。不同的分解有不同的用途，用途也没有统一的法则。因此只学某某分解是无用的，必须学习某某分解的用途。

矩阵变换

矩阵变换有很多形式，当然最常见的是乘以一个变换矩阵B。例如AB或者BA就是一个变换。矩阵变换有什么用？我们把矩阵看成数字，假设$A=3$，$B=2$，那么$AB=6$。这个变换可以认为是把A增加了2倍，也可以认为是把A变换为偶数。所以矩阵变换的作用**与变换矩阵的性质有关**。
回到矩阵，假设

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

变换矩阵

$$B=\begin{bmatrix}0.5 & 0\\ 0 & 0.5\end{bmatrix}$$

则

$$BA=\begin{bmatrix}1*0.5 & 2*0.5 & 3*0.5\\ 4*0.5 & 5*0.5 & 6*0.5\end{bmatrix}$$

相当于A的每一个元素都乘以0.5。这种变换就是缩放变换。
当然还有其他很多变换，有些变换还相当复杂，例如PCA变换，傅里叶变换（严格来说，这些变换不能归在矩阵变换，但是我这个文档不那么严谨）。
总之，矩阵变换就是把矩阵A经过处理变成A'，变换的性质和用途都是由变换本身决定的。与矩阵分解类似，学习矩阵变换也必须和用途结合起来，脱离了用途，矩阵变换就是无用的。

总结

数学是自然科学的基础，自然科学是工程应用的基础。线性代数近代的数学，故此但凡是有一点市场价值的工程应用，没有不用到线性代数的。遗憾的是线性代数的教材只涉及理论，并不涉及应用。这种现象是如何导致的？数学在产生之初，铁定是从应用中来的。商贩因为

需要交易，所以发明了算术。地主为了收租，所以发明了土地面积的测定方法。但是随着数学方法的增多，数学成为了一个庞杂的逻辑体系。渐渐的数学家发现，某些数学性质可以仅仅依赖逻辑推理，而不用对客观世界进行观测，就可以推导出来。这里面最著名的恐怕要数黎曼几何在爱因斯坦相对论中的应用啦。黎曼的几何空间无法直接用肉眼观测到，故此黎曼的几何是完全通过数学推导得出的。这种与常识相违背的几何，却和物理现象吻合得很好（真理往往不是肉眼可见的）。

因此从某个时代开始，数学和应用的关系发生了转变。过去先有应用再有数学，现有先有数学再有应用。例如**奇异值分解**有广泛的应用（例如数据压缩、人脸识别、最小二乘法解线性方程、主成分分析等等），但如果只看到奇异值分解的公式，你是很难理解它有如此广泛应用的。**奇异值分解**就是先有数学后有应用的典型例子。时至今日，奇异值分解的应用仍然在不断的被发掘。线性代数的应用也不仅仅局限于此。

我揣测线性代数教材只重理论不重实践的可能原因：**应用问题太多，课时有限，多讲不利于考试，学生没有这个需求**。我也不知道这么说对不对。