Von Neumann Stability Analysis

In finite difference methods for initial value problems, we will typically introduce some truncation error due to our inability to write exact expressions for derivatives. We will also commonly have roundoff error, although this tends to be subdominant. Due to these errors, there will be some difference between a "true," analytic solution $f_{\rm true}$, and an approximate numerical solution $f_{\rm num}$,

$$f_{\rm num} = f_{\rm true} + \epsilon\,. \tag{1}$$

In general, a method for solving initial value PDEs is considered to be stable if the magnitude of the error $|\epsilon|$ does not increase during integration. It suffices to examine the behavior of the error when we substitute the error into a finite difference equation.

We will first consider the 1+1 advection equation using the most simple example of a time integration method–Euler's method–and right-handed differencing. We can generalize these ideas later. Our differential equation of interest is

$$\partial_t f = \partial_x f\,. \tag{2}$$

Given knowledge of a function a time t, or $f(x,t)$, Euler's method allows us to determine function values at a later time $t+\Delta t$,

$$f(x, t+\Delta t) = f(x,t) + \Delta t \partial_x f\,, \tag{3}$$

while right-handed finite differencing then allows us to replace the spatial derivative expression with a discrete form,

$$f(x, t+\Delta t) = f(x,t) + \frac{\Delta t}{\Delta x}\left[f(x+\Delta x, t) - f(x,t)\right]\,. \tag{4}$$

While the numerical solution is determined directly by this expression, the true solution will obey this form as well up to truncation error,

$$\begin{aligned} f_{\rm num}(x,t+\Delta t) =& f_{\rm num}(x,t) + \frac{\Delta t}{\Delta x}\left[f_{\rm num}(x+\Delta x,t) - f_{\rm num}(x,t)\right] \\ f_{\rm true}(x,t+\Delta t) =& f_{\rm true}(x,t) + \frac{\Delta t}{\Delta x}\left[f_{\rm true}(x+\Delta x,t) - f_{\rm true}(x,t)\right] + \mathcal{O}(\Delta x^2, \Delta t^2)\,. \end{aligned} \tag{5}$$

Combining these expressions with Eq. 1, we obtain an expression for the error,

$$\epsilon(x, t+\Delta t) = \epsilon(x,t) + \frac{\Delta t}{\Delta x}\left[\epsilon(x+\Delta x, t) - \epsilon(x,t)\right]\,. \tag{6}$$

At this point, we see that the error itself will obey the advection equation; in fact, for Eulerian integration of systems linear in f, the error will generally follow the original differential equation. For higher-order methods, or nonlinear equations, this may not be true.

While this expression provides us with point-wise information about the error, it does not tell us in a more global sense if the error will be bounded. To get a handle on this latter idea, it is useful to consider the finite difference expression in Fourier space. Suppose we have a function $f(x,t)$. We can Fourier transform one coordinate of this function,

$$f(x,t) = \int \frac{dk}{2\pi}\tilde{f}(k,t)e^{ikx}. \tag{7}$$

For functions of many variables, we can perform an analogous transformation in each coordinate we wish. For now, transforming the spatial variable in Eq. 6 gives us

$$\int \frac{dk}{2\pi}e^{ikx}\tilde{\epsilon}(k,t+\Delta t) = \int \frac{dk}{2\pi}e^{ikx}\tilde{\epsilon}(k,t) + \frac{\Delta t}{\Delta x}\left[\int \frac{dk}{2\pi}e^{ik(x+\Delta x)}\tilde{\epsilon}(k,t) - \int \frac{dk}{2\pi}e^{ikx}\tilde{\epsilon}(k,t)\right]\,.$$

We can simplify this expression by dropping the integral and defining $E_k(t) \equiv \tilde{\epsilon}(k,t)$. E_k is just the (time-dependent) Fourier coefficient of the mode with wavevector k.

$$e^{ikx}E_k(t+\Delta t) = e^{ikx}E_k(t) + \frac{\Delta t}{\Delta x}\left[e^{ik(x+\Delta x)}E_k(t) - e^{ikx}E_k(t)\right]\,.$$

We can further consider this equation for a single (arbitrary) mode k, and drop the index k on the function E_k. Considering the behavior of a single mode can also provides us with a way to analyze solutions more generally, where we might have a nonlinear system and cannot perform a similar procedure.

For the case at hand, the expression simplifies to

$$E_k(t+\Delta t) = E_k(t)\left(1 + \frac{\Delta t}{\Delta x}\left[e^{ik\Delta x} - 1\right]\right).$$

For stability, we would like to know if we can choose values of Δt and/or Δx such that the error decreases for all values of k and c (which we can restrict to be a finite number),

$$|E_k(t+\Delta t)| \leq |E_k(t)|.$$

This expression reduces to

$$\left|1 + \frac{\Delta t}{\Delta x}\left[e^{ik\Delta x} - 1\right]\right| \leq 1.$$

For a thorough stability assessment, we should consider all possible values k (and the combination $k\Delta x$) can take. In this case, it suffices to consider a "worse case" scenario. When $e^{ik\Delta x} = -1$, the inequality becomes

$$\begin{aligned} \left|1 - 2\frac{\Delta t}{\Delta x}\right| &\leq 1 \\ 0 < \frac{\Delta t}{\Delta x} &\leq 1, \end{aligned}$$

so through a careful choice of Δt and Δx we can ensure stability.