29 January 2025 – PHYS 6260
Prof. John Wise
Newman Chapter 9: Partial Differential Equations

Numerical stability

The FTCS method works well for the diffusion equation, but there are other equations in which it is stable only under certain conditions. But for the wave equation, the FTCS method can fail spectacularly for interesting reasons.

The wave equation can be written as

$$\frac{\partial^2\phi}{\partial x^2} - \frac{1}{v^2}\frac{\partial^2\phi}{\partial t^2} = 0 \tag{1}$$
$$\frac{\partial^2\phi}{\partial t^2} = v^2\frac{\partial^2\phi}{\partial x^2}$$

and can be used to describe a wave traveling along a string, for instance. To solve the equation, using the FTCS method, we would replace the spatial derivative with a center difference, just like before, resulting in

$$\frac{\partial^2\phi}{\partial t^2} = \frac{v^2}{a^2}\left[\phi(x+a,t) + \phi(x-a,t) - 2\phi(x,t)\right] \tag{2}$$

We can write this second-order ODE as two first-order ODEs by changing variables,

$$\frac{\partial\phi}{\partial t} = \psi(x,t), \qquad \frac{\partial\psi}{\partial t} = \frac{v^2}{a^2}\left[\phi(x+a,t) + \phi(x-a,t) - 2\phi(x,t)\right] \tag{3}$$

Applying Euler's method to both variables, we have two FTCS equations

$$\phi(x,t+h) = \phi(x,t) + h\psi(x,t) \tag{4}$$
$$\psi(x,t+h) = \psi(x,t) + h\frac{v^2}{a^2}\left[\phi(x+a,t) + \phi(x-a,t) - 2\phi(x,t)\right]$$

Figure 1 shows the solution at times $t = (2, 50, 100)$ ms for some particular timestep h. One can see that the solution becomes noisier (errors) with time, and this solution is **numerically unstable**. This is not rounding error, but the errors would continue to accumulate, eventually resulting in floating point overflows.

We can perform a **von Neumann stability analysis** in which the solution is expressed as a Fourier series: $\phi(x,t) = \sum_k c_k(t)e^{ikx}$ for some suitable set of wave vectors k and time-varying and (potentially) complex coefficients $c_k(t)$. Given such an expression, one can evaluate how each term changes in the next timestep. As an example, we plug this term into

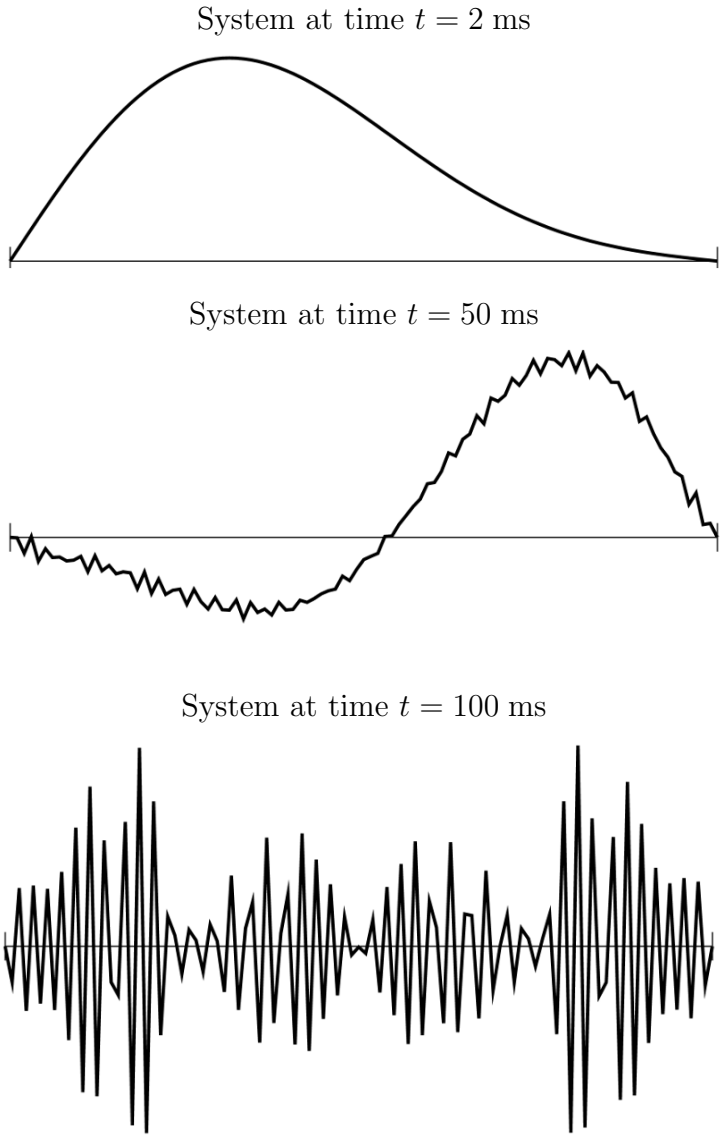

Figure 1: Example of numerically unstable solution to the wave equation.

the FTCS diffusion equation, resulting in

$$\begin{aligned}\phi(x, t+h) &= c_k(t)e^{ikx} + h\frac{D}{a^2}\left[e^{ik(x+a)} + e^{ik(x-a)} - 2e^{ikx}\right] \\ &= \left[1 + h\frac{D}{a^2}\left(e^{ika} + e^{-ika} - 2\right)\right] c_k(t)e^{ikx} \\ &= \left[1 - h\frac{4D}{a^2}\sin^2\left(\frac{ka}{2}\right)\right] c_k(t)e^{ikx},\end{aligned} \tag{5}$$

where we have used $e^{i\theta} + e^{-i\theta} = 2\cos\theta$ and $1 - \cos\theta = 2\sin^2(\theta/2)$.

We can see that each Fourier coefficient is independent of each other (not dependent on

x or t) and evolve with time as

$$c_k(t+h) = \left[1 - h\frac{4D}{a^2}\sin^2\left(\frac{ka}{2}\right)\right] c_k(t). \tag{6}$$

The solution will be unstable if these coefficients grow with time, otherwise it is *stable*. Stability happens when the magnitude of the bracket factor is less than one, $h(4D/a^2)\sin^2(ka/2) \leq 2$ for all k. The largest the $\sin^2$ term can be is 1, so the solution will be stable for all k when

$$\boxed{h \leq \frac{a^2}{2D}} \tag{7}$$

If the timestep h is larger than this limit, then the solution can diverge. If it is smaller than this limit, all of the Fourier modes will converge to zero, leaving the first ($k = 0$) term, which converges to 1. This solution makes physical sense in the diffusion equation because any perturbations will diffuse away, leaving behind a constant quantity (i.e. temperature, displacement, etc).

It should be noted that von Neumann analysis does not work on nonlinear equations, but we can apply it to the FTCS wave equation to determine some stability criteria. Because it has a second derivative, we must perform the analysis for the two first-order equations:

$$\begin{pmatrix}\phi(x,t)\\ \psi(x,t)\end{pmatrix} = \begin{pmatrix}c_\phi(t)\\ c_\psi(t)\end{pmatrix} e^{ikx} \tag{8}$$

We can substitute this Fourier representation into Equation (16) to find the Fourier coefficients:

$$c_\phi(x+h) = c_\phi(t) + hc_\psi(t), \tag{9}$$

$$c_\psi(x+h) = c_\psi(t) - hc_\phi(t)\frac{4v^2}{a^2}\sin^2\frac{ka}{2}. \tag{10}$$

$$\tag{11}$$

This can be written in vector form as $\mathbf{c}(t+h) = \mathbf{A}\mathbf{c}(t)$, where

$$\mathbf{A} = \begin{pmatrix}1 & h\\ -hr^2 & 1\end{pmatrix} \qquad \text{with} \qquad r = \frac{2v}{a}\sin\frac{ka}{2} \tag{12}$$

We can write $\mathbf{c}(t)$ as a linear combination of the two eigenvectors of $\mathbf{A}$, which we will call $\mathbf{v}_1$ and $\mathbf{v}_2$, so that $\mathbf{c}(t) = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2$ with $\alpha_{1,2}$ are constants. Plugging this form into Equation (9), we find that

$$\mathbf{c}(t+h) = \mathbf{A}(\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2) = \alpha_1\lambda_1\mathbf{v}_1 + \alpha_2\lambda_2\mathbf{v}_2, \tag{13}$$

where $\lambda_{1,2}$ are the eigenvalues to the two eigenvectors, which are the multiplicative factor in which the Fourier modes change. After m timesteps, we would have

$$\mathbf{c}(t+h) = \alpha_1\lambda_1^m\mathbf{v}_1 + \alpha_2\lambda_2^m\mathbf{v}_2, \tag{14}$$

meaning that the solution is stable if the both eigenvalues are ≤ 1. The eigenvalues are given by the determinant equation, $\mathbf{A} - \lambda\mathbf{I} = 0$. By using $\mathbf{A}$ from Equation (12), we find that $\lambda = 1 \pm ihr$. Both solutions have the same magnitude,

$$|\lambda| = \sqrt{1 + h^2 r^2} = \sqrt{1 + \frac{4h^2 v^2}{a^2} \sin^2 \frac{ka}{2}}, \tag{15}$$

which is always greater than unity! This means that even for the small choices of h, the numerical solution will diverge over a long enough time. Thus, we need to find another method to solve the wave equation accurately.

The Implicit method

One solution to the inherently unstable problems, such as the wave equation, is the *implicit method*. The derivation of this method starts with substituting $h \to -h$ into the wave equation,

$$\phi(x, t - h) = \phi(x, t) - h\psi(x, t) \tag{16}$$

$$\psi(x, t - h) = \phi(x, t) - h\frac{v^2}{a^2} \left[\phi(x + a, t) + \phi(x - a, t) - 2\phi(x, t)\right]$$

that tell us how to go backwards in time. But if we make a second substitution $t \to t + h$ and rearrange, we find that

$$\phi(x, t) = \phi(x, t + h) - h\psi(x, t + h) \tag{17}$$

$$\psi(x, t) = \psi(x, t + h) - h\frac{v^2}{a^2} \left[\phi(x + a, t + h) + \phi(x - a, t + h) - 2\phi(x, t + h)\right] \tag{18}$$

These two equations give the solution at $t + h$ indirectly. We can view these equations as a set of simultaneous equations in the values of ϕ and ψ at each grid point. We can solve these with standard linear algebra methods, such as Gaussian elimination.

We can perform a *von Neumann* stable analysis on the implicit method to see the differences from FTCS. First, we have

$$\mathbf{B}\mathbf{c}(t + h) = \mathbf{c}(t), \qquad \text{with} \qquad \mathbf{B} = \begin{pmatrix} 1 & -h \\ hr^2 & 1 \end{pmatrix} \tag{19}$$

and $r = (2v/a)\sin(ka/2)$. We can solve for $\mathbf{c}(t + h)$ by multiplying both sides with $\mathbf{B}^{-1}$. This determinant of the inverse of $\mathbf{B}$ gives eigenvalues

$$\lambda = \frac{1 \pm ihr}{1 + h^2 r^2} \tag{20}$$

$$|\lambda| = (1 + h^2 r^2)^{-1/2}. \tag{21}$$

This value is *always less than one* and is thus **unconditionally stable**. However, being stable does not mean the solution is always correct. We have just shown that the solution converges to one (i.e. all modes except for $k = 0$ one decay), which we know is not correct because an ideal wave will propagate forever. The implicit method over-corrects the errors and can produce an incorrect result.

The Crank-Nicolson method

An optimal method would lie between the FTCS and implicit methods that doesn't exponentially diverge or decay. Such a method is called the *Crank-Nicolson method*, which is derived by taking the average of the FTCS (previous lecture notes) and Implicit methods (Equation 17),

$$\phi(x,t+h) - \tfrac{1}{2}h\psi(x,t+h) = \phi(x,t) + \tfrac{1}{2}\psi(x,t), \tag{22}$$

$$\psi(x,t+h) - h\frac{v^2}{2a^2}\left[\phi(x+a,t+h) + \phi(x-a,t+h) - 2\phi(x,t+h)\right] \tag{23}$$

$$= \psi(x,t) + h\frac{v^2}{2a^2}\left[\phi(x+a,t) + \phi(x-a,t) - 2\phi(x,t)\right]. \tag{24}$$

These equations are indirect, just like the implicit method, so we have to solve them as a set of simultaneous equations.

But we should check whether the Crank-Nicolson method is formally stable. This method has the solution in the form: $\mathbf{Bc}(t+h) = \mathbf{Ac}(t)$, where $\mathbf{A}$ and $\mathbf{B}$ are the same as before, but with $r = (v/a)\sin(ka/2)$. We can rearrange this equation to have only the time-update on the left-hand side to read

$$\mathbf{c}(t+h) = \mathbf{B}^{-1}\mathbf{Ac}(t), \tag{25}$$

where

$$\mathbf{B}^{-1}\mathbf{A} = \frac{1}{1+h^2r^2}\begin{pmatrix} 1 & h \\ -hr^2 & 1 \end{pmatrix}\begin{pmatrix} 1 & h \\ -hr^2 & 1 \end{pmatrix} = \frac{1}{1+h^2r^2}\begin{pmatrix} 1-h^2r^2 & 2h \\ -2hr^2 & 1-h^2r^2 \end{pmatrix} \tag{26}$$

that has eigenvalues of

$$\lambda = \frac{1-h^2r^2 \pm 2ihr}{1+h^2r^2} \tag{27}$$

These eigenvalues have the same magnitude and are *exactly* one, right on the border of unstable (FTCS) and stable (implicit). This suits the wave equation very well, where its solution is neither amplified or suppressed.

Although the Crank-Nicolson method is more complicated than FTCS, it is still relatively fast and only depends on the neighboring grid points. Therefore, one would have a tridiagonal matrix, which can be solved quickly with Gaussian elimination and other optimized methods. This method will be used in a homework problem (Exercise 9.8).

Spectral methods

Until now, we have been focusing on finite differencing methods, but there are other another sets of methods that have fewer stability problems and can give more accurate solutions in particular cases. One method is called the *finite element method* that solves the PDEs in question in small elements of space and time and then stitches together these solutions at the

boundaries of these elements to form a complete solution. However, this method is highly complex, and we will not cover it in this class.

We will focus in this last section of the PDE chapter on *spectral methods* that is less complex and usually gives better accuracy than finite element and differencing methods in particular situations. Let's consider again the wave equation,

$$\frac{\partial^2 \phi}{\partial t^2} = v^2 \frac{\partial^2 \phi}{\partial x^2} \tag{28}$$

that can describe a wave on a string of length L, fixed on both ends ($x = 0$ and $x = L$). Now we can consider a *trial solution* to this PDE,

$$\phi_k(x,t) \sin\left(\frac{\pi k x}{L}\right) e^{ikx}. \tag{29}$$

Assuming that the solution ϕ is real, we could just take the real part of the solution. But in the long run, it is more mathematically convenient to carry around the full complex solution.

As long as k is an integer, this solution satisfies the wave equation at the boundaries provided that

$$\omega = \frac{\pi v k}{L}, \tag{30}$$

where v is velocity. Let's now divide the domain into N equal intervals by $N+1$ grid points with positions

$$x_n = \frac{n}{N} L. \tag{31}$$

The value of our solution at these points is

$$\phi_k(x_n,t) = \sin\left(\frac{\pi k n}{N}\right) \exp\left(i \frac{\pi v k t}{L}\right). \tag{32}$$

Since the wave equation is linear, any linear combination of solutions like these for different values of k is also a solution. Therefore,

$$\phi_k(x_n,t) = \frac{1}{N} \sum_{k=1}^{N-1} b_k \sin\left(\frac{\pi k n}{N}\right) \exp\left(i \frac{\pi v k t}{L}\right) \tag{33}$$

is a solution for any choice of complex coefficients b_k. Note that the summation starts at $k = 1$ because the $k = 0$ term vanishes.

We can use the solution at $t = 0$ to understand how the coefficients behave as a function of time. First, we can express them as real and complex parts: $b_k = \alpha_k + i\eta_k$ and inspect the real part:

$$\phi(x_n,0) = \frac{1}{N} \sum_{k=1}^{N-1} \alpha_k \sin\left(\frac{\pi k n}{N}\right), \tag{34}$$

which is a Fourier sine series. This series can represent any set of solutions at the grid points $\phi(x_n)$ at an arbitrary time. Given this solution (Equation 33), we can also inspect its time derivative,

$$\frac{\partial \phi}{\partial t} = -\left(\frac{\pi v}{L}\right) \frac{1}{N} \sum_{k=1}^{N-1} k\eta_k \sin\left(\frac{\pi k n}{N}\right), \tag{35}$$

which is another sine series with different coefficients. Thus, we can match the initial values of the time derivatives with some choice of η_k.

After we have determined the coefficients b_k, and thus α_k and η_k, we can calculate the solution at **any time t without knowing the solution between the initial time and current time**. Furthermore, the solutions are described with a Fourier series, so we can use FFTs to first calculate the coefficients α_k and η_k, which is much faster than evaluating the sums in the above equations. Once these coefficients are known, we can calculate the solution

$$\phi(x_n, t) = \frac{1}{N} \sum_{k=1}^{N-1} \left[\alpha \cos\left(\frac{\pi v k t}{L}\right) - \eta_k \sin\left(\frac{\pi v k t}{L}\right)\right] \sin\left(\frac{\pi k n}{N}\right). \tag{36}$$

at some time t with an inverse FFT. In principle, the spectral method is much slower per timestep. But since we don't have to "step through" the time evolution, it can be faster than FTCS methods. Spectral methods do have limitations in that (1) it only works well when the boundary conditions are simple, such as constant values and well-defined geometry (walls, spheres, etc.), and (2) it only works with linear differential equations.