The Matrix Exponential (with exercises)
by Dan Klain

Version 2018.09.28
Corrections and comments are welcome.

The Matrix Exponential

For each $n \times n$ complex matrix A, define the *exponential* of A to be the matrix

$$e^A \quad = \quad \sum_{k=0}^{\infty} \frac{A^k}{k!} \quad = \quad I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \cdots \tag{1}$$

It is not difficult to show that this sum converges for all complex matrices A of any finite dimension. But we will not prove this here.

If A is a 1×1 matrix $[t]$, then $e^A = [e^t]$, by the Maclaurin series formula for the function $y = e^t$. More generally, if D is a diagonal matrix having diagonal entries $d_1, d_2, \ldots, d_n$, then we have

$$e^D \quad = \quad I + D + \frac{1}{2!}D^2 + \cdots$$

$$\cdots = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \cdots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix} + \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \cdots & \ddots & 0 \\ 0 & \cdots & 0 & d_n \end{bmatrix} + \begin{bmatrix} \frac{d_1^2}{2!} & 0 & \cdots & 0 \\ 0 & \frac{d_1^2}{2!} & \cdots & 0 \\ \vdots & \cdots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{d_1^2}{2!} \end{bmatrix} + \cdots$$

$$\cdots = \begin{bmatrix} e^{d_1} & 0 & \cdots & 0 \\ 0 & e^{d_2} & \cdots & 0 \\ \vdots & \cdots & \ddots & 0 \\ 0 & \cdots & 0 & e^{d_n} \end{bmatrix}$$

The situation is more complicated for matrices that are not diagonal. However, if a matrix A happens to be *diagonalizable*, there is a simple algorithm for computing e^A, a consequence of the following lemma.

Lemma 1. *Let A and P be complex $n \times n$ matrices, and suppose that P is invertible. Then*

$$e^{P^{-1}AP} = P^{-1}e^AP$$

Proof. Recall that, for all integers $m \geq 0$, we have $(P^{-1}AP)^m = P^{-1}A^mP$. The definition (1) then yields

$$\begin{aligned} e^{P^{-1}AP} \quad &= \quad I + P^{-1}AP + \frac{(P^{-1}AP)^2}{2!} + \cdots \\ &= \quad I + P^{-1}AP + P^{-1}\frac{A^2}{2!}P + \cdots \\ &= \quad P^{-1}(I + A + \frac{A^2}{2!} + \cdots)P \quad = \quad P^{-1}e^AP \end{aligned}$$

$\square$

If a matrix A is diagonalizable, then there exists an invertible P so that $A = PDP^{-1}$, where D is a diagonal matrix of eigenvalues of A, and P is a matrix having eigenvectors of A as its columns. In this case, $e^A = Pe^DP^{-1}$.

Example: Let A denote the matrix

$$A = \begin{bmatrix} 5 & 1 \\ -2 & 2 \end{bmatrix}$$

The reader can easily verify that 4 and 3 are eigenvalues of A, with corresponding eigenvectors $w_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and $w_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$. It follows that

$$A = PDP^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}$$

so that

$$e^A = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} e^4 & 0 \\ 0 & e^3 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} 2e^4 - e^3 & e^4 - e^3 \\ 2e^3 - 2e^4 & 2e^3 - e^4 \end{bmatrix}$$

❧

The definition (1) immediately reveals many other familiar properties. The following proposition is easy to prove from the definition (1) and is left as an exercise.

Proposition 2. *Let A be a complex square $n \times n$ matrix.*

(1) *If $\mathbf{0}$ denotes the zero matrix, then $e^{\mathbf{0}} = I$, the identity matrix.*
(2) *$A^m e^A = e^A A^m$ for all integers m.*
(3) *$(e^A)^T = e^{(A^T)}$*
(4) *If $AB = BA$ then $Ae^B = e^B A$ and $e^A e^B = e^B e^A$.*

Unfortunately not all familiar properties of the scalar exponential function $y = e^t$ carry over to the matrix exponential. For example, we know from calculus that $e^{s+t} = e^s e^t$ when s and t are numbers. However this is often *not true* for exponentials of matrices. In other words, it is possible to have $n \times n$ matrices A and B such that $e^{A+B} \neq e^A e^B$. See, for example, Exercise 10 at the end of this section. Exactly when we have equality, $e^{A+B} = e^A e^B$, depends on specific properties of the matrices A and B that will be explored later on. Meanwhile, we can at least verify the following limited case:

Proposition 3. *Let A be a complex square matrix, and let $s, t \in \mathbb{C}$. Then*

$$e^{A(s+t)} = e^{As} e^{At}.$$

Proof. From the definition (1) we have

$$\begin{aligned} e^{As} e^{At} &= \left(I + As + \frac{A^2 s^2}{2!} + \cdots \right) \left(I + At + \frac{A^2 t^2}{2!} + \cdots \right) \\ &= \left(\sum_{j=0}^{\infty} \frac{A^j s^j}{j!} \right) \left(\sum_{k=0}^{\infty} \frac{A^k t^k}{k!} \right) \\ &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \frac{A^{j+k} s^j t^k}{j!k!} \qquad (*) \end{aligned}$$

Let $n = j + k$, so that $k = n - j$. It now follows from $(*)$ that

$$e^{As}e^{At} = \sum_{j=0}^{\infty}\sum_{n=j}^{\infty} \frac{A^n s^j t^{n-j}}{j!(n-j)!} = \sum_{n=0}^{\infty} \frac{A^n}{n!} \sum_{j=0}^{n} \frac{n!}{j!(n-j)!} s^j t^{n-j} = \sum_{n=0}^{\infty} \frac{A^n (s+t)^n}{n!} = e^{A(s+t)}$$

□

Setting $s = 1$ and $t = -1$ in Proposition 3, we find that

$$e^A e^{-A} = e^{A(1+(-1))} = e^{\mathbf{0}} = I.$$

In other words, regardless of the matrix A, the exponential matrix e^A is *always* invertible, and has inverse e^{-A}.

We can now prove a fundamental theorem about matrix exponentials. Both the statement of this theorem and the method of its proof will be important for the study of differential equations in the next section.

Theorem 4. *Let A be a complex square matrix, and let t be a real scalar variable. Let $f(t) = e^{tA}$. Then $f'(t) = Ae^{tA}$.*

Proof. Applying Proposition 3 to the limit definition of derivative yields

$$f'(t) = \lim_{h\to 0} \frac{e^{A(t+h)} - e^{At}}{h} = e^{At}\left(\lim_{h\to 0} \frac{e^{Ah} - I}{h}\right)$$

Applying the definition (1) to $e^{Ah} - I$ then gives us

$$f'(t) = e^{At}\left(\lim_{h\to 0} \frac{1}{h}\left[Ah + \frac{A^2h^2}{2!} + \cdots\right]\right) = e^{At}A = Ae^{At}.$$

□

Theorem 4 is the fundamental tool for proving important facts about the matrix exponential and its uses. Recall, for example, that there exist $n \times n$ matrices A and B such that $e^A e^B \neq e^{A+B}$. The following theorem provides a condition for when this identity does hold.

Theorem 5. *Let A, B be $n \times n$ complex matrices. If $AB = BA$ then $e^{A+B} = e^A e^B$.*

Proof. If $AB = BA$, it follows from the formula (1) that $Ae^{Bt} = e^{Bt}A$, and similarly for other combinations of $A, B, A + B$, and their exponentials.

Let $g(t) = e^{(A+B)t}e^{-Bt}e^{-At}$, where t is a real (scalar) variable. By Theorem 4, and the product rule for derivatives,

$$\begin{aligned} g'(t) &= (A+B)e^{(A+B)t}e^{-Bt}e^{-At} + e^{(A+B)t}(-B)e^{-Bt}e^{-At} + e^{(A+B)t}e^{-Bt}(-A)e^{-At} \\ &= (A+B)g(t) - Bg(t) - Ag(t) \\ &= \mathbf{0}. \end{aligned}$$

Here $\mathbf{0}$ denotes the $n \times n$ zero matrix. Note that it was only possible to factor $(-A)$ and $(-B)$ out of the terms above because we are *assuming* that $AB = BA$.

Since $g'(t) = \mathbf{0}$ for all t, it follows that $g(t)$ is an $n \times n$ matrix of *constants*, so $g(t) = C$ for some constant matrix C. In particular, setting $t = 0$, we have $C = g(0)$. But the definition of $g(t)$ then gives

$$C = g(0) = e^{(A+B)0}e^{-B0}e^{-A0} = e^0 e^0 e^0 = I,$$

the identity matrix. Hence,

$$I = C = g(t) = e^{(A+B)t}e^{-Bt}e^{-At}$$

for all t. After multiplying by $e^{At}e^{Bt}$ on both sides we have

$$e^{At}e^{Bt} = e^{(A+B)t}.$$

□

Exercises:

1. If $A^2 = \mathbf{0}$, the zero matrix, prove that $e^A = I + A$.

2. Use the definition (1) of the matrix exponential to prove the basic properties listed in Proposition 2. (Do not use any of the theorems of the section! Your proofs should use only the definition (1) and elementary matrix algebra.)

3. Show that $e^{cI+A} = e^c e^A$, for all numbers c and all square matrices A.

4. Suppose that A is a real $n \times n$ matrix and that $A^T = -A$. Prove that e^A is an orthogonal matrix (i.e. Prove that, if $B = e^A$, then $B^T B = I$.)

5. If $A^2 = A$ then find a nice simple formula for e^A, similar to the formula in the first exercise above.

6. Compute e^A for each of the following examples:

(**a**) $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ (**b**) $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ (**c**) $A = \begin{bmatrix} a & b \\ 0 & a \end{bmatrix}$

7. Compute e^A for each of the following examples:

(**a**) $A = \begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix}$ (**b**) $A = \begin{bmatrix} a & 0 \\ b & 0 \end{bmatrix}$

8. If $A^2 = I$, show that

$$2e^A = \left(e + \frac{1}{e}\right) I + \left(e - \frac{1}{e}\right) A.$$

9. Suppose $\lambda \in \mathbb{C}$ and $X \in \mathbb{C}^n$ is a non-zero vector such that $AX = \lambda X$.

Show that $e^A X = e^\lambda X$.

10. Let A and B denote the matrices

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \qquad B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

Show by direct computation that $e^{A+B} \neq e^A e^B$.

11. The *trace* of a square $n \times n$ matrix A is defined to be the sum of its diagonal entries:

$$\text{trace(A)} = a_{11} + a_{22} + \cdots + a_{nn}.$$

Show that, if A is diagonalizable, then $\det(e^A) = e^{\text{trace}(A)}$.
Note: Later it will be seen that this is true for all square matrices.

Selected Answers and Solutions

4. Since $(e^A)^T = e^{A^T}$, when $A^T = -A$ we have

$$(e^A)^T e^A = e^{A^T} e^A = e^{-A} e^A = e^{A-A} = e^{\mathbf{0}} = I$$

5. If $A^2 = A$ then $e^A = I + (e-1)A$.

6.

$(\mathbf{a})\ e^A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ $(\mathbf{b})\ e^A = \begin{bmatrix} e & e \\ 0 & e \end{bmatrix}$ $(\mathbf{c})\ A = \begin{bmatrix} e^a & e^a b \\ 0 & e^a \end{bmatrix}$

7.

$(\mathbf{a})\ e^A = \begin{bmatrix} e^a & \frac{b}{a}(e^a - 1) \\ 0 & 1 \end{bmatrix}$ $(\mathbf{b})\ e^A = \begin{bmatrix} e^a & 0 \\ \frac{b}{a}(e^a - 1) & 1 \end{bmatrix}$

(Replace $\frac{b}{a}(e^a - 1)$ by 1 in each case if $a = 0$.)

Linear Systems of Ordinary Differential Equations

Suppose that $y = f(x)$ is a differentiable function of a real (scalar) variable x, and that $y' = ky$, where k is a (scalar) constant. In calculus this differential equation is solved by separation of variables:

$$\frac{y'}{y} = k \qquad \Longrightarrow \qquad \int \frac{y'}{y}\, dx = \int k\, dx$$

so that

$$\ln y = kx + c, \qquad \text{and} \qquad y = e^c e^{kx},$$

for some constant $c \in \mathbb{R}$. Setting $x = 0$ we find that $y_0 = f(0) = e^c$, and conclude that

$$y = y_0 e^{kx}. \tag{2}$$

Instead, let us solve the same differential equation $y' = ky$ in a slightly different way.

Let $F(x) = e^{-kx}y$. Differentiating both sides, we have

$$F'(x) = -ke^{-kx}y + e^{-kx}y' = -ke^{-kx}y + e^{-kx}ky = 0,$$

where the second identity uses the assumption that $y' = ky$. Since $F'(x) = 0$ for all x, the function $F(x)$ must be a constant, $F(x) = a$, for some $a \in \mathbb{R}$. Setting $x = 0$, we find that $a = F(0) = e^{-k0}y(0) = y_0$, where we again let y_0 denote $y(0)$. We conclude that $y = e^{kx}y_0$ as before. Moreover, this method proves that (2) describes *all* solutions to $y' = ky$.

The second point of view will prove valuable for solving a more complicated *linear system* of ordinary differential equations (ODEs). For example, suppose $Y(t)$ is a differentiable vector-valued function:

$$Y = \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix}$$

satisfying the differential equations

$$\begin{aligned} y_1' &= 5y_1 + y_2 \\ y_2' &= -2y_1 + 2y_2 \end{aligned}$$

and initial condition $Y_0 = Y(0) = \begin{bmatrix} -3 \\ 8 \end{bmatrix}$. In other words,

$$Y'(t) = \begin{bmatrix} 5 & 1 \\ -2 & 2 \end{bmatrix} Y = AY,$$

where A denotes the matrix $\begin{bmatrix} 5 & 1 \\ -2 & 2 \end{bmatrix}$.

To solve this system of ODEs, set $F(t) = e^{-At}Y$, where e^{-At} is defined using the matrix exponential formula (1) of the previous section. Differentiating (using the product rule) and applying Theorem 4 then yields

$$F'(t) = -Ae^{-At}Y + e^{-At}Y' = -Ae^{-At}Y + e^{-At}AY = \vec{0},$$

where the second identity uses the assumption that $Y' = AY$. Since $F'(t) = \vec{0}$ (the zero vector), for all t, the function F must be equal to a constant vector $\vec{v}$; that is, $F(t) = \vec{v}$ for all t. Evaluating at $t = 0$ gives

$$\vec{v} = F(0) = e^{-A0}Y(0) = Y_0,$$

where we denote the value $Y(0)$ by the symbol Y_0. In other words,

$$Y_0 = \vec{v} = F(t) = e^{-At}Y,$$

for *all* values of t. Hence,

$$Y = e^{At}Y_0 = e^{At}\begin{bmatrix} -3 \\ 8 \end{bmatrix},$$

and the differential equation is solved! Assuming, of course, that we have a formula for e^{At}.

In the previous section we observed that the eigenvalues of the matrix $A = \begin{bmatrix} 5 & 1 \\ -2 & 2 \end{bmatrix}$ are 4 and 3, with corresponding eigenvectors $w_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and $w_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$. Therefore, for all scalar values t,

$$At = PDtP^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}\begin{bmatrix} 4t & 0 \\ 0 & 3t \end{bmatrix}\begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}$$

so that

$$e^{At} = Pe^{Dt}P^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}\begin{bmatrix} e^{4t} & 0 \\ 0 & e^{3t} \end{bmatrix}\begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}.$$

It follows that

$$Y(t) = e^{At}Y_0 = e^{At}\begin{bmatrix} -3 \\ 8 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}\begin{bmatrix} e^{4t} & 0 \\ 0 & e^{3t} \end{bmatrix}\begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}\begin{bmatrix} -3 \\ 8 \end{bmatrix},$$

so that

$$Y(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} 2e^{4t} - 5e^{3t} \\ -2e^{4t} + 10e^{3t} \end{bmatrix} = e^{4t}\begin{bmatrix} 2 \\ -2 \end{bmatrix} + e^{3t}\begin{bmatrix} -5 \\ 10 \end{bmatrix} = 2e^{4t}\begin{bmatrix} 1 \\ -1 \end{bmatrix} - 5e^{3t}\begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

More generally, if

$$Y'(t) = AY(t),$$

is a linear system of ordinary differential equations, then the arguments above imply that

$$Y = e^{At}Y_0$$

If, in addition, we can diagonalize A, so that

$$A = PDP^{-1} = P\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \cdots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}P^{-1}$$

then

$$e^{At} = Pe^{Dt}P^{-1} = P\begin{bmatrix} e^{\lambda_1 t} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2 t} & \cdots & 0 \\ \vdots & \cdots & \ddots & 0 \\ 0 & \cdots & 0 & e^{\lambda_n t} \end{bmatrix}P^{-1}$$

and $Y(t) = Pe^{Dt}P^{-1}Y_0$. If the columns of P are the eigenvectors $v_1, \ldots, v_n$ of A, where each $Av_i = \lambda_i v_i$, then

$$Y(t) = Pe^{Dt}P^{-1}Y_0 = \left[\begin{array}{c|c|c|c} v_1 & v_2 & \cdots & v_n \end{array}\right]\begin{bmatrix} e^{\lambda_1 t} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2 t} & \cdots & 0 \\ \vdots & \cdots & \ddots & 0 \\ 0 & \cdots & 0 & e^{\lambda_n t} \end{bmatrix}\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$$

where

$$\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = P^{-1}Y_0. \tag{3}$$

Hence,

$$Y(t) = \left[\begin{array}{c|c|c|c} e^{\lambda_1 t}v_1 & e^{\lambda_2 t}v_2 & \cdots & e^{\lambda_n t}v_n \end{array}\right] \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \cdots$$

$$\cdots = c_1e^{\lambda_1 t}v_1 + c_2e^{\lambda_2 t}v_2 + \cdots + c_ne^{\lambda_n t}v_n.$$

These arguments are summarized as follows.

Theorem 6. *Suppose that $Y(t) : \mathbb{R} \longrightarrow \mathbb{R}^n$ (or $\mathbb{C}^n$) is a differentiable function of t such that*

$$Y'(t) = AY(t),$$

and initial value $Y(0) = Y_0$. where A is a diagonalizable matrix, having eigenvalues $\lambda_1, \ldots, \lambda_n$ and corresponding eigenvectors $v_1, \ldots, v_n$. Then

$$Y(t) = c_1e^{\lambda_1 t}v_1 + c_2e^{\lambda_2 t}v_2 + \cdots + c_ne^{\lambda_n t}v_n. \tag{4}$$

If P is the matrix having columns $v_1, \ldots, v_n$ then the constants c_i are given by the identity (3).

If one is given a different initial value of Y, say $Y(t_0)$ at time t_0, then the equation (4) still holds, where

$$\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = e^{-Dt_0}P^{-1}Y(t_0).$$

For exercises on differential equations, please consult the textbook.