Chapter 2

Linear Differential Equations

2.1. First-Order Linear ODE

Two differential equations that students usually meet very early in their mathematical careers are the first-order "equation of exponential growth", $\frac{dx}{dt} = ax$, with the explicit solution $x(t) = x(0)e^{at}$, and the second-order "equation of simple harmonic motion", $\frac{d^2x}{dt^2} = -\omega^2 x$, whose solution can also be written down explicitly: $x(t) = x(0)\cos(\omega t) + \frac{x'(0)}{\omega}\sin(\omega t)$. The interest in these two equations goes well beyond the fact that they have simple and explicit solutions. Much more important is the fact that they can be used to model successfully many real-world situations. Indeed, they are so important in both pure and applied mathematics that we will devote this and the next several sections to studying various generalizations of these equations and their applications to building models of real-world phenomena. Let us start by looking at (and behind) the property that gives these two equations their special character.

One of the most obvious features common to both of these equations is that their right-hand sides are linear functions. Now, in many real-world situations the response of a system to an influence is well approximated by a linear function of that influence, so granting that the dynamics of such problems can be described by an ODE, it should be no surprise that the dynamical equations for such systems are linear. In particular, if x measures the deviation of some system from an equilibrium configuration, then there will usually be a restoring force driving the system back towards equilibrium, the magnitude of which is linear in x—this is the general formulation of Hooke's Law that "stress is proportional to strain". From a mathematical point of view, there is nothing mysterious about this—the restoring force is actually only approximately linear, with the approximation getting better as we approach equilibrium. If we assume only that the restoring force is a differentiable function of the deviation from equilibrium, then, since it vanishes at the equilibrium, we see that the approximate linearity of the force near equilibrium is just a manifestation of Taylor's Theorem with Remainder. This observation points to a further reason for why linear equations play such a central role. Suppose we have a nonlinear differential equation $\frac{dx}{dt} = V(x)$. At an "equilibrium point" p, i.e., a point where $V(p) = 0$, define A to be the differential of V at p. Then for small x, Ax is a good approximation of $V(p+x)$, so we can hope to approximate solutions of the nonlinear equation near p with solutions of the linear equation $\frac{dx}{dt} = Ax$ near 0. In fact this technique of "linearization" is one of the most powerful tools for analyzing nonlinear differential equations and one that we shall return to repeatedly.

The most natural generalization of the equation of exponential growth to an n-dimensional system is an equation of the form $\frac{dx}{dt} = Ax$, where now x represents a point of $\mathbf{R}^n$ and $A : \mathbf{R}^n \to \mathbf{R}^n$ is a linear operator, or equivalently an $n \times n$ matrix. Such an equation is called an autonomous, first-order, linear ordinary differential equation.

▷ **Exercise 2–1. The Principle of Superposition.** Show that any linear combination of solutions of such a system is again a solution, so that if as usual σ_p denotes the solution of the initial value problem with initial condition p, then $\sigma_{p_1+p_2} = \sigma_{p_1} + \sigma_{p_2}$.

When $n = 1$, A is just a scalar, and we know that $\sigma_p(t) = e^{tA}p$, or in other words, the flow ϕ_t generated by the differential equation is just multiplication by e^{tA}. What we shall see below is that for $n > 1$ we can still make good sense out of e^{tA}, and this same formula still gives the flow.

We saw very early that in one-dimensional space successive approximations worked particularly well for the linear case, so we will begin by attempting to repeat that success in higher dimensions.

Denote by $C(\mathbf{R}, \mathbf{R}^n)$ the continuous maps of $\mathbf{R}$ into $\mathbf{R}^n$, and as earlier let $F = F^{A,x_0}$ be the map of $C(\mathbf{R}, \mathbf{R}^n)$ to itself defined by $F(x)(t) := x_0 + \int_0^t A(x(s))\,ds$. Since A is linear, this can also be written as $F(x)(t) := x_0 + A\int_0^t x(s)\,ds$. We know that the solution of the IVP with initial value x_0 is just the unique fixed point of F, so let's try to find it by successive approximations starting from the constant path $x^0(t) = x_0$. If we recall that the sequence of successive approximations, x^n, is defined recursively by $x^{n+1} = F(x^n)$, then an elementary induction gives $x^n(t) = \sum_{k=0}^n \frac{1}{k!}(tA)^k x_0$, suggesting that the solution to the initial value problem should be given by the limit of this sequence, namely the infinite series $\sum_{k=0}^\infty \frac{1}{k!}(tA)^k x_0$. Now (for obvious reasons) given a linear operator T acting on $\mathbf{R}^n$, the limit of the infinite series of operators $\sum_{k=0}^\infty \frac{1}{k!}T^k$ is denoted by e^T or $\exp(T)$, so we can also say that the solution to our IVP should be $e^{tA}x_0$.

The convergence properties of the series for $e^T x$ follow easily from the Weierstrass M-test. If we define $M_k = \frac{1}{k!}\|T\|^k r$, then $\sum M_k$ converges to $e^{\|T\|}r$, and since $\left\|\frac{1}{k!}T^k x\right\| < M_k$ when $\|x\| < r$, it follows that $\sum_{k=0}^\infty \frac{1}{k!}T^k x$ converges absolutely and uniformly to a limit, $e^T x$, on any bounded subset of $\mathbf{R}^n$.

▷ **Exercise 2–2.** Provide the details for the last statement. (Hint: Since the sequence of partial sums $\sum_{k=0}^n M_k$ converges, it is Cauchy; i.e., given $\epsilon > 0$, we can choose N large enough that $\sum_m^{m+k} M_k < \epsilon$ provided $m > N$. Now if $\|x\| < r$, $\left\|\sum_{k=0}^{m+k} \frac{1}{k!}T^k x - \sum_{k=0}^m \frac{1}{k!}T^k x\right\| < \sum_m^{m+k} M_k < \epsilon$, proving that the infinite series defining $e^T x$ is uniformly Cauchy and hence uniformly convergent in $\|x\| < r$.)

Since the partial sums of the series for $e^T x$ are linear in x, so is their limit, so e^T is indeed a linear operator on $\mathbf{R}^n$.

Next observe that since a power series in t can be differentiated term by term, it follows that $\frac{d}{dt}e^{tA}x_0 = Ae^{tA}x_0$; i.e., $x(t) = e^{tA}x_0$ is a solution of the ODE $\frac{dx}{dt} = Ax$. Finally, substituting zero for t in the power series gives $e^{0A}x_0 = x_0$. This completes the proof of the following proposition.

2.1.1. Proposition. *If A is a linear operator on $\mathbf{R}^n$, then the solution of the linear differential equation $\frac{dx}{dt} = Ax$ with initial condition x_0 is $x(t) = e^{tA}x_0$.*

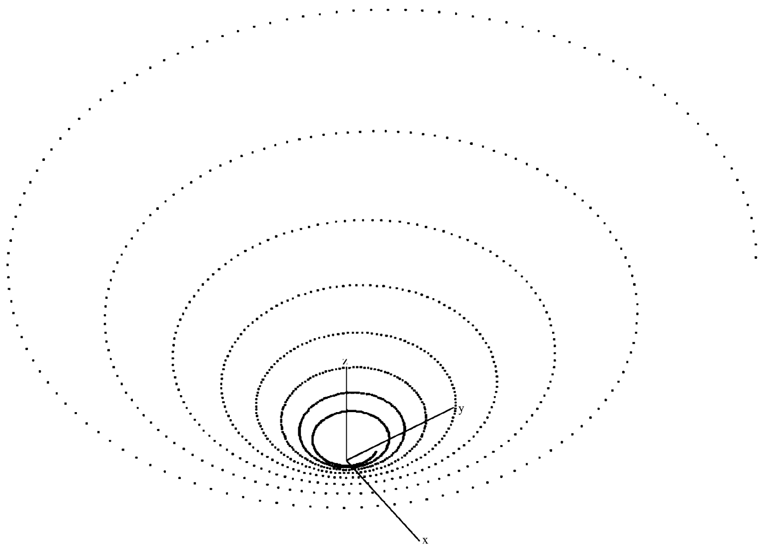

Figure 2.1. A typical solution of a first-order linear ODE in $\mathbf{R}^3$. Note: The dots are placed along the solution at fixed time intervals. This gives a visual clue to the speed at which the solution is traversed.

As a by-product of the above discussion we see that a linear ODE $\frac{dx}{dt} = Ax$ is complete, and the associated flow ϕ_t is just e^{tA}. By a general fact about flows it follows that $e^{(s+t)A} = e^{sA}e^{tA}$ and $e^{-A} = (e^A)^{-1}$, so $\exp : A \mapsto e^A$ is a map of the vector space $\mathbf{L}(\mathbf{R}^n)$ of all linear maps of $\mathbf{R}^n$ into the group $\mathbf{GL}(\mathbf{R}^n)$ of invertible elements of $\mathbf{L}(\mathbf{R}^n)$ and for each $A \in \mathbf{L}(\mathbf{R}^n)$, $t \mapsto e^{tA}$ is a homomorphism of the additive group of real numbers into $\mathbf{GL}(\mathbf{R}^n)$.

▷ **Exercise 2–3.** Show more generally that if A and B are commuting linear operators on $\mathbf{R}^n$, then $e^{A+B} = e^A e^B$. (Hint: Since A and B commute, the Binomial Theorem is valid for $(A+B)^k$, and since the series defining e^{A+B} is absolutely convergent, it is permissible to rearrange terms in the infinite sum. For a different proof, show that $e^{tA}e^{tB}x_0$ satisfies the initial value problem $\frac{dx}{dt} = (A+B)x$, $x(0) = x_0$, and use the Uniqueness Theorem.)

At first glance it might seem hopeless to attempt to solve the linear ODE $\frac{dx}{dt} = Ax$ by computing the power series for e^{tA}—if A is a 10×10 matrix, then computing just the first dozen powers of A will already be pretty time consuming. However, suppose that v is an eigenvector of A belonging to the eigenvalue λ, i.e., $Av = \lambda v$. Then $A^n v = \lambda^n v$, so that in this case $e^{tA}v = e^{t\lambda}v$! If we combine this fact with the Principle of Superposition, then we see that we are in good shape whenever the operator A is diagonalizable. Recall that this just means that there is a basis of $\mathbf{R}^n$, $e_1, \dots, e_n$, consisting of eigenvectors of A, so that $Ae_i = \lambda_i e_i$. We can expand an arbitrary initial condition $x_0 \in \mathbf{R}^n$ in this basis, i.e., $x_0 = \sum_i a_i e_i$, and then $e^{tA}x_0 = \sum_i a_i e^{t\lambda_1} e_i$ is the explicit solution of the initial value problem (a fact we could have easily verified without introducing the concept of the exponential of a matrix).

Nothing in this section has depended on the fact that we were dealing with real rather than complex vectors and matrices. If $A : \mathbf{C}^n \to \mathbf{C}^n$ is a complex linear map (or a complex $n \times n$ matrix), then the same argument as above shows that the power series for $e^{tA}z$ converges absolutely for all z in $\mathbf{C}^n$ (and for all t in $\mathbf{C}$).

If A is initially given as an operator on $\mathbf{R}^n$, it can be useful to "extend" it to an operator on $\mathbf{C}^n$ by a process called complexification. The inclusion of $\mathbf{R}$ in $\mathbf{C}$ identifies $\mathbf{R}^n$ as a real subspace of $\mathbf{C}^n$, and $\mathbf{C}^n$ is the direct sum (as a real vector space) $\mathbf{C}^n = \mathbf{R}^n \oplus i\mathbf{R}^n$. If $z = (z_1, \dots, z_n) \in \mathbf{C}^n$, then we project on these subspaces by taking the real and imaginary parts of z (i.e., the real vectors x and y whose components x_i and y_i are the real and imaginary parts of z_i). This is clearly the unique decomposition of z in the form $z = x + iy$ with both x and y in $\mathbf{R}^n$. We extend A to $\mathbf{C}^n$ by defining $Az = Ax + iAy$, and it is easy to see that this extended map is complex linear. (Hint: It is enough to check that $Aiz = iAz$.)

$\triangleright$ **Exercise 2–4.** Show that if we complexify an operator A on $\mathbf{R}^n$ as above and if a curve $z(t)$ in $\mathbf{C}^n$ is a solution of $\frac{dz}{dt} = Az$, then its real and imaginary parts are also solutions of this equation.

What is the advantage of complexification? As the following example shows, a nondiagonalizable operator A on $\mathbf{R}^n$ may become diagonalizable after complexification, allowing us to solve $\frac{dz}{dt} = Az$ easily in $\mathbf{C}^n$. Moreover, we can then apply the preceding exercise to solve the initial value problem in $\mathbf{R}^n$ from the solution in $\mathbf{C}^n$.

$\bullet$ **Example 2–1.** We can write the system $\frac{dx_1}{dt} = x_2$, $\frac{dx_2}{dt} = -x_1$ as $\frac{dx}{dt} = Ax$, where A is the linear operator on $\mathbf{R}^2$ that is defined by $A(x_1, x_2) = (x_2, -x_1)$. Since A^2 is minus the identity, A has no real eigenvalues and so is not diagonalizable. But, if we complexify A, then the vectors $e_1 = (1, i)$ and $e_2 = (1, -i)$ in $\mathbf{C}^2$ satisfy $Ae_1 = ie_1$ and $Ae_2 = -ie_2$, so they are an eigenbasis for the complexification of A, and we have diagonalized A in $\mathbf{C}^2$. The solution of $\frac{dz}{dt} = Az$ with initial value $e_1 = (1, i)$ is $e^{it}e_1 = (e^{it}, ie^{it})$. Taking real parts, we find that the solution of the initial value problem for $\frac{dx}{dt} = Ax$ with initial condition $(1, 0)$ is $(\cos(t), -\sin(t))$, while taking imaginary parts, we see that the solution with initial condition $(0, 1)$ is $(\sin(t), \cos(t))$. By the Principle of Superposition the solution $\sigma_{(a,b)}(t)$ with initial condition (a, b) is $(a\cos(t) + b\sin(t), -a\sin(t) + b\cos(t))$.

Next we will analyze in more detail the properties of the flow e^{tA} on $\mathbf{C}^n$ generated by a linear differential equation $\frac{dz}{dt} = Az$. We have seen that this flow is transparent for the case that A is diagonalizable, but we want to treat the general case, so we will not assume this. Our approach is based on the following elementary consequence of the Principle of Superposition.

2.1.2. Reduction Principle. *Let $\mathbf{C}^n$ be the direct sum of subspaces V_i, each of which is mapped into itself by the operator A, and let $v \in \mathbf{C}^n$ and $v = v_1 + \cdots + v_k$, with $v_i \in V_i$. If σ_p denotes the solution of $\frac{dz}{dt} = Az$ with initial condition p, then $\sigma_{v_i}(t) \in V_i$ for all t and $\sigma_v(t) = \sigma_{v_1}(t) + \cdots + \sigma_{v_k}(t)$.*