Matrix algebra for beginners, Part III
the matrix exponential

Jeremy Gunawardena

Department of Systems Biology
Harvard Medical School
200 Longwood Avenue, Cambridge, MA 02115, USA
jeremy@hms.harvard.edu

October 21, 2006

Contents

1 Introduction

Matrices, which represent *linear* transformations, also arise in the study of *nonlinear* differential equations. Suppose that we have an n-dimensional system of nonlinear differential equations

$$\frac{dx}{dt} = f(x) \;, \tag{1}$$

where $x = (x_1, \cdots, x_n)$ is an n-dimensional vector and $f(x) = (f_1(x), \cdots, f_n(x))$ is an n-dimensional function of x. If x^{st} is a steady state of the system, so that $f(x^{st}) = 0$, then the dynamical behaviour of (1) can, under the right condition, be approximated in the immediate vicinity of x^{st} by the linear system of differential equations

$$\frac{dy}{dt} = Df(x^{st})y \tag{2}$$

where $y = x - x^{st}$ and $Df(x)$ is the $n \times n$ Jacobian matrix of f,

$$Df(x) = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \vdots & \vdots & \vdots \\ \dfrac{\partial f_n}{\partial x_1} & \cdots & \dfrac{\partial f_n}{\partial x_n} \end{pmatrix} .$$

This approximation is known as the Hartman-Grobman Theorem. The "right condition" is that the steady state x^{st} is *hyperbolic*, which means that none of the eigevalues of $Df(x^{st})$ have zero real part. For a 1-dimensional nonlinear dynamical system, non-hyperbolicity corresponds to $(df/dx)|_{x=x^{st}} = 0$, a situation in which the behaviour in the vicinity of x^{st} cannot be determined without examining higher derivatives of f. It is also true that the eigenvalues of $Df(x^{st})$ can tell us about the stability or instability of the steady state x^{st}: if all the eigenvalues of $Df(x^{st})$ have negative real part then x^{st} is stable; if at least one of the eigenvalues has positive real part then x^{st} is unstable. We are not going to prove either of these claims here but they give us the motivation to try and understand the behaviour of linear systems of differential equations like (2). This will need much of what we have learned in Parts I and II.

There is one result about systems of nonlinear differential equations like (1) which is helpful to keep in mind. It was first proved by the French mathematician Emil Pickard around 1890. Provided the function f is well-behaved—in practice the functions that we normally encounter in modelling biological systems are "well-behaved"—then a solution to (1) always exists and is unique. More precisely, if $a = (a_1, \cdots, a_n)$ is any n-dimensional vector that describes an initial condition at some time $t = t_0$, then there exists an n-dimensional function $\phi(t)$ of one (time) variable, which is differentiable as a function of t, which satisfies the initial condition at $t = t_0$ and satisfies the differential equation in some interval around $t = t_0$:

$$\begin{aligned} \phi(t_0) &= a \\ d\phi(t)/dt &= f(\phi(t)) \qquad \text{for } t_0 - h < t < t_0 + h \,. \end{aligned} \tag{3}$$

Just because a solution exists does not mean that we can always write it down in terms of more familiar functions, like exp, sin or cosh. Another way to think about Pickard's existence theorem is that it provides a way of constructing new functions as solutions to an appropriate differential equation. We will see this at work in the next section.

Not only do solutions of (1) exist, they are also unique. Suppose that $\theta(t)$ is another n-dimensional function of one (time) variable, which also satisfies (3). Then, $\phi(t) = \theta(t)$ in some interval around $t = 0$. This will come in useful below.

2 Solving a linear differential equation in 1 dimension

In Part I we exploited the idea that matrices were like numbers and that some things that are easy to do with numbers, like solving the equation $ax = b$ to get $x = a^{-1}b$, can also be done with matrices. We are going to use the same idea here to solve a linear system of differential equations. First, we have to remember how to do this for numbers.

You should know that the 1-dimensional linear differential equation, $dx/dt = ax$, is solved by the exponential function $x(t) = \exp(at)v$, where v is the initial condition, or value of x at time 0. But what is the definition of the exponential function? It is instructive to try and work out what $\exp(at)$ must be in order for it to satisfy the differential equation $dx/dt = ax$.

We need to find a function $x(t)$ with the property that when it is differentiated it gives a times itself. We could try to build this up using simple functions that are easy to differentiate, like t^n, whose derivative is nt^{n-1}. We might start with something like $x_1(t) = at$. If we differentiate this we get a, which is not ax_1 but is $a.1$. So perhaps we should start instead with $x_0(t) = 1$ and $x_1(t) = x_0(t)+at$. If we do that then $dx_1/dt = ax_0$, which is not quite what we want because it has different x's on either side of the equation. However, it is a start and it seems to be going in the right direction, so perhaps it may lead us somewhere. Can we find $x_2(t)$, such that $dx_2/dt = ax_1 = a + a^2t$? We might try $x_2(t) = 1 + at + a^2t^2$ but this does not quite work because $d(a^2t^2)/dt = 2a^2t$ and what we really want is a^2t. So instead, let us try

$$x_2(t) = 1 + at + \frac{1}{2}(at)^2$$

from which it is easy to see that, indeed, $dx_2/dt = ax_1$. We can now try and keep going and find $x_3(t)$ such that

$$dx_3/dt = ax_2 = a + a^2t + \frac{1}{2}a^3t^2 \,.$$

Let us do this by adding to $x_2(t)$ the term $u(t) = c(at)^3$, where we choose the constant c so that

$$\frac{du}{dt} = 3ca(at)^2 = \frac{1}{2}a^3t^2 \,.$$

We see that $c = 1/2.3$. It is not difficult to extrapolate from this. If we define $x_k(t)$ so that

$$x_k(t) = 1 + at + \frac{(at)^2}{2} + \frac{(at)^3}{2.3} + \frac{(at)^4}{2.3.4} + \cdots + \frac{(at)^k}{2.3.\cdots.(k-1).k} \qquad (4)$$

then it is not hard to check that $dx_k/dt = ax_{k-1}$.

The appearance of the product of successive integers in this formula leads to the definition of k *factorial*, denoted $k!$, which is the product of the first k integers, $k! = 2.3.\cdots.(k-1).k$. We can rewrite (4) more succinctly as

$$x_k(t) = \sum_{i=0}^{k} \frac{(at)^i}{i!} \,.$$

Does this get any closer to a solution of $dx/dt = ax$? We seem to have merely generated an infinite sequence of solutions, none of which are quite right. However, suppose we could *"let k go to infinity"* and use an infinite number of terms in the sum, so that we could write

$$x(t) = 1 + at + \frac{(at)^2}{2} + \frac{(at)^3}{2.3} + \cdots + \frac{(at)^k}{2.3.\cdots.(k-1).k} + ... \qquad (5)$$

and suppose that we can differentiate this just as we did before, on a term by term basis. It is then easy to see by differentiating each term in (5) and adding them up that

$$\frac{dx(t)}{dt} = 0 + a + a^2t + \frac{a^3t^2}{2} + \cdots + \frac{a^kt^{k-1}}{2.3.\cdots.(k-1)} + \cdots = ax(t) \,.$$

So we have found a solution to $dx/dt = ax$ provided that we can add up an infinite number of terms. In this case we can but we cannot always do it. We will discuss when you can and cannot in the next section. Although the Pickard existence theorem only guarantees the existence of a solution locally (meaning in some neighbourhood of the initial condition), (5) works globally for all values of t.

We can summarise what we have worked out in the following definition of the exponential function (taking $t = 1$ in (5)):

$$\exp(a) = \sum_{i=0}^{\infty} \frac{a^i}{i!}\,. \tag{6}$$

We should now be able to prove some of the properties of the exponential function. One of the simplest is that $\exp(0) = 1$. This follows immediately from (6). A more interesting property is the addition formula for exponentials,

$$\exp(a + b) = \exp(a)\exp(b)\,, \tag{7}$$

which really characterises what it means to be "exponential". To prove this, let us consider $\exp(at)\exp(bt)$ as a function of t. What differential equation does this function satisfy? If we use the rule for differentiating a product of two functions, we find that

$$\frac{d}{dt}(\exp(at)\exp(bt)) = \left(\frac{d}{dt}\exp(at)\right)\exp(bt) + \exp(at)\left(\frac{d}{dt}\exp(bt)\right)\,.$$

Since, $\exp(at)$ satisfies $dx/dt = ax$, we can simplify this to

$$a\exp(at)\exp(bt) + \exp(at)b\exp(bt) = (a+b)\exp(at)\exp(bt)\,.$$

In other words, $\exp(at)\exp(bt)$ satisfies the differential equation $dx/dt = (a + b)x$ and, furthermore, $\exp(a0)\exp(b0) = 1$. Since we already know that $\exp((a + b)t)$ is a solution of this differential equation which satisfies the same initial condition, we can conclude from the uniqueness of solutions to differential equations that

$$\exp((a+b)t) = \exp(at)\exp(bt)\,.$$

Taking $t = 1$, we recover (7). We really have defined the exponential function!

3 Convergence and divergence

Adding up infinitely many terms is a very useful technique in mathematics and lets you construct things that would be hard to do otherwise. However, you have to be very careful about it. You already have some familiarity with infinite sums from ordinary arithmetic. The decimal expansion of $1/3$ is the "recurring" decimal $0.3333\cdots$, which actually means

$$\frac{3}{10} + \frac{3}{10^2} + \cdots + \frac{3}{10^k} + \cdots\,. \tag{8}$$

How do we add this up? We can easily rewrite it as

$$\frac{3}{10}\left(1 + \frac{1}{10} + \frac{1}{10^2} + \cdots + \frac{1}{10^k} + \cdots\right)\,,$$

in which the infinite sum within the brackets is a *geometric progression*. You should remember how to add this up but let us recall it anyway. For any a,

$$1 + a + a^2 + \cdots + a^k = \frac{1 - a^{k+1}}{1 - a}\,.$$

(This is easy to confirm by multiplying across by $(1-a)$.) If $0 < a < 1$ (for instance, $a = 3/10$) then powers of a like a^{k+1} become smaller and smaller as k gets larger. Hence, for very large values of k, we can ignore the contribution from a^{k+1}. In the *limit* as k gets very large, as we add up an infinite number of terms, we should find that

$$1 + a + a^2 + \cdots + a^k + \cdots = \frac{1}{1-a}\,. \tag{9}$$

When $0 < a < 1$, we say that the sum *converges*. What happens if $1 < a$? In this case, the powers of a like a^{k+1} get larger and larger as k increases (take a = 2, for instance) and it clear that the sum will eventually become larger than any number we can think of. In this case, the sum *diverges*. We cannot assign any consistent mathematical meaning to the sum of infinitely many terms when the sum diverges.

For the recurring decimal $a = 3/10 < 1$, so we can use (9) to see that (8) is

$$\frac{3}{10}\left(\frac{1}{1-1/10}\right) = \frac{1}{3}$$

as expected.

You might think from this example that convergence and divergence are easy to tell apart. For the geometric progression, each term in the sum either gets smaller and smaller and the sum converges, or gets larger and larger and the sum diverges. However, it is more subtle than that. Consider the following infinite sum

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{k} + \cdots\,. \tag{10}$$

Each successive term is smaller than the preceding. However, let us group the sum as follows:

$$1 + \underbrace{\frac{1}{2}}_{\geq 1/2} + \underbrace{\frac{1}{3} + \frac{1}{4}}_{\geq 2(1/4)=1/2} + \underbrace{\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}}_{\geq 4(1/8)=1/2} + \underbrace{\frac{1}{9} + \frac{1}{10} + \frac{1}{11} + \frac{1}{12} + \frac{1}{13} + \frac{1}{14} + \frac{1}{15} + \frac{1}{16}}_{\geq 8(1/16)=1/2} + \cdots\,.$$

Each term in a group is larger than the last term, which is $1/2^k$ for some k. There are 2^{k-1} terms in that group. Hence, the sum of the terms in each group is always at least 1/2. By taking sufficiently many groups, the sum can be made as large as we please. We see that the series (10) does not converge. In general, successive terms of a series have to decrease sufficiently fast for the series to converge.

Back in the seventeenth and eigteenth centuries, as the calculus was being developed, mathematicians got themselves very confused about when infinite sums converged and when they did not. When some of d'Alembert's students complained to him that they were going round in circles trying to understand this, he is reported to have said, *"Go on, my sons, and faith will come to you!"*. (Daughters were otherwise occupied in those days.) A little later in the early nineteenth century Cauchy, Weierstrass and others found out how to do it without requiring religious guidance.

We will avoid these issues here. For the exponential function, the numerators in each term, $(at)^k$, can increase very quickly with k but the denominators, $k!$, increase even faster, as you will see if you calculate a few of them. The ratio, $(at)^k/k!$, decreases fast enough that (5) always converges. From now on, we will assume that we can operate on the infinite series (5) as if it were a finite sum. That works for (5) but it is important to keep in mind that it does not always work.

4 Defining the matrix exponential

Now consider the linear system of differential equations

$$\frac{dx}{dt} = Ax \tag{11}$$

where x is an n-dimensional vector and A is an $n \times n$ matrix. We are going to follow the analogy between matrices and numbers and try the same formula as in (6) to define the matrix exponential, $\exp(A)$:

$$\exp(A) = I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots = \sum_{i=0}^{\infty} \frac{A^i}{i!}\,. \tag{12}$$

This infinite sum turns out to converge for much the same reasons that (5) converges. We are going to assume that fact here and take it for granted that we can operate with (12) as we would with a finite sum.

With this definition, I claim that it is easy to check that $x(t) = \exp(At)$ satisfies (11). Note first that At (or tA) is a shorthand for the matrix in which each entry in A is multiplied by t. Another way of writing this is $A.(tI)$ (or $(tI).A$), where tI is the identify matrix multiplied by t or, equivalently, the matrix with entry t on the main diagonal and 0 elsewhere. Since the identity matrix commutes with every matrix, it is easy to see that

$$(At)^k = A(tI)\underbrace{\cdots}_{k}A(tI) = A\underbrace{\cdots}_{k}A(tI)\underbrace{\cdots}_{k}(tI) = A^k t^k\,.$$

Hence,

$$\exp(At) = \sum_{i=0}^{\infty} A^i \frac{t^i}{i!}\,. \tag{13}$$

Furthermore, differentiating a matrix-valued function, just like a vector-valued function, just means differentiating each entry of the matrix individually. It should now be easy to see that if we differentiate the matrix function $\exp(At)$ with respect to t we get, for exactly the same reasons as before, $A.\exp(At)$. The matrix terms A^i in (13) are just constants, so the calculation of derivatives works exactly as it did previously for numbers. To relate these $n \times n$ matrices back to the linear system (11) we need to take into account the initial condition at time $t = 0$, which is given by some n-dimensional vector v. Let $x(t) = \exp(At).v$. This is now an n-dimensional vector. It is easy to see that $x(t)$ is a solution of $dx/dt = Ax$ and that, because $\exp(0) = I$, $x(0) = v$. Hence, we have found the (unique) solution to the system of linear differential equations (11).

Note that when $n = 1$, the definition of the matrix exponential reverts to the usual definition of the exponential for numbers. Does the matrix exponential satisfy the same properties as the number exponential? We know that $\exp(0) = I$. Let us try to prove the addition formula as we did before. Consider the matrix function $\exp(At)\exp(Bt)$. What differential equation does this satisfy? It is not immediately obvious that we can use the usual formula for differentiating a product of two functions, when the functions are matrix valued. However, we can and it is a nice exercise to convince yourself that this is true. Accordingly, we find that,

$$\frac{d}{dt}(\exp(At)\exp(Bt)) = \left(\frac{d}{dt}\exp(At)\right)\exp(Bt) + \exp(At)\left(\frac{d}{dt}\exp(Bt)\right)\,.$$

Since $d(\exp(At)/dt = A\exp(At)$, we can simplify this to

$$A\exp(At)\exp(Bt) + \exp(At)B\exp(Bt)\,. \tag{14}$$

So far, the argument is identical to the one given above but at this point we have to use the commutativity of multiplication to rewrite the second term. We cannot do this in general for matrices. However, suppose that A commutes with B, so that $AB = BA$. In that case it is easy to check that B commutes with all powers of At, so that $(At)^k B = B(At)^k$. It follows from (12) that B must then commute with $\exp(A)$. Hence, when A and B commute, we can rewrite the second term in (14) to get

$$A\exp(At)\exp(Bt) + B\exp(At)\exp(Bt) = (A + B)\exp(At)\exp(Bt)\,.$$

It follows, as before, from the uniqueness of solutions to differential equations, that

$$\exp(A+B) = \exp(A)\exp(B) \quad \text{if } AB = BA\,. \tag{15}$$

The analogy between numbers and matrices only works up to a point, as we have already seen a few times. You will get some very strange results if you forget that A and B need to commute before you can use the addition formula!

5 Calculating the matrix exponential

If you take a general 2×2 matrix A and try working out its matrix exponential using (12), you will quickly come to the conclusion that you do not want to do it that way! The matrix powers A^i that enter into (12) become complicated functions of the entries of A and it is not easy to see how to relate them to anything more familiar. Except in special cases, like when A is a diagonal or triangular matrix—see below, (12) is not a good way to calculate $\exp(A)$.

We learned in Part II that we can use a change of basis to transform A into a simpler form. If T is the change of basis matrix, then A is transformed into TAT^{-1}. How does this affect $\exp(A)$? It is easy to see that, because $T^{-1}T = I$,

$$TA^2T^{-1} = TAAT^{-1} = TA(T^{-1}T)AT^{-1} = (TAT^{-1})(TAT^{-1}) = (TAT^{-1})^2\,,$$

and using the same reasoning we can conclude that

$$TA^iT^{-1} = (TAT^{-1})^i\,,$$

for any integer $i > 0$. It is now easy to see that

$$T\exp(A)T^{-1} = T\left(\sum_{i=0}^{\infty}\frac{A^i}{i!}\right)T^{-1} = \sum_{i=0}^{\infty}\frac{TA^iT^{-1}}{i!} = \sum_{i=0}^{\infty}\frac{(TAT^{-1})^i}{i!} = \exp(TAT^{-1})\,,$$

Transforming A by a change of basis results in precisely the same transformation being applied to $\exp(A)$. This provides us with a systematic method for calculating $\exp(A)$. We use a change of basis to transform A into a simpler form, calculate the exponential of the simpler form and then use the reverse transformation to return to the original basis:

$$\exp(A) = T^{-1}\exp(TAT^{-1})T\,. \tag{16}$$

In Part II we learned that any 2×2 matrix can be transformed by a change of basis into one of three classes, depending on whether the characteristic equation has two complex conjugate roots (class I), two real roots with two distinct eigenvectors (class II) or two equal real roots with only one eigenvector (class III). Let us look at each of these in turn in increasing order of difficulty.

Class II A 2×2 matrix with 2 distinct eigenvectors can always be diagonalised (see Part II). The powers of a diagonal matrix are easy to work out. The entries on the diagonal are simply raised to the same power:

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}^k = \begin{pmatrix} a^k & 0 \\ 0 & b^k \end{pmatrix}\,.$$

It should then be easy to see from (12) that

$$\exp\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} = \begin{pmatrix} \exp(a) & 0 \\ 0 & \exp(b) \end{pmatrix} \tag{17}$$

In fact, this works for any n. Diagonal matrices are easy and, consequently, it is easy to compute $\exp(A)$ for any matrix A which can be diagonalised.

Class III A 2×2 matrix with two equal real roots but only one eigenvector cannot be diagonalised but it can be transformed into an upper triangular matrix (see Part II). We can always write an upper triangular matrix as the sum of a diagonal matrix and whatever is left over

$$\begin{pmatrix} a & b \\ 0 & a \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} + \begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix}.$$

Because the eigenvalues of the matrix are assumed to be equal, the diagonal matrix is just a scalar times the identity matrix. That is very convenient because the identity matrix commutes with any matrix. (It is not true that an arbitrary diagonal matrix commutes with any matrix. The entries on the diagonal need to be the same.) In this case, we can legitimately use (15) to get

$$\exp\begin{pmatrix} a & b \\ 0 & a \end{pmatrix} = \exp\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} \exp\begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix}. \tag{18}$$

We already know how to work out the first factor, as it is the exponential of a diagonal matrix. As for the second, it is easy to check that

$$\begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix}^2 = 0\,.$$

Matrices for which some power is 0 are called *nilpotent*. If you take away the diagonal terms from any triangular matrix (upper or lower) then what is left is always nilpotent (why?). The exponential of a nilpotent matrix can often be calculated directly from the definition (12) because the infinite series gets truncated to only a finite number of terms. In this case it is particularly easy:

$$\exp\begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix} = I + \begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}.$$

Substituting into (18), we find that

$$\exp\begin{pmatrix} a & b \\ 0 & a \end{pmatrix} = \begin{pmatrix} \exp(a) & 0 \\ 0 & \exp(a) \end{pmatrix}\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \exp(a) & b\exp(a) \\ 0 & \exp(a) \end{pmatrix}. \tag{19}$$

Notice the appearance of an algebraic, rather than an exponential, function of b in one of the terms. This is a small but important difference between the matrix exponential of Class II and Class III matrices.

Class I A matrix in Class I has two complex conjugate eigenvalues and can always be transformed by a change of basis to

$$A = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

If you do not already know it, you should quickly check that the eigenvalues of this matrix are just $a \pm ib$. Matrices of this form are in 1-to-1 correspondence with complex numbers of the form $z = a + ib$ and we showed in Part I that this correspondence is an *isomorphism*: it preserves the algebraic operations of addition and multiplication. Now just imagine that we take the definition of the matrix exponential (12) and apply the one-to-one correspondence to each term, using the fact that the correspondence preserves the algebraic operations. It is easy to see that we would get

$$1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots = \sum_{i=0}^{\infty} \frac{z^i}{i!}\,.$$

This is precisely the definition of the exponential for complex numbers, $\exp(z)$. By now you should be accustomed to the idea that we can define an exponential in this way and that the function $\exp(zt)$ will be the solution to the complex differential equation $dx/dt = zx$, where $x(t)$ is now a complex-valued function of time. Moreover, just as we did above, we can show that the complex exponential satisfies the usual addition formula. Because multiplication of complex numbers, unlike matrices, is always commutative, $\exp(w+z) = \exp(w)\exp(z)$, where w and z are arbitray complex numbers.

We see from this that $\exp(A)$ is the matrix corresponding to the complex number $\exp(z)$. What is that complex number? We need to rewrite it in terms of a and b. Applying the addition formula we see that

$$\exp(z) = \exp(a+ib) = \exp(a)\exp(ib)\,. \tag{20}$$

We can now use one of the most famous formulas in mathematics, which relates i, e and π:

$$\exp(ib) = \cos(b) + i\sin(b)\,.$$

You should have seen this before. If you have not then you need to brush up your trigonometry. We are going to assume it here. Substituting back in (20) and converting back into a matrix, we find that

$$\exp(A) = \exp(a)\begin{pmatrix} \cos(b) & -\sin(b) \\ \sin(b) & \cos(b) \end{pmatrix}\,. \tag{21}$$

We have now worked out how to calculate the exponential for any 2×2 matrix (up to a transformation).