APR 29TH, 2014

LAS 2-Matrix

线性代数那些事 Things of Linear Algebra
逸夫图书馆，2014/4/28

矩阵

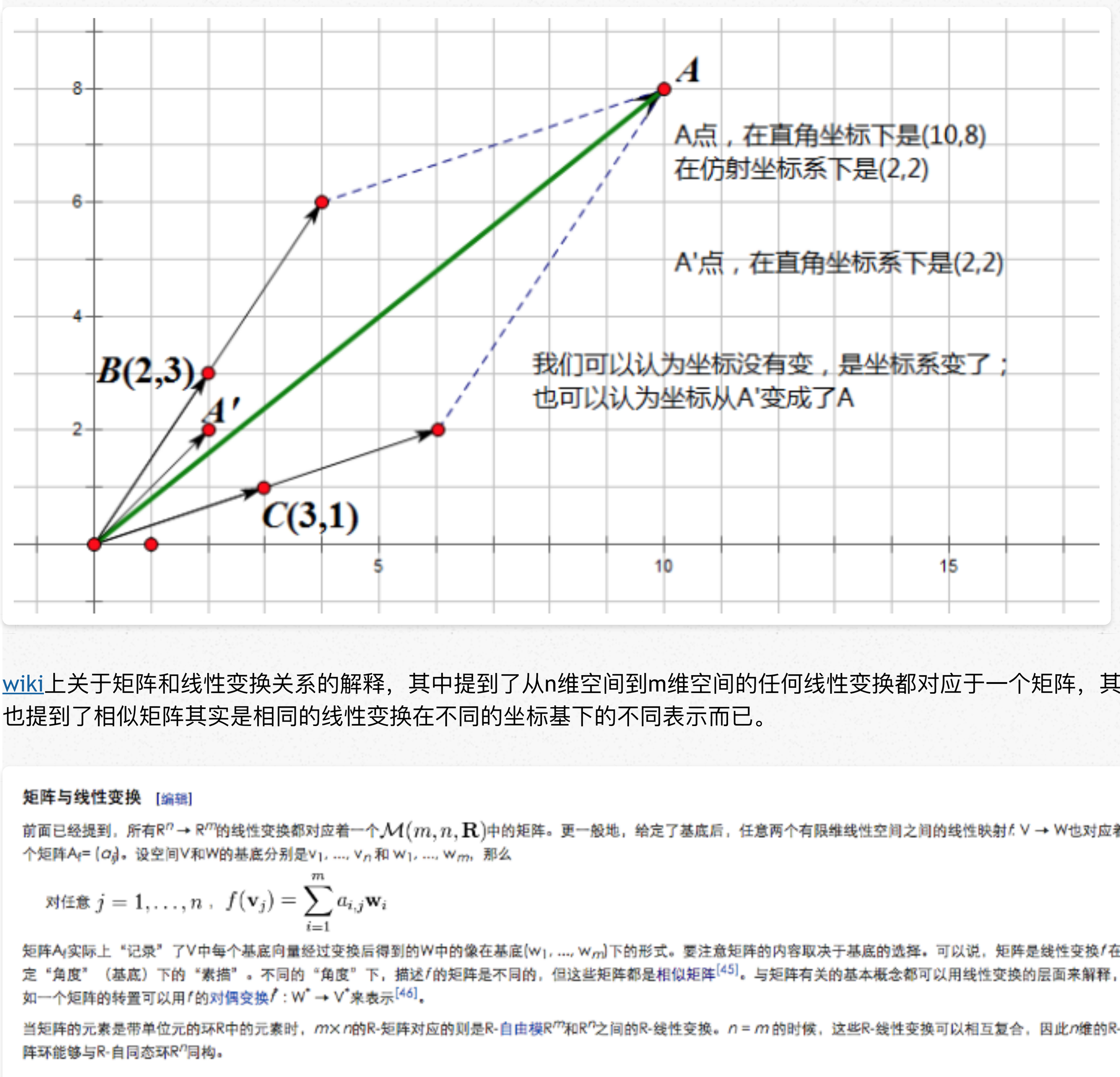
1.什么是矩阵?

这个问题很多人进行过探讨，在网上也比较火，比如孟岩的三篇《理解矩阵》，知乎上如何直观理解矩阵和线代，还有其他人对矩阵的理解，例如新理解矩阵1和新理解矩阵2等等。

那，到底什么是矩阵呢?

总结起来，我觉得，矩阵就是线性变换，作用在一个点上就是将这个点移动到该空间的另一个点，作用在向量上就是对这个向量进行放缩或者旋转或者反射等一系列的线性变换，作用在矩阵上那就是对矩阵中的每一个列向量进行线性变换之后然后进行叠加结果(这就是为什么矩阵的乘法有些奇妙的原因)。因为运动是相对的，你可以理解为坐标系没有变，被作用对象发生了变化，也可以理解为被作用对象没有变，变的是坐标系(也就是空间的基)。

> *这里借用下小苏的图和解释。矩阵$A=\begin{bmatrix} a11 & a12 \\ a21 & a22 \end{bmatrix}$事实上是由两个向量$[a11,a21]^T$和$[a12,a22]^T$（这里的向量都是列向量）组成，它描述了一个平面（仿射）坐标系。换句话说，这两个向量其实是这个坐标系的两个基，而运算$y=Ax$则是告诉我们，在A这个坐标系下的x向量，在I坐标系下是怎样的。这里的I坐标系就是我们最常用的直角坐标系，也就是说，任何向量（包括矩阵里边的向量），只要它前面没有矩阵作用于它，那么它都是在直角坐标系下度量出来的。下图所用的矩阵$A=\begin{bmatrix} 3 & 2 \\ 1 & 3 \end{bmatrix}$构成了一个仿射坐标系，在这个坐标系下，有一个向量$x=[2,2]^T$，它在直角坐标系下测得的坐标为$[10,8]^T$，现在我们不难发现，直接用矩阵乘法来计算，有 $Ax=[3*2+2*2,1*2+3*2]^T=[10,8]^T$小苏对此展开讨论了它和矩阵乘法之间的联系*



wiki上关于矩阵和线性变换关系的解释，其中提到了从n维空间到m维空间的任何线性变换都对应于一个矩阵，其中也提到了相似矩阵其实是相同的线性变换在不同的坐标基下的不同表示而已。

矩阵与线性变换 [编辑]

前面已经提到，所有$R^n\to R^m$的线性变换都对应着一个$\mathcal{M}(m,n,\mathbf{R})$中的矩阵。更一般地，给定了基底后，任意两个有限维线性空间之间的线性映射$f:V\to W$也对应着一个矩阵$A_f=(\alpha_{ij})$。设空间V和W的基底分别是$v_1,\ldots,v_n$和$w_1,\ldots,w_m$，那么

对任意 $j=1,\ldots,n$，$f(\mathbf{v}_j)=\sum_{i=1}^{m}a_{i,j}\mathbf{w}_i$

矩阵A_f实际上"记录"了V中每个基底向量经过变换后得到的W中的像在基底$(w_1,\ldots,w_m)$下的形式。要注意矩阵的内容取决于基底的选择。可以说，矩阵是线性变换f在特定"角度"（基底）下的"素描"。不同的"角度"下，描述f的矩阵是不同的，但这些矩阵都是相似矩阵[45]。与矩阵有关的基本概念都可以用线性变换的层面来解释，比如一个矩阵的转置可以用f的对偶变换$f^*:W^*\to V^*$来表示[46]。

当矩阵的元素是带单位元的环R中的元素时，$m\times n$的R-矩阵对应的则是R-自由模R^m和R^n之间的R-线性变换。$n=m$的时候，这些R-线性变换可以相互复合，因此n维的R-矩阵环能够与R-自同态环R^n同构。

2.线性变换

好吧，矩阵是线性变换，那什么是线性变换呢?

wiki中对线性变换的解释，这些变换其实主要包括缩放、旋转、反射等，注意不包括平移，线性变换需要满足下面两个条件：

$$f(\alpha v)=\alpha v \quad f(u+v)=f(u)+f(v)$$

线性变换 [编辑]

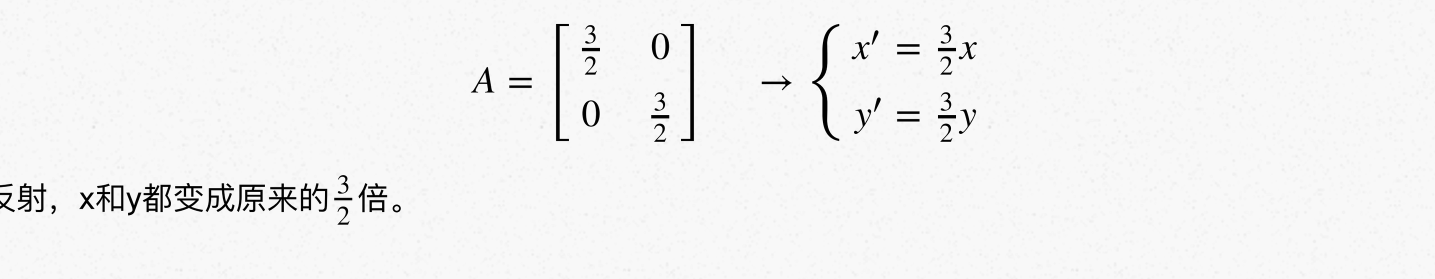
主条目：线性变换

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连系：以$\mathbb{R}^n$表示所有长度为n的行向量的集合。每个$m\times n$的矩阵A都代表了一个从$\mathbb{R}^n$射到$\mathbb{R}^m$的线性变换。反过来，对每个线性变换$f:\mathbb{R}^n\to\mathbb{R}^m$，都存在唯一$m\times n$矩阵$A_f$使得对所有$\mathbb{R}^n$中的元素$x$，$f(x)=A_f x$。这个矩阵$A_f$第$i$行第$j$列上的元素是正则基向量$\mathbf{e}_j=(0,\cdots,0,1,0,\cdots0)^T$（第$j$个元素是1，其余元素是0的向量）在$f$映射后的向量$f(\mathbf{e}_j)$的第$i$个元素。

也就是说，从$\mathbb{R}^n$射到$\mathbb{R}^m$的线性变换构成的向量空间$\mathcal{L}(\mathbb{R}^n,\mathbb{R}^m)$上存在一个到$\mathcal{M}(m,n,\mathbb{R})$的一一映射：$f\mapsto A_f$

以下是一些典型的2维实平面上的线性变换对平面向量（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。

水平错切变换，幅度m=1.25.	水平反射变换	"挤压"变换，压缩程度r=3/2	放缩变换，3/2倍	旋转变换，左转30°
$\begin{bmatrix}1 & 1.25\\0 & 1\end{bmatrix}$	$\begin{bmatrix}-1 & 0\\0 & 1\end{bmatrix}$	$\begin{bmatrix}3/2 & 0\\0 & 2/3\end{bmatrix}$	$\begin{bmatrix}3/2 & 0\\0 & 3/2\end{bmatrix}$	$\begin{bmatrix}\cos(\pi/6) & -\sin(\pi/6)\\\sin(\pi/6) & \cos(\pi/6)\end{bmatrix}$



设有$k\times m$的矩阵B代表线性变换$g:R^m\to R^k$，则矩阵积BA代表了线性变换的复合$g\circ f$[9]，因为

$(g\circ f)(x)=g(f(x))=g(Ax)=B(Ax)=(BA)x$

矩阵的秩是指矩阵中线性无关的行（列）向量的最大个数[10]，同时也是矩阵对应的线性变换的像空间的维度[11]。秩－零化度定理说明矩阵的列数量等于矩阵的秩与零空间维度之和[12]。

(1)放缩变换

$$A=\begin{bmatrix}\frac{3}{2} & 0\\ 0 & \frac{3}{2}\end{bmatrix} \quad \to \begin{cases} x'=\frac{3}{2}x \\ y'=\frac{3}{2}y \end{cases}$$

放缩反射，x和y都变成原来的$\frac{3}{2}$倍。

(2)反射变换

$$A=\begin{bmatrix}-1 & 0\\ 0 & 1\end{bmatrix} \quad \to \begin{cases} x'=-x \\ y'=y \end{cases}$$

水平反射，x变成对应的相反数，y不变。

什么是反射?

反射 on wiki上的解释是：**反射是把一个物体变换成它的镜像的映射。对于二维空间中的反射，需要使用一条直线（反射轴）作为"镜子"，对于三维空间中的反射就要使用平面作为镜子。**

最常用的反射变换就是Householder变换 on wiki了，这一变换将一个向量变换为由一个超平面反射的镜像，是一种线性变换。Householder变换可以将向量的某些元素置零，同时保持该向量的范数不变。关于Householder的内部原理以及代码实现请请参考后面的矩阵分解总结部分，或者参考我写的另一份总结《Numerical Methods Using Matlab》第三章 矩阵特征值和奇异值分解

(3)旋转变换

$$A=\begin{bmatrix}\cos(\frac{\pi}{6}) & -\sin(\frac{\pi}{6})\\ \sin(\frac{\pi}{6}) & \cos(\frac{\pi}{6})\end{bmatrix} \quad \to \begin{cases} x'=\frac{\sqrt{3}}{2}x-\frac{1}{2}y \\ y'=\frac{1}{2}x+\frac{\sqrt{3}}{2}y \end{cases}$$

旋转变换，一般性的证明请看这张图，wiki中对二维空间旋转的解释。

二维空间 [编辑]

在讨论旋转的时候理解参照系是重要的。从一个观点，你可以保持坐标轴固定旋转向量。从另一个观点，你可以保持向量固定旋转坐标系。

在第一种观点看来，坐标或向量关于原点的逆时针旋转；或者从第二种观点看来，平面或轴关于原点的顺时针旋转。这里的(x,y)被旋转了θ并希望知道旋转后的坐标(x',y')：

$$\begin{bmatrix}x'\\y'\end{bmatrix}=\begin{bmatrix}\cos\theta & -\sin\theta\\ +\sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}.$$

或

$$x'=x\cos\theta-y\sin\theta$$
$$y'=x\sin\theta+y\cos\theta$$

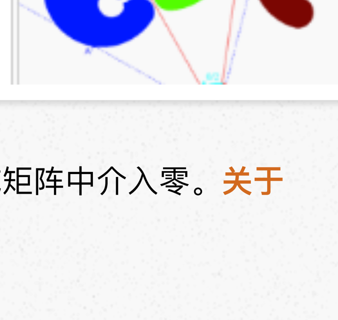
平面或轴关于原点的逆时针旋转，在新平面中的坐标将顺时针旋转到新坐标。在这种情况下，如果在旧平面中的坐标是(x,y)，同一个向量在新平面中的坐标是(x',y')，则：

$$\begin{bmatrix}x'\\y'\end{bmatrix}=\begin{bmatrix}\cos\theta & +\sin\theta\\ -\sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}.$$

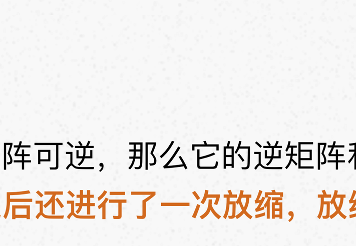
或

$$x'=x\cos\theta+y\sin\theta$$
$$y'=-x\sin\theta+y\cos\theta$$

向量(x, y)的大小同于向量(x', y')的大小。



在绕一个点旋转之后绕另一个不同的点的平面旋转导致要么是旋转（如本图）要么是平移的一个总和运动。



最常用的旋转矩阵就是Givens旋转。Givens 旋转在数值线性代数中主要的用途是在向量或矩阵中介入零。**关于Givens旋转的内部原理以及代码实现请请参考后面的矩阵分解总结部分。**

3.逆矩阵和伴随矩阵

理解了矩阵就是线性变换之后，那么就很容易明白逆矩阵就是将被作用对象从变换后的位置变换回来!

那伴随矩阵又是什么呢?

wiki在线性代数中，一个方形矩阵的伴随矩阵A*是一个类似于逆矩阵A^{-1}的概念。如果矩阵可逆，那么它的逆矩阵和它的伴随矩阵之间只差一个系数($A^{-1}=\frac{A^*}{det(A)}$)。也就是说，**伴随矩阵其实是变换回来之后还进行了一次放缩，放缩的大小与矩阵的行列式值有关**。

设R是一个交换环，A是一个以R中元素为系数的 n×n 的矩阵。A的伴随矩阵可按如下步骤定义：

- 定义：A关于第i行第j列的余子式（记作M_{ij}）是去掉A的第i行第j列之后得到的(n − 1)×(n − 1)矩阵的行列式。
- 定义：A关于第i行第j列的代数余子式是：

  $\mathbf{C}_{ij}=(-1)^{i+j}\mathbf{M}_{ij}$。
- 定义：A的余子矩阵是一个n×n的矩阵C，使得其第i行第j列的元素是A关于第i行第j列的代数余子式。

引入以上的概念后，可以定义：矩阵A的伴随矩阵是A的余子矩阵的转置矩阵：

$\mathrm{adj}(\mathbf{A})=\mathbf{C}^T$。

也就是说，A的伴随矩阵是一个n×n的矩阵（记作adj(A)），使得其第i行第j列的元素是A关于第j行第i列的代数余子式：

$[\mathrm{adj}(\mathbf{A})]_{ij}=\mathbf{C}_{ji}$。

举例说明伴随矩阵的计算，伴随矩阵其实就是原矩阵的代数余子式矩阵的转置!

3x3矩阵 [编辑]

对于3 × 3的矩阵，情况稍微复杂一点：

$$\mathbf{A}=\begin{pmatrix}A_{11} & A_{12} & A_{13}\\ A_{21} & A_{22} & A_{23}\\ A_{31} & A_{32} & A_{33}\end{pmatrix}.$$

其伴随矩阵是：

$$\mathrm{adj}(\mathbf{A})=\begin{pmatrix} +\begin{vmatrix}A_{22} & A_{23}\\A_{32} & A_{33}\end{vmatrix} & -\begin{vmatrix}A_{12} & A_{13}\\A_{32} & A_{33}\end{vmatrix} & +\begin{vmatrix}A_{12} & A_{13}\\A_{22} & A_{23}\end{vmatrix} \\ -\begin{vmatrix}A_{21} & A_{23}\\A_{31} & A_{33}\end{vmatrix} & +\begin{vmatrix}A_{11} & A_{13}\\A_{31} & A_{33}\end{vmatrix} & -\begin{vmatrix}A_{11} & A_{13}\\A_{21} & A_{23}\end{vmatrix} \\ +\begin{vmatrix}A_{21} & A_{22}\\A_{31} & A_{32}\end{vmatrix} & -\begin{vmatrix}A_{11} & A_{12}\\A_{31} & A_{32}\end{vmatrix} & +\begin{vmatrix}A_{11} & A_{12}\\A_{21} & A_{22}\end{vmatrix} \end{pmatrix}$$

其中

$$\begin{vmatrix}A_{im} & A_{in}\\A_{jm} & A_{jn}\end{vmatrix}=\det\begin{pmatrix}A_{im} & A_{in}\\A_{jm} & A_{jn}\end{pmatrix}.$$

要注意伴随矩阵是余子矩阵的转置，第3行第2列的系数应该是A关于第2行第3列的代数余子式。

伴随矩阵的一些性质

性质 [编辑]

对n×n的矩阵A和B，有：

1. $\mathrm{adj}(\mathbf{I})=\mathbf{I}$
2. $\mathrm{adj}(\mathbf{AB})=\mathrm{adj}(\mathbf{B})\,\mathrm{adj}(\mathbf{A})$
3. $\mathrm{adj}(\mathbf{A}^T)=\mathrm{adj}(\mathbf{A})^T$
4. $\det\big(\mathrm{adj}(\mathbf{A})\big)=\det(\mathbf{A})^{n-1}$
5. $\mathrm{adj}(k\mathbf{A})=k^{n-1}\,\mathrm{adj}(\mathbf{A})$
6. 当n>2时，$\mathrm{adj}(\mathrm{adj}(\mathbf{A}))=(\det\mathbf{A})^{n-2}\mathbf{A}$
7. 如果A可逆，那么$\mathrm{adj}(\mathbf{A}^{-1})=\mathrm{adj}(\mathbf{A})^{-1}=\frac{A}{\det A}$
8. 如果A是对称矩阵，那么其伴随矩阵也是对称矩阵；如果A是反对称矩阵，那么当n为偶数时，A的伴随矩阵也是反对称矩阵，n为奇数时则是对称矩阵。
9. 如果A是（半）正定矩阵，那么其伴随矩阵也是（半）正定矩阵。
10. 如果矩阵A和B相似，那么$\mathrm{adj}(\mathbf{A})$和$\mathrm{adj}(\mathbf{B})$也相似。
11. 如果n>2，那么非零矩阵A是正交矩阵当且仅当$\mathrm{adj}(\mathbf{A})=\pm A^T$

伴随矩阵的秩 [编辑]

当矩阵A可逆时，它的伴随矩阵也可逆，因此两者的秩一样，都是n。当矩阵A不可逆时，A的伴随矩阵的秩通常并不与A相同。当A的秩为n-1 时，其伴随矩阵的秩为1，当A的秩小于n-1 时，其伴随矩阵为零矩阵。

伴随矩阵的特征值 [编辑]

设矩阵A在复域中的特征值为$\lambda_1,\lambda_2\cdots\lambda_n$（即为特征多项式的n个根），则A的伴随矩阵的特征值为$\lambda_2\lambda_3\cdots\lambda_n,\ \lambda_1\lambda_3\cdots\lambda_n,\cdots,\lambda_1\lambda_2\cdots\lambda_{n-1}$。

还需要注意的是，**逆矩阵是对于方阵来说的，只有方阵才有逆矩阵的概念，那要不是方阵呢？那么就是广义的逆矩阵**！广义逆矩阵在最小二乘法中有重要的应用。关于逆矩阵的求解以及代码实现请参考我写的另一份总结《Numerical Methods Using Matlab》第一章 线性方程组求解，最小二乘问题请参考第四章 曲线拟合和多项式插值。关于矩阵逆矩阵的求解可以看后面的矩阵分解部分。

4.秩

什么是矩阵的秩?

wiki中的解释在线性代数中，一个矩阵A的列秩是A的线性独立的纵列的极大数目。类似地，行秩是A的线性独立的横行的极大数目。矩阵的列秩和行秩总是相等的，因此它们可以简单地称作矩阵A的秩。通常表示为r(A)，rk(A)或rank A。m × n矩阵的秩最大为m和n中的较小者，表示为 min(m,n)。有尽可能大的秩的矩阵被称为有满秩；类似的，否则矩阵是秩不足（或称为"欠秩"）的。

矩阵的行秩与列秩相等，是线性代数基本定理的重要组成部分. 其基本证明思路是，矩阵可以看作线性映射的变换矩阵，列秩为像空间的维度，行秩为非零原像空间的维度，因此列秩与行秩相等，即像空间的维度与非零原像空间的维度相等（这里的非零原像空间是指约去了零空间后的商空间：原像空间）。这从矩阵的奇异值分解就可以看出来。证明可以参见wiki

黄老师的总结中还给出了**初等变换不改变矩阵的行秩和列秩**的证明，此外还有，以下四个表述是等价的：

• A 为满秩矩阵. • A 为可逆矩阵. • A 为非奇异矩阵. • $|A|\neq 0$.

说了这么多，那到底矩阵的秩对于矩阵表示的这个线性变换来说意味着什么?

我的理解是，矩阵的秩其实就是至少需要几个基向量就能完全表示该矩阵(线性变换)。

好吧，其实还有转置矩阵对不对？转置on wiki是从m×n矩阵的向量空间到所有n×m矩阵的向量空间的线性映射。这个我还不太理解，读者若明白了请留言告知，谢谢!

OK，矩阵的理解就到这里吧，下节介绍矩阵的特征值和特征向量，也可以直接看看后面关于其他矩阵的理解，例如相似矩阵和正交矩阵等等。

Posted by hujiawei • Apr 29th, 2014 • math

« LAS 1-Determinant

LAS 3-EigenVectors and Eigenvalues »