第二十二讲：对角化和A的幂

对角化矩阵

上一讲我们提到关键方程$Ax=\lambda x$，通过$\det(A-\lambda I)=0$得到特征向量λ，再带回关键方程算出特征向量x。

在得到特征值与特征向量后，该如何使用它们？我们可以利用特征向量来对角化给定矩阵。

有矩阵A，它的特征向量为$x_1,x_2,\cdots,x_n$，使用特征向量作为列向量组成一个矩阵$S=[x_1x_2\cdots x_n]$，即特征向量矩阵， 再使用公式$$S^{-1}AS=\Lambda\tag{1}$$将A对角化。注意到公式中有S^{-1}，也就是说特征向量矩阵S必须是可逆的，于是我们需要n个线性无关的特征向量。

现在，假设A有n个线性无关的特征向量，将它们按列组成特征向量矩阵S，则$AS=A[x_1x_2\cdots x_n]$，当我们分开做矩阵与每一列相乘的运算时，易看出Ax_1就是矩阵与自己的特征向量相乘，其结果应该等于λ_1x_1。那么$AS=[(\lambda_1x_1)(\lambda_2x_2)\cdots(\lambda_nx_n)]$。可以进一步化简原式，使用右乘向量按列操作矩阵的方法，将特征值从矩阵中提出来，得到$[x_1x_2\cdots x_n]\begin{bmatrix}\lambda_1&0&\cdots&0\\0&\lambda_2&\cdots&0\\\vdots&\vdots&\ddots&\vdots\\0&0&\cdots&\lambda_n\end{bmatrix}=S\Lambda$。

于是我们看到，从AS出发，得到了$S\Lambda$，特征向量矩阵又一次出现了，后面接着的是一个对角矩阵，即特征值矩阵。这样，再继续左乘S^{-1}就得到了公式(1)。当然，所以运算的前提条件是特征向量矩阵S可逆，即矩阵A有n个线性无关的特征向量。这个式子还要另一种写法，$A=S\Lambda S^{-1}$。

我们来看如何应用这个公式，比如说要计算A^2。

- 先从$Ax=\lambda x$开始，如果两边同乘以A，有$A^2x=\lambda Ax=\lambda^2x$，于是得出结论，对于矩阵$A^2$，其特征值也会取平方，而特征向量不变。
- 再从$A=S\Lambda S^{-1}$开始推导，则有$A^2=S\Lambda S^{-1}S\Lambda S^{-1}=S\Lambda^2S^{-1}$。同样得到特征值取平方，特征向量不变。

两种方法描述的是同一个现象，即对于矩阵幂运算A^2，其特征向量不变，而特征值做同样的幂运算。对角矩阵$\Lambda^2=\begin{bmatrix}\lambda_1^2&0&\cdots&0\\0&\lambda_2^2&\cdots&0\\\vdots&\vdots&\ddots&\vdots\\0&0&\cdots&\lambda_n^2\end{bmatrix}$。

特征值和特征向量给我们了一个深入理解矩阵幂运算的方法，$A^k=S\Lambda^kS^{-1}$。

再来看一个矩阵幂运算的应用：如果$k\to\infty$，则$A^k\to 0$（趋于稳定）的条件是什么？从$S\Lambda^kS^{-1}$易得，$|\lambda_i|<1$。再次强调，所有运算的前提是矩阵A存在n个线性无关的特征向量。如果没有n个线性无关的特征向量，则矩阵就不能对角化。

关于矩阵可对角化的条件：

- 如果一个矩阵有n个互不相同的特征值（即没有重复的特征值），则该矩阵具有n个线性无关的特征向量，因此该矩阵可对角化。

- 如果一个矩阵的特征值存在重复值，则该矩阵可能具有n个线性无关的特征向量。比如取10阶单位矩阵，I_{10}具有10个相同的特征值1，但是单位矩阵的特征向量并不短缺，每个向量都可以作为单位矩阵的特征向量，我们很容易得到10个线性无关的特征向量。当然这里例子中的I_{10}的本来就是对角矩阵，它的特征值直接写在矩阵中，即对角线元素。

  同样的，如果是三角矩阵，特征值也写在对角线上，但是这种情况我们可能会遇到麻烦。矩阵$A=\begin{bmatrix}2&1\\0&2\end{bmatrix}$，计算行列式值$\det(A-\lambda I)=\begin{vmatrix}2-\lambda&1\\0&2-\lambda\end{vmatrix}=(2-\lambda)^2=0$，所以特征值为$\lambda_1=\lambda_2=2$，带回$Ax=\lambda x$得到计算$\begin{bmatrix}0&1\\0&0\end{bmatrix}$的零空间，我们发现$x_1=x_2=\begin{bmatrix}1\\0\end{bmatrix}$，代数重度（algebraic multiplicity，计算特征值重复次数时，就用代数重度，就是它作为多项式根的次数，这里的多项式就是$(2-\lambda)^2$）为2，这个矩阵无法对角化。这就是上一讲的退化矩阵。

我们不打算深入研究有重复特征值的情形。

求$u_{k+1}=Au_k$

从$u_1=Au_0$开始，$u_2=A^2u_0$，所有$u_k=A^ku_0$。下一讲涉及微分方程（differential equation），会有求导的内容，本讲先引入简单的差分方程（difference equation）。本例是一个一阶差分方程组（first order system）。

要解此方程，需要将u_0展开为矩阵A特征向量的线性组合，即$u_0=c_1x_1+c_2x_2+\cdots+c_nx_n=[x_1x_2\cdots x_n]\begin{bmatrix}c_1\\c_2\\\vdots\\c_n\end{bmatrix}=Sc$。于是$Au_0=c_1Ax_1+c_2Ax_2+\cdots+c_nAx_n=c_1\lambda_1x_1+c_2\lambda_2x_2+\cdots+c_n\lambda_nx_n$。继续化简原式，$Au_0=[x_1x_2\cdots x_n]\begin{bmatrix}\lambda_1&0&\cdots&0\\0&\lambda_2&\cdots&0\\\vdots&\vdots&\ddots&\vdots\\0&0&\cdots&\lambda_n\end{bmatrix}\begin{bmatrix}c_1\\c_2\\\vdots\\c_n\end{bmatrix}=S\Lambda c$。用矩阵的方式同样可以得到该式：$Au_0=S\Lambda S^{-1}u_0=S\Lambda S^{-1}Sc=S\Lambda c$。

那么如果我们要求$A^{100}u_0$，则只需要将λ变为λ^{100}，而系数c与特征向量x均不变。

当我们真的要计算$A^{100}u_0$时，就可以使用$S\Lambda^{100}c=c_1\lambda_1^{100}x_1+c_2\lambda_2^{100}x_2+\cdots+c_n\lambda_n^{100}x_n$。

接下来看一个斐波那契数列（Fibonacci sequence）的例子：

$0,1,1,2,3,5,8,13,\cdots,F_{100}=?$，我们要求第一百项的公式，并观察这个数列是如何增长的。可以想象这个数列并不是稳定数列，因此无论如何该矩阵的特征值并不都小于一，这样才能保持增长。而他的增长速度，则有特征值来决定。

已知$F_{k+2}=F_{k_1}+F_k$，但这不是$u_{k+1}=Au_k$的形式，而且我们只要一个方程，而不是方程组，同时这是一个二阶差分方程（就像含有二阶导数的微分方程，希望能够化简为一阶倒数，也就是一阶差分）。

使用一个**小技巧**，令$u_k=\begin{bmatrix}F_{k+1}\\F_k\end{bmatrix}$，再追加一个方程组成方程组：$\begin{cases}F_{k+2}&=F_{k+1}+F_k\\F_{k+1}&=F_{k+1}\end{cases}$，再把方程组用矩阵表达得到$\begin{bmatrix}F_{k+2}\\F_{k+1}\end{bmatrix}=\begin{bmatrix}1&1\\1&0\end{bmatrix}\begin{bmatrix}F_{k+1}\\F_k\end{bmatrix}$，于是我们得到了$u_{k+1}=$

$Au_k, A=\begin{bmatrix}1&1\\1&0\end{bmatrix}$。我们把二阶标量方程（second-order scalar problem）转化为一阶向量方程组（first-order system）。

我们的矩阵$A=\begin{bmatrix}1&1\\1&0\end{bmatrix}$是一个对称矩阵，所以它的特征值将会是实数，且他的特征向量将会互相正交。因为是二阶，我们可以直接利用迹与行列式解方程组$\begin{cases}\lambda_1+\lambda_2&=1\\\lambda_1\cdot\lambda_2&=-1\end{cases}$。

在求解之前，我们先写出一般解法并观察$|A-\lambda I|=\begin{vmatrix}1-\lambda&1\\1&-\lambda\end{vmatrix}=\lambda^2-\lambda-1=0$，与前面斐波那契数列的递归式$F_{k+2}=F_{k+1}+F_k\longrightarrow F_{k+2}-F_{k+1}-F_k=0$比较，我们发现这两个式子在项数与幂次上非常相近。

- 用求根公式解特征值得$\begin{cases}\lambda_1=\frac{1}{2}\left(1+\sqrt{5}\right)\approx 1.618\\\lambda_2=\frac{1}{2}\left(1-\sqrt{5}\right)\approx -0.618\end{cases}$，得到两个不同的特征值，一定会有两个线性无关的特征向量，则该矩阵可以被对角化。

我们先来观察这个数列是如何增长的，数列增长由什么来控制？——特征值。哪一个特征值起决定性作用？——较大的一个。

$F_{100}=c_1\left(\frac{1+\sqrt{5}}{2}\right)^{100}+c_2\left(\frac{1-\sqrt{5}}{2}\right)^{100}\approx c_1\left(\frac{1+\sqrt{5}}{2}\right)^{100}$，由于$-0.618$在幂增长中趋近于$0$，所以近似的忽略该项，剩下较大的项，我们可以说数量增长的速度大约是1.618。可以看出，这种问题与求解$Ax=b$不同，这是一个动态的问题，A的幂在不停的增长，而问题的关键就是这些特征值。

- 继续求解特征向量，$A-\lambda I=\begin{bmatrix}1-\lambda&1\\1&1-\lambda\end{bmatrix}$，因为有根式且矩阵只有二阶，我们直接观察$\begin{bmatrix}1-\lambda&1\\1&1-\lambda\end{bmatrix}\begin{bmatrix}?\\?\end{bmatrix}=0$，由于$\lambda^2-\lambda-1=0$，则其特征向量为$\begin{bmatrix}\lambda\\1\end{bmatrix}$，即$x_1=\begin{bmatrix}\lambda_1\\1\end{bmatrix}, x_2=\begin{bmatrix}\lambda_2\\1\end{bmatrix}$。

最后，计算初始项$u_0=\begin{bmatrix}F_1\\F_0\end{bmatrix}=\begin{bmatrix}1\\0\end{bmatrix}$，现在将初始项用特征向量表示出来$\begin{bmatrix}1\\0\end{bmatrix}=c_1x_1+c_2x_2$，计算系数得$c_1=\frac{\sqrt{5}}{5}, c_2=-\frac{\sqrt{5}}{5}$。

来回顾整个问题，对于动态增长的一阶方程组，初始向量是u_0，关键在于确定A的特征值及特征向量。特征值将决定增长的趋势，发散至无穷还是收敛于某个值。接下来需要找到一个展开式，把u_0展开成特征向量的线性组合。

- 再下来就是套用公式，即A的k次方表达式$A^k=S\Lambda^kS^{-1}$，则有$u_{99}=Au_{98}=\cdots=A^{99}u_0=S\Lambda^{99}S^{-1}Sc=S\Lambda^{99}c$，代入特征值、特征向量得$u_{99}=\begin{bmatrix}F_{100}\\F_{99}\end{bmatrix}=\begin{bmatrix}\frac{1+\sqrt5}{2}&\frac{1-\sqrt5}{2}\\1&1\end{bmatrix}\begin{bmatrix}\left(\frac{1+\sqrt5}{2}\right)^{99}&0\\0&\left(\frac{1-\sqrt5}{2}\right)^{99}\end{bmatrix}\begin{bmatrix}\frac{\sqrt5}{5}\\-\frac{\sqrt5}{5}\end{bmatrix}=\begin{bmatrix}c_1\lambda_1^{100}+c_2\lambda_2^{100}\\c_1\lambda_1^{99}+c_2\lambda_2^{99}\end{bmatrix}$，最终结果为$F_{100}=c_1\lambda_1^{100}+c_2\lambda_2^{100}$。
- 原式的通解为$u_k=c_1\lambda^kx_1+c_2\lambda^kx_2$。

下一讲将介绍求解微分方程。