Hand Computation of Matrix Exponentials

Richard T. Bumby
Rutgers University

Introduction. It is now common for textbooks for a first course in Differential Equations to describe the solution of the first-order, linear, constant-coefficient system

$$\frac{dY}{dt} = MY, \tag{1}$$

where M is an n by n matrix, in terms of a matrix exponential e^{Mt}. However, there is only a limited discussion of techniques for finding the matrix exponential. The student is left with the impression that it is necessary to find eigenvalues and eigenvectors as the first step in employing this technique. This impression is reinforced by giving examples that are limited to $n = 2$ and a matrix M with integer entries whose characteristic polynomial factors to reveal distinct integer eigenvalues, for which the determination of eigenvalues is easy.

Examples may also given in which the eigenvalues are complex (usually Gaussian integers), and the solution is interpreted in this case. However, converting the solution to a useful form is tedious and rarely performed accurately. Thus, no interesting example can be solved quickly enough to be useful for lectures or examinations.

This appears to be an oversight, or perhaps *collective amnesia*, since there is *an easily remembered formula* (a phrase that was used as a title when part of this article was used as course notes) for the solution. The uniqueness theorem shows that once any proposed answer is verified, it is the only correct answer. Thus, it is not necessary to perfect the underlying theory before proposing a method of solution. Any expression that is easily checked may be seen to solve the equation — however it was discovered. There is an advantage to having a obtaining solutions indirectly since it shifts the emphasis to verifying that the answer is correct rather than merely expecting the student to echo (some of) the steps in one method of computing the answer. In the same way that methods used in computer solutions are not just translations of a method used in proof, but are optimized for that environment, methods for hand computation should show human solvers the same respect.

Instead of trying to mechanize the process of solving differential equations, or the related process of indefinite integration, more emphasis should be given to a “guess and check” approach (which is usually called by the misleading name of “trial and error”). Although this makes the subject seem more of an art than a science, the results will sometimes allow new principles for guessing answers to be formulated. This approach will be used for $n = 2$ with complex eigenvalues. After finding the solution, a direct way of obtaining it will be explored. This approach will then be generalized to apply to examples with $n = 3$ and $n = 4$.

The matrix exponential. To begin, let us first ask: what is Y? The usual answer is, “a vector of n real (or complex) functions of t”. An *initial condition* for (1) is a vector of n numbers which give Y when $t = 0$. It is convenient to consider Y as a vector function (rather than a vector *of* functions) and denote this initial value by $Y(0)$. The columns of e^{Mt} are then described as the solutions whose initial conditions are 1 in one coordinate and 0 in all others. Routine linear algebra says that once e^{Mt} is known, the solution of an initial value problem for the equation (1) is

$$Y(t) = e^{Mt}Y(0). \tag{2}$$

However, both (1) and (2) may be interpreted, and are correct, if Y is any matrix with n rows. Taking $Y(0)$ to be an identity matrix in (2) gives us the

Working Definition. *The matrix exponential e^{Mt} is the unique solution $Y(t)$ to (1) with $Y(0) = I_n$, the n by n identity matrix.*

The characteristic polynomial. Even the usual solution of (1) starts with a guess of the solution containing some parameters and then finds values of these parameters giving a solution. That guess is $Y = ve^{\lambda t}$, where v is a constant vector and λ is a number. Substituting into (1) shows that we have a solution if $Mv = \lambda v$. Since it is now common for students to meet linear algebra early in their studies, we may freely use the following

Standard Terminology. *The polynomial $\det(M - \lambda I)$ is called the characteristic polynomial of M; a zero of the characteristic polynomial is called an eigenvalue; the nonzero v with $Mv = \lambda v$ are called eigenvectors.*

When n distinct eigenvalues can be found, it is not difficult to find an eigenvector for each eigenvalue, and the n by n matrix S whose columns are the eigenvectors is an invertible matrix that satisfies $MS = S\Lambda$, where Λ is a diagonal matrix whose entries are the eigenvalues. This leads to

$$e^{Mt} = Se^{\Lambda t}S^{-1}, \tag{3}$$

which looks like a formula for the solution. However, implementing this formula requires that S and S^{-1} be computed, and this is not always easy (and is only possible if there is a basis of eigenvectors). Complex eigenvalues are possible and the algebra needed to compute e^{Mt} is not familiar enough to have a high likelihood of leading to correct answers. Unless enough time is spent to gain fluency with linear algebra over $\mathbb{C}$, this formula is likely to be painful to use. It would be better to refine our method of guessing. Furthermore, an initial value problem formulated over the real numbers will have a solution that is a real function. It would be desirable to get such solutions directly.

Complex exponentials. To interpret complex exponentials, the formula $e^{it} = \cos t + i \sin t$ is used.

Suppose now that $n = 2$ and the eigenvalues of M are $r \pm si$ with $s \neq 0$. Then equation (3) applies. If the resulting expression is expanded and the complex exponentials $e^{(r \pm si)t}$ converted to real exponentials and trigonometric functions, then

$$e^{Mt} = Pe^{rt} \cos st + Qe^{rt} \sin st,$$

where P and Q are 2 by 2 matrices of real numbers. It turns out to be much easier to identify P and Q from this equation than to calculate them from (3). In particular, putting $t = 0$ in this expression leads (immediately!) to $P = I$.

To find Q, we can differentiate e^{Mt}. For the discovery of the solution, it suffices to consider only the value at $t = 0$. This should be M, and direct calculation shows it to be $rP + sQ$. Since we have $P = I$, knowing r and s allows us to obtain Q as $(1/s)(M - rI)$. The Cayley-Hamilton Theorem, which can be verified by direct computation for 2 by 2 matrices, shows that the square of Q is $-I$. A change to more suggestive notation gives

Theorem 1. *If $n = 2$ and M has eigenvalues $r \pm si$, then*

$$e^{Mt} = Ie^{rt} \cos st + Je^{rt} \sin st. \tag{E}$$

where I is the identity and J is characterized by $M = rI + sJ$. Furthermore, $J^2 = -I$ so that it plays the role of the number i in the algebra $\mathbb{R}[M]$ generated by M over $\mathbb{R}$.

Example 1. Let

$$M = \begin{pmatrix} 7 & -13 \\ 2 & -3 \end{pmatrix}.$$

One eigenvalues is $2 + i$, so

$$J = M - 2I = \begin{pmatrix} 5 & -13 \\ 2 & -5 \end{pmatrix};$$

$$e^{Mt} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} e^{2t} \cos t + \begin{pmatrix} 5 & -13 \\ 2 & -5 \end{pmatrix} e^{2t} \sin t.$$

A straightforward check shows that this is correct.

General properties of the matrix exponential. The key to finding e^{Mt} in Example 1 was to consider the exponentials of all matrices in $\mathbb{R}[M]$ and to express the desired exponential in terms of others that are easier to compute. The validity of this approach is based on the following

Lemma. *If $NM = MN$, then $Ne^{Mt} = e^{Mt}N$. Also, $Ne^{Mt} = e^{Mt}N$ implies that $NM = MN$.*

Proof. The first part could be proved using the power series representation of e^{Mt}, but it is more in keeping with the present approach to note that $F(t) = e^{Mt}N$ satisfies (1) and $F(0) = N$. Now, let $G(t) = Ne^{Mt}$. Clearly, $G(0) = N$ and $G'(t) = NMe^{Mt}$. Thus $G(t) = F(t)$ if and only if $G(t)$ satisfies (1), which holds if and only if $NM = MN$.

Corollary. *If $AB = BA$, then $e^{(A+B)t} = e^{At}e^{Bt}$*

Proof. It suffices to show that $H(t) = e^{At}e^{Bt}$ satisfies (1) with $M = A + B$. We have $H'(t) = Ae^{At}e^{Bt} + e^{At}Be^{Bt}$, and the lemma shows that the second term is $Be^{At}e^{Bt}$, allowing the distributive law to apply to give (1).

Taking $A = rI$ and $B = sJ$ allows (E) to be recovered from special cases.

More two by two matrices. For any 2 by 2 matrix M, if $M = rI + Q$, then $e^{Mt} = e^{rt}e^{Qt}$. We expect that matrices Q of trace zero will play a special role, and r can be chosen to reduce to this special case. Since $e^{Bt} = I \cos t + B \sin t$ when $\det B = 1$, we can guess that $e^{Bt} = I \cosh t + B \sinh t$ when $\det B = -1$, and this is easily verified. As before, other negative determinants are covered by taking a suitable constant multiple of t in this expression. The definition of e^{Bt} as a series also shows that $e^{Bt} = I + Bt$ when $\det B = 0$. Again, however one guesses this solution, a proof consists of showing that it satisfies the (1) and reduces to I when $t = 0$. This use of *leading special cases* seems much more robust than the traditional solution.

Laplace transforms. If a course includes Laplace transforms, they may be used to solve problems with given initial conditions. It is worth noting that this method applies to matrix solutions as well as the customary vector solutions. Thus $Y(s) = \mathcal{L}(e^{Mt})$ can be found directly from

$$sY(s) - I = MY(s)$$

so that

$$Y(s) = (sI - M)^{-1}.$$

For 2 by 2 matrices, the ability to express the entries of the inverse directly gives another approach to Theorem 1. Moreover, the reduction to the case of matrices of trace zero and the use of hyperbolic functions reflect familiar methods for efficiently recognizing inverse Laplace transforms.

A triumph of abstraction. By expression the solution in terms of a matrix J found in $\mathbb{R}[M]$, we have shifted emphasis from the matrix M to its corresponding linear transformation. That is, we are looking at the way that M acts on all vectors in $\mathbb{R}^n$ instead of emphasizing its action on one particular basis. This approach is also present in the use of eigenvectors, but the goal there only seems to find a better basis. The selection of the matrix J was based instead on abstract considerations — it represented the number i in $\mathbb{R}[M]$.

In fact, much more was shown. Only the fact that the minimal polynomial of M was of degree 2 was needed to obtain the expression for e^{Mt} and to verify that it was correct. This means that it is degree of the minimal polynomial rather than the size of the matrix that determines the structure of exponential. A description of $\mathbb{R}[M]$ using matrices that satisfy equations of low degree will lead to a simple computation of e^{Mt}.

An example with repeated complex roots.. Let

$$M = \begin{pmatrix} 2 & -2 & 1 & 0 \\ 2 & 2 & 0 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & -5 & 1 \end{pmatrix}.$$

Since this is block-triangular, its characteristic polynomial is easily recognized to be $(\lambda^2 - 4\lambda + 8)^2$. The eigenvalues are thus $2 \pm 2i$. Since

$$M^2 - 4M + 8I = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

the characteristic polynomial is also the minimal polynomial. While this can be factored into relatively prime polynomials over $\mathbb{C}$, there is no factorization into relatively prime polynomials with real coefficients. Since we are aiming to avoid algebra over the complex numbers, we seek a different approach.

The ring generated by M over $\mathbb{R}$ may be identified with the ring of polynomials in an indeterminate x modulo the primary ideal generated by $(x^2 - 4x + 8)^2$. Call this ring S. In S, the ideal I generated by $x^2 - 4x + 8$ is nilpotent, and S/I is isomorphic to $\mathbb{C}$. A key result for computation is that the ring homomorphism $S \to S/I$ has a left inverse. In particular, S has a subring isomorphic to $\mathbb{C}$, and elements of S can be written as a sum of a nilpotent element and an element of this subring. If M is written as a sum of an element of this subring and an element of I, e^{Mt} will be the product of the exponentials of two elements of S that satisfy equations of degree 2.

We will identify the subring isomorphic to $\mathbb{C}$ by producing an element of the form $j = (x-2)/2 + (x^2+4x+8)y$, with $y \in S$ that plays the role of i in the sense that $j^2 = -1$ in S. Direct computation gives

$$j^2 + 1 \equiv \frac{1}{4}(x^2 + 4x + 8)(4xy - 8y + 1) \text{ mod } (x^2 + 4x + 8)^2.$$

We get the value of j that we seek if $(x-2)y \equiv -1/4 \pmod{x^2+4x+8}$. Since $(x-2)^2 \equiv -4 \pmod{x^2+4x+8}$, the unique solution modulo x^2+4x+8 is $y=(x-2)/16$. Computing the matrix corresponding to j gives

$$J = \frac{1}{8}\begin{pmatrix} 0 & -8 & 2 & 0 \\ 8 & 0 & 1 & 1 \\ 0 & 0 & 4 & 4 \\ 0 & 0 & -20 & -4 \end{pmatrix}$$

Inverting the definition of J gives

$$M = 2 + 2J + N$$

with

$$N = -\frac{1}{8}(M-2)(M^2-4M+8I) = \frac{1}{4}\begin{pmatrix} 0 & 0 & 2 & 0 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Since J and N belong to $\mathbb{R}(M)$, $LN = NJ$ and we get

$$\begin{aligned} e^{Mt} &= e^{2t}e^{2Jt}e^{Nt} \\ &= e^{2t}(I\cos 2t + J\sin 2t)(I+Nt) \\ &= Ie^{2t}\cos 2t + Je^{2t}\sin 2t + Nt\cos 2t + JNt\sin 2t \end{aligned}$$

The matrix coefficients have all been shown except for

$$JN = \frac{1}{4}\begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \frac{M^2-4M+8I}{4}.$$

The process of solving for y is exactly Newton's Method in S (sometimes called Hensel's Lemma in this context). We are seeking a root of the separable polynomial $p(x) = x^2+1$, and if we have already found a root of $p(x)$ modulo I^k, then $p(x+y) \equiv p(x) + p'(x)y \pmod{I^{2k}}$. Since $p(x) \in I$ and $p'(x)$ is relatively prime to $p(x)$, $p'(x)$ is invertible modulo I. This allows us to find y such the $p(x+y) \in I^{2k}$. Since we are working in a ring for which some power of I is zero, iterating this leads to an exact solution.

Splitting by projections. We have described exponentials of matrices rather than of linear transformations, but the exponentials that we have found involved products of scalar functions of t with matrices in $\mathbb{R}[M]$. Such results could be expressed in a coordinate-free manner. However, matrices will be used in both proofs and examples, although the proofs will contain some matrices that need never be found in practice. This is because some constructions require transformations acting on a subspace of dimension m with $m < n$. To find an m by m matrix representing this action, one chooses a basis for the subspace. This basis is useful in the proof, but there will never be any need to compute it. The n by n matrices that appear in an expression for e^{Mt} will all be found directly, and not in terms of any factorization that may be used in theoretical discussions.

The traditional solution when $n=2$ and M has distinct real eigenvalues may be written in the form $e^{Mt} = E_1e^{\lambda_1 t} + E_2e^{\lambda_2 t}$ for some matrices E_1 and E_2. The *Spectral Decomposition* of M identifies E_i as the projection onto the λ_i-eigenspace of M whose kernel is the other eigenspace.

We have already seen a multiplicative factorization of the matrix exponential, but this is an *additive* splitting. With suitable modification, such a splitting can be found for any idempotent in $\mathbb{R}[M]$. Since basic expressions have been found that use spaces other than the one-dimensional space spanned by an eigenvector, it is more useful to have a means of using splittings inductively than to aim for a universal formula for the exponential.

To study this splitting, fix a matrix E such that $E^2 = E$ and $EM = ME$. Let m be the rank of E. Choose a basis for the column space of E and let B be an n by m matrix whose columns are this basis. Since M takes this column space to itself, there is a matrix M' such that $MB = BM'$. This M' is unique since the columns of B are linearly independent. Furthermore, there is a unique matrix B' such that $E = BB'$. Since $B = EB = (BB')B = B(B'B)$, the independence of the columns of B shows that $B'B = I$. Thus, $M^n E = B(M')^n B'$ for all n. More generally, for any polynomial p, $p(M)E = Bp(M')B'$.

Proposition. *If* $EM = ME$, $E^2 = E$, *and* $MB = BM'$, *then*

$$e^{Mt}E = Be^{M't}B'. \tag{P}$$

Hence, if

$$e^{M't} = \sum_i f_i(t)p_i(M'),$$

then

$$e^{Mt}E = \sum_i f_i(t)p_i(M)E = \sum_i f_i(t)p_i(ME)E. \tag{S}$$

Proof. First, note that $e^{Mt}B$ and $Be^{M't}$ both satisfy (1) and evaluate to B when $t = 0$. The uniqueness theorem of differential equations then shows that they are equal. Now, multiply on the right by B' to obtain (P). The remaining statements follow from the discussion that preceded the statement of the Proposition.

If $E^2 = E$, $e^{Mt} = e^{Mt}E + e^{Mt}(I - E)$ and each of these terms is given by the action of M on the range of E or $I - E$ if $EM = ME$. The terms will be evaluated using one of the sums in (S) (or the corresponding statement for $I - E$). The matrices M' and B' and equation (P) are used in the proof, but do not need to be found.

A three dimensional example.. Markov matrices are a good source of examples allowing robust calculation, so let

$$M = \frac{1}{10}\begin{pmatrix} 3 & 1 & 5 \\ 3 & 3 & 1 \\ 4 & 6 & 4 \end{pmatrix}.$$

The column sums are all 1, so this has 1 as an eigenvalue, and its eigenvector is easily found to be (18, 11, 23). The projection on this subspace that commutes with M is a matrix with all columns equal to the multiple of this vector with sum of entries equal to 1. Thus,

$$E = \frac{1}{52}\begin{pmatrix} 18 & 18 & 18 \\ 11 & 11 & 11 \\ 23 & 23 & 23 \end{pmatrix} \qquad I - E = \frac{1}{52}\begin{pmatrix} 34 & -18 & -18 \\ -11 & 41 & -11 \\ -23 & -23 & 29 \end{pmatrix}.$$

Since E is constructed from eigenvectors of M, $ME = E$. To get the action on the two dimensional space of vectors with column sums zero, which is the range of the projection $(I - E)$, form

$$M(I - E) = \frac{1}{260}\begin{pmatrix} -12 & -64 & 40 \\ 23 & 23 & -29 \\ -11 & 41 & -11 \end{pmatrix}.$$

This matrix has trace zero and one of its eigenvalues is known to be zero, so the sum of the other two eigenvalues is zero. The square of this matrix is seen to be $-(I-E)/25$, so $(I-E)M$ acts like $i/5$ on the range of $(I-E)$. Thus, using the known exponentials of the action of M on the subspaces, we have

$$e^{Mt} = \frac{e^t}{52}\begin{pmatrix} 18 & 18 & 18 \\ 11 & 11 & 11 \\ 23 & 23 & 23 \end{pmatrix} + \frac{\cos(t/5)}{52}\begin{pmatrix} 34 & -18 & -18 \\ -11 & 41 & -11 \\ -23 & -23 & 29 \end{pmatrix} + \frac{\sin(t/5)}{52}\begin{pmatrix} -12 & -64 & 40 \\ 23 & 23 & -29 \\ -11 & 41 & -11 \end{pmatrix}.$$

The product of M with the matrices in this expression have already been determined, so the verification that this is e^{Mt} is easy.

Expressing the projections in terms of M. It was remarked in passing that the projection E commutes with M because it can be expressed as a polynomial in M. Although we found E in the Markov example by determining the eigenvector, an approach that applies to more general matrices would to compute the characteristic polynomial using the method of Leverrier (or Faddeev's modification, see [Faddeeva] for details). This method is very robust for hand computation, which led to its frequent rediscovery (see [Householder] for more information). If this characteristic polynomial can be factored into relatively prime factors, the Euclidean algorithm can be used to construct idempotents. In the Markov example, this gives $E = (25M^2 + I)/26$. This has been called an "application of the Chinese remainder theorem" (see [Oberst]). Any factorization of the minimal polynomial of M into relatively prime factors reduces the determination of e^{Mt} to finding the exponentials of matrices whose minimal polynomials are those factors.

References..

[Faddeeva] V. N. Faddeeva, *Computational Methods of Linear Algebra*, Dover, New York, 1959. (ISBN 0-486-60424-1)

[Householder] Alton S. Householder, *The Theory of Matrices in Numerical Analysis*, Dover, New York, 1975. (ISBN 0-486-61781-5)

[Oberst] Ulrich Oberst, "Anwendungen des chinesischen Restsatz", *Exposition. Math.* **3** (1985), 97–148, MR 87d:13002.