

Lecture 33: Stiff Differential Equations

5.2.5. Stiff Differential Equations.

Thus far we have mainly considered scalar ODEs. Both one-step and linear multistep methods readily generalize to *systems* of ODEs, where the scalar $x(t)$ is replaced by a vector $\mathbf{x}(t)$. In these notes, we shall focus upon *linear systems* of ODEs. (In applications one often encounters nonlinear ODEs, but behavior of such a system near a steady state can often be understood by examining the *linearization* of the equation about that steady state.)

Consider the linear system of differential equations

$$\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t), \quad \mathbf{x}(0) = \mathbf{x}_0,$$

for $\mathbf{A} \in \mathbb{C}^{n \times n}$ and $\mathbf{x}(t) \in \mathbb{C}^n$. We wish to see how the scalar linear stability theory discussed in the last lecture applies to such systems. Assume that the matrix \mathbf{A} is diagonalizable, so that it can be written $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$ for the diagonal matrix $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$. Premultiplying the differential equation by \mathbf{V}^{-1} yields

$$\mathbf{V}^{-1}\mathbf{x}'(t) = \mathbf{\Lambda}\mathbf{V}^{-1}\mathbf{x}(t), \quad \mathbf{V}^{-1}\mathbf{x}(0) = \mathbf{V}^{-1}\mathbf{x}_0.$$

Now let $\mathbf{y}(t) = \mathbf{V}^{-1}\mathbf{x}(t)$, which can be thought of as the vector $\mathbf{x}(t)$ represented in a transformed coordinate system. In these new coordinates, the matrix equation decouples into a system of n linear independent scalar equations, as the above equation takes the form

$$\mathbf{y}'(t) = \mathbf{\Lambda}\mathbf{y}(t), \quad \mathbf{y}(0) = \mathbf{y}(0).$$

This is equivalent to

$$\begin{aligned} y_1'(t) &= \lambda_1 y_1(t), & y_1(0) &= [\mathbf{V}^{-1}\mathbf{x}_0]_1; \\ &\vdots \\ y_n'(t) &= \lambda_n y_n(t), & y_n(0) &= [\mathbf{V}^{-1}\mathbf{x}_0]_n, \end{aligned}$$

and each of these equations has the simple solution

$$y_j(t) = e^{\lambda_j t} y_j(0).$$

Now we can use the relationship $\mathbf{x}(t) = \mathbf{V}\mathbf{y}(t)$ to transform back to the original coordinates. Define

$$e^{\mathbf{A}t} := \begin{bmatrix} e^{t\lambda_1} & & \\ & \ddots & \\ & & e^{t\lambda_n} \end{bmatrix}.$$

Then we can write

$$\mathbf{x}(t) = \mathbf{V}\mathbf{y}(t) = \mathbf{V}e^{\mathbf{\Lambda}t}\mathbf{y}(0) = \mathbf{V}e^{\mathbf{A}t}\mathbf{V}^{-1}\mathbf{x}_0, \quad (33.1)$$

which motivates the definition of the *matrix exponential*,

$$e^{t\mathbf{A}} := \mathbf{V}e^{\mathbf{\Lambda}t}\mathbf{V}^{-1},$$

in which case the solution $\mathbf{x}(t)$ has the convenient form

$$\mathbf{x}(t) = e^{\mathbf{A}t} \mathbf{x}_0.$$

What can be said of the magnitude of the solution $\mathbf{x}(t)$? We can bound the solution using norm inequalities,

$$\|\mathbf{x}(t)\|_2 \leq \|\mathbf{V}\|_2 \|e^{\mathbf{A}t}\|_2 \|\mathbf{V}^{-1}\|_2 \|\mathbf{x}_0\|_2.$$

Since $e^{\mathbf{A}t}$ is a diagonal matrix, its 2-norm is the largest magnitude of its entries:

$$\|e^{\mathbf{A}t}\|_2 = \max_{1 \leq j \leq n} |e^{t\lambda_j}|,$$

and hence

$$\frac{\|\mathbf{x}(t)\|_2}{\|\mathbf{x}_0\|_2} \leq \|\mathbf{V}\|_2 \|\mathbf{V}^{-1}\|_2 \max_{1 \leq j \leq n} |e^{t\lambda_j}|. \quad (33.2)$$

Thus the asymptotic decay rate of $\|\mathbf{x}(t)\|_2$ is controlled by the rightmost eigenvalue of \mathbf{A} in the complex plane. If all eigenvalues of \mathbf{A} have negative real part, then $\|\mathbf{x}(t)\|_2 \rightarrow 0$ as $t \rightarrow \infty$. Note that when $\|\mathbf{V}\|_2 \|\mathbf{V}^{-1}\|_2 > 1$, it is possible that $\|\mathbf{x}(t)\|_2 / \|\mathbf{x}_0\|_2 > 1$ for small $t > 0$, even if this ratio must eventually decay to zero as $t \rightarrow \infty$.[†]

Note that the definition $e^{t\mathbf{A}} = \mathbf{V}e^{t\mathbf{\Lambda}}\mathbf{V}^{-1}$ is consistent with the more general definition obtained by substituting $t\mathbf{A}$ into the same Taylor series that defines the scalar exponential:

$$e^{t\mathbf{A}} = \mathbf{I} + t\mathbf{A} + \frac{1}{2!}t^2\mathbf{A}^2 + \frac{1}{3!}t^3\mathbf{A}^3 + \frac{1}{4!}t^4\mathbf{A}^4 + \cdots.$$

If we set $\mathbf{x}(t) = e^{t\mathbf{A}}\mathbf{x}_0$, then we have

$$\begin{aligned} \mathbf{x}'(t) &= \frac{d}{dt} \left(e^{t\mathbf{A}} \mathbf{x}_0 \right) \\ &= \frac{d}{dt} \left(\mathbf{I} + t\mathbf{A} + \frac{t^2}{2!}\mathbf{A}^2 + \frac{t^3}{3!}\mathbf{A}^3 + \cdots \right) \mathbf{x}_0 \\ &= \left(\mathbf{A} + t\mathbf{A}^2 + \frac{t^2}{2!}\mathbf{A}^3 + \frac{t^3}{3!}\mathbf{A}^4 + \cdots \right) \mathbf{x}_0 \\ &= \mathbf{A} \left(\mathbf{I} + t\mathbf{A} + \frac{t^2}{2!}\mathbf{A}^2 + \frac{t^3}{3!}\mathbf{A}^3 + \cdots \right) \mathbf{x}_0 \\ &= \mathbf{A} e^{t\mathbf{A}} \mathbf{x}_0 \\ &= \mathbf{A} \mathbf{x}(t). \end{aligned}$$

Hence $\mathbf{x}(t) = e^{t\mathbf{A}}\mathbf{x}_0$ solves the equation $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$, and satisfies the initial condition $\mathbf{x}(0) = \mathbf{x}_0$.

What can be said of the behavior of a linear multistep method applied to this equation? Euler's method, for example, takes the form

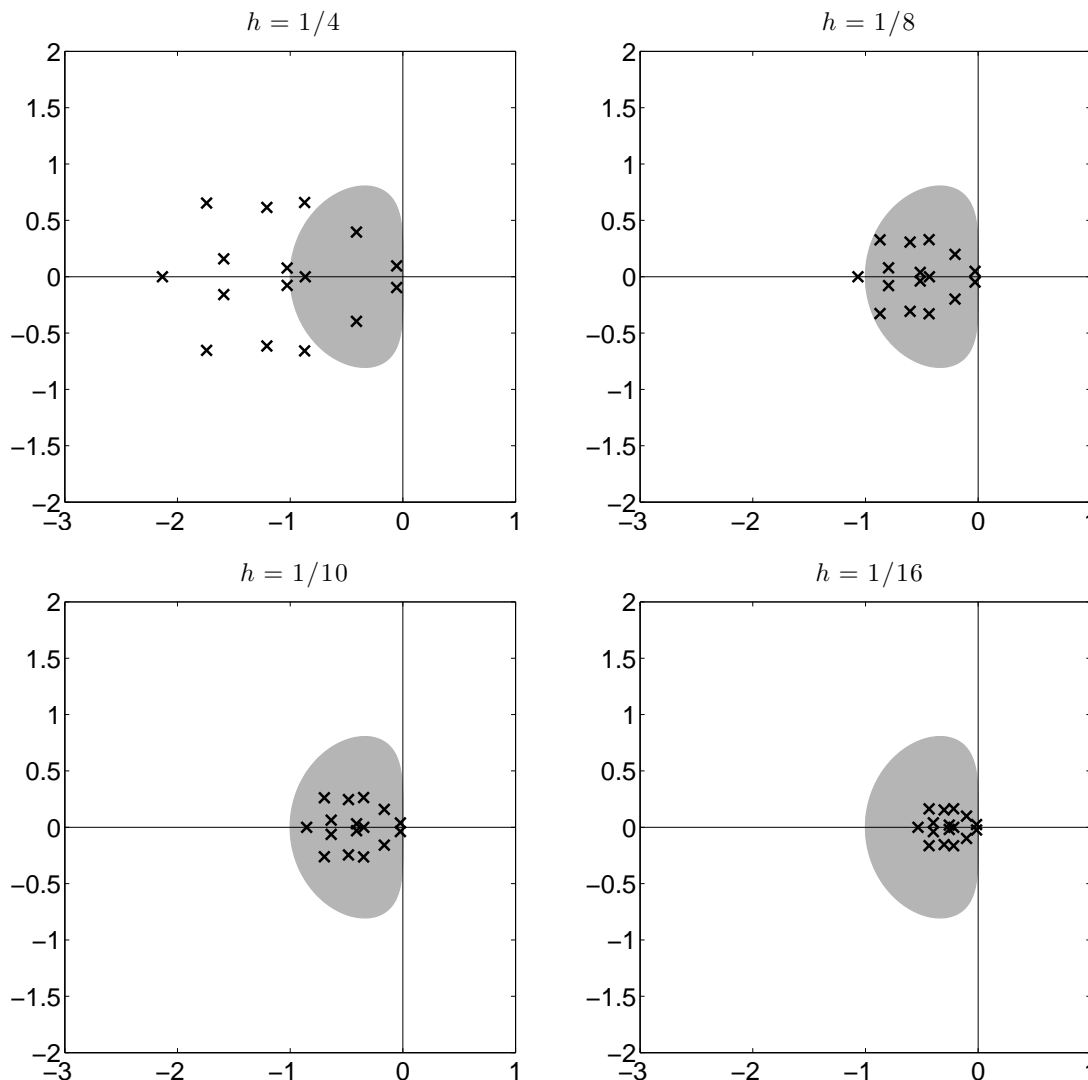
$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{x}_k + h\mathbf{A}\mathbf{x}_k \\ &= (\mathbf{I} + h\mathbf{A})\mathbf{x}_k, \end{aligned}$$

and hence $\mathbf{x}_k = (\mathbf{I} + h\mathbf{A})^k \mathbf{x}_0$.

[†]The possibility of this *transient growth* complicates the analysis of dynamical systems with non-Hermitian coefficient matrices, and turns out to be closely related to the sensitivity of the eigenvalues of \mathbf{A} to perturbations. This behavior is both fascinating and physically important, but regrettably beyond the scope of these lectures.

We can understand the *asymptotic* behavior of $(\mathbf{I}+h\mathbf{A})^k$ by examining the eigenvalues of $(\mathbf{I}+h\mathbf{A})^k$: the quantity $(\mathbf{I}+h\mathbf{A})^k \rightarrow \mathbf{0}$ if and only if all the eigenvalues of $\mathbf{I}+h\mathbf{A}$ are less than one in modulus. The *spectral mapping theorem* ensures that if $(\lambda_j, \mathbf{v}_j)$ is an eigenvalue-eigenvector pair for \mathbf{A} , then $(1+h\lambda_j, \mathbf{v}_j)$ is an eigenpair of $\mathbf{I}+h\mathbf{A}$. This is easy to verify by a direct computation: If $\mathbf{A}\mathbf{v}_j = \lambda_j\mathbf{v}_j$, then $(\mathbf{I}+h\mathbf{A})\mathbf{v}_j = \mathbf{v}_j+h\mathbf{A}\mathbf{v}_j = (1+h\lambda_j)\mathbf{v}_j$. Hence, the numerical solution \mathbf{x}_k computed by Euler's method will decay to zero if $|1+h\lambda_j| < 1$ for *all* eigenvalues λ_j of \mathbf{A} . In the language of the last lecture, we need $h\lambda_j$ to fall in the absolute stability region for the forward Euler method for all eigenvalues λ_j of \mathbf{A} .

For a general linear multistep method, this criterion generalizes to the requirement that $h\lambda_j$ be located in the method's absolute stability region for all eigenvalues λ_j of \mathbf{A} . This is illustrated in the following example. Here \mathbf{A} is a 16×16 matrix with all its eigenvalues in the left half of the complex plane. We wish to solve $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$ using the second-order Adams-Bashforth method, whose stability region was plotted in the last lecture. The plots below show $h\lambda_j$ as crosses for the eigenvalues $\lambda_1, \dots, \lambda_{16}$ of \mathbf{A} . If *any* value of $h\lambda_j$ is outside the stability region (shown in gray), then the iteration will *grow exponentially!* If h is sufficiently small that $h\lambda_j$ is in the stability region for all eigenvalues λ_j , then $\mathbf{x}_k \rightarrow \mathbf{0}$ as $k \rightarrow \infty$, consistent with the fact that $\mathbf{x}(t) \rightarrow \mathbf{0}$ as $t \rightarrow \infty$.



It is worth looking at this example a little bit closer. Suppose \mathbf{A} is diagonalizable, so we can write $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$. Thus,

$$\begin{aligned}\mathbf{x}_k &= (\mathbf{I} + h\mathbf{A})^k \mathbf{x}_0 \\ &= (\mathbf{I} + h\mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1})^k \mathbf{x}_0 \\ &= (\mathbf{V}\mathbf{V}^{-1} + h\mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1})^k \mathbf{x}_0 \\ &= \mathbf{V}(\mathbf{I} + h\mathbf{\Lambda})^k \mathbf{V}^{-1} \mathbf{x}_0.\end{aligned}$$

Compare this last expression to the formula (33.1) for the true solution $\mathbf{x}(t)$ in terms of the matrix exponential. As we did in that case, we can bound \mathbf{x}_k as follows:

$$\begin{aligned}\|\mathbf{x}_k\|_2 &= \|\mathbf{V}(\mathbf{I} + h\mathbf{\Lambda})^k \mathbf{V}^{-1} \mathbf{x}_0\|_2 \\ &= \|\mathbf{V}(\mathbf{I} + h\mathbf{\Lambda})^k \mathbf{V}^{-1}\|_2 \|\mathbf{x}_0\|_2 \\ &= \|\mathbf{V}\|_2 \|\mathbf{V}^{-1}\|_2 \|(\mathbf{I} + h\mathbf{\Lambda})^k\|_2 \|\mathbf{x}_0\|_2.\end{aligned}$$

Since $\mathbf{I} + h\mathbf{\Lambda}$ is a diagonal matrix, we have

$$(\mathbf{I} + h\mathbf{\Lambda})^k = \begin{bmatrix} (1 + h\lambda_1)^k & & & \\ & (1 + h\lambda_2)^k & & \\ & & \ddots & \\ & & & (1 + h\lambda_n)^k \end{bmatrix},$$

giving

$$\|(\mathbf{I} + h\mathbf{\Lambda})^k\|_2 = \max_{1 \leq j \leq n} |1 + h\lambda_j|^k.$$

Thus, we arrive at the bound

$$\frac{\|\mathbf{x}_k\|_2}{\|\mathbf{x}_0\|_2} \leq \|\mathbf{V}\|_2 \|\mathbf{V}^{-1}\|_2 \max_{1 \leq j \leq n} |1 + h\lambda_j|^k,$$

which is analogous to the bound (33.2) for the exact solution.

We can glean just a bit more from our analysis of \mathbf{x}_k . Since \mathbf{A} is diagonalizable, its eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ form a basis for \mathbb{C}^n . Expand the initial condition \mathbf{x}_0 in this basis:

$$\mathbf{x}_0 = \sum_{j=1}^n c_j \mathbf{v}_j = \mathbf{V}\mathbf{c}.$$

Now, our earlier expression for \mathbf{x}_k gives

$$\mathbf{x}_k = \mathbf{V}(\mathbf{I} + h\mathbf{\Lambda})^k \mathbf{V}^{-1} \mathbf{x}_0 = \mathbf{V}(\mathbf{I} + h\mathbf{\Lambda})^k \mathbf{V}^{-1} \mathbf{V}\mathbf{c} = \mathbf{V}(\mathbf{I} + h\mathbf{\Lambda})^k \mathbf{c}.$$

Since

$$\begin{bmatrix} (1 + h\lambda_1)^k & & & \\ & (1 + h\lambda_2)^k & & \\ & & \ddots & \\ & & & (1 + h\lambda_n)^k \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} (1 + h\lambda_1)^k c_1 \\ (1 + h\lambda_2)^k c_2 \\ \vdots \\ (1 + h\lambda_n)^k c_n \end{bmatrix},$$

we have

$$\mathbf{x}_k = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_n] \begin{bmatrix} (1 + h\lambda_1)^k c_1 \\ (1 + h\lambda_2)^k c_2 \\ \vdots \\ (1 + h\lambda_n)^k c_n \end{bmatrix} = \sum_{j=1}^n c_j (1 + h\lambda_j)^k \mathbf{v}_j.$$

Thus as $k \rightarrow \infty$, the approximate solution \mathbf{x}_k will start to look more and more like (a scaled version of) the vector \mathbf{v}_ℓ , where ℓ is the index that maximizes $|1 + h\lambda_j|$:

$$|1 + h\lambda_\ell| = \max_{1 \leq j \leq n} |1 + h\lambda_j|.$$

In our last example plotted above, the step size did not need to be very small in order for all $h\lambda_j$ to be contained within the stability region. However, most practical examples in science and engineering yield matrices \mathbf{A} whose eigenvalues span multiple orders of magnitude – and in this case, the stability requirement is far more difficult to satisfy. Consider the following simple example. Let

$$\mathbf{A} = \begin{bmatrix} -1999 & -1998 \\ 999 & 998 \end{bmatrix},$$

which has the diagonalization

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} -100 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}.$$

The eigenvalues are $\lambda_1 = -100$ and $\lambda_2 = -1$, and the exact solution takes the form

$$\mathbf{x}(t) = e^{t\mathbf{A}}\mathbf{x}_0 = \mathbf{V} \begin{bmatrix} e^{-100t} & 0 \\ 0 & e^{-t} \end{bmatrix} \mathbf{V}^{-1}\mathbf{x}_0.$$

If the initial condition has the form

$$\mathbf{x}_0 = \mathbf{V} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = c_1 \begin{bmatrix} 2 \\ -1 \end{bmatrix} + c_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix},$$

then the solution can be written as

$$\mathbf{x}(t) = \mathbf{V} \begin{bmatrix} e^{-100t} & 0 \\ 0 & e^{-t} \end{bmatrix} \mathbf{V}^{-1}\mathbf{x}_0 = \mathbf{V} \begin{bmatrix} e^{-100t} & 0 \\ 0 & e^{-t} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = c_1 e^{-100t} \begin{bmatrix} 2 \\ -1 \end{bmatrix} + c_2 e^{-t} \begin{bmatrix} -1 \\ 1 \end{bmatrix},$$

and so we see that $\mathbf{x}(t) \rightarrow \mathbf{0}$ as $t \rightarrow \infty$. The eigenvalue $\lambda_1 = -100$ corresponds to a *fast transient*, a component of the solution that decays very rapidly; the eigenvalue $\lambda_2 = -1$ corresponds to a *slow transient*, a component of the solution that decays much more slowly.

Suppose we wish to obtain a solution with the forward Euler method. To obtain a numerical solution $\{\mathbf{x}_k\}$ that mimics the asymptotic behavior of the true solution, $\mathbf{x}(t) \rightarrow \mathbf{0}$, we must choose h sufficiently small that $|1 + h\lambda_1| = |1 - 100h| < 1$ and $|1 + h\lambda_2| = |1 - h| < 1$. The first condition requires $h \in (0, 1/50]$, which the second condition is far less restrictive: $h \in (0, 2)$. The more restrictive condition describes the values of h that will give $\mathbf{x}_k \rightarrow \mathbf{0}$ for all \mathbf{x}_0 .

Take note of this phenomenon: *the faster a component decays from the true solution (like e^{-100t} in our example), the smaller the time step must be for the forward Euler method (and other explicit schemes).*

Problems for which \mathbf{A} has eigenvalues with significantly different magnitudes are called *stiff differential equations*. For such problems, implicit methods – which generally have much larger stability regions – are generally favored.

Thus far we have only sought $\mathbf{x}_k \rightarrow \mathbf{0}$ as $k \rightarrow \infty$. In some cases, we merely wish for \mathbf{x}_k to be bounded. (Such examples are seen in the method of lines problems on Problem Set 6.) In this case, it is acceptable to have an eigenvalue $h\lambda_j$ on the boundary of the absolute stability region of a method, provided it is not a repeated eigenvalue (more precisely, provided it is associated with 1×1 Jordan blocks, i.e., it is not defective).