On the Solutions of Autonomous Systems of Linear ODEs

R. C. Daileda

An *autonomous system of n first order linear ordinary differential equations* has the form

$$\begin{aligned} x_1' &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n, \\ x_2' &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n, \\ &\vdots \\ x_n' &= a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n, \end{aligned} \tag{1}$$

where the x_i are all functions of a single common independent variable, say t, and the *coefficients* a_{ij} are all constants. If we let

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix},$$

and remember how matrices and vectors are multiplied, then our system takes on the more compact form

$$\mathbf{x}' = A\mathbf{x}.$$

The matrix A is called the *coefficient matrix* of the given system, and its true utility far exceeds that of mere notational simplification. In most courses on ODEs one learns how to use the eigenvalues and eigenvectors of A to determine the general solution to the system (1). But this is usually completely unmotivated. The alleged solution is given, and the instructor/textbook then simply checks that it works. We can actually gain more insight into the eigenvalue/eigenvector technique by proceeding more naïvely.

For example, suppose we are faced with the two-dimensional system

$$x' = 5x - 6y, \tag{2}$$

$$y' = 3x - 4y. \tag{3}$$

The equations in this system are *coupled*, in the sense that the ODE for x depends on y and vice versa. One way to solve this system is to *decouple* it by increasing the order of each equation. Indeed, if we differentiate both sides of (2) and then substitute in y' from (3) we obtain

$$x'' = 5x' - 6y' = 5x' - 6(3x - 4y) = 5x' - 18x + 24y. \tag{4}$$

This equation is still coupled to y, but if we multiply (2) by 4 we get

$$4x' = 20x - 24y \;\;\Rightarrow\;\; 24y = 20x - 4x'.$$

Using this in equation (4) we then have

$$x'' = 5x' - 18x + 24y = 5x' - 18x + 20x - 4x' = x' + 2x \quad \Rightarrow \quad x'' - x' - 2x = 0.$$

We now solve this second order equation using the characteristic equation

$$\lambda^2 - \lambda - 2 = 0 \quad \Leftrightarrow \quad (\lambda - 2)(\lambda + 1) = 0.$$

The factorization on the right tells us that $\lambda = -1, 2$, so that

$$x = c_1 e^{-t} + c_2 e^{2t}.$$

To find y we could rewrite the first equation in our system as $6y = 5x - x'$, substitute in our formula for x, and then divide by 6. We could also implement the decoupling procedure starting from (3) instead. This eventually tells us that

$$y'' - y' = 2y = 0.$$

Notice that this is *exactly the same* second order ODE satisfied by x! It turns out that this is no mere coincidence.

One way to see this is to recall that e^{-t} and e^{2t} are the so-called *fundamental solutions* of $x'' - x' - 2x = 0$. Since x is a linear combination of these two functions, so is x'. And since y is a linear combination of x and x', it follows that y is also a linear combination of e^{-t} and e^{2t}. It now follows from the principle of superposition that y must be a solution of $x'' - x' - 2x = 0$, too.

Although the technique we employed above ultimately produced the solutions of our system, it is tedious and somewhat ad hoc. Moreover, it's not entirely clear how it could be generalized to higher dimensional systems. And although the decoupled second order equation $x'' - x' - 2x = 0$ can be used to find both x and y, it appears to be completely mysterious. It turns out that decoupling the system is actually the *source* of all of this confusion, and we gain much more insight into our system by studying its matrix-vector form.

Let's return to the general system $\mathbf{x}' = A\mathbf{x}$. Because differentiation is a linear operation we have

$$\mathbf{x}'' = (\mathbf{x}')' = (A\mathbf{x})' = A\mathbf{x}' = A(A\mathbf{x}) = A^2\mathbf{x}.$$

Likewise, this implies

$$\mathbf{x}''' = (\mathbf{x}'')' = (A^2\mathbf{x})' = A^2\mathbf{x}' = A^2(A\mathbf{x}) = A^3\mathbf{x},$$

and in a similar way we obtain

$$\mathbf{x}^{(k)} = A^k\mathbf{x}, \tag{5}$$

where the superscript on the left represents k differentiations of the entires of $\mathbf{x}$. Now recall that the *characteristic polynomial* of A is

$$p_A(\lambda) = \det(\lambda I - A) = |\lambda I - A|.$$

The characteristic polynomial of A is important because its roots are precisely the eigenvalues of A. It also has another useful feature. An important result in advanced linear algebra known as the *Cayley-Hamilton theorem* asserts that every matrix is a root of its characteristic polynomial:

$$p_A(A) = 0,$$

where the zero on the right-hand side is the $n \times n$ zero matrix.

For example, if

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix},$$

then

$$p_A(\lambda) = \left| \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} - \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \right| = \begin{vmatrix} \lambda - 1 & -2 \\ -3 & \lambda - 4 \end{vmatrix} = (\lambda - 1)(\lambda - 4) - 6 = \lambda^2 - 5\lambda - 2.$$

We then have

$$p_A(A) = A^2 - 5A - 2I = \begin{pmatrix} 7 & 10 \\ 15 & 22 \end{pmatrix} - \begin{pmatrix} 5 & 10 \\ 15 & 20 \end{pmatrix} - \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

in agreement with Cayley-Hamilton.

Let's see what the Cayley-Hamilton theorem has to say about our linear system of ODEs. Write

$$p_A(\lambda) = \sum_{j=0}^{n} c_j \lambda^j.$$

Then Cayley-Hamilton and equation (5) yield

$$p_A(A) = \sum_{i=0}^{n} c_j A^j = 0 \quad \Rightarrow \quad \mathbf{0} = 0\mathbf{x} = \sum_{j=0}^{n} c_j A_j \mathbf{x} = \sum_{j=0}^{n} c_j \mathbf{x}^{(j)}.$$

Since the ith entry of the vector on the far right-hand side is $c_0 x_i + c_1 x_i' + c_2 x_i'' + \cdots c_n x_i^{(n)}$, we conclude that for any i we have

$$c_0 x_i + c_1 x_i' + c_2 x_i'' + \cdots c_n x_i^{(n)} = 0.$$

That is, the individual functions that solve the original system are all solutions of the *same* nth order linear ODE

$$c_0 x + c_1 x' + c_2 x'' + \cdots c_n x^{(n)} = 0, \tag{6}$$

whose characteristic equation is precisely $p_A(\lambda) = 0$. Another way to write equation (6) is

$$p_A\left(\frac{d}{dx}\right)(x) = 0,$$

where we interpret "powers" of $\frac{d}{dx}$ to be higher order derivatives:

$$\left(\frac{d}{dx}\right)^j = \frac{d^j}{dx^j}.$$

Let's go back to our earlier example. The coefficient matrix is

$$A = \begin{pmatrix} 5 & -6 \\ 3 & -4 \end{pmatrix}$$

with characteristic polynomial

$$p_A(\lambda) = |\lambda I - A| = \begin{vmatrix} \lambda - 5 & 6 \\ -3 & \lambda + 4 \end{vmatrix} = (\lambda - 5)(\lambda + 4) + 18 = \lambda^2 - \lambda - 2.$$

So the solution functions x and y must *both* solve

$$p_A\left(\frac{d}{dx}\right)(x) = \left(\frac{d^2}{dx^2} - \frac{d}{dx} - 2\right)(x) = x'' - x' - 2x = 0,$$

as we saw above.

The upshot is that if one seeks to solve the original system (1) using as little linear algebra as possible, but still somewhat efficiently and systematically, simply compute the characteristic polynomial p_A of the coefficient matrix A, and deal directly with the *single* nth order ODE $p_A(\frac{d}{dx})(x) = 0$ satisfied by *all* of the coordinate functions x_i.

That being said, let's think a bit more about the vector approach to solving the system (1). Let A be its coefficient matrix. A constant $\alpha \in \mathbb{R}$ is called an *eigenvalue* of A provided there is a vector $\mathbf{v} \neq \mathbf{0}$ so that $A\mathbf{v} = \alpha\mathbf{v}$. Any such (nonzero!) vector $\mathbf{v}$ is called an *eigenvector* of A associated to α. Notice that if $A\mathbf{v} = \alpha\mathbf{v}$, then

$$\mathbf{0} = \alpha\mathbf{v} - A\mathbf{v} = (\alpha I - A)\mathbf{v}.$$

Since $\mathbf{v} \neq \mathbf{0}$, this implies that $\det(\alpha I - A) = 0$. Conversely, if $\det(\alpha I - A) = 0$, then $\alpha I - A$ is not invertible, which is equivalent to saying that $(\alpha I - A)\mathbf{v} = \mathbf{0}$ (or $A\mathbf{v} = \alpha\mathbf{v}$) for some $\mathbf{v} \neq \mathbf{0}$. That is, the eigenvalues of A are precisely the roots of the characteristic polynomial $p_A(\lambda) = \det(\lambda I - A)$!

We have seen that the solutions x_i of the system (1) all solve the nth order ODE $p_A(\frac{d}{dx}) = 0$. The characteristic equation of this system is just $p_A(\lambda) = 0$, and its solutions are the eigenvalues of A. So if the eigenvalues of A are $\lambda_1, \lambda_2, \ldots, \lambda_n$ (which we assume to be real and distinct, for simplicity), then each x_i can be written

$$x_i(t) = \sum_{j=1}^{n} b_{ij} e^{\lambda_j t},$$

for some constants b_{ij}. It follows that

$$\mathbf{x}(t) = \sum_{j=1}^{n} e^{\lambda_j t} \begin{pmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{pmatrix} = \sum_{j=1}^{n} e^{\lambda_j t} \mathbf{b}_j,$$

where

$$\mathbf{b}_j = \begin{pmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{pmatrix}.$$

Now, because $\mathbf{x}$ solves $\mathbf{x}' = A\mathbf{x}$, we find that

$$\sum_{j=1}^{n} \lambda_j e^{\lambda_j t}\mathbf{b}_j = \mathbf{x}' = A\mathbf{x} = A\left(\sum_{j=1}^{n} e^{\lambda_j t}\mathbf{b}_j\right) = \sum_{j=1}^{n} e^{\lambda_j t} A\mathbf{b}_j.$$

Subtracting the final expression from the far left-hand side yields

$$\mathbf{0} = \sum_{j=1}^{n} \left(\lambda_j e^{\lambda_j t}\mathbf{b}_j - e^{\lambda_j t} A\mathbf{b}_j\right) = \sum_{j=1}^{n} e^{\lambda_j t}\left(\lambda_j \mathbf{b}_j - A\mathbf{b}_j\right). \tag{7}$$

This is an identity of vector *functions*, so it holds for all t. The ith entry in the vector on the right is a linear combination of the functions $e^{\lambda_j t}$ whose coefficients are just the ith entries of the vectors $\lambda_j \mathbf{b}_j - A\mathbf{b}_j$. But according to (7), all of these linear combinations are equal to 0. But the functions $\mathbf{e}^{\lambda_j t}$ are linearly independent, since we have assumed the eigenvalues λ_j are all distinct. This means all of the coefficients in the ith entries of the right-hand side of (7) are also zero. So we must have $\lambda_j \mathbf{j} - A\mathbf{b}_j = \mathbf{0}$, or $A\mathbf{b}_j = \lambda_j \mathbf{b}_j$, for every j. That is, each (nonzero) vector $\mathbf{b}_j$ is an eigenvector of A with eigenvalue λ_j!

We have therefore *deductively* arrived at the following well-known result.

Theorem 1. *Let A be an $n \times n$ matrix with n distinct real eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then the general solution to the linear system of ODEs $\mathbf{x}' = A\mathbf{x}$ is given by*

$$\mathbf{x}(t) = \sum_{j=1}^{n} e^{\lambda_j t}\mathbf{b}_j, \tag{8}$$

where each $\mathbf{b}_j$ is either $\mathbf{0}$ or an eigenvector of A with eigenvalue λ_j.

Remarks.

a. Strictly speaking, we have only shown that *if* $\mathbf{x}(t)$ solves $\mathbf{x}' = A\mathbf{x}$, then it must have the form given in the theorem. It still remains to prove the converse, namely that *any* vector function of the form (8) satisfies $\mathbf{x}' = A\mathbf{x}$. But this is easy, and is left to the reader.

b. The statement "$\mathbf{b}_j$ is either $\mathbf{0}$ or an eigenvector of A with eigenvalue λ_j" can be written a bit more succinctly as "$\mathbf{b}_j$ belongs to the λ_j-eigenspace of A."