4 Linear Systems

We [he and Paul Halmos] share a philosophy about linear algebra: we think basis free, we write basis free, but when the chips are down we close the office door and compute with matrices like fury.
—Irving Kaplansky

In this chapter, we consider linear systems, that is, ODEs of the form

$$\dot{\mathbf{x}}(t) = A(t)\mathbf{x} + \mathbf{b}(t), \tag{4.1}$$

where $A : \mathbb{R} \to M_n(\mathbb{F})$ is continuous and $\mathbf{b} : \mathbb{R} \to \mathbb{F}^n$ is a continuous vector-valued function; see Definition 2.1.9(ii).

We spend a significant amount of time working with linear systems before considering fully nonlinear ODEs. There are a variety of reasons for this. First, linear systems are easier to analyze and the solutions are much easier to compute. More importantly though, the most powerful tool for analyzing nonlinear ODEs is to linearize them and investigate those linearizations.

4.1 Homogeneous Linear Systems

Recall from Definition 2.1.9 that a homogeneous linear system is one where the term $\mathbf{b}(t)$ in (4.1) is zero. That is

$$\dot{\mathbf{x}}(t) = A(t)\mathbf{x}(t). \tag{4.2}$$

Homogeneous systems are important not only for their own sake, but also because the difference $\mathbf{x}(t) = \mathbf{x}_1(t) - \mathbf{x}_2(t)$ between any two solutions $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ of (4.1) is itself a solution of the homogeneous system (4.2). That means to find all solutions of the inhomogeneous equation (4.1), it suffices to find one such solution and then add to it all the solutions of the homogeneous system (4.2).

4.1.1 Basic Properties

The function $A(t)$ is continuous, thus for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{F}^n$ on a compact interval $\overline{I}$ we have

$$\begin{aligned}\|A(t)\mathbf{x}_1 + \mathbf{b}(t) - A(t)\mathbf{x}_2 - \mathbf{b}(t)\| &= \|A(t)(\mathbf{x}_1 - \mathbf{x}_2)\| \\ &\leq \|A(t)\|\|\mathbf{x}_1 - \mathbf{x}_2\| \\ &\leq \sup_{t\in\overline{I}} \|A(t)\|\|\mathbf{x}_1 - \mathbf{x}_2\|.\end{aligned}$$

By continuity of A and the compactness of $\overline{I}$ the quantity

$$L = \sup_{t\in\overline{I}} \|A(t)\|$$

is finite, and the function $\mathbf{f}(\mathbf{x}, t) = A(t)\mathbf{x} + \mathbf{b}(t)$ is locally Lipschitz in $\mathbf{x}$. Hence the Existence and Uniqueness Theorem guarantees that (4.1) has a unique solution on an open interval containing the initial point t_0.

Theorem 4.1.1. *If $A(t)$ is continuous at all $t \in \mathbb{R}$, then for any solution of* (4.2)*, the maximal interval of existence for that solution is the full real line* $(-\infty, \infty)$.

Remark 4.1.2. We show at the end of this chapter (Theorem 4.4.1) that Theorem 4.1.1 holds also for the inhomogeneous case, with $\mathbf{b}(t)$ not necessarily zero (but also continuous).

Proof. By the Unique Extension Theorem the open interval I extends to a maximal open interval J; so it suffices to prove that $J = (-\infty, \infty)$. By way of contradiction suppose that the right end point of J is $\beta < \infty$ (the argument for a finite left endpoint is similar). By the Finite Time Blowup Theorem it follows that $\|\mathbf{x}(t)\| \to \infty$ as $t \to \beta$. We also know that the solution $\mathbf{x}(t)$ is given by

$$\mathbf{x}(t) = \mathbf{x}(t_0) + \int_{t_0}^{t} A(s)\mathbf{x}(s)\, ds, \quad t \in [t_0, \beta).$$

It follows that

$$\|\mathbf{x}(t)\| \leq \|\mathbf{x}_0\| + \int_{t_0}^{t} \sup_{t_0 \leq s \leq \beta} (\|A(s)\|)\, \|\mathbf{x}(s)\|\, ds,$$

for $t \in [t_0, \beta)$. Because $A(t)$ is continuous on the compact interval $[t_0, \beta]$ then

$$L = \sup_{t_0 \leq t \leq \beta} \|A(t)\| < \infty,$$

and hence

$$\|\mathbf{x}(t)\| \leq \|\mathbf{x}_0\| + \int_{t_0}^{t} L\|\mathbf{x}(s)\|\, ds, \quad t \in [t_0, \beta).$$

Gronwall's inequality gives

$$\|\mathbf{x}(t)\| \leq \|x_0\| \exp\left(L(t - t_0)\right), \quad t \in [t_0, \beta).$$

The quantity on the right hand side of this estimate stays bounded as $t \to \beta$, which is a contradiction ($\|x(t)\|$ does not blow up as $t \to \beta$). $\square$

4.1.2 Fundamental Matrix Solutions

It is straightforward to check that if $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ are both solutions to the homogeneous linear system (4.2), then so is any linear combination $c_1\mathbf{x}_1(t)+c_2\mathbf{x}_2(t)$. This implies that the set of solutions is a vector space. Moreover, the map $G : \mathbb{R}^n \to C(I; \mathbb{R}^n)$ between initial conditions and solutions given by

$$\mathbf{v} \mapsto \mathbf{x}(t) \text{ such that } \mathbf{x}(0) = \mathbf{v} \text{ and (4.2) holds}$$

is a linear map because if $G(\mathbf{v}_1) = \mathbf{x}_1$ and $G(\mathbf{v}_2) = \mathbf{x}_2$, then for any $c_1, c_2 \in \mathbb{R}$ the function $\mathbf{x} = c_1\mathbf{x}_1 + c_2\mathbf{x}_2$ satisfies (4.2) because

$$\begin{aligned}\dot{\mathbf{x}}(t) &= c_1\dot{\mathbf{x}}_1(t) + c_2\dot{\mathbf{x}}_2(t)\\ &= c_1A(t)\mathbf{x}_1(t) + c_2A(t)\mathbf{x}_2(t)\\ &= A(t)\mathbf{x}(t),\end{aligned}$$

and

$$\mathbf{x}(0) = c_1\mathbf{x}_1(0) + c_2\mathbf{x}_2(0) = c_1\mathbf{v}_1 + c_2\mathbf{v}_2. \tag{4.3}$$

That is to say, $c_1G(\mathbf{v}_1) + c_2G(\mathbf{v}_2)$ is the unique solution of the IVP defined by (4.2) and (4.3), and therefore

$$G(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1G(\mathbf{v}_1) + c_2G(\mathbf{v}_2).$$

Definition 4.1.3. *Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis for $\mathbb{R}^n$. Denote by $\mathbf{x}_1(t), \ldots, \mathbf{x}_n(t)$ the solutions of (4.2) with each $\mathbf{x}_k(t_0) = \mathbf{v}_k$ (each $\mathbf{x}_k$ is uniquely determined by the initial value, by the Existence and Uniqueness theorem. The matrix of column vectors $\Phi(t) = \begin{bmatrix}\mathbf{x}_1(t) & \cdots & \mathbf{x}_n(t)\end{bmatrix}$ is called the* fundamental matrix of solutions *of (4.2) corresponding to the basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. If $\mathbf{v}_k = \mathbf{e}_k$ are the standard basis vectors, so that $\Phi(t_0) = \begin{bmatrix}\mathbf{e}_1 & \ldots & \mathbf{e}_n\end{bmatrix}$, then the matrix Φ is called the* principal fundamental matrix*.*

If $\Phi(t)$ is the fundamental matrix of solutions of (4.2) corresponding to the basis $\{\mathbf{v}_1 \ldots, \mathbf{v}_n\}$, then

$$\dot{\Phi}(t) = A(t)\Phi(t), \quad \text{for all } t \in \mathbb{R} \tag{4.4}$$

and

$$\Phi(t_0) = \begin{bmatrix}\mathbf{v}_1 & \ldots & \mathbf{v}_n\end{bmatrix}.$$

Because the map G from initial conditions to solutions is linear, given any initial condition $\mathbf{x}(t_0) = \mathbf{v}$, the solution to the corresponding IVP (satisfying the initial condition and (4.2)) can be found from the principal fundamental matrix of solutions $\Phi(t)$ by a simple matrix-vector product[22] $\mathbf{x}(t) = \Phi(t)\mathbf{v}$.

Proposition 4.1.4. *The columns of a fundamental matrix $\Phi(t)$ are linearly independent solutions of $\dot{\mathbf{x}} = A(t)\mathbf{x}$.*

[22] We have used $\mathbf{v}$ to mean both a vector in $\mathbb{F}^n$ and its representation in the standard basis. If the vector were expressed in a different basis, then instead of the principal fundamental matrix solution Φ, we'd need to use the fundamental matrix solution corresponding to that other basis.

Proof. If there were a nontrivial relation $\sum_i a_k \mathbf{x}_k = \mathbf{0}$ among the columns, then the corresponding IVP with initial value $\mathbf{x}(t_0) = \sum_i a_k \mathbf{v}_k$ would have the same solution as the IVP with initial value $\mathbf{x}(t_0) = \mathbf{0}$. But $\sum_i a_k \mathbf{v}_k \neq \mathbf{0}$ because $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis. This contradicts the Uniqueness Theorem; and therefore no such relation can exist. ∎

Example 4.1.5. Consider the ODE (4.2), where

$$A = \begin{bmatrix} 0 & -1 \\ 0 & 1/2 \end{bmatrix}. \tag{4.5}$$

A straightforward check shows that the principal fundamental matrix is

$$\Phi(t) = \begin{bmatrix} 1 & 2(1 - e^{t/2}) \\ 0 & e^{t/2} \end{bmatrix}.$$

Thus, $\Phi(t)$ satisfies (4.4) with $\Phi(0) = I$.

Example 4.1.6. For the nonautonomous system

$$A(t) = \begin{bmatrix} -1 + \frac{3}{2}\cos^2 t & 1 - \frac{3}{2}\sin t \cos t \\ -1 - \frac{3}{2}\sin t \cos t & -1 + \frac{3}{2}\sin^2 t \end{bmatrix},$$

two linearly independent solutions of the linear homogeneous system

$$\dot{\mathbf{x}} = A(t)\mathbf{x}$$

are

$$\mathbf{x}_1(t) = e^{t/2} \begin{bmatrix} \cos t \\ -\sin t \end{bmatrix}, \quad \mathbf{x}_2(t) = e^{-t} \begin{bmatrix} \sin t \\ \cos t \end{bmatrix}.$$

To verify that these are solutions you will need the unfamiliar "cubic" trig identities

$$\cos^3(t) = \frac{3\cos t + \cos 3t}{4}, \quad \sin^3 t = \frac{3\sin t - \sin 3t}{4}$$

the more familiar "quadratic" trig identities

$$\cos^2 t = \frac{1 + \cos 2t}{2}, \quad \sin^2 t = \frac{1 - \cos 2t}{2}$$

and the less familiar "product to sum" trig identities

$$\cos\alpha\cos\beta = \frac{\cos(\alpha - \beta) + \cos(\alpha + \beta)}{2}, \quad \sin\alpha\cos\beta = \frac{\sin(\alpha + \beta) + \sin(\alpha - \beta)}{2}.$$

With $t_0 = 0$ the principal fundmental matrix is

$$\Phi(t) = \begin{bmatrix}\mathbf{x}_1(t) & \mathbf{x}_2(t)\end{bmatrix} = \begin{bmatrix} e^{t/2}\cos t & e^{-t}\sin t \\ -e^{t/2}\sin(t) & e^{-t}\cos t\end{bmatrix}.$$

4.1.3 The Exponential of a Matrix

Recall from Volume 1, Example 5.7.2 and Exercise 6.13 that for any $A \in M_n(\mathbb{C})$ (and indeed for any bounded linear operator A on a Banach space X) the function $e^{tA} : \mathbb{R} \to M_n(\mathbb{C})$ given by

$$e^{tA} = I + tA + \frac{A^2}{2}t^2 + \frac{A^3}{6}t^3 + \cdots = \sum_{k=0}^{\infty} \frac{A^k}{k!}t^k \tag{4.6}$$

converges absolutely for all $t \in \mathbb{R}$, and it converges uniformly on any bounded interval. When $A(t)$ doesn't depend on t, so $A(t)$ is a constant matrix function $A \in M_n(\mathbb{C})$, then a solution of the autonomous linear homogeneous equation

$$\dot{\mathbf{x}} = A\mathbf{x}. \tag{4.7}$$

is given by the matrix exponential, as shown below.

Remark 4.1.7. Recall that the Cayley-Hamilton Theorem (Volume 1, Corollary 12.4.8) guarantees that for any matrix $A \in M_n(\mathbb{F})$, the characteristic polynomial $p(z) = \det(zI - A)$ has the property that $p(A) = 0$. Since the characteristic polynomial has degree n, this means that A^n (and all higher powers of A) can all be rewritten as polynomials of degree at most $n-1$ in A. In particular, that means although the matrix exponential 4.6 involves infinitely many powers of A, it can rewritten as a polynomial of degree no more than $n-1$ in A.

Theorem 4.1.8. *If $A, B \in M_n(\mathbb{F})$ and $s, t \in \mathbb{R}$, then the following hold:*

(i) $e^0 = I$.

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

(iii) $(e^{At})^{-1} = e^{-At}$. *In particular, e^{At} is invertible.*

(iv) $\frac{d}{dt}e^{At} = Ae^{At} = e^{At}A$.

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

(vi) *If $A = PDP^{-1}$ for a matrix D and some invertible matrix P, then $e^{tA} = Pe^{Dt}P^{-1}$.*

Proof. The proof is Exercise 4.3. ∎

Corollary 4.1.9. *Given $A \in M_n(\mathbb{F})$, consider the autonomous ODE*

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t). \tag{4.8}$$

The principal fundamental solution is $\Phi(t) = e^{(t-t_0)A}$.

Proof. The matrix function $\Phi(t) = \exp((t-t_0)A)$ converges uniformly on any bounded interval, so it is differentiable at all $t \in \mathbb{R}$. Its derivative is $\frac{d}{dt}\Phi = A\exp((t-t_0)A) = A\Phi$ by Theorem 4.1.8.(iv), so it satisfies (4.8). Moreover, $\Phi(t_0) = I$, by Theorem 4.1.8.(i), so it is the principal fundamental solution. ∎

It usually isn't a good idea to compute the matrix exponential via (4.6). If $A \in M_n$ is dense, then each matrix multiplication typically costs at least $O(n^{2.3})$ (usually more like $O(n^3)$), and assuming that it takes p iterations for (4.6) to converge then it costs about $O(pn^{2.3})$ to compute a single matrix exponential, but the solution to (4.8) requires computing e^{tA} for all t.

If A is semisimple (diagonalizable) then Theorem 4.1.8.(vi) leads to a much better method to compute the matrix exponential. Specifically, if Λ is diagonal with diagonal entries $\lambda_1, \ldots, \lambda_n$, then a straightforward calculation shows that

$$e^{t\Lambda} = \begin{bmatrix} e^{\lambda_1 t} & 0 & 0 & \cdots & 0 \\ 0 & e^{\lambda_2 t} & 0 & \cdots & 0 \\ 0 & 0 & e^{\lambda_3 t} & \cdots & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \cdots & e^{\lambda_n t} \end{bmatrix},$$

so E^{tA} can be computed just by finding the eigenvalues Λ and the eigenvectors P and computing the product $e^{tA} = Pe^{t\Lambda}P^{-1}$.

Example 4.1.10. Consider the ODE (4.8) where

$$A = \begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix}$$

and $t_0 = 0$. The matrix A is diagonalizable by the matrix of eigenvectors

$$P = \begin{bmatrix} 2 & -2 \\ 1 & 1 \end{bmatrix} \qquad \text{and} \qquad P^{-1} = \frac{1}{4}\begin{bmatrix} 1 & 2 \\ -1 & 2 \end{bmatrix},$$

with corresponding eigenvalues $\lambda = 3, -1$. This implies that

$$\Phi(t) = e^{tA} = Pe^{t\Lambda}P^{-1} = \frac{1}{4}\begin{bmatrix} 2(e^{3t} + e^{-t}) & 4(e^{3t} - e^{-t}) \\ e^{3t} - e^{-t} & 2(e^{3t} + e^{-t}) \end{bmatrix}$$

is the principal fundamental matrix solution.

Example 4.1.11. Consider the linear first-order form of the damped harmonic oscillator (2.5) with values $m = 1$, $k = 1000$, and $c = 1001$. This corresponds to the situation where Bob is a lightweight fellow on a very stiff spring, sliding on a highly frictional surface like sandpaper.

The model is, therefore,

$$\dot{\mathbf{y}} = A\mathbf{y}, \qquad A = \begin{bmatrix} 0 & 1 \\ -1000 & -1001 \end{bmatrix}. \tag{4.9}$$

It is straightforward to check that the eigenvalues of A are $\lambda = -1, -1000$, and the corresponding matrix of eigenvectors is $P = \left[\begin{smallmatrix} 1 & 1 \\ 1 & 1000 \end{smallmatrix}\right]$. Thus, the principle fundamental matrix solution is

$$\begin{aligned}\Phi(t) &= \exp(tA) \\ &= P \begin{bmatrix} e^{-t} & 0 \\ 0 & e^{-1000t} \end{bmatrix} P^{-1} \\ &= \frac{1}{999} \begin{bmatrix} 1000e^{-t} - e^{-1000t} & e^{-t} - e^{-1000t} \\ -1000e^{-t} + 1000e^{-1000t} & -e^{-t} + 1000e^{-1000t} \end{bmatrix}.\end{aligned}$$

Unexample 4.1.12. It is an unhappy fact that the relation $e^{tA+tB} = e^{tA}e^{tB}$ only holds when $AB = BA$. For example if $A = \left[\begin{smallmatrix} 1 & 0 \\ 1 & 0 \end{smallmatrix}\right]$ and $B = \left[\begin{smallmatrix} 2 & 2 \\ 0 & 0 \end{smallmatrix}\right]$, then

$$AB = \left[\begin{smallmatrix} 4 & 0 \\ 0 & 0 \end{smallmatrix}\right] \neq \left[\begin{smallmatrix} 2 & 2 \\ 2 & 2 \end{smallmatrix}\right] = BA.$$

The matrix A is simple, with $A = P\Lambda P^{-1}$, where $\Lambda = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right]$ and $P = \begin{bmatrix} 1 & -\frac{\sqrt{2}}{2} \\ 0 & \frac{\sqrt{2}}{2} \end{bmatrix}$. Similarly B is simple with $B = QDQ^{-1}$, where $D = \left[\begin{smallmatrix} 0 & 0 \\ 0 & 2 \end{smallmatrix}\right]$ and $Q = \begin{bmatrix} 0 & \frac{\sqrt{2}}{2} \\ 1 & \frac{\sqrt{2}}{2} \end{bmatrix}$. And finally there is a matrix of eigenvectors R such that $A + B = R \left[\begin{smallmatrix} 3.561 & 0 \\ 0 & -0.562 \end{smallmatrix}\right] R^{-1}$. Taking $t = 1$ and using the same method as Example 4.1.10 gives

$$e^{A+B} = R \begin{bmatrix} e^{3.56155281} & 0 \\ 0 & e^{-0.56155281} \end{bmatrix} R^{-1} = \begin{bmatrix} 20.08 & e \\ 7.39 & 1 \end{bmatrix}$$

but

$$e^A e^B = Pe^{\Lambda}P^{-1}Qe^{D}Q^{-1} = \begin{bmatrix} 40.16 & 2e \\ 14.78 & 2 \end{bmatrix} = 2e^{A+B}.$$

If the matrix A is semisimple (diagonalizable), then the diagonalizing approach of Example 4.1.10 is often a good way to compute the matrix exponential. But if it is not semisimple, we need other methods. If the matrix A is nilpotent, then the infinite series in (4.6) for e^{tA} is finite, which often makes it relatively easy to compute.

Example 4.1.13. Consider the ODE (4.8) where $A = \left[\begin{smallmatrix} -1 & 1 \\ -1 & 1 \end{smallmatrix}\right]$. A straightforward calculation shows that $A^2 = 0$, which implies that

$$\Phi(t) = e^{tA} = I + At = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + t \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1-t & t \\ -t & 1+t \end{bmatrix}$$

is the principal fundamental matrix solution (assuming $t_0 = 0$).

Example 4.1.14. The matrix $A = \left[\begin{smallmatrix} 1 & 2 \\ 0 & 1 \end{smallmatrix}\right]$ can be written as $A = I + B$ with $I = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$ and $B = \left[\begin{smallmatrix} 0 & 2 \\ 0 & 0 \end{smallmatrix}\right]$. Since $IB = BI = B$, the identity $e^{tA} = e^{It}e^{tB}$ holds. The matrix B is nilpotent, with $B^2 = 0$, so $e^{tB} = I + tB$. This gives

$$e^{tA} = e^{tI}e^{tB} = \begin{bmatrix} e^t & 0 \\ 0 & e^t \end{bmatrix} (I + tB) = \begin{bmatrix} e^t & 2te^t \\ 0 & e^t \end{bmatrix}.$$

4.1.4 Numerical Computation of Matrix Exponentials

Computing the Taylor series expansion is usually not the best way to find $\exp(tA)$. There is a host of algorithms devoted to rapid computation of the matrix exponential, most of which, unfortunately, we are unable to cover here in any detail.

For a semisimple matrix A, if the matrix P of eigenvectors is well conditioned (the ratio of largest singular value to smallest singular value is not too large), then the diagonalization method (see Example 4.1.10) works very well. This is especially true when A is normal, which implies that it can be orthonormally diagonalized. In that case the condition of P is best possible and P^{-1} is just P^{H}. But if P is ill conditioned, for example when two eigenvectors are nearly parallel, then the diagonalization method does not work well, because the computation of P^{-1} is ill conditioned, which makes the product $Pe^{tA}P^{-1}$ very sensitive to small changes in the inputs (tiny errors in A lead to huge errors in the value of e^{tA}). So even when A is semisimple, other methods are needed.

One valuable approach is to use Schur's Lemma which decomposes $A = UTU^{\mathsf{H}}$, where T is upper triangular and U is orthonormal. There are several numerical algorithms for computing $\exp(T)$ when T is upper triangular, and Theorem 4.1.8.(vi) gives $\exp(tA) = U\exp(tT)U^{\mathsf{H}}$. In this case $U^{\mathsf{H}} = U^{-1}$), and there is no issue with the condition of computing U^{-1}.

Remark 4.1.15. Most matrices are semisimple. In fact, for both $\mathbb{F} = \mathbb{R}$ and $\mathbb{F} = \mathbb{C}$, the matrices in $M_n(\mathbb{F})$ are generically[23] simple. That means a slight perturbation of any nonsemisimple matrix (which must have repeated eigenvalues) gives a matrix with distinct (albeit close) eigenvalues, where the distance between the original and the perturbed matrix is small in whatever matrix norm you wish to use (recall that all norms on finite-dimensional vector spaces are equivalent).

This might tempt you to think that the "problematic" or "defective" matrices, those which are not diagonalizable, are not important, but the lack of diagonalizability may be essential to obtain an accurate model. The "problematic" or "defective" matrices are precisely the places where bifurcations—fundamental changes in behavior—may occur in a model as parameters are changed, and these are often exactly the phenomena you want or need to describe in the model. In particular these fundamental changes in behavior are often exactly the circumstances that are most meaningful and relevant to a physical system.

4.1.5 Complex Eigenvalues of a Real-Valued Matrix

Recall that for any $A \in M_n(\mathbb{R})$, complex-valued eigenvalues must come in conjugate pairs, that is, if $\lambda \in \sigma(A)$, then $\overline{\lambda} \in \sigma(A)$. We begin the study of these complex eigenvalues by focusing on a single conjugate pair of complex eigenvalues that are simple (their algebraic multiplicity is one), and we restrict attention to the subspace spanned by the two corresponding eigenvectors. The subspace spanned by the two eigenvectors is invariant under A.

Assume that the two eigenvalues are $\lambda = a + ib$ and $\overline{\lambda} = a - ib$ for $b \neq 0$. In addition, suppose that the eigenvectors of A (which can be chosen to be complex conjugates of each other) are $\mathbf{u} \pm i\mathbf{v}$ where $\mathbf{u}$ and $\mathbf{v}$ are both real n-dimensional vectors. You will find in the homework exercises that if $Q = \begin{bmatrix}\mathbf{u} & \mathbf{v}\end{bmatrix} \in M_2(\mathbb{R})$, then

$$Q^{-1}AQ = \begin{bmatrix} a & b \\ -b & a \end{bmatrix} = C. \tag{4.10}$$

The matrix C can be written as the sum of two matrices

$$C = \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix} + \begin{bmatrix} 0 & b \\ -b & 0 \end{bmatrix}$$

which commute because the first one is a multiple of the identity matrix.

Thus

$$\begin{aligned} \exp(tC) &= \exp\left(t\begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix} + t\begin{bmatrix} 0 & b \\ -b & 0 \end{bmatrix}\right) \\ &= \exp\left(t\begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}\right)\exp\left(t\begin{bmatrix} 0 & b \\ -b & 0 \end{bmatrix}\right) \end{aligned}$$

The first exponential is a multiple of the identity matrix and yields $\exp(ta)I$.

[23]We say that a property of points in a topological space X is a *generic* property if there is an open, dense subset $U \subset X$ such that all points of U have that property. Thus saying that the property of being a simple matrix is generic (alternatively one might say that matrices in $M_n(\mathbb{F})$ are generically simple) means that there is an open dense set $U \subset M_n(\mathbb{F})$ such that any matrix in U is simple. The fact that U is dense means that any matrix is arbitrarily close to a simple matrix.

For the second exponential note that if $B = \left[\begin{smallmatrix} 0 & b \\ -b & 0 \end{smallmatrix}\right]$, then $B^2 = -bI$, $B^3 = -bB$, $B^{2k} = (-b)^k I$ for all $k \geq 0$ and $B^{2k+1} = (-b)^k B$. Plugging these into the Taylor expansion (4.6) and recognizing the Taylor expansions of $\cos(tb)$ and $\sin(bt)$ gives

$$\exp(tB) = \begin{bmatrix} \cos(tb) & \sin(tb) \\ -\sin(tb) & \cos(tb) \end{bmatrix}.$$

You should verify that the columns of $\exp(tB)$ form a fundamental set of solutions for the ODE

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & b \\ -b & 0 \end{bmatrix} \mathbf{x}.$$

Putting this all together gives

$$\exp\left(t \begin{bmatrix} a & b \\ -b & a \end{bmatrix}\right) = \exp(ta) I \begin{bmatrix} \cos(tb) & \sin(tb) \\ -\sin(tb) & \cos(tb) \end{bmatrix} = \begin{bmatrix} e^{at}\cos(bt) & e^{at}\sin(bt) \\ -e^{at}\sin(bt) & e^{at}\cos(bt) \end{bmatrix}.$$

This says that a simple complex conjugate pair of eigenvalues $a \pm bi$ implies the existence of oscillatory solutions which expand as $t \to \infty$ if $a > 0$, contract if $a < 0$, and are bounded if $a = 0$.

The matrix exponential above can be computed in at least two other ways. First, since the eigenvalue pair is simple, we could diagonalize:

$$\begin{aligned}
\exp\left(t \begin{bmatrix} a & b \\ -b & a \end{bmatrix}\right) &= \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} \begin{bmatrix} e^{t(a+bi)} & 0 \\ 0 & e^{t(a-bi)} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}^{-1} \\
&= \frac{1}{2} \begin{bmatrix} (e^{t(a+bi)} + e^{t(a-bi)}) & -i(e^{t(a+bi)} - e^{t(a-bi)}) \\ i(e^{t(a+bi)} - e^{t(a-bi)}) & (e^{t(a+bi)} + e^{t(a-bi)}) \end{bmatrix} \\
&= e^{ta} \begin{bmatrix} \cos(tb) & \sin(tb) \\ -\sin(tb) & \cos(tb) \end{bmatrix}.
\end{aligned}$$

Here the last equality follows from Euler's identity $e^{i\theta} = \cos(\theta) + i\sin(\theta)$ for all real values of θ. Finally, this could be calculated using spectral calculus, as described in the next subsection.

4.2 Matrix Exponentials via Spectral Calculus

An important way to compute matrix exponential is to use the spectral decomposition of A, multiply it by t, and use Theorem 12.7.6 (Mapping the Spectral Decomposition) in Volume 1. This is not especially well suited to numerical calculation, but it has significant theoretical power for proving that solutions of ODEs have particular properties.

We briefly review spectral calculus (also known as *Riesz functional calculus*) from Volume 1. Recall first that the spectrum $\sigma(A)$ of a matrix $A \in M_n(\mathbb{C})$ is the set of eigenvalues of A, and the resolvent $R_A : \mathbb{C} \to M_n(\mathbb{C})$ of A is the function (see Volume 1, Section 12.3)

$$R(z) = (zI - A)^{-1}. \tag{4.11}$$

Remark 4.2.1. Each entry of the resolvent is a rational function by Cramer's Rule:

$$R_A(z) = (zI - A)^{-1} = \frac{1}{\det(zI - A)} \operatorname{adj}(zI - A)$$

where $\det(zI - A)$ is the characteristic polynomial of A, and $\operatorname{adj}(zI - A)$ is the adjugate matrix, namely, the transpose of the matrix of signed minors of $zI - A$, which satisfies $(zI - A)\operatorname{adj}(zI - A) = \det(zI - A)I$.

One of the reasons the resolvent is so important is that it plays a fundamental role in the Spectral Resolution Formula.

Theorem 4.2.2 (Spectral Resolution Formula (Volume 1 Theorem 12.4.6)). *Suppose that $f(z)$ has a power series at $z = 0$ with radius of convergence $b > r(A)$. For any positively-oriented simple closed contour Γ containing $\sigma(A)$, we have*

$$f(A) = \frac{1}{2\pi i}\oint_\Gamma f(z)R(z)dz. \tag{4.12}$$

As a special case, the Spectral Resolution Formula gives another formula for the matrix exponential

$$e^{At} = \frac{1}{2\pi i}\oint_\Gamma e^{zt}R(z)\,dz.$$

Example 4.2.3. If $A = \left[\begin{smallmatrix}1 & 2\\ 0 & 1\end{smallmatrix}\right]$, then

$$e^{zt}R(z) = e^{zt}(zI - A)^{-1} = \frac{e^{zt}}{(z-1)^2}\begin{bmatrix} z-1 & 2\\ 0 & z-1\end{bmatrix} = \begin{bmatrix} \frac{e^{zt}}{(z-1)} & \frac{2e^{zt}}{(z-1)^2}\\ 0 & \frac{e^{zt}}{(z-1)}\end{bmatrix}.$$

There is only one eigenvalue of A, namely 1, so if Γ is a positively oriented circle in the complex plane around 1 then

$$\begin{aligned} e^{At} &= \frac{1}{2\pi i}\oint_\Gamma e^{zt}R(z)\,dt \\ &= \begin{bmatrix} \frac{1}{2\pi i}\oint_\Gamma \frac{e^{zt}}{(z-1)}\,dz & \frac{1}{2\pi i}\oint_\Gamma \frac{2e^{zt}}{(z-1)^2}\,dz \\ 0 & \frac{1}{2\pi i}\oint_\Gamma \frac{e^{zt}}{(z-1)}\,dz \end{bmatrix} \end{aligned}$$

These integrals can be evaluated by Cauchy's integral formula (Volume 1, Section 11.4). Specifically, the Taylor expansion of e^{zt} around $z = 1$ is $e^t + te^t(z-1) + \frac{t^2}{2!}e^t(z-1)^2 + \cdots$, which means that the residue of

$$\frac{e^{zt}}{(z-1)} = \frac{e^t}{z-1} + te^t + \frac{t^2}{2!}e^t(z-1) + \cdots$$

is e^t, and the residue of

$$\frac{2e^{zt}}{(z-1)^2} = 2\frac{e^t}{(z-1)^2} + 2\frac{te^t}{z-1} + t^2e^t + \cdots$$

is $2te^t$. This gives

$$e^{At} = \begin{bmatrix} e^t & 2te^t \\ 0 & e^t \end{bmatrix},$$

which agrees with the result in Example 4.1.14.

In this special case the method of Example 4.1.14 is a more efficient way to compute the matrix exponential because A can easily be composed into a diagonal and nilpotent part that commute. Such a decomposition is always possible in exact arithmetic—it is called the *Jordan decomposition* (see Volume 1, Section 12.10)—but finding that decomposition is generally not a well-conditioned problem, so in floating point arithmetic the Jordan decomposition is not really computable.

Definition 4.2.4. *For a simple, positively oriented contour Γ that encircles $\lambda \in \sigma(A)$ and no other points of $\sigma(A)$ (this is always possible because the spectrum of a matrix A is always discrete), define*

$$P_\lambda = \frac{1}{2\pi i}\oint_\Gamma R(z)dz = \operatorname{Res}(R;\lambda) \tag{4.13}$$

$$D_\lambda = \frac{1}{2\pi i}\oint_\Gamma (z-\lambda)R(z)dz. \tag{4.14}$$

The operator P_λ is called the eigenprojection, *and D_λ is the* eigennilpotent *of A. The linear operator D_λ being nilpotent means there is a positive integer m_λ such that $D_\lambda^{m_\lambda} = 0$.*

Remark 4.2.5. The linear operators P_λ and D_λ are related by $D_\lambda = (A - \lambda I)P_\lambda$.

The eigenprojections and eigennilpotents reveal many important properties of a matrix and are very useful for computing functions of matrices.

Theorem 4.2.6 (Spectral Decomposition Theorem). *For $A \in M_n(\mathbb{C})$, and $\lambda \in \sigma(A)$, let P_λ be the eigenprojection of A associated to λ, and let D_λ be the eigennilpotent of A associated to λ with its order m_λ. The resolvent of A takes the form*

$$R_A(z) = \sum_{\lambda\in\sigma(A)} \left[\frac{P_\lambda}{z-\lambda} + \sum_{k=1}^{m_\lambda - 1} \frac{D_\lambda^k}{(z-\lambda)^{k+1}}\right],$$

and the following spectral decomposition *holds:*

$$A = \sum_{\lambda\in\sigma(A)} \big(\lambda P_\lambda + D_\lambda\big). \tag{4.15}$$

Theorem 4.2.7 (Spectral Mapping Theorem). *For $A \in M_n(\mathbb{C})$, if f is holomorphic on an open disk containing $\sigma(A)$, then*

$$\sigma(f(A)) = f(\sigma(A)).$$

Moreover, if $\mathbf{x} \in \mathbb{C}^n$ *is an eigenvector of* A *corresponding to* $\lambda \in \sigma(A)$*, then* $\mathbf{x}$ *is an eigenvector of* $f(A)$ *corresponding to* $f(\lambda)$*.*

Theorem 4.2.8 (Mapping the Spectral Decomposition Theorem). *Let* $A \in M_n(\mathbb{C})$ *and* f *be holomorphic on a simply connected open set* U *containing* $\sigma(A)$*. If for each* $\lambda \in \sigma(A)$ *the Taylor series*

$$f(z) = f(\lambda) + \sum_{k=1}^{\infty} a_{n,\lambda}(z-\lambda)^k, \quad a_{n,\lambda} = \frac{f^n(\lambda)}{n!}$$

converges in a neighborhood of λ*, then*

$$f(A) = \sum_{\lambda \in \sigma(A)} \left(f(\lambda)P_\lambda + \sum_{k=1}^{m_\lambda - 1} a_{k,\lambda} D_\lambda^k \right).$$

We now show how to use spectral calculus to compute $\exp(tA)$ for any $A \in M_n(\mathbb{F})$.

Corollary 4.2.9. *For* $A \in M_n(\mathbb{C})$ *and* $t \in \mathbb{R}$ *the matrix exponential* e^{tA} *can be computed as*

$$\exp(tA) = \sum_{\lambda \in \sigma(A)} e^{t\lambda} \left(P_\lambda + \sum_{k=1}^{m_\lambda - 1} \frac{t^k D_\lambda^k}{k!} \right). \tag{4.16}$$

Proof. Each $A \in M_n(\mathbb{F})$ has the spectral decomposition $A = \sum_{\lambda \in \sigma(A)} (\lambda P_\lambda + D_\lambda)$. For each $t \in \mathbb{R}$ the map $f(z) = \exp(tz)$ is entire, so the Spectral Mapping Theorem gives $\sigma(\exp(tA)) = \exp(t\sigma(A))$; that is, for each $\lambda \in \sigma(A)$ the quantity $e^{t\lambda}$ is an eigenvalue of $\exp(tA)$.

By the Mapping the Spectral Decomposition Theorem we have

$$\exp(tA) = \sum_{\lambda \in \sigma(A)} \left(e^{t\lambda} P_\lambda + \sum_{k=1}^{m_\lambda - 1} a_{k,\lambda} D_\lambda^k \right),$$

where

$$a_{0,\lambda} = f(\lambda) = e^{t\lambda}, \qquad a_{1,\lambda} = f'(\lambda) = te^{t\lambda},$$

$$a_{2,\lambda} = \frac{f''(\lambda)}{2!} = \frac{t^2 e^{t\lambda}}{2!}, \qquad a_{3,\lambda} = \frac{f'''(\lambda)}{3!} = \frac{t^3 e^{t\lambda}}{3!},$$

and so forth. Hence

$$\exp(tA) = \sum_{\lambda \in \sigma(A)} \left(e^{t\lambda} P_\lambda + \sum_{k=1}^{m_\lambda - 1} \frac{t^k e^{t\lambda}}{k!} D_\lambda^k \right) = \sum_{\lambda \in \sigma(A)} e^{t\lambda} \left(P_\lambda + \sum_{k=1}^{m_\lambda - 1} \frac{t^k D_\lambda^k}{k!} \right). \qquad \square$$

If $A \in M_n(\mathbb{R})$ and $\sigma(A) \subset \mathbb{R}$, then $\exp(tA) \in M_n(\mathbb{R})$ for all $t \in \mathbb{R}$.

Example 4.2.10. Find $\exp(tA)$ for

$$A = \begin{bmatrix} -1 & 11 & -3 \\ -2 & 8 & -1 \\ -1 & 5 & 0 \end{bmatrix}.$$

The resolvent is $R_A(z) = \frac{\operatorname{adj}(zI-A)}{\det(zI-A)}$, where $\det(zI - A)$ is the characteristic polynomial

$$\det(zI - A) = z^3 - 7z^2 + 16z - 12 = (z-2)^2(z-3),$$

and $\operatorname{adj}(zI - A)$ is the adjugate

$$\operatorname{adj}(zI - A) = \begin{bmatrix} z^2 - 8z + 5 & 11z - 15 & -3z + 13 \\ -2z + 1 & z^2 + z - 3 & -z + 5 \\ -z - 2 & 5z - 6 & z^2 - 7z + 14 \end{bmatrix}.$$

To compute the integrals in Definition 4.2.4, it is useful to compute the (nine) partial fraction decompositions (one for each entry of the resolvent):

$$R_A(z) = \frac{1}{z-2}\begin{bmatrix} 11 & -18 & -4 \\ 5 & -8 & -2 \\ 5 & -9 & -1 \end{bmatrix} + \frac{1}{(z-2)^2}\begin{bmatrix} 7 & -7 & -7 \\ 3 & -3 & -3 \\ 4 & -4 & -4 \end{bmatrix} + \frac{1}{z-3}\begin{bmatrix} -10 & 18 & 4 \\ -5 & 9 & 2 \\ -5 & 9 & 2 \end{bmatrix}.$$

This decomposition and (4.13) and (4.14) give

$$P_2 = \begin{bmatrix} 11 & -18 & -4 \\ 5 & -8 & -2 \\ 5 & -9 & -1 \end{bmatrix}, \quad D_2 = \begin{bmatrix} 7 & -7 & -7 \\ 3 & -3 & -3 \\ 4 & -4 & -4 \end{bmatrix}, \quad P_3 = \begin{bmatrix} -10 & 18 & 4 \\ -5 & 9 & 2 \\ -5 & 9 & 2 \end{bmatrix}.$$

We check that $D_2^2 = 0$, so $m_2 = 2$, and the spectral decomposition is

$$A = 2P_2 + D_2 + 3P_3.$$

Corollary 4.2.9 yields

$$\begin{aligned}
\exp(tA) &= \sum_{\lambda \in \sigma(A)} e^{t\lambda}\left(P_\lambda + \sum_{k=1}^{m_\lambda - 1} \frac{t^k}{k!} D_\lambda^k\right) \\
&= \exp(2t)(P_2 + tD_2) + \exp(3t)P_3 \\
&= \exp(2t)\left(\begin{bmatrix} 11 & -18 & -4 \\ 5 & -8 & -2 \\ 5 & -9 & -1 \end{bmatrix} + t\begin{bmatrix} 7 & -7 & -7 \\ 3 & -3 & -3 \\ 4 & -4 & -4 \end{bmatrix}\right) + \exp(3t)\begin{bmatrix} -10 & 18 & 4 \\ -5 & 9 & 2 \\ -5 & 9 & 2 \end{bmatrix} \\
&= \begin{bmatrix} (11+7t)e^{2t} - 10e^{3t} & -(18+7t)e^{2t} + 18e^{3t} & -(4+7t)e^{2t} + 4e^{3t} \\ (5+3t)e^{2t} - 5e^{3t} & -(8+3t)e^{2t} + 9e^{3t} & -(2+3t)e^{2t} + 2e^{3t} \\ (5+4t)e^{2t} - 5e^{3t} & -(9+4t)e^{2t} + 9e^{3t} & -(1+4t)e^{2t} + 2e^{3t} \end{bmatrix}.
\end{aligned}$$

As a basic check, evaluate this expression at $t = 0$, which gives the identity matrix. That's good because $\exp(0A) = I$ should always be true. A better (but messier) check is to verify that this matrix function $\Phi(t) = \exp(tA)$ satisfies $\frac{d}{dt}\Phi = A\Phi$.

Example 4.2.11. If

$$A = \begin{bmatrix} 1 & 3 & 0 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 7 \end{bmatrix},$$

then some work shows that

$$R(z) = \begin{bmatrix} (z-1)^{-1} & 3(z-1)^{-2} & 9(z-1)^{-3} & 27(z-1)^{-3}(z-7)^{-1} \\ 0 & (z-1)^{-1} & 3(z-1)^{-2} & 9(z-1)^{-2}(z-7)^{-1} \\ 0 & 0 & (z-1)^{-1} & 3(z-1)^{-1}(z-7)^{-1} \\ 0 & 0 & 0 & (z-7)^{-1} \end{bmatrix}$$

Computing partial fractions and using (4.13) and (4.14) gives

$$P_1 = \begin{bmatrix} 1 & 0 & 0 & -\frac{1}{8} \\ 0 & 1 & 0 & -\frac{1}{4} \\ 0 & 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad P_7 = \begin{bmatrix} 0 & 0 & 0 & \frac{1}{8} \\ 0 & 0 & 0 & \frac{1}{4} \\ 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$D_1 = \begin{bmatrix} 0 & 3 & 0 & -\frac{3}{4} \\ 0 & 0 & 3 & -\frac{3}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad D_1^2 = \begin{bmatrix} 0 & 0 & 9 & -\frac{9}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$D_1^3 = 0, \qquad D_7 = 0.$$

Thus $m_7 = 1$ and $m_1 = 3$, so that

$$\begin{aligned} e^{tA} &= e^t\left(P_1 + tD_1 + \frac{t^2 D_1^2}{2}\right) + e^{7t}P_7 \\ &= \begin{bmatrix} e^t & 3te^t & \frac{9}{2}t^2e^t & \left(-\frac{1}{8} - \frac{3}{4}t - \frac{9}{2}t^2\right)e^t + \frac{1}{8}e^{7t} \\ 0 & e^t & 3te^t & \left(-\frac{1}{4} - \frac{3}{2}t\right)e^t + \frac{1}{4}e^{7t} \\ 0 & 0 & e^t & -\frac{1}{2}e^t + \frac{1}{2}e^{7t} \\ 0 & 0 & 0 & e^{7t} \end{bmatrix}. \end{aligned}$$

Remark 4.2.12. As the examples illustrate, computing $R(z) = (zI - A)^{-1}$ is computationally costly. These methods have a lot of power for describing general properties of solutions, but for practical computation, the numerical techniques discussed earlier are generally more useful.

4.2.1 *Generalized eigenvectors and the matrix exponential

We now decompose the unique solution $\mathbf{x}(t) = \exp(tA)\mathbf{x}_0$ in terms of the generalized eigenspaces of A and how this relates to the eigenprojections P_λ and the eigennilpotents D_λ of A.

Suppose $A \in M_n(\mathbb{C})$ has k distinct eigenvalues $\lambda_1, \dots, \lambda_k$ with corresponding algebraic multiplicities $n_1, \dots, n_k$.

The generalized eigenspace of A corresponding to λ_j is

$$\mathscr{N}((A - \lambda_j I)^{n_j}).$$

The generalized eigenspaces of A form a direct sum decomposition

$$\mathbb{C}^n = \bigoplus_{j=1}^{k} \mathscr{N}((A - \lambda_j I)^{n_j}).$$

An initial condition $\mathbf{x}_0 \in \mathbb{C}^n$ then has a unique decomposition

$$\mathbf{x}_0 = I\mathbf{x}_0 = \sum_{j=1}^{k} P_{\lambda_j}(\mathbf{x}_0) = \sum_{j=1}^{k} \mathbf{v}_j,$$

where (as we learned in Volume 1) $P_{\lambda_j}^2 = P_{\lambda_j}$, $I = \sum_{j=1}^{k} P_{\lambda_j}$, $\mathscr{R}(P_{\lambda_j}) = \mathscr{N}((A - \lambda_j I)^{n_j})$, and $m_j \leq \dim(\mathscr{R}(P_{\lambda_j})) = n_j$.

Then

$$\begin{aligned}
\mathbf{x}(t) = \exp(tA)\mathbf{x}_0 = \exp(tA)\sum_{j=1}^{k}\mathbf{v}_j &= \sum_{j=1}^{k} \exp(tA)\mathbf{v}_j \\
&= \sum_{j=1}^{k} \exp\big(t(A + \lambda_j I - \lambda_j I)\big)\mathbf{v}_j \\
&= \sum_{j=1}^{k} \exp(t\lambda_j I)\exp\big(t(A - \lambda_j I)\big)\mathbf{v}_j,
\end{aligned}$$

where we have used that $(t\lambda_j I)(t(A - \lambda_j I)) = (t(A - \lambda_j I))(t\lambda_j I)$ implies the third equality (a scalar multiple of the identity matrix commutes with any matrix).

For each j the resolvent of the matrix $t\lambda_j I$ is

$$R_{t\lambda_j I}(z) = (zI - t\lambda_j I)^{-1} = ((z - t\lambda_j)I)^{-1} = \frac{1}{z - t\lambda_j} I.$$

Thus $P_{t\lambda_j} = I$ and so

$$\exp(t\lambda_j I) = \exp(t\lambda_j) I.$$

Returning to the calculation we have

$$\begin{aligned}\mathbf{x}(t) &= \sum_{j=1}^{k} \exp(t\lambda_j I) \exp\big(t(A-\lambda_j I)\big)\mathbf{v}_j \\ &= \sum_{j=1}^{k} \exp(t\lambda_j) I \exp\big(t(A-\lambda_j I)\big)\mathbf{v}_j \\ &= \sum_{j=1}^{k} \exp(t\lambda_j) \exp\big(t(A-\lambda_j I)\big)\mathbf{v}_j .\end{aligned}$$

Since

$$\exp\big(t(A-\lambda_j I)\big) = \sum_{k=0}^{\infty} \frac{t^k (A-\lambda_j I)^k}{k!}$$

and $\mathbf{v}_j \in \mathscr{N}((A-\lambda_j I)^{n_j})$ there holds for each $j = 1, 2, \ldots, k$,

$$\exp\big(t(A-\lambda_j I)\big)\mathbf{v}_j = \left(\sum_{k=0}^{\infty} \frac{t^k (A-\lambda_j I)^k}{k!}\right)\mathbf{v}_j = \left(\sum_{k=0}^{n_j-1} \frac{t^k (A-\lambda_j I)^k}{k!}\right)\mathbf{v}_j .$$

Thus the decomposition of the unique solution is

$$\begin{aligned}\mathbf{x}(t) &= \sum_{j=1}^{k} \exp(t\lambda_j) \left(\sum_{k=0}^{n_j-1} \frac{t^k (A-\lambda_j I)^k}{k!}\right)\mathbf{v}_j \\ &= \sum_{j=1}^{k} \exp(t\lambda_j) \left(I + \sum_{k=1}^{n_j-1} \frac{t^k (A-\lambda_j I)^k}{k!}\right) P_{\lambda_j}(\mathbf{x}_0) \\ &= \sum_{j=1}^{k} \exp(t\lambda_j) \left(P_{\lambda_j} + \sum_{k=1}^{n_j-1} \frac{t^k (A-\lambda_j I)^k P_{\lambda_j}}{k!}\right)\mathbf{x}_0 \\ &= \sum_{j=1}^{k} \exp(t\lambda_j) \left(P_{\lambda_j} + \sum_{k=1}^{n_j-1} \frac{t^k (A-\lambda_j I)^k P_{\lambda_j}^k}{k!}\right)\mathbf{x}_0 \\ &= \sum_{j=1}^{k} \exp(t\lambda_j) \left(P_{\lambda_j} + \sum_{k=1}^{n_j-1} \frac{t^k D_\lambda^k}{k!}\right)\mathbf{x}_0 ,\end{aligned}$$

where we have used $P_{\lambda_j}^k = P_{\lambda_j}^2 = P_{\lambda_j}$ and $D_{\lambda_j} = (A-\lambda_j I)P_{\lambda_j}$.

Comparing this with the result from spectral calculus for the solution of the IVP,

$$\exp(tA)\mathbf{x}_0 = \sum_{\lambda\in\sigma(A)} e^{t\lambda}\left(P_\lambda + \sum_{k=1}^{m_\lambda-1} \frac{t^k D_\lambda^k}{k!}\right)\mathbf{x}_0 ,$$

and noting the $m_j \leq \dim(\mathscr{R}(P_{\lambda_j})) = n_j$, we see that we have agreement.

4.3 Nonautonomous Homogeneous Linear Systems

In this section we consider a linear homogeneous equations that are not autonomous, that is, equations of the form

$$\dot{\mathbf{x}} = A(t)\mathbf{x}, \tag{4.17}$$

where $A(t)$ need not be constant. The exponential solution from the autonomous case no longer works in this setting.

4.3.1 Nonautonomous Homogeneous Linear Systems

For a *scalar* nonautonomous linear homogeneous equation $\dot{x}(t) = a(t)x$, we have already seen that the general solution can be written as $x(t) = \exp\left(\int_{t_0}^t a(s)\,ds\right)$. Moreover, in the autonomous case (4.7), the principal fundamental matrix solution can be written as $\Phi(t) = e^{tA} = \exp(\int_0^t A\,ds)$. You might also guess that the exponential

$$\Phi(t) = \exp\left(\int_{t_0}^t A(s)\,ds\right)$$

should be the principal fundamental matrix solution for the general nonautonomous case, because that works in the scalar case (and in the autonomous case). But it does not work in higher dimensions because the matrices involved don't usually commute like we need them to. Specifically, we have

$$\frac{d}{dt}\Phi(t) = \exp\left(\int_{t_0}^t A(s)\,ds\right)A(t) \neq A(t)\exp\left(\int_{t_0}^t A(s)\,ds\right). \tag{4.18}$$

In special cases the two sides might be equal, but not for general $A(t)$. One special case where the equality holds is when $A(s)$ commutes with $A(t)$ for all $s, t \in \mathbb{R}$, as shown in the next proposition.

Proposition 4.3.1. *If $A(s)$ commutes with $A(t)$ for all $s, t \in \mathbb{R}$, that is, $A(s)A(t) = A(t)A(s)$ for all $s, t \in \mathbb{R}$, then*

$$\Phi(t) = \exp\left(\int_{t_0}^t A(s)\,ds\right) \tag{4.19}$$

is the principal fundamental matrix solution of $\dot{\mathbf{x}} = A(t)\mathbf{x}$.

Proof. First observe that $\Phi(t_0) = \exp(0I) = I$, so if we can show that Φ satisfies

$$\frac{d}{dt}\Phi(t) = A(t)\Phi(t), \tag{4.20}$$

then Φ is the principal fundamental matrix solution. Let

$$B(t) = \int_{t_0}^t A(s)\,ds,$$

so that $\Phi(t) = \exp(B(t))$. It suffices to show that

$$\exp(B(t))A(t) = A(t)\exp(B(t))$$

because in that case $\frac{d}{dt}\Phi(t) = \exp(B(t))A(t) = A(t)\exp(B(t) = A(t)\Phi(t)$, as desired.

Observe that

$$B(t)A(t) = \int_{t_0}^{t} A(s)A(t)\,ds = A(t)\int_{t_0}^{t} A(s)\,ds = A(t)B(t),$$

which gives

$$\begin{aligned}\exp\left(B(t)\right)A(t) &= \sum_{k=0}^{\infty}\frac{(B(s))^k A(t)}{k!} = \sum_{k=0}^{\infty}\frac{A(t)(B(s))^k}{k!}\\ &= A(t)\exp\left(B(t)\right),\end{aligned}$$

as required. $\square$

Unexample 4.3.2. The matrix from Example 4.1.6,

$$A(t) = \begin{bmatrix} -1+\frac{3}{2}\cos^2 t & 1-\frac{3}{2}\sin t\cos t \\ -1-\frac{3}{2}\sin t\cos t & -1+\frac{3}{2}\sin^2 t\end{bmatrix},$$

doesn't satisfy $A(s)A(t) = A(t)A(s)$ for all $s,t\in\mathbb{R}$ because for

$$A(0) = \begin{bmatrix} \frac{1}{2} & 1 \\ -1 & -1\end{bmatrix} \quad\text{and}\quad A(\tfrac{\pi}{2}) = \begin{bmatrix} -1 & 1 \\ -1 & \frac{1}{2}\end{bmatrix}$$

we have

$$A(0)A(\tfrac{\pi}{2}) = \begin{bmatrix} \frac{1}{2} & 1 \\ -1 & -1\end{bmatrix}\begin{bmatrix} -1 & 1 \\ -1 & \frac{1}{2}\end{bmatrix} = \begin{bmatrix} -\frac{3}{2} & 1 \\ 2 & -\frac{3}{2}\end{bmatrix}$$

and

$$A(\tfrac{\pi}{2})A(0) = \begin{bmatrix} -1 & 1 \\ -1 & \frac{1}{2}\end{bmatrix}\begin{bmatrix} \frac{1}{2} & 1 \\ -1 & -1\end{bmatrix} = \begin{bmatrix} -\frac{3}{2} & -2 \\ -1 & -\frac{3}{2}\end{bmatrix}.$$

4.3.2 The Abel–Jacobi–Liouville Formula

In general the integral (4.19) is not the fundamental matrix solution for $\dot{\mathbf{x}} = A(t)\mathbf{x}$ (as in the case of Unexample 4.3.2). But the determinant $\det(\Phi(t))$ of the principal fundamental matrix solution can be computed as an integral of the trace of $A(t)$ (the determinant is called the *Wronskian*).

Theorem 4.3.3 (Abel–Jacobi–Liouville (AJL) Formula). *If $\Phi(t)$ is a fundamental matrix of a linear homogeneous system, that is, $\dot{\Phi}(t) = A(t)\Phi(t)$, then*

$$\det\Phi(t) = \det\Phi(t_0)\exp\left(\int_{t_0}^{t}\operatorname{tr} A(s)\,ds\right). \tag{4.21}$$

The proof is given after a few examples.

Example 4.3.4. Consider the matrix $A = \left[\begin{smallmatrix} 1 & 4 \\ 1 & 1 \end{smallmatrix}\right]$ and corresponding principal fundamental matrix

$$\Phi(t) = \frac{1}{4}\begin{bmatrix} 2(e^{3t}+e^{-t}) & 4(e^{3t}-e^{-t}) \\ e^{3t}-e^{-t} & 2(e^{3t}+e^{-t}) \end{bmatrix}.$$

of in Example 4.1.10. A direct computation shows that $\det \Phi(t) = e^{2t}$; and $\operatorname{tr}(A) = 2$, so Theorem 4.3.3 gives an alternative route to the same conclusion, that $\det \Phi(t) = e^{2t}$.

so

Example 4.3.5. Consider the matrix $A = \left[\begin{smallmatrix} -1 & 1 \\ -1 & 1 \end{smallmatrix}\right]$ and corresponding principal fundamental matrix

$$\Phi(t) = \begin{bmatrix} 1-t & t \\ -t & 1+t \end{bmatrix}.$$

of Example 4.1.13. A direct computation gives $\det \Phi(t) = 1$; and $\operatorname{tr}(A) = 0$, Theorem 4.3.3 gives that same answer: $\det \Phi(t) = e^0 = 1$.

Example 4.3.6. Abel–Jacobi–Liouville can sometimes be used to find new solutions from known solutions. Assume we already know the solution $\mathbf{x} = (1, t)$ of the system

$$\dot{\mathbf{x}} = \begin{bmatrix} 1 & -\frac{1}{t} \\ 1+t & -1 \end{bmatrix}\mathbf{x}$$

on the domain $t \in (0, \infty)$. The solution $\mathbf{x}$ corresponds to one column of a fundamental matrix solution Φ (not the principal fundamental matrix solution). We wish to find another column of Φ; denote it by $\mathbf{z}(t) = (z_1(t), z_2(t))$ so that

$$\Phi = \begin{bmatrix} z_1(t) & 1 \\ z_2(t) & t \end{bmatrix}.$$

The AJL formula guarantees that

$$\det(\Phi(t)) = \det(\Phi(t_0)) \exp(\int_{t_0}^{t} \operatorname{tr}(A(s))\, ds = \det(\Phi(t_0))$$

for some choice of t_0, because the trace $\operatorname{tr}(A(s))$ vanishes for all s. This implies that $\det(\Phi(t))$ is constant for all t in the domain of definition, which gives

$$c = tz_1(t) - z_2(t) \tag{4.22}$$

for some constant c. The differential equation for $\mathbf{z}$ gives

$$\begin{bmatrix} \dot{z}_1(t) \\ \dot{z}_2(t) \end{bmatrix} = A\mathbf{z} = \begin{bmatrix} z_1(t) - \frac{z_2}{t} \\ z_1(t)(1+t) - z_2(t) \end{bmatrix}$$
$$= \begin{bmatrix} \frac{tz_1(t)-z_2}{t} \\ z_1(t)(1+t) - z_2(t) \end{bmatrix} = \begin{bmatrix} \frac{c}{t} \\ z_1(t)(1+t) - z_2(t) \end{bmatrix}.$$

In particular, we have $\dot{z}_1(t) = \frac{c}{t}$. Integrating gives $z_1(t) = c\log(t) + d$ for some constant d. Substituting back into (4.22) gives $z_2(t) = ct\ln(t) + dt - c$. This is the general solution to the original system. For any choice of initial value $\mathbf{v}$ except $(1, t_0)$, the corresponding solution gives a column of the corresponding fundamental matrix solution Φ.

Proof. (of Theorem 4.3.3) Denote the entries of $\Phi(t)$ by $\phi_{ij}(t)$, that is,

$$\Phi(t) = \begin{bmatrix} \phi_{11}(t) & \phi_{12}(t) & \cdots & \phi_{1n}(t) \\ \phi_{21}(t) & \phi_{22}(t) & \cdots & \phi_{2n}(t) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{n1}(t) & \phi_{n2}(t) & \cdots & \phi_{nn}(t) \end{bmatrix}.$$

Let E_{ij} be the matrix with zero entries except 1 in the (i,j).

The product $E_{jj}A$ is the matrix whose entries are zero except for the jth row which is the jth row of A. Starting with Exercise 4.13, we see that

$$\frac{d}{dt}\left(\det \Phi(t)\right) = \det \begin{bmatrix} \dot{\phi}_{11}(t) & \cdots & \dot{\phi}_{1n}(t) \\ \vdots & & \vdots \\ \phi_{n1}(t) & \cdots & \phi_{nn}(t) \end{bmatrix} + \cdots + \det \begin{bmatrix} \phi_{11}(t) & \phi_{12}(t) & \cdots & \phi_{1n}(t) \\ \dot{\phi}_{21}(t) & \dot{\phi}_{22}(t) & \cdots & \dot{\phi}_{2n}(t) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{n1}(t) & \phi_{n2}(t) & \cdots & \phi_{nn}(t) \end{bmatrix}$$
$$+ \cdots + \det \begin{bmatrix} \phi_{11}(t) & \phi_{12}(t) & \cdots & \phi_{1n}(t) \\ \phi_{21}(t) & \phi_{22}(t) & \cdots & \phi_{2n}(t) \\ \vdots & \vdots & \ddots & \vdots \\ \dot{\phi}_{n1}(t) & \dot{\phi}_{n2}(t) & \cdots & \dot{\phi}_{nn}(t) \end{bmatrix}.$$

Recognize that each of these determinants can be expressed as

$$\det\left[(I - E_{jj})\Phi(t) + E_{jj}\dot{\Phi}(t)\right].$$

Thus

$$\begin{aligned}\frac{d}{dt}\det(\Phi(t)) &= \sum_{j=1}^{n}\det\big[(I-E_{jj})\Phi(t)+E_{jj}\dot{\Phi}(t)\big]\\ &= \sum_{j=1}^{n}\det\big[I\Phi(t)-E_{jj}\Phi(t)+E_{jj}A(t)\Phi(t)\big]\\ &= \sum_{j=1}^{n}\det\big[(I-E_{jj}+E_{jj}A)\Phi(t)\big]\\ &= \sum_{j=1}^{n}\det(I-E_{jj}+E_{jj}A)\det(\Phi(t))\\ &= \left(\sum_{j=1}^{n}\det(I-E_{jj}+E_{jj}A)\right)\det(\Phi(t)).\end{aligned}$$

The matrix

$$I-E_{jj}+E_{jj}A=\begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 & 0\\ 0 & 1 & \cdots & 0 & \cdots & 0 & 0\\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots\\ a_{j1} & a_{j2} & \cdots & a_{jj} & \cdots & a_{j,n-1} & a_{jn}\\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & \cdots & 0 & \cdots & 1 & 0\\ 0 & 0 & \cdots & 0 & \cdots & 0 & 1\end{bmatrix}$$

is row equivalent to an upper triangle matrix by the row operation of adding a scalar multiple of one row to another row (which doesn't change the determinant) which removes the entries a_{jl} for $l=1,2,\ldots,j-1$. This shows that

$$\det\big[I-E_{jj}+E_{jj}A\big]=a_{jj}.$$

Hence we obtain

$$\begin{aligned}\frac{d}{dt}\det(\Phi(t)) &= \left(\sum_{j=1}^{n}\det(I-E_{jj}+E_{jj}A)\right)\det(\Phi(t))\\ &= \left(\sum_{j=1}^{n}a_{jj}\right)\det(\Phi(t))\\ &= \big(\operatorname{tr}A(t)\big)\det(\Phi(t)).\end{aligned}$$

Solving this first-order separable equation over the interval with endpoints t_0 and t gives

$$\det(\Phi(t))=\det(\Phi(t_0))\exp\left(\int_{t_0}^{t}\operatorname{tr}A(s)\,ds\right),$$

giving the result □

Example 4.3.7. Recall that

$$\Phi(t) = \begin{bmatrix} e^{t/2}\cos t & e^{-t}\sin t \\ -e^{t/2}\sin t & e^{-t}\cos t \end{bmatrix}$$

is a principle fundamental matrix solution for the non-autonomous linear homogeneous system $\dot{\mathbf{x}} = A(t)\mathbf{x}$ where

$$A(t) = \begin{bmatrix} -1 + \frac{3}{2}\cos^2 t & 1 - \frac{3}{2}\sin t\cos t \\ -1 - \frac{3}{2}\sin t\cos t & -1 + \frac{3}{2}\sin^2 t \end{bmatrix}.$$

Here

$$\det(\Phi(0)) = 1$$

and the AJL formula gives

$$\det(\Phi(t)) = \det(\Phi(0))\exp\left(\int_0^t \operatorname{tr} A(s)\,ds\right) = \exp\left(\int_0^t -\frac{1}{2}\,ds\right) = e^{-t/2}.$$

We can verify this by directly computing the determinant:

$$\det(\Phi(t)) = \det\begin{bmatrix} e^{t/2}\cos t & e^{-t}\sin t \\ -e^{t/2}\sin t & e^{-t}\cos t \end{bmatrix} = e^{-t/2}\cos^2 t + e^{-t/2}\sin^2 t = e^{-t/2}.$$

As a corollary of the AJL formula, we get another proof that the fundamental matrix solution is nonsingular (see Proposition 4.1.4).

Corollary 4.3.8. *Any fundamental matrix solution $\Phi(t)$ is nonsingular for all $t \in \mathbb{R}$.*

Proof. By assumption a fundamental matrix solution has $\det(\Phi(t_0)) \neq 0$. The AJL formula gives $\det(\Phi(t)) \neq 0$ for all $t \in \mathbb{R}$. □

Corollary 4.3.9 (Jacobi's Formula). *For any constant matrix $A \in M_n(\mathbb{F})$ we have*

$$\det(\exp(tA)) = \exp(t\operatorname{tr} A), \quad \forall t \in \mathbb{R}.$$

Proof. When A is a constant matrix, the matrix function $\exp(tA)$ is a principal fundamental matrix solution of $\dot{\mathbf{x}} = A\mathbf{x}$ with $t_0 = 0$. Applying the AJL formula to $\Phi(t) = \exp(tA)$ gives

$$\begin{aligned} \det(\exp(tA)) &= \det(\exp(0A))\exp\left(\int_0^t \operatorname{tr} A\,ds\right) \\ &= \det(I)\exp\left(\operatorname{tr} A\int_0^t ds\right) = \exp(t\operatorname{tr} A) \end{aligned}$$ □

Corollary 4.3.10. *The equality*

$$\frac{d}{dt}\det\big(I+tA\big)\Big|_{t=0} = \operatorname{tr} A$$

holds for any $A \in M_n(\mathbb{F})$.

Proof. Differentiating the exponential of tA and evaluating at $t=0$ gives

$$\begin{aligned}
\frac{d}{dt}\det(\exp(tA))\Big|_{t=0} &= (D_B \det(B))\Big|_{B=I}\frac{d}{dt}\exp(tA)\Big|_{t=0} \\
&= (D_B \det(B))\Big|_{B=I} A \\
&= (D_B \det(B))\Big|_{B=I}\frac{d}{dt}(I+tA)\Big|_{t=0} \\
&= \frac{d}{dt}\det(I+tA)\Big|_{t=0}.
\end{aligned} \tag{4.23}$$

Here D_B indicates the derivative with respect to the matrix B. Differentiating the exponential of the trace gives

$$\begin{aligned}
\frac{d}{dt}(\exp(t\operatorname{tr}(A)))\Big|_{t=0} &= \operatorname{tr}(A)\exp(t\operatorname{tr}(A))\Big|_{t=0} \\
&= \operatorname{tr}(A).
\end{aligned} \tag{4.24}$$

Differentiating Jacobi's Formula gives

$$\frac{d}{dt}\det(\exp(tA))\Big|_{t=0} = \frac{d}{dt}\exp(t\operatorname{tr} A)\Big|_{t=0},$$

which implies that the final expressions in (4.23) and (4.24) are also equal. ∎

4.4 Nonhomogeneous Systems and Numerical Approximation of Linear Systems

4.4.1 Nonhomogeneous Linear Systems

In this section, we show how to extend the homogeneous linear case to the nonhomogeneous linear case, where the solution $\mathbf{x}(t)$ satisfies the equation

$$\begin{aligned}
\dot{\mathbf{x}}(t) &= A(t)\mathbf{x}(t) + \mathbf{b}(t) \\
\mathbf{x}(t_0) &= \mathbf{x}_0.
\end{aligned} \tag{4.25}$$

Throughout this section, assume that the corresponding homogeneous linear system

$$\dot{\mathbf{x}}(t) = A(t)\mathbf{x}(t) \tag{4.26a}$$

$$\mathbf{x}(t_0) = \mathbf{x}_0, \tag{4.26b}$$

has the fundamental matrix solution $\Phi(t)$ satisfying $\Phi(t_0) = I$ and $\mathbf{x}(t) = \Phi(t)\mathbf{x}_0$.

The first important result guarantees that the maximal interval of existence for any solution of (4.25) is the full real line. This is a straightforward extension of Theorem 4.1.1. Much of the proof below is nearly identical to the proof of Theorem 4.1.1. We have retained these details because they illustrate some ideas that are relevant to topics covered later in this book.

Theorem 4.4.1. *The interval of existence of the solution of the IVP* (4.25) *is* $\mathbb{R}$.

Proof. The vector field

$$\mathbf{f}(\mathbf{x},t) = A(t)\mathbf{x} + \mathbf{b}(t),\ \ \mathbf{x} \in \mathbb{R}^n,\ \ t \in \mathbb{R}$$

satisfies

$$\begin{aligned}
\|\mathbf{f}(\mathbf{x}_1,t) - \mathbf{f}(\mathbf{x}_2,t)\| &= \|A(t)\mathbf{x}_1 + \mathbf{b}(t) - A(t)\mathbf{x}_2 - \mathbf{b}(t)\| \\
&= \|A(t)(\mathbf{x}_1 - \mathbf{x}_2)\| \\
&\le \|A(t)\|\,\|\mathbf{x}_1 - \mathbf{x}_2\| \\
&\le \sup_{t\in\mathbb{R}} \|A(t)\|\,\|\mathbf{x}_1 - \mathbf{x}_2\|.
\end{aligned}$$

Continuity of $A(t)$ and $\mathbf{b}(t)$ on $\mathbb{R}$ implies continuity of $\mathbf{f}(\mathbf{x},t)$ on $U \times I$ for U open in $\mathbb{R}^n$ and I an open bounded interval in $\mathbb{R}$, and

$$L = \sup_{t\in\overline{I}} \|A(t)\| < \infty$$

gives the uniform Lipschitz property in the variable $\mathbf{x}$ for $\mathbf{f}$.

By the Existence and Uniqueness Theorem (Theorem 2.2.2) there exists an open interval I containing t_0 and a C^1 function $\mathbf{x} : I \to \mathbb{F}^n$ such that

$$\begin{aligned}
\dot{\mathbf{x}}(t) &= A(t)\mathbf{x}(t) + \mathbf{b}(t) \quad \text{for} \quad \text{all} \quad t \in I, \\
\mathbf{x}(t_0) &= \mathbf{x}_0.
\end{aligned}$$

By the Unique Extension Theorem (Theorem 2.2.7) this solution extends uniquely to a maximal open interval of existence J.

If the right endpoint β of J is not ∞ (similar argument for the left endpoint), then the Finite Time Blowup Theorem (Theorem 2.2.9) guarantees

$$\lim_{t\to\beta} \|\mathbf{x}(t)\| \to \infty.$$

Applying the Fundamental Theorem of Calculus to the IVP gives

$$\mathbf{x}(t) = \mathbf{x}(t_0) + \int_{t_0}^{t} \big(A(s)\mathbf{x}(s) + \mathbf{b}(s)\big)\, ds,\ \ s \in J.$$

Applying the norm to both sides of this gives

$$\begin{aligned}
\|\mathbf{x}(t)\| &\le \|\mathbf{x}_0\| + \int_{t_0}^{t} \|A(s)\mathbf{x}(s) + \mathbf{b}(s)\|\, ds,\ \ s \in [t_0,\beta) \\
&\le \|\mathbf{x}_0\| + \int_{t_0}^{t} \|A(s)\mathbf{x}(s)\|\, ds + \int_{t_0}^{t} \|\mathbf{b}(s)\|\, ds \\
&\le \|\mathbf{x}_0\| + \int_{t_0}^{t} \|A(s)\mathbf{x}(s)\|\, ds + \int_{t_0}^{\beta} \|\mathbf{b}(s)\|\, ds
\end{aligned}$$

Continuity of $\mathbf{b}$ on the compact interval $[t_0, \beta]$ implies that

$$C = \int_{t_0}^{\beta} \|\mathbf{b}(s)\| \, ds < \infty.$$

Thus we obtain

$$\|\mathbf{x}(t)\| \leq \|\mathbf{x}_0\| + \int_{t_0}^{t} \|A(s)\mathbf{x}(s)\| \, ds + C = \|\mathbf{x}_0\| + C + \int_{t_0}^{t} \|A(s)\mathbf{x}(s)\| \, ds. \quad (4.27)$$

There is a generalization of Gronwall's Inequality that states that if for $a < b$ there are continuous nonnegative functions α, ϕ, and ψ on $[a, b]$ for which α is C^1 on (a, b) with nonnegative derivative and

$$\phi(t) \leq \alpha(t) + \int_a^t \psi(s)\phi(s) \, ds \text{ for all } t \in [a, b],$$

then

$$\phi(t) \leq \alpha(t) \exp\left(\int_a^t \psi(s) \, ds \right) \quad \text{for all } t \in [a, b].$$

Applying this generalization of Gronwall's inequality to the inequality (4.27) with $\alpha(t) = \|\mathbf{x}_0\| + C$, $\phi(t) = \|\mathbf{x}(t)\|$, and $\psi(t) = \|A(s)\|$, we obtain

$$\|\mathbf{x}(t)\| \leq (\|\mathbf{x}_0\| + C) \exp\left(\int_{t_0}^{t} \|A(s)\| \, ds \right) \quad \text{for all} \quad t \in [t_0, \beta].$$

This implies that $\|\mathbf{x}(t)\|$ is bounded as $t \to \beta$, a contradiction. $\square$

4.4.2 Duhamel's Principle

One of the main ways to solve the IVP (4.25) is a special case of a method called *Duhamel's Principle*, which applies to a class of IVPs that includes linear nonhomogeneous IVPs as a special case. When Duhamel's principle is applied to linear IVPs, it is often called *variation of parameters*.

Theorem 4.4.2 (Duhamel's Principle). *Consider the IVP*

$$\dot{\mathbf{x}} = A\mathbf{x} + Q(\mathbf{x}, t), \qquad \mathbf{x}(0) = \mathbf{x}_0, \quad (4.28)$$

where $Q(\mathbf{x}, t)$ is a (potentially) nonlinear term. Let $\Phi(t)$ be the principle fundamental matrix solution for the corresponding homogeneous system $\dot{\mathbf{y}} = A\mathbf{y}$. The solution of the full nonlinear system (4.28) *satisfies the implicit equation*

$$\mathbf{x}(t) = \Phi(t)\mathbf{x}(0) + \Phi(t) \int_0^t \Phi(s)^{-1} Q(\mathbf{x}(s), s) \, ds$$

for t in the maximal open interval I of existence for $\mathbf{x}$ (where I contains 0).

Proof. We use the ansatz[24] $\mathbf{x}(t) = \Phi(t)\mathbf{h}(t)$ for a C^1 function $\mathbf{h}$. Differentiation of the ansatz gives

$$\dot{\mathbf{x}}(t) = \dot{\Phi}(t)\mathbf{h}(t) + \Phi(t)\dot{\mathbf{h}}(t).$$

Substitution of the ansatz into (4.28) gives

$$\dot{\Phi}(t)\mathbf{h}(t) + \Phi(t)\dot{\mathbf{h}}(t) = A\Phi(t)\mathbf{h}(t) + Q(\mathbf{x}(t), t).$$

Since $\Phi(t)$ is a fundamental matrix solution then $A\Phi(t) = \dot{\Phi}(t)$ and so

$$\dot{\Phi}(t)\mathbf{h}(t) + \Phi(t)\dot{\mathbf{h}}(t) = \dot{\Phi}(t)\mathbf{h}(t) + Q(\mathbf{x}(t), t).$$

Cancellation of the common term gives

$$\Phi(t)\dot{\mathbf{h}}(t) = Q(\mathbf{x}(t), t).$$

Since $\det(\Phi(0)) = 1$, Abel's Theorem implies that $\det(\Phi(t)) \neq 0$ for all $t \in I$, so that $\Phi(t)$ is invertible for all $t \in I$.

Hence we obtain

$$\dot{\mathbf{h}}(t) = \Phi(t)^{-1}Q(\mathbf{x}(t)\mathbf{t}).$$

Integrating this over the interval with endpoints 0 and $t \in I$ gives

$$\mathbf{h}(t) = \mathbf{h}(0) + \int_0^t \Phi(s)^{-1}Q(\mathbf{x}(s), s)\, ds,\ t \in I.$$

Putting this expression into the ansatz gives

$$\mathbf{x}(t) = \Phi(t)\mathbf{h}(0) + \Phi(t)\int_0^t \Phi(s)^{-1}Q(\mathbf{x}(s), s)\, ds,\ t \in I.$$

Setting $t = 0$ in this implicit expression that $\mathbf{x}(t)$ satisfies gives

$$\mathbf{x}(0) = \Phi(0)\mathbf{h}(0).$$

Since $\Phi(0) = I$ we obtain $\mathbf{h}(0) = \mathbf{x}(0)$.

Thus the solution $\mathbf{x}(t)$ satisfies the implicit expression

$$\mathbf{x}(t) = \Phi(t)\mathbf{x}(0) + \Phi(t)\int_0^t \Phi(s)^{-1}Q(\mathbf{x}(s), s)\, ds,\ t \in I.$$

This gives the result. □

For the purposes of the current discussion with regard to nonhomogeneous linear systems we don't need to consider a truly nonlinear remainder $Q(\mathbf{x}(t), t)$. Instead we restrict our attention to the case when $Q(\mathbf{x}(t), t) = \mathbf{b}(t)$ which leads to the following corollary.

[24] In the setting of an IVP there is only one solution, so if the ansatz leads you to a solution, then you have found that one solution. But if ansatz doesn't work out (it either didn't help you solve the problem, or there is no solution that satisfies the ansatz), then that does not mean that there is no solution—it just means your ansatz was not a good one.

Corollary 4.4.3 (Linear Duhamel, a.k.a. Variation of Parameters). *The solution of* (4.25) *is given by*

$$\mathbf{x}(t) = \Phi(t)\mathbf{x}_0 + \Phi(t)\int_{t_0}^{t} \Phi(s)^{-1}\,\mathbf{b}(s)\,ds \tag{4.29}$$

where $\Phi(t)$ *is a fundamental matrix for* (4.26).

Remark 4.4.4. Another way of looking at this (that is commonly advocated by introductory ODE textbooks) is to find a solution to the homogeneous problem where $\mathbf{b}(t) = 0$, and then to combine this with a particular solution that likely has the same form (or similar at least) as $\mathbf{b}(t)$ itself. For instance, if $\mathbf{b}(t)$ is a trigonometric function then the particular solution would be a trigonometric function of the same type. We do not pursue this approach here, as Duhamel's Principle always suffices and there is no need to resort to this method of inspired guessing. In this situation, the fundamental matrix $\Phi(t)$ is the homogeneous solution.

Remark 4.4.5. For autonomous systems, since the fundamental matrix solution is $\Phi(t) = e^{At}$, the linear case of Duhamel's principle (Corollary 4.4.3) implies that the solution is

$$\mathbf{x}(t) = e^{A(t-t_0)}\mathbf{x}_0 + \int_{t_0}^{t} e^{A(t-s)}\mathbf{b}(s)\,ds. \tag{4.30}$$

Example 4.4.6. Consider the ODE

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\mathbf{x} + \begin{bmatrix} t \\ 1 \end{bmatrix} \tag{4.31}$$

subject to

$$\mathbf{x}(0) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

A by-now-routine calculation shows that the exponential function for the matrix A is

$$e^{tA} = \begin{bmatrix} e^t & te^t \\ 0 & e^t \end{bmatrix}.$$

The linear case of Duhamel's principle (Corollary 4.4.3) implies the particular solution we are seeking is given by

$$\begin{aligned}
\mathbf{x}(t) &= \begin{bmatrix} e^t & te^t \\ 0 & e^t \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \int_0^t \begin{bmatrix} e^{t-s} & (t-s)e^{t-s} \\ 0 & e^{t-s} \end{bmatrix} \begin{bmatrix} s \\ 1 \end{bmatrix} ds \\
&= \begin{bmatrix} (1+t)e^t \\ e^t \end{bmatrix} + \int_0^t \begin{bmatrix} te^{t-s} \\ e^{t-s} \end{bmatrix} ds \\
&= \begin{bmatrix} (1+t)e^t \\ e^t \end{bmatrix} + \begin{bmatrix} -t \\ -1 \end{bmatrix} - \begin{bmatrix} -te^t \\ -e^t \end{bmatrix} \\
&= \begin{bmatrix} 2te^t + e^t - t \\ 2e^t - 1 \end{bmatrix}
\end{aligned}$$

Example 4.4.7. Find the solution of the nonhomogeneous linear system

$$\dot{\mathbf{x}} = A\mathbf{x} + \mathbf{b}(t),\ \mathbf{x}(0) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \text{where } A = \begin{bmatrix} 1 & 1 \\ 4 & 1 \end{bmatrix} \text{ and } \mathbf{b}(t) = \begin{bmatrix} 2e^t \\ -e^t \end{bmatrix}.$$

The principal fundamental matrix for the associated linear homogeneous system is

$$\Phi(t) = \begin{bmatrix} \frac{1}{2}e^{3t} + \frac{1}{2}e^{-t} & \frac{1}{4}e^{3t} - \frac{1}{4}e^{-t} \\ e^{3t} - e^{-t} & \frac{1}{2}e^{3t} + \frac{1}{2}e^{-t} \end{bmatrix}.$$

The inverse of this is

$$\begin{aligned}
\Phi^{-1}(t) &= \frac{1}{e^{2t}} \begin{bmatrix} \frac{1}{2}e^{3t} + \frac{1}{2}e^{-t} & -\frac{1}{4}e^{3t} + \frac{1}{4}e^{-t} \\ -e^{3t} + e^{-t} & \frac{1}{2}e^{3t} + \frac{1}{2}e^{-t} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{2}e^{t} + \frac{1}{2}e^{-3t} & -\frac{1}{4}e^{t} + \frac{1}{4}e^{-3t} \\ -e^{t} + e^{-3t} & \frac{1}{2}e^{t} + \frac{1}{2}e^{-3t} \end{bmatrix}.
\end{aligned}$$

The linear case of Duhamel's principle (Corollary 4.4.3) guarantees the solution of the IVP is

$$
\begin{aligned}
\mathbf{x}(t) &= \Phi(t)\mathbf{x}_0 + \Phi(t)\int_0^t \Phi^{-1}(s)\mathbf{b}(s)\,ds \\
&= \begin{bmatrix} \frac{1}{2}e^{3t} + \frac{1}{2}e^{-t} & \frac{1}{4}e^{3t} - \frac{1}{4}e^{-t} \\ e^{3t} - e^{-t} & \frac{1}{2}e^{3t} + \frac{1}{2}e^{-t} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\
&\quad + \begin{bmatrix} \frac{1}{2}e^{3t} + \frac{1}{2}e^{-t} & \frac{1}{4}e^{3t} - \frac{1}{4}e^{-t} \\ e^{3t} - e^{-t} & \frac{1}{2}e^{3t} + \frac{1}{2}e^{-t} \end{bmatrix} \int_0^t \begin{bmatrix} \frac{1}{2}e^{s} + \frac{1}{2}e^{-3s} & -\frac{1}{4}e^{s} + \frac{1}{4}e^{-3s} \\ -e^{s} + e^{-3s} & \frac{1}{2}e^{s} + \frac{1}{2}e^{-3s} \end{bmatrix} \begin{bmatrix} 2e^{s} \\ -e^{s} \end{bmatrix} ds \\
&= \begin{bmatrix} \frac{3}{4}e^{3t} + \frac{1}{4}e^{t} \\ \frac{3}{2}e^{3t} - \frac{1}{2}e^{-t} \end{bmatrix} + \begin{bmatrix} \frac{1}{2}e^{3t} + \frac{1}{2}e^{-t} & \frac{1}{4}e^{3t} - \frac{1}{4}e^{-t} \\ e^{3t} - e^{-t} & \frac{1}{2}e^{3t} + \frac{1}{2}e^{-t} \end{bmatrix} \int_0^t \begin{bmatrix} \frac{5}{4}e^{2s} + \frac{3}{4}e^{-2s} \\ -\frac{5}{2}e^{2s} + \frac{3}{2}e^{-2s} \end{bmatrix} ds \\
&= \begin{bmatrix} \frac{3}{4}e^{3t} + \frac{1}{4}e^{t} \\ \frac{3}{2}e^{3t} - \frac{1}{2}e^{-t} \end{bmatrix} + \begin{bmatrix} \frac{1}{2}e^{3t} + \frac{1}{2}e^{-t} & \frac{1}{4}e^{3t} - \frac{1}{4}e^{-t} \\ e^{3t} - e^{-t} & \frac{1}{2}e^{3t} + \frac{1}{2}e^{-t} \end{bmatrix} \begin{bmatrix} \frac{5}{8}e^{2t} - \frac{5}{8} - \frac{3}{8}e^{-2t} + \frac{3}{8} \\ -\frac{5}{4}e^{2t} + \frac{5}{4} - \frac{3}{4}e^{-2t} + \frac{3}{4} \end{bmatrix} \\
&= \begin{bmatrix} \frac{3}{4}e^{3t} + \frac{1}{4}e^{t} \\ \frac{3}{2}e^{3t} - \frac{1}{2}e^{-t} \end{bmatrix} + \begin{bmatrix} \frac{1}{2}e^{3t} + \frac{1}{2}e^{-t} & \frac{1}{4}e^{3t} - \frac{1}{4}e^{-t} \\ e^{3t} - e^{-t} & \frac{1}{2}e^{3t} + \frac{1}{2}e^{-t} \end{bmatrix} \begin{bmatrix} \frac{5}{8}e^{2t} - \frac{3}{8}e^{-2t} - \frac{1}{4} \\ -\frac{5}{4}e^{2t} - \frac{3}{4}e^{-2t} + 2 \end{bmatrix} \\
&= \begin{bmatrix} \frac{3}{4}e^{3t} + \frac{1}{4}e^{t} + \frac{5}{16}e^{5t} - \frac{3}{16}e^{t} + \frac{1}{8}e^{3t} \\ \frac{3}{2}e^{3t} - \frac{1}{2}e^{-t} + \frac{5}{8}e^{5t} - \frac{3}{8}e^{t} + \frac{1}{4}e^{3t} - \frac{5}{8}e^{t} \end{bmatrix} \\
&\quad + \begin{bmatrix} \frac{5}{16}e^{t} - \frac{3}{16}e^{-3t} + \frac{1}{8}e^{-t} - \frac{5}{16}e^{5t} - \frac{3}{16}e^{t} + \frac{1}{2}e^{3t} + \frac{5}{16}e^{t} + \frac{3}{16}e^{-3t} - \frac{1}{2}e^{-t} \\ \frac{3}{8}e^{-3t} - \frac{1}{4}e^{-t} - \frac{5}{8}e^{5t} - \frac{3}{8}e^{t} + e^{3t} - \frac{5}{8}e^{t} - \frac{3}{8}e^{-3t} + e^{-t} \end{bmatrix} \\
&= \begin{bmatrix} \frac{3}{4}e^{3t} + \frac{1}{4}e^{t} \\ \frac{3}{2}e^{3t} - \frac{1}{2}e^{-t} \end{bmatrix} \\
&\quad + \begin{bmatrix} -\frac{3}{16}e^{t} + \frac{1}{8}e^{3t} + \frac{5}{16}e^{t} + \frac{1}{8}e^{-t} - \frac{3}{16}e^{t} + \frac{1}{2}e^{3t} + \frac{5}{16}e^{t} - \frac{1}{2}e^{-t} \\ -\frac{3}{8}e^{t} + \frac{1}{4}e^{3t} - \frac{5}{8}e^{t} - \frac{1}{4}e^{-t} - \frac{3}{8}e^{t} + e^{3t} - \frac{5}{8}e^{t} + e^{-t} \end{bmatrix} \\
&= \begin{bmatrix} \frac{3}{4}e^{3t} + \frac{1}{4}e^{t} \\ \frac{3}{2}e^{3t} - \frac{1}{2}e^{-t} \end{bmatrix} + \begin{bmatrix} \frac{4}{16}e^{t} + \frac{5}{8}e^{3t} - \frac{3}{8}e^{-t} \\ -2e^{t} + \frac{5}{4}e^{3t} + \frac{3}{4}e^{-t} \end{bmatrix}
\end{aligned}
$$

The reader should check that the first vector function is $(1, 1)$ when $t = 0$ and the second vector function is zero when $t = 0$.

4.4.3 Numerical Stability

Most of the techniques of this chapter are not well suited to numerical computation, especially when the system has a high dimension, or the matrix A is poorly conditioned. In the nonhomogeneous case, the inverse $\Phi(t)^{-1}$ is often computationally expensive to compute and integrate (and may also be ill conditioned). But even in the very nice setting of an autonomous homogeneous linear system with a semisimple matrix A, finding the explicit principle fundamental matrix solution $\Phi(t) = \exp(tA)$ by exponentiating A is not easy to do accurately when A has eigenvalues that are very close together or A is very large. But doing this computation is also not always necessary.

The solution to a specific IVP with initial value $\mathbf{x}(t_0) = \mathbf{v}$ is not a matrix, but rather the vector $\Phi(t)\mathbf{v}$, which may be easier to compute than $\Phi(t)$ alone. Often this can be estimated well with a numerical method, such as those introduced in Chapter 3, to get quantitative estimates of solutions of IVPs over relative short time intervals. If extremely long time behavior is sought, then numerically integrating the system of ODEs may take too long to be practical and is more likely to suffer from numerical instability.

If the eigenvalues of the matrix A all have negative real part, then in the long term (as $t \to \infty$) the solutions will all converge to 0.

Example 4.4.8. Consider the situation of Example 4.1.11, with lightweight Bob on a stiff spring, sliding on sandpaper, modeled by

$$\dot{\mathbf{y}} = A\mathbf{y}, \qquad A = \begin{bmatrix} 0 & 1 \\ -1000 & -1001 \end{bmatrix}. \tag{4.32}$$

with principle fundamental matrix solution

$$\begin{aligned}
\Phi(t) &= \exp(tA) \\
&= P \begin{bmatrix} e^{-t} & 0 \\ 0 & e^{-1000t} \end{bmatrix} P^{-1} \\
&= \frac{1}{999} \begin{bmatrix} 1000e^{-t} - e^{-1000t} & e^{-t} - e^{-1000t} \\ -1000e^{-t} + 1000e^{-1000t} & -e^{-t} + 1000e^{-1000t} \end{bmatrix}.
\end{aligned}$$

For any $\mathbf{x}_0 \in \mathbb{R}^2$ the unique solution $\mathbf{x}(t) = \Phi(t)\mathbf{x}_0 = \exp(tA)\mathbf{x}_0$ of the IVP $\dot{\mathbf{x}} = A\mathbf{x}, \quad \mathbf{x}(0) = \mathbf{x}_0$, converges to 0 as $t \to \infty$ because

$$\begin{aligned}
\|\mathbf{x}(t)\| &= \|\exp(tA)\mathbf{x}_0\| \\
&\le \|\exp(tA)\|\,\|x_0\| \\
&\le \|P\| \left\| e^{-t} \begin{bmatrix} 1 & 0 \\ 0 & e^{-999t} \end{bmatrix} \right\| \|P^{-1}\|\|x_0\| \\
&= e^{-t}\|P\|\|P^{-1}\|\|x_0\|
\end{aligned}$$

for $t > 0$. This implies $\|\mathbf{x}(t)\| \to 0$ as $t \to \infty$. Note that the negative eigenvalues of A are the determining factor in this asymptotic behavior of all solutions.

But even when the solutions are known to converge to 0 in the long term, that does not say much about their short-term behavior.

Example 4.4.9. Consider the particular solution $\mathbf{x}(t) = \exp(tA)\mathbf{x}_0$ of the ODE (4.32) in the previous example (Example 4.4.8) with the initial condition $\mathbf{x}_0 = (1, 1)$. This solution $\mathbf{x}(t) = \exp(tA)\mathbf{x}_0$ is

$$\begin{aligned}\mathbf{x}(t) &= \frac{1}{999}\begin{bmatrix} 1000e^{-t} - e^{-1000t} & e^{-t} - e^{-1000t} \\ -1000e^{-t} + 1000e^{-1000t} & -e^{-t} + 1000e^{-1000t} \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= \frac{1}{999}\begin{bmatrix} 1001e^{-t} - 2e^{-1000t} \\ -1001e^{-t} + 2000e^{-1000t} \end{bmatrix}.\end{aligned}$$

It *eventually* decays exponentially to 0 as $t \to \infty$, but in the short term the first component $x_1(t)$ of $\mathbf{x}(t)$ goes up slightly before converging to 0 (the derivative $\dot{x}_1(0)$ is 1), and the second component $x_2(t)$ of $\mathbf{x}(t)$ moves rapidly downward before converging to 0. The first derivative $\dot{x}_2(t)$ of the second component has a very large negative derivative for t close to 0 (this is the realm of *singular perturbation theory*). This is illustrated in Figure 4.1, where the first component is plotted in green and the second in blue.

The plot of this solution in phase space (the graph of parameterized curve $\mathbf{x}(t) = (x_1(t), x_2(t))$ in the plane) is given in Figure 4.2 shows the solution initially moving slightly toward the origin, then farther away, and then finally converging towards the origin. The sudden change in direction of the solution in phase space is a consequence of the huge difference in the two negative eigenvalues. The two directions seen here correspond to the eigenvectors of A which are

$$\begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -1/1000 \\ 1 \end{bmatrix}$$

corresponding to the eigenvalues -1 and -1000 of A respectively. The first eigenspace corresponds to "slow" convergence to the origin while the second eigenspace corresponds to "fast" convergence to the "slow" eigenspace.

Unfortunately, numerical schemes do not always accurately reflect the rapidly changing behavior of solutions such as the one in Example 4.4.9. The time step in the forward Euler method in vector form for the linear system $\dot{\mathbf{x}} = A\mathbf{x}$ is

$$\mathbf{x}(t + \Delta t) = \mathbf{x}(t) + (\Delta t)A\mathbf{x}(t).$$

We numerically execute the forward Euler method on the interval $[0, 2]$ with varying choices of Δt to capture this exponential convergence to 0, Table 4.1 shows the values of the two entries of $\mathbf{x}(t)$ at $t = 2$. Something very strange is occurring when $\Delta t \geq 0.003$, and also when $\Delta t = 0.002$ where the second component is positive when it should be negative.

The first time step of the forward Euler method gives glimpse of the problem: for $\Delta t = 0.1$ we have

$$\mathbf{x}(0.1) = \mathbf{x}_0 + (0.1)A\mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + (0.1)\begin{bmatrix} 0 & 1 \\ -1000 & -1001 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1.1 \\ -199.1 \end{bmatrix}.$$

This looks more like exponential expansion (eigenvalues with positive real part) than decay—the complete opposite of what the true solution is doing.

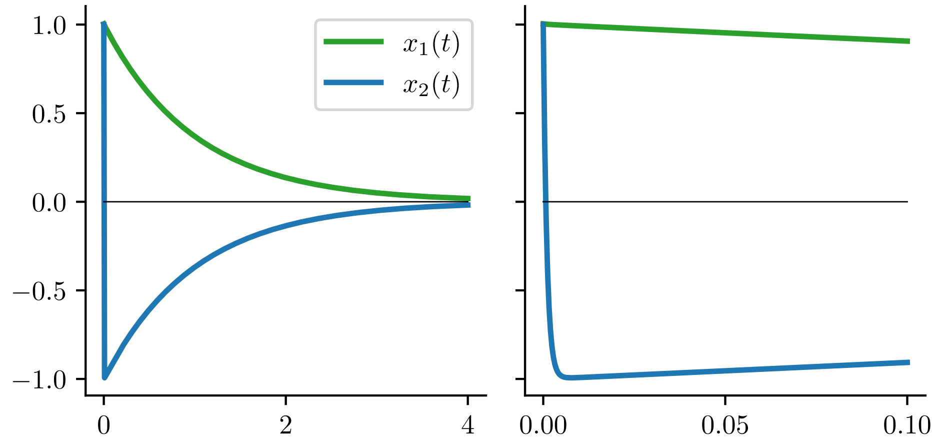

Figure 4.1: Plot of the two components (x_1 in green and x_2 in blue) of the particular solution of the IVP in Example 4.4.9. The left panel shows the solution over $t \in [0, 4]$, while the right panel is zoomed in, showing only $t \in [0, 0.1]$ The solution eventually decays exponentially to 0 as $t \to \infty$, but in the short term the first component $x_1(t)$ goes up slightly before converging to 0, and the second component $x_2(t)$ moves rapidly downward before converging back toward 0.

Δt	$x_1(2)$	$x_2(2)$
0.1	-1.64×10^{37}	1.64×10^{40}
0.01	-1.41×10^{188}	1.41×10^{191}
0.003	-6.13×10^{197}	6.13×10^{200}
0.002	0.13333292	1.86666708
0.001	0.1354706	−0.1354706

Table 4.1: Table of values at $t = 2$ of the result of the forward Euler method on the interval $[0, 2]$ for $\mathbf{x}(t)$ with varying choices of Δt. A small change in Δt between 0.002 and 0.003 leads to a huge change in the computed final value of $\mathbf{x}(2)$. This is an example of extreme numerical instability.

The backward Euler method

$$\mathbf{x}(t + \Delta t) = \mathbf{x}(t) + (\Delta t)A\mathbf{x}(t + \Delta t) \tag{4.33}$$

behaves better. Solving (4.33) for $\mathbf{x}(t + \Delta t)$ shows that

$$\begin{aligned}
\mathbf{x}(t + \Delta t) - (\Delta t)A\mathbf{x}(t + \Delta t) &= \mathbf{x}(t) \\
[I - (\Delta t)A]\mathbf{x}(t + \Delta t) &= \mathbf{x}(t),
\end{aligned}$$

and thus

$$\mathbf{x}(t + \Delta t) = [I - (\Delta t)A]^{-1}\mathbf{x}(t). \tag{4.34}$$

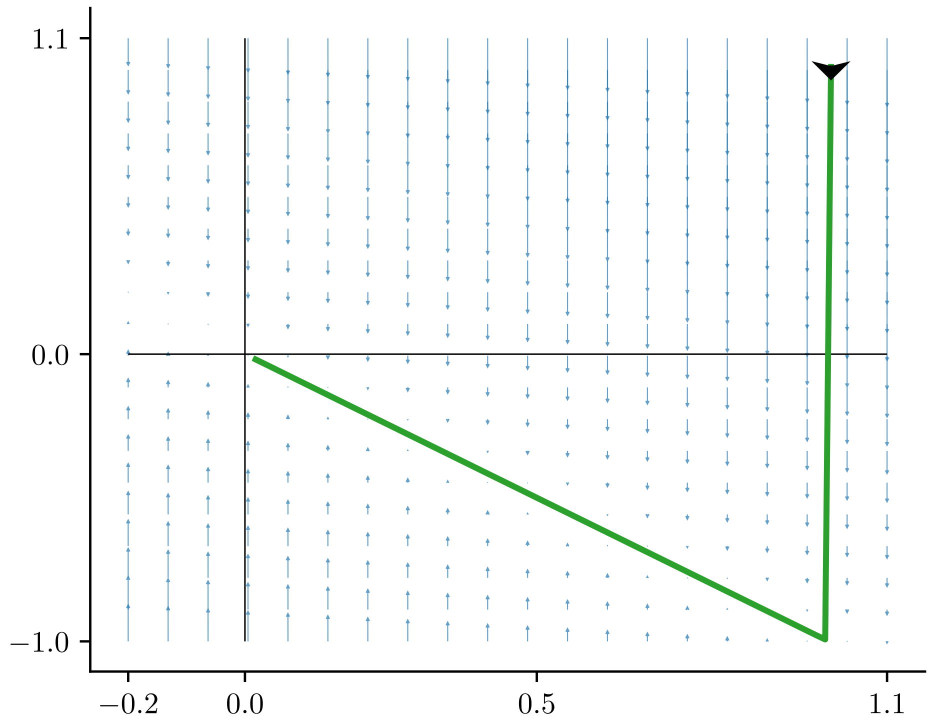

Figure 4.2: Phase plot (graph of parameterized curve $\mathbf{x}(t) = (x_1(t), x_2(t)$ in the plane) of the solution of the IVP in Example 4.4.9. The solution begins at $(1, 1)$ (green star) and initially moves almost vertically downward and then takes a sudden turn and heads toward the origin. Also plotted is a representation (blue arrows) of the vector field $\mathbf{f}(\mathbf{x}) = A\mathbf{x}$ on a grid of points.

Provided that the matrix $I - (\Delta t)A$ is invertible, we have $I - (\Delta t)A \to I$ as $\Delta t \to 0$, so $\mathbf{x}(t + \Delta t)$ is close to $\mathbf{x}(t)$ for sufficiently small Δt. Moreover, since I is invertible and the set of invertible matrices is open in the set of all matrices[25] it follows that $I - (\Delta t)A$ is invertible for small enough Δt. This means that for sufficiently small Δt, the backward Euler method should perform fairly well on homogeneous linear IVPs.

In the particular case of light Bob on a stiff spring, sliding on sandpaper (Example 4.4.9), the first time step of backward Euler with $\Delta t = 0.1$ gives

$$\mathbf{x}(0.1) = [I - (0.1)A]^{-1}\mathbf{x}_0 = \begin{bmatrix} 0.91899190 \\ -0.89108911 \end{bmatrix}.$$

This is a more reasonable approximation of the actual solution at time $t = 0.1$.

[25] Recall that the determinant function is continuous, so the preimage of the closed set $\{0\}$ in $\mathbb{R}$ is closed in $M_n(\mathbb{R})$, meaning the complement of that preimage is open

Remark 4.4.10. This example motivates why implicit methods are used in practice, even if they are more difficult to code up. For linear systems the cost of coding up an implicit method isn't that substantial, but for a nonlinear system it can be significant.

4.4.4 *Harmonic Oscillators with Both Damping and Forcing

As an example of a linear non-homogeneous system we will consider a harmonic oscillator that is damped, and has an external force applied. As we saw earlier, we anticipate the solution to be periodic (possibly decaying or growing in amplitude, but still periodic). In this instance though, the solution will of course depend highly on the type of forcing put into place. The most boring force to consider is when Bob is pulled on his spring and held at a given position. This is of course quite un-interesting and we are not prone to dwell on un-interesting cases, so we instead consider when the forcing itself is periodic. This physically may be if Bob is pulled one direction then pushed in the opposite in a periodic fashion. If Bob is coupled to another spring that has a mass several orders of magnitude larger than Bob's mass (or just has a much stiffer spring) then we can safely assume that Bob's motion won't affect the larger object, but the larger object will influence Bob's motion, i.e. provide a driving force. Of course there are more practical situations where this may be of interest, but I think you get the idea.

Rather than finding the solution via Duhamel's principle, we will use the method of inspired guessing, i.e. we are going to assume there is a particular solution roughly the same form as the forcing function.

Consider the damped, forced harmonic oscillator (TODO: include image)

$$m\ddot{y}(t) = \sum F = -ky(t) - b\dot{y}(t) + F_0\cos(\omega_f t).$$

Rewriting this we get

$$\ddot{y}(t) + \frac{b}{m}\dot{y}(t) + \omega_0^2 y(t) = \frac{F_0}{m}\cos(\omega_f t). \tag{4.35}$$

The method of inspired guessing (otherwise known as method of undetermined coefficients) involves solving the homogeneous system and then finding a particular solution that satisfies the non-homogeneous term. Since the problem is linear then we note that the particular solution will likely look like the forcing itself. We write the solution as a sum of the homogeneous solution and a particular solution, that is $y(t) = y_h(t) + y_p(t)$. The homogeneous solution is given by

$$y_h(t) = C\exp\left(\frac{-bt}{2m}\right)\cos(\omega_0 t + \phi_0),$$

where C is the amplitude and ϕ_0 is the phase, both determined by the initial data. The particular solution is

$$y_p(t) = A\cos(\omega_f t + \phi). \tag{4.36}$$

Plugging (4.36) into (4.35), we find that

$$-\omega_f^2 A\cos(\omega_f t + \phi) - \omega_f\frac{b}{\omega_f}A\sin(\omega_f t + \phi) + w_0^2 A\cos(\omega_f t + \phi) = \frac{F_0}{m}\cos(\omega_f t)$$

Using the trigonometric identities

$$\cos(\alpha \pm \beta) = \cos(\alpha)\cos(\beta) \mp \sin(\alpha)\sin(\beta)$$
$$\sin(\alpha \pm \beta) = \cos(\alpha)\sin(\beta) \pm \sin(\alpha)\cos(\beta),$$

we find that

$$\begin{aligned}
&-\omega_f^2 A\cos(\omega_f t)\cos(\phi) + \omega_f^2 A\sin(\omega_f t)\sin(\phi) \\
&\quad -\omega_f \frac{b}{m} A\cos(\omega_f t)\sin(\phi) - \omega_f \frac{b}{m} A\sin(\omega_f t)\cos(\phi) \\
&\quad + w_0^2 A\cos(\omega_f t)\cos(\phi) - w_0^2 A\sin(\omega_f t)\sin(\phi) = \frac{F_0}{m}\cos(\omega_f t).
\end{aligned}$$

Since this holds for all $t \in \mathbb{R}$, we can separate out parts and write it as the linear system

$$\begin{bmatrix} \omega_0^2 - \omega_f^2 & -\omega_f \frac{b}{m} \\ \omega_f \frac{b}{m} & \omega_0^2 - \omega_f^2 \end{bmatrix} \begin{bmatrix} A\cos(\phi) \\ A\sin(\phi) \end{bmatrix} = \begin{bmatrix} F_0/m \\ 0 \end{bmatrix}.$$

Solving this system gives

$$\begin{bmatrix} A\cos(\phi) \\ A\sin(\phi) \end{bmatrix} = \frac{1}{(\omega_0^2 - \omega_f^2)^2 + \omega_f^2 \frac{b^2}{m^2}} \begin{bmatrix} \omega_0^2 - \omega_f^2 & \omega_f \frac{b}{m} \\ -\omega_f \frac{b}{m} & \omega_0^2 - \omega_f^2 \end{bmatrix} \begin{bmatrix} F_0/m \\ 0 \end{bmatrix}.$$

Thus, we have

$$A\cos(\phi) = \frac{\frac{F_0}{m}(\omega_0^2 - \omega_f^2)}{(\omega_0^2 - \omega_f^2)^2 + \omega_f^2 \frac{b^2}{m^2}}$$
$$A\sin(\phi) = \frac{-\frac{F_0}{m}\omega_f \frac{b}{m}}{(\omega_0^2 - \omega_f^2)^2 + \omega_f^2 \frac{b^2}{m^2}}$$

It follows from the trig identity $\cos^2(\phi) + \sin^2(\phi) = 1$ that

$$A = \frac{\frac{F_0}{m}}{\sqrt{(\omega_0^2 - \omega_f^2)^2 + \omega_f^2 \frac{b^2}{m^2}}}. \tag{4.37}$$

In other words we find that the amplitude of the particular solution depends not only on the strength of the forcing but on its frequency, and on the frequency of the corresponding homogeneous solution, the strength of the damping and the mass of the object. We also find that the phase is described by the relation:

$$\tan(\phi) = \frac{\sin(\phi)}{\cos(\phi)} = \frac{-\omega_f \frac{b}{m}}{w_0^2 - w_f^2}. \tag{4.38}$$

As much fun as this calculation was, we note that memorizing trigonometric relations is not everyone's favorite past time (I know there are some of you that are shocked by this statement but it is true nonetheless), so we may hope that Duhamel's principle will work out better. It doesn't necessarily work out 'better' in terms of an analytical solution we can write down, but it does provide us with a clear algorithmic way of calculating the answer via the solution of the homogeneous system first.

Remark 4.4.11. When it's not forced, the dynamics converge to zero. When forced, it oscillates. Different frequencies have different amplitudes. When the forcing amplitude is close to the natural frequency, the resulting output has the highest amplitude. Interpreting the phase shift is a bit more complicated. Most can't look at (4.38) and get a clear picture of what this means, but it is clear that something very unpleasant will happen if the oscillator is forced at the natural frequency of the system.

Exercises

Note to the student: Each section of this chapter has several corresponding exercises, all collected here at the end of the chapter. The exercises between the first and second line are for Section 1, the exercises between the second and third lines are for Section 2, and so forth.

You should **work every exercise** (your instructor may choose to let you skip some of the advanced exercises marked with *). We have carefully selected them, and each is important for your ability to understand subsequent material. Many of the examples and results proved in the exercises are used again later in the text. Exercises marked with ⚠ are especially important and are likely to be used later in this book and beyond. Those marked with † are harder than average, but should still be done.

Although they are gathered together at the end of the chapter, we strongly recommend you do the exercises for each section as soon as you have completed the section, rather than saving them until you have finished the entire chapter.

4.1. Find e^{tA} for the following matrices. Hint: Check if the matrix A is diagonalizable, idempotent, nilpotent, or can be written as a sum of commuting matrices B and C, with B diagonal and C nilpotent.

(i)
$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

(ii)
$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

(iii)
$$A = \begin{bmatrix} -2 & 1 \\ 0 & -2 \end{bmatrix}.$$

(iv)
$$A = \begin{bmatrix} 3 & -6 \\ 1 & -2 \end{bmatrix}$$

4.2. Using the definition of the matrix exponential, show that for the matrix A, if $\mathbf{v}$ is an eigenvector with corresponding eigenvalue λ then $e^{tA}\mathbf{v} = e^{\lambda t}\mathbf{v}$.

4.3. Prove Theorem 4.1.8. Hint: It may be helpful to prove (v) before proving (ii) and (iii).

4.4. The 2×2 matrix A has complex conjugate eigenvalues $\lambda = a \pm ib$ and corresponding eigenvectors $\mathbf{u} \pm i\mathbf{v}$ where a, b are real values and $\mathbf{u}$ and $\mathbf{v}$ are real vectors. Show that if $Q = [\mathbf{u}\ \mathbf{v}]$, then

$$Q^{-1}AQ = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}.$$

4.5. For $\alpha, \omega, \delta \in \mathbb{R}$, consider the matrix

$$A = \begin{bmatrix} -\delta & -\omega & 0 \\ \omega & -\delta & 0 \\ 0 & 0 & \alpha \end{bmatrix}.$$

(i) Find the eigenvalues and eigenvectors of A.

(ii) Compute e^{tA} when $\delta \neq 0$ and $\omega = 0$.

(iii) Compute e^{tA} when $\alpha \neq 0$ and $\omega \neq 0$.

4.6. Compute all the matrix exponentials in Exercise 4.1 using spectral calculus (that is, using resolvents).

4.7. If A is semisimple (diagonalizable), with $A = Q\Lambda Q^{-1}$ and Λ diagonal, show that the eigenprojection P_λ corresponding to eigenvalue λ can be written as

$$Q \begin{bmatrix} 0 & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & 0 \end{bmatrix} Q^{-1},$$

where the identity I in the middle block is located in the same place that the eigenvalue λ occurs in the diagonal matrix Λ, and the dimension of I is the number of times that λ repeats in the corresponding block of Λ.

4.8. Use the result of Exercise 4.7 to give an alternative proof of the spectral decomposition formula (4.15) in the special case that A is semisimple.

4.9. Use the result of Exercise 4.7 to give an alternative proof of the Mapping the Spectral Decomposition Theorem (Theorem 4.2.8) in the special case that A is semisimple.

4.10. For a matrix $A \in Mn(\mathbb{F})$, define $\sin(tA)$ and $\cos(tA)$ by their Taylor series

$$\sin(tA) = \sum_{k=0}^{\infty} (-1)^k \frac{(tA)^{2k+1}}{(2k+1)!}$$
$$\cos(tA) = \sum_{k=0}^{\infty} (-1)^k \frac{(tA)^{2k}}{(2k)!}.$$

These also can be written in terms of the spectral resolution formula (4.12). For both $\sin(tA)$ and $\cos(tA)$ use the Mapping the Spectral Decomposition Theorem (Theorem 4.2.8) to write a formula in terms of the eigenprojections P_λ and eigennilpotents D_λ of A, analogous to the formula for the exponential (4.16).

4.11. Let

$$A = \begin{bmatrix} t & 0 \\ 0 & 3t^2 \end{bmatrix}.$$

(i) Verify that A satisfies the conditions of Proposition 4.3.1 and

(ii) Find the principal fundamental matrix solution $\Phi(t)$ of the ODE $\dot{\mathbf{x}}(t) = A(t)\mathbf{x}(t)$.

4.12. Consider the ODE

$$y'''(x) - \frac{2}{x}y''(x) + y'(x) - \frac{2}{x}y = 0 \tag{4.39}$$

(i) Verify that $y = \sin(x)$ and $y = \cos(x)$ are both solutions.

(ii) Rewrite this ODE as a three-dimensional first-order linear system $\mathbf{z}' = A(x)\mathbf{z}(x)$. Each solution of the third-order system corresponds to a solution of the three-dimensional first-order system. Write the three-dimensional solutions to the first-order system that correspond to the solutions $y = \sin(x)$ and $y = \cos(x)$.

(iii) For an unknown solution $y_3(x)$ that is linearly independent of $\sin(x)$ and $\cos(x)$, write down the fundamental matrix solution $\Phi(x)$ corresponding to these three functions sin, cos, and y_3, and compute the determinant of that matrix as a function of y_3, y_3', and y_3''.

(iv) Use the AJL formula to compute the determinant of a fundamental matrix solution $\Phi(x)$, expressed in terms of the initial value $\Phi(\pi/2)$.

(v) Equating the two expressions for the determinant gives a new second-order (nonhomogeneous) ODE that the function y must satisfy.

(vi) Verify that there is a constant c so that $y(x) = c(x^2 - 2)$ is a solution of the new second-order equation.

(vii) Explain why this means that the general solution to (4.39) is of the form $y(x) = c_1 \sin(x) + c_2 \cos(x) + c_3(x^2 - 2)$.

4.13.* Prove that the derivative $\det(A(t)) = |A(t)|$ of an $n \times n$ matrix $A(t) =$ with ijth entry given by $a_{ij}(t)$ is

$$\frac{d}{dt}\det(A(t)) = \begin{vmatrix} \dot{a}_{11}(t) & \dot{a}_{12}(t) & \cdots & \dot{a}_{1n}(t) \\ a_{21}(t) & a_{22}(t) & \cdots & a_{2n}(t) \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}(t) & a_{n2}(t) & \cdots & a_{nn}(t) \end{vmatrix} + \begin{vmatrix} a_{11}(t) & a_{12}(t) & \cdots & a_{1n}(t) \\ \dot{a}_{21}(t) & \dot{a}_{22}(t) & \cdots & \dot{a}_{2n}(t) \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}(t) & a_{n2}(t) & \cdots & a_{nn}(t) \end{vmatrix}$$
$$+ \cdots + \begin{vmatrix} a_{11}(t) & a_{12}(t) & \cdots & a_{1n}(t) \\ a_{21}(t) & a_{22}(t) & \cdots & a_{2n}(t) \\ \vdots & \vdots & \ddots & \vdots \\ \dot{a}_{n1}(t) & \dot{a}_{n2}(t) & \cdots & \dot{a}_{nn}(t) \end{vmatrix}.$$

4.14.* Let $X(t)$ be an $n \times n$ matrix and suppose its entries are differentiable functions on the interval $a < t < b$. If $\det(X(t)) \neq 0$ for $a < t < b$, prove there is an $A(t)$ such that $\frac{d}{dt}X = A(t)X$.

4.15. Write code that will solve any linear, homogeneous system of the form $\dot{\mathbf{x}} = A\mathbf{x},\ \mathbf{x}(0) = \mathbf{x}_0$ using the midpoint method (you may want to use forward Euler on the first step).

4.16. Using your code from the previous problem, solve this system for the following cases. In each case, plot the phase portrait of the solution, that is, generate plots where x_1 is the horizontal axis, and x_2 is the vertical axis. You should run this out to a final time of $T = 3.0$, with time-step $\Delta t = 0.01$.

(i) $A = \begin{bmatrix} 0 & 3 \\ 3 & 0 \end{bmatrix}$ with $\mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

(ii) $A = \begin{bmatrix} 0 & 3 \\ -3 & 0 \end{bmatrix}$ with $\mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

How does the solution change for the two different matrices?

4.17. Modify your code from the previous two problems to work for the trapezoidal method. Generate the same plots. How does the solution compare? How do you handle the implicit methods?

4.18. Use Duhamel's method to find the solution to

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} \sin(t) \\ \cos(t) \end{bmatrix} \\ \mathbf{x}(0) &= \begin{bmatrix} 1 \\ -1 \end{bmatrix}. \end{aligned}$$

4.19. Suppose $p(t)$, $q(t)$, and $h(t)$ are continuous real-valued functions defined on an open interval I and that

$$\ddot{x} + p(t)\dot{x} + q(t)x = 0$$

has linearly independent solutions $\phi_1(t)$ and $\phi_2(t)$. Find an explicit formula, using Duhamel's method for the solution to

$$\ddot{x} + p(t)\dot{x} + q(t)x = h(t).$$

Do not simplify.

4.20. *Use Duhamel's Principle to compute the solution to the damped, forced harmonic oscillator.

4.21. *Find the general solution to the differential equation

$$2\dot{x} + 18x = 6\tan(3t).$$

Notes