Matrix Exponentials

Suppose A is a square matrix. We then make the following definition:

$$e^A = I + A + \frac{1}{2}A^2 + \frac{1}{6}A^3 + \cdots = \sum_{n=0}^{\infty} \frac{1}{n!}A^n.$$

Note that this is a matrix of the same size as A (let's not worry about convergence...).

`Example:` `If` $A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$, `then`

$$e^A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 2^2 & 0 \\ 0 & 3^2 \end{bmatrix} + \frac{1}{6}\begin{bmatrix} 2^3 & 0 \\ 0 & 3^3 \end{bmatrix} = \cdots$$

$$= \begin{bmatrix} 1 + 2 + \frac{1}{2}2^2 + \frac{1}{6}2^3 + \cdots & 0 \\ 0 & 1 + 3 + \frac{1}{2}3^2 + \frac{1}{6}3^3 + \cdots \end{bmatrix} = \begin{bmatrix} e^2 & 0 \\ 0 & e^3 \end{bmatrix}.$$

`This sort of equality is true for any diagonal matrix.`

Now suppose A is not diagonal. We have two problems:

(1) In general, it is quite time consuming to compute high powers of A;

(2) Even if we knew all the high powers of A, we'd still have to put corresponding entries together into a series and determine its sum in order to find the entries of e^A.

The key to the problem of explicitly calculating e^A is the Cayley-Hamilton theorem.

Theorem (Cayley-Hamilton). *A matrix satisfies its own characteristic equation.*

`Example:` `The matrix` $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ `has characteristic equation`

$$\lambda^2 - 2\lambda + 1 = 0.$$

`The Cayley-Hamilton theorem implies that`

$$A^2 - 2A + I = O.$$

`To verify this, we have`

$$A^2 - 2A + I = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} - 2\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

How does this help? Suppose A is 2×2. Then the characteristic equation for A looks like

$$\lambda^2 + b\lambda + c = 0,$$

so, by the Cayley-Hamilton theorem, we have

$$A^2 + bA + cI = O, \qquad \text{or} \qquad A^2 = -bA - cI.$$

This allows us to replace the A^2 term in the series for e^A with a linear combination of A and I. Similarly we have that

$$A^3 = (A^2)A = (-bA - cI)A = -bA^2 - cA = -b(-bA - cI) - cA = (b^2 - c)A - bcI,$$

so that, again, we can replace A^3 with a linear combination of A and I. This argument continues for higher powers of A, so that eventually we find that we may write e^A as a linear combination of A and I, rather than as an infinite linear combination of I and all positive powers of A. In other words, the expression simplifies to

$$e^A = I + A + \frac{1}{2}A^2 + \frac{1}{6}A^3 + \cdots = a_1 A + a_0 I,$$

for some constants a_0 and a_1. So to find the matrix e^A, we just need to find these two constants (for an $n \times n$ matrix, there will be n constants).

How do we find these constants?

Suppose λ is an eigenvalue for A. We know that

$$e^\lambda = 1 + \lambda + \frac{1}{2}\lambda^2 + \frac{1}{6}\lambda^3 + \cdots.$$

On the other hand, λ also satisfies the characteristic equation for A. So we can do exactly the same substitutions we did for the series for e^A, and write e^λ as a linear combination of λ and 1. Moreover, *it will be exactly the same linear combination.* So we also have

$$e^\lambda = a_1\lambda + a_0.$$

But this is the sort of equation we can solve (it's linear in the unknowns a_0 and a_1).

So here's the general strategy (for 2×2):

Case 1 (This is the only case of relevance for the final.): A has two distinct real eigenvalues, λ_1 and λ_2. Then we have two equations

$$e^{\lambda_1} = a_1\lambda_1 + a_0 \qquad e^{\lambda_2} = a_1\lambda_2 + a_0.$$

This is a linear system of two equations in two unknowns, so we can easily solve for a_1 and a_0, and then set $e^A = a_1 A + a_0 I$.

Example: The matrix $A = \begin{bmatrix} -5 & 2 \\ -6 & 2 \end{bmatrix}$ has eigenvalues $\lambda_1 = -1$ and $\lambda_2 = -2$ (verify this). So we have

$$\begin{cases} -a_1 + a_0 = e^{-1} \\ -2a_1 + a_0 = e^{-2} \end{cases} .$$

Row-reducing, we have

$$\left[\begin{array}{cc|c} -1 & 1 & e^{-1} \\ -2 & 1 & e^{-2} \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 1 & -1 & -e^{-1} \\ 0 & 1 & 2e^{-1} - e^{-2} \end{array}\right]$$

so the solution is

$$a_1 = e^{-1} - e^{-2} \qquad a_0 = 2e^{-1} - e^{-2}$$

It follows that

$$\begin{aligned} e^A = a_1 A + a_0 I &= (e^{-1} - e^{-2}) \begin{bmatrix} -5 & 2 \\ -6 & 2 \end{bmatrix} + (2e^{-1} - e^{-2}) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} -3e^{-1} + 4e^{-2} & 2e^{-1} - 2e^{-2} \\ 6e^{-2} - 6e^{-1} & 4e^{-1} - 3e^{-2} \end{bmatrix} \end{aligned}$$

Everything that follows is just for your personal information. It will not be on the test.

Case 2: Assume A has a repeated eigenvalue λ. We still have two unknowns, but we seem only to have one equation. To get a second equation, we differentiate* the first with respect to λ. Thus our two equations are

$$e^\lambda = a_1 \lambda + a_0 \qquad e^\lambda = a_1.$$

Case 3: Assume A has complex eigenvalues. These are necessarily conjugates of one another, and therefore not equal. This gives two equations in two (complex) unknowns. Setting real and imaginary parts equal gives four equations in four unknowns, that can easily be solved.

* Why are we allowed to do this? Just because two equations in λ agree at a certain point (when λ equals an eigenvalue for A) doesn't mean they have the same derivative there. So what's going on? In general, the expression on the left, as a function of λ, is the standard exponential curve, while the one on the right is a straight line. Points of intersection of the line with the curve correspond to eigenvalues. If there are two distinct eigenvalues, there are two distinct places where the line crosses the curve. If the eigenvalues are complex, the line and the curve are disjoint. If there is only one eigenvalue, with multiplicity two (as we are assuming in this case), then the line and the curve intersect in exactly one point. But this means the line and the curve are tangent at that point, and so must have the same derivative there. By the way, in the application to differential equations that we did in class, you should denote the eigenvalues by λt, so that when differentiating with respect to λ, a factor of t comes out in front of the exponential.