第二十三讲：微分方程和e^{At}

微分方程$\frac{\mathrm{d}u}{\mathrm{d}t}=Au$

本讲主要讲解解一阶方程（first-order system）一阶倒数（first derivative）常系数（constant coefficient）线性方程，上一讲介绍了如何计算矩阵的幂，本讲将进一步涉及矩阵的指数形式。我们通过解一个例子来详细介绍计算方法。

有方程组$\begin{cases}\frac{\mathrm{d}u_1}{\mathrm{d}t}&=-u_1+2u_2\\\frac{\mathrm{d}u_2}{\mathrm{d}t}&=u_1-2u_2\end{cases}$，则系数矩阵是$A=\begin{bmatrix}-1&2\\1&-2\end{bmatrix}$，设初始条件为在0时刻$u(0)=\begin{bmatrix}u_1\\u_2\end{bmatrix}=\begin{bmatrix}1\\0\end{bmatrix}$。

- 这个初始条件的意义可以看做在开始时一切都在u_1中，但随着时间的推移，将有$\frac{\mathrm{d}u_2}{\mathrm{d}t}>0$，因为$u_1$项初始为正，$u_1$中的事物会流向$u_2$。随着时间的发展我们可以追踪流动的变化。

- 根据上一讲所学的知识，我们知道第一步需要找到特征值与特征向量。$A=\begin{bmatrix}-1&2\\1&-2\end{bmatrix}$，很明显这是一个奇异矩阵，所以第一个特征值是$\lambda_1=0$，另一个特征向量可以从迹得到$tr(A)=-3$。当然我们也可以用一般方法计算$\left|A-\lambda I\right|=\begin{vmatrix}-1-\lambda&2\\1&-2-\lambda\end{vmatrix}=\lambda^2+3\lambda=0$。

  （教授提前剧透，特征值$\lambda_2=-3$将会逐渐消失，因为答案中将会有一项为e^{-3t}，该项会随着时间的推移趋近于0。答案的另一部分将有一项为e^{0t}，该项是一个常数，其值为1，并不随时间而改变。通常含有0特征值的矩阵会随着时间的推移达到稳态。）

- 求特征向量，$\lambda_1=0$时，即求A的零空间，很明显$x_1=\begin{bmatrix}2\\1\end{bmatrix}$；$\lambda_2=-3$时，求$A+3I$的零空间，$\begin{bmatrix}2&2\\1&1\end{bmatrix}$的零空间为$x_2=\begin{bmatrix}1\\-1\end{bmatrix}$。

- 则方程组的通解为：$u(t)=c_1e^{\lambda_1t}x_1+c_2e^{\lambda_2t}x_2$，通解的前后两部分都是该方程组的纯解，即方程组的通解就是两个与特征值、特征向量相关的纯解的线性组合。我们来验证一下，比如取$u=e^{\lambda_1t}x_1$带入$\frac{\mathrm{d}u}{\mathrm{d}t}=Au$，对时间求导得到$\lambda_1e^{\lambda_1t}x_1=Ae^{\lambda_1t}x_1$，化简得$\lambda_1x_1=Ax_1$。

  对比上一讲，解$u_{k+1}=Au_k$时得到$u_k=c_1\lambda^kx_1+c_2\lambda^kx_2$，而解$\frac{\mathrm{d}u}{\mathrm{d}t}=Au$我们得到$u(t)=c_1e^{\lambda_1t}x_1+c_2e^{\lambda_2t}x_2$。

- 继续求c_1,c_2，$u(t)=c_1\cdot 1\cdot\begin{bmatrix}2\\1\end{bmatrix}+c_2\cdot e^{-3t}\cdot\begin{bmatrix}1\\-1\end{bmatrix}$，已知$t=0$时，$\begin{bmatrix}1\\0\end{bmatrix}=c_1\begin{bmatrix}2\\1\end{bmatrix}+c_2\begin{bmatrix}1\\-1\end{bmatrix}$（$Sc=u(0)$），所以$c_1=\frac{1}{3},c_2=\frac{1}{3}$。

- 于是我们写出最终结果，$u(t)=\frac{1}{3}\begin{bmatrix}2\\1\end{bmatrix}+\frac{1}{3}e^{-3t}\begin{bmatrix}1\\-1\end{bmatrix}$。

稳定性：这个流动过程从$u(0)=\begin{bmatrix}1\\0\end{bmatrix}$开始，初始值1的一部分流入初始值0中，经过无限的时间最终达到稳态$u(\infty)=\begin{bmatrix}\frac{2}{3}\\\frac{1}{3}\end{bmatrix}$。所以，要使得$u(t)\to 0$，则需要负的特征值。但如果特征值为复数呢？如$\lambda=-3+6i$，我们来计算$\left|e^{(-3+6i)t}\right|$，其中的$\left|e^{6it}\right|$部分为$|\cos 6t+i\sin 6t|=1$，因为这部分的模为$\cos^2\alpha+\sin^2\alpha=1$，这个虚部就在单位圆上转悠。所以只有实数部分才是重要的。所以我们可以把前面的结论改为**需要实部为负数的特征值**。实部会决定最终结果趋近于0或∞，虚部不过是一些小杂音。

收敛态：需要其中一个特征值实部为0，而其他特征值的实部皆小于0。

发散态：如果某个特征值实部大于0。上面的例子中，如果将A变为$-A$，特征值也会变号，结果发散。

再进一步，我们想知道如何从直接判断任意二阶矩阵的特征值是否均小于零。对于二阶矩阵$A=\begin{bmatrix}a&b\\c&d\end{bmatrix}$，矩阵的迹为$a+d=\lambda_1+\lambda_2$，如果矩阵稳定，则迹应为负数。但是这个条件还不够，有反例迹小于0依然发散：$\begin{bmatrix}-2&0\\0&1\end{bmatrix}$，迹为$-1$但是仍然发散。还需要加上一个条件，因为$\det A=\lambda_1\cdot\lambda_2$，所以还需要行列式为正数。

总结：原方程组有两个相互耦合的未知函数，u_1,u_2相互耦合，而特征值和特征向量的作则就是解耦，也就是对角化（diagonalize）。回到原方程组$\frac{\mathrm{d}u}{\mathrm{d}t}=Au$，将$u$表示为特征向量的线性组合$u=Sv$，代入原方程有$S\frac{\mathrm{d}v}{\mathrm{d}t}=ASv$，两边同乘以$S^{-1}$得$\frac{\mathrm{d}v}{\mathrm{d}t}=S^{-1}ASv=\Lambda v$。以特征向量为基，将$u$表示为$Sv$，得到关于$v$的对角化方程组，新方程

组不存在耦合，此时$\begin{cases}\frac{\mathrm{d}v_1}{\mathrm{d}t} &= \lambda_1 v_1\\ \frac{\mathrm{d}v_2}{\mathrm{d}t} &= \lambda_2 v_2\\ \vdots & \vdots\\ \frac{\mathrm{d}v_n}{\mathrm{d}t} &= \lambda_n v_n\end{cases}$，这是一个各未知函数间没有联系的方程组，它们的解的一般形式为$v(t)=e^{\Lambda t}v(0)$，则原方程组的解的一般形式为$u(t)=e^{At}u(0)=Se^{\Lambda t}S^{-1}u(0)$。这里引入了指数部分为矩阵的形式。

指数矩阵e^{At}

在上面的结论中，我们见到了e^{At}。这种指数部分带有矩阵的情况称为指数矩阵（exponential matrix）。

理解指数矩阵的关键在于，将指数形式展开称为幂基数形式，就像$e^x=1+\frac{x^2}{2}+\frac{x^3}{6}+\cdots$一样，将$e^{At}$展开成幂级数的形式为：

$$e^{At}=I+At+\frac{(At)^2}{2}+\frac{(At)^3}{6}+\cdots+\frac{(At)^n}{n!}+\cdots$$

再说些题外话，有两个极具美感的泰勒级数：$e^x=\sum\frac{x^n}{n!}$与$\frac{1}{1-x}=\sum x^n$，如果把第二个泰勒级数写成指数矩阵形式，有$(I-At)^{-1}=I+At+(At)^2+(At)^3+\cdots$，这个式子在$t$非常小的时候，后面的高次项近似等于零，所以可以用来近似$I-At$的逆矩阵，通常近似为$I+At$，当然也可以再加几项。第一个级数对我们而言比第二个级数好，因为第一个级数总会收敛于某个值，所以e^x总会有意义，而第二个级数需要A特征值的绝对值小于1（因为涉及矩阵的幂运算）。我们看到这些泰勒级数的公式对矩阵同样适用。

回到正题，我们需要证明$Se^{\Lambda t}S^{-1}=e^{At}$，继续使用泰勒级数：

$$\begin{gathered}e^{At}=I+At+\frac{(At)^2}{2}+\frac{(At)^3}{6}+\cdots+\frac{(At)^n}{n!}+\cdots\\ e^{At}=SS^{-1}+S\Lambda S^{-1}t+\frac{S\Lambda^2S^{-1}}{2}t^2+\frac{S\Lambda^3S^{-1}}{6}t^3+\cdots+\frac{S\Lambda^nS^{-1}}{n!}t^n+\cdots\\ e^{At}=S\left(I+\Lambda t+\frac{\Lambda^2t^2}{2}+\frac{\Lambda^3t^3}{3}+\cdots+\frac{\Lambda^nt^n}{n}+\cdots\right)S^{-1}\\ e^{At}=Se^{\Lambda t}S^{-1}\end{gathered}$$

需要注意的是，e^{At}的泰勒级数展开是恒成立的，但我们推出的版本却需要**矩阵可对角化**这个前提条件。

最后，我们来看看什么是$e^{\Lambda t}$，我们将e^{At}变为对角矩阵就是因为对角矩阵简单、没有耦合，$e^{\Lambda t}=\begin{bmatrix}e^{\lambda_1 t}&0&\cdots&0\\0&e^{\lambda_2 t}&\cdots&0\\\vdots&\vdots&\ddots&\vdots\\0&0&\cdots&e^{\lambda_n t}\end{bmatrix}$。

有了$u(t)=Se^{\Lambda t}S^{-1}u(0)$，再来看矩阵的稳定性可知，所有特征值的实部均为负数时矩阵收敛，此时对角线上的指数收敛为0。如果我们画出复平面，则要使微分方程存在稳定解，则特征值存在于复平面的左侧（即实部为负）；要使矩阵的幂收敛于0，则特征值存在于单位圆内部（即模小于1），这是幂稳定区域。（上一讲的差分方程需要计算矩阵的幂。）

同差分方程一样，我们来看二阶情况如何计算，有$y''+by'+k=0$。我们也模仿差分方程的情形，构造方程组$\begin{cases}y''&=-by'-ky\\y'&=y'\end{cases}$，写成矩阵形式有$\begin{bmatrix}y''\\y'\end{bmatrix}=\begin{bmatrix}-b&-k\\1&0\end{bmatrix}\begin{bmatrix}y'\\y\end{bmatrix}$，令$u'=\begin{bmatrix}y''\\y'\end{bmatrix}, u=\begin{bmatrix}y'\\y\end{bmatrix}$。

继续推广，对于5阶微分方程$y'''''+by''''+cy'''+dy''+ey'+f=0$，则可以写作$\begin{bmatrix}y'''''\\y''''\\y'''\\y''\\y'\end{bmatrix}=\begin{bmatrix}-b&-c&-d&-e&-f\\1&0&0&0&0\\0&1&0&0&0\\0&0&1&0&0\\0&0&0&1&0\end{bmatrix}\begin{bmatrix}y''''\\y'''\\y''\\y'\\y\end{bmatrix}$，这样我们就把一个五阶微分方程化为$5\times 5$一阶方程组了，然后就是求特征值、特征向量了步骤了。