Finite Diference Numerical Methods

July 22, 2017

Partial differential equations are often classified. Equations with the same classification have qualitatively similar mathematical and physical properties. The heat equation $\left(\dfrac{\partial u}{\partial t} = k\dfrac{\partial^2 u}{\partial x^2}\right)$ is an example of a parabolic partial differential equation. Solutions usually exponentially decay in time and approach an equilibrium solution. Information and discontinuities propagate at an infinite velocity. The wave equation $\left(\dfrac{\partial^2 u}{\partial t^2} = c^2\dfrac{\partial^2 u}{\partial x^2}\right)$ typifies hyperbolic partial differential equations. There are modes of vibration. Information propagates at a finite velocity and thus discontinuities persist. Laplace's equation $\left(\dfrac{\partial^2 u}{\partial x^2} + \dfrac{\partial^2 u}{\partial y^2} = 0\right)$ is an example of an elliptic partial differential equation. Solutions usually satisfy maximum principles. The terminology parabolic, hyperbolic, and elliptic result from transformation properties of the conic sections.

1 Finite Differences and Truncated Taylor Series

The fundamental technique for finite difference numerical calculations is based on polynomial approximations to $f(x)$ near $x = x_0$. Let $x = x_0 + \Delta x$, so $\Delta x = x - x_0$,

$$f(x) \approx f(x_0) + (x - x_0)f'(x_0) + \frac{(x - x_0)^2}{2!}f''(x_0) = f(x_0) + \Delta x f'(x_0) + \frac{(\Delta x)^2}{2!}f''(x_0) \ , \quad (1)$$

A formula for the error in these polynomial approximations is obtained from

$$f(x) = f(x_0) + \Delta x f'(x_0) + \cdots + \frac{(\Delta x)^n}{n!}f^{(n)}(x_0) + R_n \ , \quad (2)$$

known as the Taylor series with remainder. The remainder R_n (called the truncation error) is known to be in the form of the next term of the series, but evaluated at a usually unknown intermediate point:

$$R_n = \frac{(\Delta x)^{(n+1)}}{(n+1)!}f^{(n+1)}(\xi_{n+1}) \quad (3)$$

where $x_0 < \xi_{n+1} < x = x_0 + \Delta x$. For this to be valid, $f(x)$ must have $n + 1$ continuous derivatives.

The error in the tangent line approximation is given

$$f(x_0 + \Delta x) = f(x_0) + \Delta x f'(x_0) + \frac{(\Delta x)^2}{2!}f''(x_0) \ , \quad (4)$$

called the extended mean value theorem. If Δx is small, then ξ_2 is contained in a small interval, and the truncation error is almost determined (provided that $\mathrm{d}^2 f/\mathrm{d}x^2$ is continuous),

$$R \approx \frac{(\Delta x)^2}{2}f''(x_0) \ .$$

It is said that the truncation error is $O(\Delta x)^2$, "order delta-x squared", meaning that

$$|R| \leqslant C(\Delta x)^2 \ ,$$

since we usually assume that $\mathrm{d}^2f/\mathrm{d}x^2$ is bounded ($|\mathrm{d}^2f/\mathrm{d}x^2| < M$). Thus, $C = M/2$.

1.1 First derivative approximations

$$\frac{\mathrm{d}f}{\mathrm{d}x}(x_0) = \frac{f(x_0+\Delta x) - f(x_0)}{\Delta x} - \frac{\Delta x}{2}\frac{\mathrm{d}^2f}{\mathrm{d}x^2}(\xi_2) \ .$$

The forward difference approximation

$$\frac{\mathrm{d}f}{\mathrm{d}x}(x_0) \approx \frac{f(x_0+\Delta x) - f(x_0)}{\Delta x} \qquad (5)$$

$$\frac{\mathrm{d}f}{\mathrm{d}x}(x_0) = \frac{f(x_0-\Delta x) - f(x_0)}{-\Delta x} + \frac{\Delta x}{2}\frac{\mathrm{d}^2f}{\mathrm{d}x^2}(\bar{\xi}_2) \ .$$

The backward difference approximation

$$\frac{\mathrm{d}f}{\mathrm{d}x}(x_0) \approx \frac{f(x_0) - f(x_0-\Delta x)}{\Delta x} \qquad (6)$$

The truncation error is $O(\Delta x)$ and nearly identical for both forward and backward difference approximations of the first derivative.

$$\begin{aligned}
f(x_0+\Delta x) &= f(x_0) + \Delta x f'(x_0) + \frac{(\Delta x)^2}{2!}f''(x_0) + \frac{(\Delta x)^3}{3!}f'''(x_0) + \cdots \ , \\
f(x_0-\Delta x) &= f(x_0) - \Delta x f'(x_0) + \frac{(\Delta x)^2}{2!}f''(x_0) - \frac{(\Delta x)^3}{3!}f'''(x_0) + \cdots \ ,
\end{aligned}$$

$$\begin{aligned}
f(x_0+\Delta x) - f(x_0-\Delta x) &= 2\Delta x f'(x_0) + \frac{2}{3!}(\Delta x)^3 f'''(x_0) \\
f'(x_0) &= \frac{f(x_0+\Delta x) - f(x_0-\Delta x)}{2\Delta x} - \frac{(\Delta x)^2}{6}f'''(\xi_3)
\end{aligned}$$

The centered difference approximation

$$\frac{\mathrm{d}f}{\mathrm{d}x}(x_0) \approx \frac{f(x_0+\Delta x)-f(x_0-\Delta x)}{2\Delta x} \, , \tag{7}$$

the truncation error is $O(\Delta x)^2$. However, it is not always better to use the centered difference formula.

1.2 Second derivative

$$\begin{aligned} f(x_0+\Delta x)+f(x_0-\Delta x) &= 2f(x_0)+(\Delta x)^2 f''(x_0)+\frac{2(\Delta x)^4}{4!}f''''(x_0)+\cdots \\ f''(x_0) &= \frac{f(x_0+\Delta x)-2f(x_0)+f(x_0-\Delta x)}{(\Delta x)^2}-\frac{(\Delta x)^2}{12}f''''(\xi) \end{aligned}$$

The centered difference approximation for the second derivative

$$\frac{\mathrm{d}^2f}{\mathrm{d}x^2}(x_0) \approx \frac{f(x_0+\Delta x)-2f(x_0)+f(x_0-\Delta x)}{(\Delta x)^2} \, , \tag{8}$$

with an $O(\Delta x)^2$ truncation error. In general, the weights must sum to zero for any finite difference approximation to any derivative.

1.3 Partial derivatives

If $u(x,y)$, for $\dfrac{\partial u}{\partial x}$

$$\frac{\partial u}{\partial x}(x_0,y_0) \approx \frac{u(x_0+\Delta x,y_0)-u(x_0-\Delta x,y_0)}{2\Delta x}$$

for $\dfrac{\partial u}{\partial y}$

$$\frac{\partial u}{\partial y}(x_0,y_0) \approx \frac{u(x_0,y_0+\Delta y)-u(x_0,y_0-\Delta y)}{2\Delta y}$$

The Laplacian $\Delta u = \dfrac{\partial^2 u}{\partial x^2} + \dfrac{\partial^2 u}{\partial y^2}$,

$$\Delta u \approx \frac{u(x_0+\Delta x, y_0) - 2u(x_0, y_0) + u(x_0 - \Delta x, y_0)}{(\Delta x)^2} + \frac{u(x_0, y_0+\Delta y) - 2u(x_0, y_0) + u(x_0, y_0 - \Delta y)}{(\Delta y)^2} , \tag{9}$$

the error is the largest of $O(\Delta x)^2$ and $O(\Delta y)^2$. Let $\Delta x = \Delta y$, the standard five-point finite difference approximation is

$$\Delta u \approx \frac{u(x_0+\Delta x, y_0) + u(x_0 - \Delta x, y_0) + u(x_0, y_0+\Delta y) + u(x_0, y_0 - \Delta y) - 4u(x_0, y_0)}{(\Delta x)^2} , \tag{10}$$

2 Heat Equation

2.1 Homogeneous Problems

Consider the one-dimensional heat equation without sources on a finite interval $0 < x < L$:

$$\frac{\partial u}{\partial t} = k \frac{\partial^2 u}{\partial x^2}$$

$$u(0,t) = 0$$

$$u(L,t) = 0$$

$$u(x,0) = f(x)$$

$$\frac{\partial u}{\partial t}(x_0, t_0) = \frac{u(x_0, t_0+\Delta t) - u(x_0, t_0)}{\Delta t} - \frac{\Delta t}{2} \frac{\partial^2 u}{\partial t^2}(x_0, \eta_1) ,$$

where $t_0 < \eta_1 < t_0 + \Delta t$.

$$\frac{\partial^2 u}{\partial x^2}(x_0, t_0) = \frac{u(x_0+\Delta x, t_0) - 2u(x_0, t_0) + u(x_0 - \Delta x, t_0)}{(\Delta x)^2} - \frac{(\Delta x)^2}{12} \frac{\partial^4 u}{\partial x^4}(\xi_1, t_0) , \tag{11}$$

where $x_0 < \xi_1 < x_0 + \Delta x$. The heat equation at any point $x = x_0, t = t_0$,

$$\frac{u(x_0, t_0 + \Delta t) - u(x_0, t_0)}{\Delta t} = k\frac{u(x_0 + \Delta x, t_0) - 2u(x_0, t_0) + u(x_0 - \Delta x, t_0)}{(\Delta x)^2} + E \ , \qquad (12)$$

where the discretization (or truncation) error is

$$E = \frac{\Delta t}{2}\frac{\partial^2 u}{\partial t^2}(x_0, \eta_1) - \frac{k(\Delta x)^2}{12}\frac{\partial^4 u}{\partial x^4}(\xi_1, t_0) \ . \qquad (13)$$

Introduce the approximation that results by ignoring the truncation error:

$$\frac{u(x_0, t_0 + \Delta t) - u(x_0, t_0)}{\Delta t} \approx k\frac{u(x_0 + \Delta x, t_0) - 2u(x_0, t_0) + u(x_0 - \Delta x, t_0)}{(\Delta x)^2} \qquad (14)$$

Introduce $\tilde{u}(x_0, t_0)$ an approximation at the point $x = x_0, t = t_0$ of the exact solution $u(x_0, t_0)$. Let the approximation $u(x_0, t_0)$ solve

$$\frac{\tilde{u}(x_0, t_0 + \Delta t) - \tilde{u}(x_0, t_0)}{\Delta t} = k\frac{\tilde{u}(x_0 + \Delta x, t_0) - 2\tilde{u}(x_0, t_0) + \tilde{u}(x_0 - \Delta x, t_0)}{(\Delta x)^2} \ , \qquad (15)$$

$\tilde{u}(x_0, t_0)$ is the exact solution of an equation that is only approximately correct. We hope that the desired solution $u(x_0, t_0)$ is accurately approximated by $\tilde{u}(x_0, t_0)$.

Introduce a uniform mesh Δx and a constant discretization time Δt. Divide the rod of length L into N equal intervals, $\Delta x = L/N$. We have $x_0 = 0, x_1 = \Delta x, x_2 = 2\Delta x, \cdots, x_N = N\Delta x = L$.

$$x_j = j\Delta x \ . \qquad (16)$$

Introduce time step sizes Δt,

$$t_m = m\Delta t \ . \qquad (17)$$

The exact temperature at the mesh point $u(x_j, t_m)$ is approximately $\tilde{u}(x_j, t_m)$. Introduce

$$\tilde{u}(x_j, t_m) \equiv u_j^{(m)} \ , \qquad (18)$$

indicating the exact solution of (**??**) at the jth mesh point at time t_m. Equation (**??**) will be satisfied at each mesh point $x_0 = x_j$ at each time $t_0 = t_m$ (excluding the space-time boundaries). $x_0 + \Delta x$ becomes $x_j + \Delta x = x_{j+1}$ and $t_0 + \Delta t$ becomes $t_m + \Delta t = t_{m+1}$.

$$\frac{u_j^{(m+1)} - u_j^{(m)}}{\Delta t} = k\frac{u_{j+1}^{(m)} - 2u_j^{(m)} + u_{j-1}^{(m)}}{(\Delta x)^2} \ , \tag{19}$$

for $j = 1, \cdots, N-1$ and m starting from 1. It is called a partial difference equation. The local truncation error is given by (**??**). It is the larger of $O(\Delta t)$ and $O(\Delta x)^2$. Since $E \to 0$ as $\Delta x \to 0$ and $\Delta t \to 0$, the approximation (**??**) is said to be consistent with the partial differential equation.

$u_j^{(m)}$ satisfies the initial conditions (at the mesh points)

$$u_j^(0) = u(x,0) = f(x) = f(x_j) \ , \tag{20}$$

where $x_j = j\Delta x$ for $j = 0, \cdots, N$. $u_j^{(m)}$ also satisfies the boundary conditions (at each time step)

$$u_0^{(m)} = u(0,t) = 0 \ , u_N^{(m)} = u(L,t) = 0 \ . \tag{21}$$

$$u_j^{(m+1)} = u_j^{(m)} + s\left(u_{j+1}^{(m)} - 2u_j^{(m)} + u_{j-1}^{(m)}\right) \tag{22}$$

where s is a dimensionless parameter,

$$s = k\frac{\Delta t}{(\Delta x)^2} \ . \tag{23}$$

$u_j^{(m+1)}$ is a linear combination of the specified three earlier values. We begin our computation using the initial condition $u_j^{(0)} = f(x_j)$, for $j = 1, \cdots, N-1$. For mesh points adjacent to the boundary (i.e., $j = 1$ or $j = N-1$), (**??**) requires the solution on the boundary points ($j = 0$ or $j = N$). We obtain these values from the boundary conditions.

Propagation speed of disturbances. Suppose that the initial conditions at the mesh points are zero except for 1 at some interior mesh point far from the boundary. At the first time step, the solution is zero everywhere except at the original nonzero mesh point and its two immediate neighbors. This process continues. The isolated initial nonzero value spreads out at a constant speed (until the boundary has been reached). This disturbance propagates at velocity $\Delta x/\Delta t$. However, for the heat equation, disturbances move at an infinite speed. In some sense the numerical scheme poorly approximates this property of the heat equation. However if the parameter s is fixed, the numerical propagation speed is

$$\frac{\Delta x}{\Delta t} = \frac{k\Delta x}{s(\Delta x)^2} = \frac{k}{s\Delta x} \tag{24}$$

As $\Delta x \to 0$ (with s fixed), this speed approaches ∞ as is desired.

2.2 Separation of Variables

Consider

$$\begin{aligned} u_j^{(m+1)} &= u_j^{(m)} + s\left(u_{j+1}^{(m)} - 2u_j^{(m)} + u_{j-1}^{(m)}\right) \\ u_j^{(0)} &= f(x_j) = f_j \\ u_0^{(m)} &= 0 \\ u_N^{(m)} &= 0 \end{aligned}$$

where $s = \dfrac{k\Delta t}{(\Delta x)^2}, x_j = j\Delta x, t = m\Delta t$. Assume the equation has special product solutions

$$u_j^{(m)} = \phi_j h_m \ , \tag{25}$$

then

$$\phi_j h_{m+1} = \phi_j h_m + s(\phi_{j+1} h_m - 2\phi_j h_m + \phi_{j-1} h_m) \ .$$

$$\frac{h_{m+1}}{h_m} = 1 + s\left(\frac{\phi_{j+1} + \phi_{j-1}}{\phi_j} - 2\right) = +\lambda$$

where λ is a separation constant.

$$h_{m+1} = +\lambda h_m \ , \tag{26}$$

$$\phi_{j+1} + \phi_{j-1} = \left(\frac{-\lambda + 1 - 2s}{s}\right)\phi_j \ , \tag{27}$$

with two homogeneous boundary conditions

$$\phi_0 = 0 \ , \ \phi_N = 0 \ . \tag{28}$$

First-order difference equations,

$$h_m = \lambda^m h_0 \ , \tag{29}$$

where h_0 is an initial condition for the first-order difference equation. Or assume that a homogeneous solution exists in the form $h_m = Q^m$. Thus $Q^{m+1} = \lambda Q^m$ or $Q = \lambda$. If $\lambda > 1$, then the solution exponentially grows [$\lambda^m = e^{m \log \lambda} = e^{(\log \lambda/\Delta t)t}$, since $m = t/\Delta t$]. If $0 < \lambda < 1$, the solution exponentially decays. If $-1 < \lambda < 0$, the solution has an oscillatory (and exponential) decay, known as a convergent oscillation. If $\lambda < -1$, the solution has a divergent oscillation. If λ is complex, $\lambda = re^{i\theta}$, $r = |\lambda|$ and $\theta = \arg \lambda$ (or angle),

$$\lambda^m = r^m e^{im\theta} = |\lambda|^m(\cos m\theta + i \sin m\theta) \ .$$

The solution grows in discrete time m if $|\lambda| > 1$ and decays if $|\lambda| < 1$.

The solution λ^m of $h_{m+1} = \lambda h_m$ remains bounded as m increases (t increases) if $|\lambda| \leqslant 1$. It grows if $|\lambda| > 1$.

Second-order difference equations, assume $\phi_j = Q^j$. The boundary conditions, $\phi_0 = \phi_N = 0$, suggest the solution may oscillate. This usually occurs if Q is complex with $|Q| = 1$,

$$\phi_j = (|Q|e^{i\theta})^j = e^{i\theta j} = e^{i\theta(x/\Delta x)} = e^{i\alpha x} \ , \tag{30}$$

where $\alpha = \theta/\Delta x = (\arg Q)/\Delta x$.

$$\begin{aligned} e^{i\alpha x} + e^{-i\alpha x} &= \frac{\lambda - 1 + 2s}{s} \\ 2\cos(\alpha\Delta) &= \frac{\lambda - 1 + 2s}{s} \end{aligned}$$

$$\phi_j = c_1 \sin \alpha x + c_2 \cos \alpha x \ . \tag{31}$$

The boundary conditions, $\phi_0 = \phi_N = 0$, imply that $c_2 = 0$ and $\alpha = n\pi/L$, where $n = 1, 2, 3, \cdots$. Thus,

$$\phi_j = \sin \frac{n\pi x}{L} = \sin \frac{n\pi j\Delta x}{L} = \sin \frac{n\pi j}{N} \tag{32}$$

2.3 Fourier-von Neumann Stability Analysis

$$u_j^{(m)} = e^{i\alpha x} Q^{t/\Delta t} = e^{i\alpha j\Delta x} Q^m \ .$$

$$Q = 1 + s(e^{i\alpha\Delta x} - 2 + e^{-i\alpha\Delta x}) = 1 - 2s\left[1 - \cos\left(\alpha\Delta x\right)\right]$$

Since

$$u_j^{(m)} = \sin \frac{n\pi x}{L} Q^{t/\Delta t} \ ,$$

where $\alpha = \dfrac{n\pi}{L}$,

$$Q = 1 - 2s\left[1 - \cos\left(\frac{n\pi\Delta x}{L}\right)\right] \tag{33}$$

and $n = 1, 2, 3, \cdots, N-1$. For partial differential equations, there are an infinite number of eigenfunctions ($\sin n\pi x/L, n = 1, 2, 3, \cdots$). However, for partial difference equation, there are only $N-1$ independent eigenfunctions ($\sin n\pi x/L, n = 1, 2, 3, \cdots, N-1$):

$$\phi_j = \sin \frac{n\pi x}{L} = \sin \frac{n\pi j \Delta x}{L} = \sin \frac{n\pi j}{N} \ . \tag{34}$$

$$u_j^{(m)} = \sum_{n=1}^{N-1} \beta_n \sin \frac{n\pi x}{L} \left[1 - 2s\left(1 - \cos \frac{n\pi}{N}\right)\right]^{t/\Delta t} \ , \tag{35}$$

where

$$s = \frac{k\Delta t}{(\Delta x)^2} \ .$$

These coefficients can be determined from the $N-1$ initial conditions, using the discrete orthogonality of the eigenfunctions $\sin \dfrac{n\pi j}{N}$.

$$u_j^{(m)} = \sin \frac{n\pi x}{L} \left[1 - 2s\left(1 - \cos \frac{n\pi}{N}\right)\right]^{t/\Delta t} \ , \quad n = 1, 2, \cdots, N-1 \tag{36}$$

$$u(x,t) = \sin \frac{n\pi x}{L} e^{-k(n\pi/L)^2 t} \ , \quad n = 1, 2, \cdots \tag{37}$$

where $s = \dfrac{k\Delta t}{(\Delta x)^2}$. For the partial differential equation, each wave exponentially decays, $e^{-k(n\pi/L)^2 t}$. For the partial difference equation, the time dependence is

$$Q^m = \left[1 - 2s\left(1 - \cos \frac{n\pi}{N}\right)\right]^{t/\Delta t} \ , \tag{38}$$

If $|Q| \leqslant 1$ for all solutions, the numerical scheme is stable. Otherwise, the scheme is unstable. If $Q > 1$, there is exponential growth in time, while exponential decay occurs if $0 < Q < 1$. The solution is constant in time if $Q = 1$. There is a convergent oscillation in time ($-1 < Q < 0$), a pure oscillation ($Q = -1$), and a divergent oscillation ($Q < -1$). Convergent oscillations do not duplicate the behavior of the partial differential equation.

$1-2s\left(1-\cos\frac{n\pi}{N}\right)\geqslant -1$, for $n=1,2,3,\cdots,N-1$ or

$$\begin{aligned} s &\leqslant \frac{1}{1-\cos\frac{n\pi}{N}} \text{ for } \ n=1,2,3,\cdots,N-1 \\ \Longrightarrow s &\leqslant \frac{1}{1-\cos\frac{(N-1)\pi}{N}} \\ \Longrightarrow s &\leqslant \frac{1}{2} < \frac{1}{1-\cos\frac{(N-1)\pi}{N}} \end{aligned}$$

Since $1-\cos(N-1)\pi/N<2$, and hence we are guaranteed that the numerical solution will be stable if $s\leqslant\frac{1}{2}$.

If $s>\frac{1}{2}$, usually $Q<-1$ (but not necesarily) for some n. Then the numerical solution will contain a divergent oscillation. It is called numerical instability. If $s>\frac{1}{2}$, the most rapidly "growing" solution corresponds to a rapid oscillation ($n=N-1$) in space. The numerical instability is characterized by divergent oscillation in time ($Q<-1$) of a rapidly oscillatory ($n=N-1$) solution in space.

$$\Delta t\leqslant\frac{(\Delta x)^2}{2k} \tag{39}$$

The time steps Δt must not be too large (otherwise the scheme becomes unstable).

When $s=\frac{1}{2}$,

$$u_j^{(m+1)}=\frac{1}{2}\left[u_{j+1}^{(m)}+u_{j-1}^{(m)}\right] \ .$$

If $n/N\ll 1$, $\cos n\pi/N\approx 1-\frac{1}{2}\left(\frac{n\pi}{N}\right)^2$,

$$Q^{t/\Delta t}\approx\left[1-s\left(\frac{n\pi}{N}\right)^2\right]^{t/\Delta t}=\left[1-k\Delta t\left(\frac{n\pi}{L}\right)^2\right]^{t/\Delta t} \quad ,$$

where $N=L/\Delta x$. As $\Delta t\to 0$,

$$Q^{t/\Delta t}\to e^{-k(n\pi/L)^2 t}$$

$$Q^m - \exp\left[-kt\frac{(n\pi)^2}{L^2}\right] = O(\Delta t) \ .$$

The relationship between convergence and stability can be generalized.

Lax equivalency Theorem

For consistent finite difference approximations of time-dependent linear partial differential equations that are well posed, the numerical scheme converges if it is stable and it is stable if it converges.

$$u_j^{(m+1)} = su_{j-1}^{(m)} + (1-2s)u_j^{(m)} + su_{j+1}^{(m)} \tag{40}$$

2.4 Matrix Notation

For fixed t, $u(x,t)$ is only a function of x. Its discretization $u_j^{(m)}$ is defined at each of the $N+1$ mesh points (at every time step). Introduce a vector $\boldsymbol{u}$ of dimension $N+1$ that changes at each time step. It is a function of m, $u^{(m)}$. The jth component of $\boldsymbol{u}^{(m)}$ is the value of $u(x,t)$ at the jth mesh point :

$$\left(\boldsymbol{u}^{(m)}\right)_j = u_j^{(m)} \tag{41}$$

The partial difference equation is

$$u_j^{(m+1)} = u_j^{(m)} + s\left(u_{j+1}^{(m)} - 2u_j^{(m)} + u_{j-1}^{(m)}\right) \ . \tag{42}$$

Apply the boundary conditions, $u_0^{(m)} = u_N^{(m)} = 0$, then

$$u_1^{(m+1)} = u_1^{(m)} + s\left(u_2^{(m)} - 2u_1^{(m)} + u_0^{(m)}\right) = (1-2s)u_1^{(m)} + su_2^{(m)} \ . \tag{43}$$

At each time step, there are $N-1$ unknowns. Introduce the $N-1 \times N-1$ tridiagonal matrix $\boldsymbol{A}$ with all entries zero except for the main diagonal (with entries $1-2s$) and neighboring diagonals (with entries s) :

$$\boldsymbol{A} = \begin{pmatrix} 1-2s & s & 0 & 0 & 0 & 0 & 0 \\ s & 1-2s & s & 0 & 0 & 0 & 0 \\ 0 & s & 1-2s & s & 0 & 0 & 0 \\ 0 & 0 & \cdots & \cdots & \cdots & 0 & 0 \\ 0 & 0 & 0 & s & 1-2s & s & 0 \\ 0 & 0 & 0 & 0 & s & 1-2s & s \\ 0 & 0 & 0 & 0 & 0 & s & 1-2s \end{pmatrix} \tag{44}$$

The partial difference equation becomes the following vector equation :

$$\boldsymbol{u}^{(m+1)} = \boldsymbol{A}\boldsymbol{u}^{(m)} \ . \tag{45}$$

$$\boldsymbol{u}^{(m)} = \boldsymbol{A}^{m}\boldsymbol{u}^{(0)} \ . \tag{46}$$

The matrix $\boldsymbol{A}$ raised to the mth power describes how the initial condition influences the solution at the mth time step ($t = m\Delta t$).

Introduce the eigenvalues μ of the matrix $\boldsymbol{A}$, the values μ such that there are nontrivial vector solutions $\boldsymbol{\xi}$:

$$\boldsymbol{A}\boldsymbol{\xi} = \mu\boldsymbol{\xi} \ . \tag{47}$$

The eigenvalues satisfy

$$\det[\boldsymbol{A} - \mu \boldsymbol{I}] = 0 \tag{48}$$

where $\boldsymbol{I}$ is the identity matrix. Nontrivial vectors $\boldsymbol{\xi}$ that satisfy (??) are called eigenvectors corresponding to μ. Since $\boldsymbol{A}$ is an $(N-1) \times (N-1)$ matrix, $\boldsymbol{A}$ has $N-1$ eigenvalues. However, some of the eigenvalues may not be distinct, there may be multiple eigenvalues (or degeneracies). For a distinct eigenvalue, there is a unique eigenvector (to within a multiplicative constant). In the case of a multiple eigenvalue (of multiplicity k), there may be at most k linearly independent eigenvectors. If for some eigenvalue there are less than k eigenvectors, the matrix is defective. If $\boldsymbol{A}$ is real and symmetric, it is known that any possible multiple eigenvalues are not defective. The matrix $\boldsymbol{A}$ has $N-1$ eigenvectors (which can be shown to be linearly independent). Furthermore, if $\boldsymbol{A}$ is real and symmetric, the eigenvalues (and consequently the eigenvectors) are real and the eigenvectors are orthogonal. Let μ_n be the nth eigenvalue and $\boldsymbol{\xi}_n$ the corresponding eigenvector.

We can solve vector equation (equivalent to the partial difference equation) using the method of eigenvector expansion.(This technique is analogous to using an eigenfunction expansion to solve the partial differential equation.) Any vector can be expanded in a series of the eigenvectors:

$$\boldsymbol{u}^{(m)} = \sum_{n=1}^{N-1} c_n^{(m)} \boldsymbol{\xi}_n \tag{49}$$

The vector changes with m(time), and the constants $c_n^{(m)}$ depend on m(time) :

$$\boldsymbol{u}^{(m+1)} = \sum_{n=1}^{N-1} c_n^{(m+1)} \boldsymbol{\xi}_n \tag{50}$$

$$\boldsymbol{u}^{(m+1)} = \boldsymbol{A}\boldsymbol{u}^{(m)} = \sum_{n=1}^{N-1} c_n^{(m)} \boldsymbol{A}\boldsymbol{\xi}_n = \sum_{n=1}^{N-1} c_n^{(m)} \mu_n \boldsymbol{\xi}_n \ , \tag{51}$$

$$c_n^{(m+1)} = \mu_n c_n^{(m)} \tag{52}$$

$$c_n^{(m+1)} = c_n^{(0)} (\mu_n)^m \ , \tag{53}$$

$$\boldsymbol{u}^{(m+1)} = \sum_{n=1}^{N-1} c_n^{(0)} (\mu_n)^m \boldsymbol{\xi}_n \tag{54}$$

$c_n^{(0)}$ can be determined from the initial condition. The growth of the solution as t increases (m increases) depends on $(u_n)^m$, where $m = t/\Delta t$. Since Δ_n is real,

$$(\mu_n)^m = \begin{cases} \text{exponential growth} & \mu_n > 1 \\ \text{exponential decay} & 0 < \mu_n < 1 \\ \text{convergent oscillation} & -1 < \mu_n < 0 \\ \text{divergent oscillation} & \mu_n < -1 \end{cases}$$

The numerical solution is unstable if any eigenvalue $\mu_n > 1$ or any $\mu_n < -1$.

Gershgorin circle Theorem

Every eigenvalue of $\boldsymbol{A}$ lies in at least one of the circles $c_1, \cdots, c_{N-1}$ in the complex plane where c_i has its center at the ith diagonal entry and its radius equal to the sum of the absolute values of the rest of that row.

If a_{ij} are the entries of $\boldsymbol{A}$, then all eigenvalues μ lie in at least one of the following circles:

$$|\mu - a_{ii}| \leqslant \sum_{j=1,\ j\neq i}^{N-1} |a_{ij}| \ . \tag{55}$$

Two circles are $|\mu - (1-2s)| < s$ and the other $N-3$ circles are

$$|\mu - (1-2s)| < 2s \tag{56}$$

Since the eigenvalues μ are also known to be real,

$$1 - 4s \leqslant \mu \leqslant 1 \ .$$

Stability is guaranted if $-1 \leqslant \mu \leqslant 1$, and the Gershgorin circle theorem implies that the numerical scheme is stable if $s \leqslant \dfrac{1}{2}$. If $s > \dfrac{1}{2}$, the Gershgorin circle theorem does not imply the scheme is unstable.

2.5 Nonhomogeneous Problems

Consider

$$\begin{aligned}
\frac{\partial u}{\partial t} &= k\frac{\partial^2 u}{\partial x^2} + Q(x,t) \\
u(0,t) &= A(t) \\
u(L,t) &= B(t) \\
u(x,0) &= f(x)
\end{aligned}$$

Using a forward difference in time and a centered difference in space

$$\begin{aligned}
\frac{u_j^{(m+1)} - u_j^{(m)}}{\Delta t} &= \frac{k}{(\Delta x)^2}\left(u_{j+1}^{(m)} - 2u_j^{(m)} + u_{j-1}^{(m)}\right) + Q(j\Delta x, m\Delta t) \\
u_0^{(m)} &= A(m\Delta t) \\
u_N^{(m)} &= B(m\Delta t) \\
u_j^{(m)} &= f(j\Delta x) \ .
\end{aligned}$$

The stability analysis for homogeneous problems is valid for nonhomogeneous problems.Thus, $s = \dfrac{k\Delta t}{(\Delta x)^2} \leqslant \frac{1}{2}$.

2.6 Other Numerical Schemes

The numerical scheme for the heat equation, which uses the centered difference in space and forward difference in time, is stable if $s = k\dfrac{\Delta t}{(\Delta x)^2} \leqslant 2$. The time step is small [being proportional to $(\Delta x)^2$].

2.6.1 Richardson's scheme

Use centered differences in both space and time

$$\frac{u_j^{(m+1)} - u_j^{(m-1)}}{\Delta t} = \frac{k}{(\Delta x)^2}\left[u_{j+1}^{(m)} - 2u_j^{(m)} + u_{j-1}^{(m)}\right] \tag{57}$$

or

$$u_j^{(m+1)} = u_j^{(m-1)} + s\left[u_{j+1}^{(m)} - 2u_j^{(m)} + u_{j-1}^{(m)}\right] \tag{58}$$

The truncation error is the sum of a $(\Delta t)^2$ and $(\Delta x)^2$ terms. This numerical method is always unstable.

2.6.2 Crank-Nicolson scheme

The forward difference in time

$$\frac{\partial u}{\partial t} \approx \frac{u(t+\Delta t) - u(t)}{\Delta t}$$

may be interpreted as the centered difference around $t+\Delta t/2$. The error in approximating $\dfrac{\partial u}{\partial t}(t+\Delta t/2)$ is $O(\Delta t)^2$. We discretize the second derivative at $t+\Delta t/2$ with a centered difference scheme. Since this involves functions evaluated at this in-between time, we take the average at t and $t+\Delta t$.

$$\frac{u_j^{(m+1)} - u_j^{(m)}}{\Delta t} = \frac{k}{2}\left[\frac{u_{j+1}^{(m)} - 2u_j^{(m)} + u_{j-1}^{(m)}}{(\Delta x)^2} + \frac{u_{j+1}^{(m+1)} - 2u_j^{(m+1)} + u_{j-1}^{(m+1)}}{(\Delta x)^2}\right] \tag{59}$$

the truncation error remains the sum of two terms, one $(\Delta x)^2$ and the other $(\Delta t)^2$. The advantage of the Crank-Nicolson method is that the scheme is stable for all $s = \dfrac{k\Delta t}{(\Delta x)^2}$. Δt can be as large as desired.

2.7 Other Types of Boundary Condition

$\dfrac{\partial u}{\partial x} = g(t)$ at $x = 0$ (rather than u being given at $x = 0$). Since the discretization of the partial differential equation has an $O(\Delta x)^2$ truncation error, introduce an equal error in the boundary condition by using a centered difference in space

$$\frac{\partial u}{\partial x} \approx \frac{u(x+\Delta x, t) - u(x-\Delta x, t)}{2\Delta x} \ . \tag{60}$$

The boundary condition $\dfrac{\partial u}{\partial x} = g(t)$ at $x = 0$ becomes

$$\frac{u_1^{(m)} - u_{-1}^{(m)}}{2\Delta x} = g(t) = g(m\Delta t) = g_m \ . \tag{61}$$

The temperature at the fictitious point $(x_{-1} = -\Delta x)$:

$$u_{-1}^{(m)} = u_1^{(m)} - 2\Delta x g_m \ . \tag{62}$$

we determine the value at the fictitious point initially, $u_{-1}^{(0)}$. This fictitious point is needed to compute the boundary temperature at later times via the partial difference equation. At $x = 0 (j = 0)$

$$\begin{aligned} u_0^{(m+1)} &= u_0^{(m)} + s\left(u_1^{(m)} - 2u_0^{(m)} + u_{-1}^{(m)}\right) \\ &= u_0^{(m)} + s\left(u_1^{(m)} - 2u_0^{(m)} + u_1^{(m)} - 2\Delta x g_m\right) \end{aligned}$$

2.8 Two-Dimensional Heat Equation

$$\frac{\partial u}{\partial t} = k\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right)$$

Introduce a two-dimensional mesh(or latice), where we asume that $\Delta x = \Delta y$,

$$\frac{u_{j,l}^{(m+1)} - u_{j,l}^{(m)}}{\Delta t} = \frac{k}{(\Delta x)^2}\left[u_{j+1,l}^{(m)} + u_{j-1,l}^{(m)} + u_{j,l+1}^{(m)} + u_{j,l-1}^{(m)} - 4u_{j,l}^{(m)}\right] \tag{63}$$

where $u_{j,l}^{(m)} \approx u(j\Delta x, l\Delta y, m\Delta t)$.

3 Partial Differential Equations

[?] Partial differential equations (PDEs) are classified into the three categories, hyperbolic, parabolic, and elliptic, on the basis of their characteristics, or curves of information propagation. The prototypical example of a hyperbolic equation is the one-dimensional wave equation

$$\frac{\partial^2 u}{\partial t^2} = v^2 \frac{\partial^2 u}{\partial x^2} \tag{64}$$

where v = constant is the velocity of wave propagation. The prototypical parabolic equation is the diffusion equation

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(D\frac{\partial u}{\partial x}\right) \tag{65}$$

where D is the diffusion coefficient. The prototypical elliptic equation is the Poisson equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \rho(x, y) \tag{66}$$

Equations (**??**) and (**??**) both define initial value or Cauchy problems: If information on u (perhaps including time derivative information) is given at some initial time t_0 for all x, then the equations describe how $u(x,t)$ propagates itself forward in time. In other words, equations (**??**) and (**??**) describe time evolution.

Equation (**??**) directs us to find a single "static" function $u(x,y)$ which satisfies the equation within some (x,y) region of interest, and which - one must also specify - has some desired behavior on the boundary of the region of interest. These problems are called boundary value problems. It is not possible stably to just "integrate in from the boundary" in the same sense that an initial value problem can be "integrated forward in time".

4 Initial Value Problems

One's principal computational concern must be the stability of the algorithm.

4.1 Flux-Conservative Initial Value Problems

The flux-conservative equation

$$\frac{\partial \boldsymbol{u}}{\partial t} = -\frac{\partial \boldsymbol{F}}{\partial x} \tag{67}$$

where $\boldsymbol{u}$ and $\boldsymbol{F}$ are vectors, and where (in some cases) $\boldsymbol{F}$ may depend not only on $\boldsymbol{u}$ but also on spatial derivatives of $\boldsymbol{u}$. The vector $\boldsymbol{F}$ is called the conserved flux.

The one-dimensional wave equation with constant velocity of propagation v is

$$\frac{\partial^2 u}{\partial t^2} = v^2 \frac{\partial^2 u}{\partial x^2} \tag{68}$$

can be rewritten as a set of two first-order equations

$$\frac{\partial r}{\partial t} = v\frac{\partial s}{\partial x} \tag{69}$$

$$\frac{\partial s}{\partial t} = v\frac{\partial r}{\partial x} \tag{70}$$

where

$$r \equiv v\frac{\partial u}{\partial x} \tag{71}$$

$$s \equiv \frac{\partial u}{\partial t} \tag{72}$$

$$\boldsymbol{F}(\boldsymbol{u}) = \begin{pmatrix} 0 & -v \\ \\ -\,v & 0 \end{pmatrix} \cdot \boldsymbol{u} \tag{73}$$

The equation for a scalar u,

$$\frac{\partial u}{\partial t} = -v\frac{\partial u}{\partial x} \tag{74}$$

with v a constant. The general solution of this equation is a wave propagating in the positive x-direction,

$$u = f(x - vt) \tag{75}$$

where f is an arbitrary function.

Choose equally spaced points along both the t- and x-axes,

$$x_j = x_0 + j\Delta x \tag{76}$$

$$t_n = t_0 + n\Delta t \tag{77}$$

Let u_j^n denote $u(t_n, x_j)$.

$$\left.\frac{\partial u}{\partial t}\right|_{j,n} = \frac{u_j^{n+1} - u_j^n}{\Delta t} + O(\Delta t) \tag{78}$$

which is called forward Euler differencing. While forward Euler is only first-order accurate in Δt, one is able to calculate quantities at timestep $n+1$ in terms of only quantities known at timestep n.

$$\left.\frac{\partial u}{\partial x}\right|_{j,n} = \frac{u_{j+1}^n - u_{j-1}^n}{2\Delta x} + O(\Delta x^2) \tag{79}$$

The FTCS representation (Forward Time Centered Space) is

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} = -v\left(\frac{u_{j+1}^n - u_{j-1}^n}{2\Delta x}\right) \tag{80}$$

The FTCS representation is an explicit scheme. This means that u_j^{n+1} for each j can be calculated explicitly from the quantities that are already known. The implicit schemes require us to solve implicit equations coupling the u_j^{n+1} for various j. The FTCS algorithm is also an example of a single-level scheme, since only values at time level n have to be stored to find values at time level $n+1$.

4.2 von Neumann Stability Analysis

The von Neumann analysis is local: Imagine that the coefficients of the difference equations are so slowly varying as to be considered constant in space and time. In that case, the independent solutions, or eigenmodes, of the difference equations are all of the form

$$u_j^n = \xi^n e^{ikj\Delta x} \tag{81}$$

where k is a real spatial wave number (which can have any value) and $\xi = \xi(k)$ is a complex number that depends on k. The time dependence of a single eigenmode is nothing more than successive integer powers of the complex number ξ. The difference equations are unstable (have exponentially growing modes) if $|\xi(k)| > 1$ for some k. The number ξ is called the amplification factor at a given wave number k.

$$\xi(k) = 1 - i\frac{v\Delta t}{\Delta x}\sin k\Delta x \tag{82}$$

whose modulus is > 1 for all k. so the FTCS scheme is unconditionally unstable.

If the velocity v were a function of t and x, then we would write v_j^n in equation (**??**). In the von Neumann stability analysis v would still be treated as a constant, the idea being that for v slowly varying the analysis is local. Even in the case of strictly constant v, the von Neumann analysis does not rigorously treat the end effects at $j = 0$ and $j = N$.

If the equation's right-hand side were nonlinear in u, a von Neumann analysis would linearize by writing $u = u_0 + \delta u$, expanding to linear order in δu. Assuming that the u_0 quantities already satisfy the difference equation exactly, the analysis would look for an unstable eigenmode of δu.

4.2.1 Lax Method

The instability in the FTCS method can be cured by a simple change due to Lax. Replaces the term u_j^n in the time derivative term by its average

$$u_j^n \to \frac{1}{2}(u_{j+1}^n + u_{j-1}^n) \tag{83}$$

$$u_j^{n+1} = \frac{1}{2}(u_{j+1}^n + u_{j-1}^n) - \frac{v\Delta t}{2\Delta x}\left(u_{j+1}^n - u_{j-1}^n\right) \tag{84}$$

The amplification factor is

$$\xi = \cos k\Delta x - i\frac{v\Delta t}{\Delta x}\sin k\Delta x \tag{85}$$

The stability condition $|\xi|^2 \leqslant 1$ leads to the requirement

$$\frac{|v|\Delta t}{\Delta x} \leqslant 1 \tag{86}$$

It is Courant-Friedrichs-Lewy stability criterion, often called simply the Courant condition.

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} = -v\left(\frac{u_{j+1}^n - u_{j-1}^n}{2\Delta x}\right) + \frac{1}{2}\left(\frac{u_{j+1}^n - 2u_j^n + u_{j-1}^n}{\Delta t}\right) \ , \tag{87}$$

which is exactly the FTCS representation of the equation

$$\frac{\partial u}{\partial t} = -v\frac{\partial u}{\partial x} + \frac{(\Delta x)^2}{2\Delta t}\nabla^2 u \ . \tag{88}$$

The Lax scheme is said to have numerical dissipation, or numerical viscosity. Unless $|v|\Delta t$ is exactly equal to Δx, $|\xi| < 1$ and the amplitude of the wave decreases spuriously.

The scales to study accurately are those that encompass many grid points, so that they have $k\Delta x \ll 1$. For these scales, the amplification factor can be seen to be very close to one, in both the stable and unstable schemes. The stable and unstable schemes are therefore about equally accurate. For the unstable scheme, however, short scales with $k\Delta x \sim 1$, which we are not interested in, will blow up and swamp the interesting part of the solution. Much better to have a stable scheme in which these short wavelengths die away innocuously. Both the stable and the unstable schemes are inaccurate for these short wavelengths, but the inaccuracy is of a tolerable character when the scheme is stable.

When the independent variable $\boldsymbol{u}$ is a vector,

$$\frac{\partial}{\partial t}\begin{bmatrix} r \\ \\ \\ s \end{bmatrix} = \frac{\partial}{\partial x}\begin{bmatrix} vs \\ \\ \\ vr \end{bmatrix} \tag{89}$$

The Lax method for this equation is

$$r_j^{n+1} = \frac{1}{2}(r_{j+1}^n + r_{j-1}^n) - \frac{v\Delta t}{2\Delta x}\left(s_{j+1}^n - s_{j-1}^n\right) \tag{90}$$

$$s_j^{n+1} = \frac{1}{2}(s_{j+1}^n + s_{j-1}^n) - \frac{v\Delta t}{2\Delta x}\left(r_{j+1}^n - r_{j-1}^n\right) \tag{91}$$

assume that the eigenmode is

$$\begin{bmatrix} r_j^n \\ \\ s_j^n \end{bmatrix} = \xi^n e^{ikj\Delta x} \begin{bmatrix} r^0 \\ \\ s^0 \end{bmatrix} \tag{92}$$

The vector on the right-hand side is a constant (both in space and in time) eigenvector, and ξ is a complex number.

$$\begin{bmatrix} (\cos k\Delta x) - \xi & i\frac{v\Delta t}{\Delta x}\sin k\Delta x \\ \\ i\frac{v\Delta t}{\Delta x}\sin k\Delta x & (\cos k\Delta x) - \xi \end{bmatrix} \cdot \begin{bmatrix} r^0 \\ \\ s^0 \end{bmatrix} = \begin{bmatrix} 0 \\ \\ 0 \end{bmatrix} \tag{93}$$

Only if the determinant of the matrix on the left vanishes, i.e.

$$\xi = \cos k\Delta x \pm i\frac{v\Delta t}{\Delta x}\sin k\Delta x \tag{94}$$

The stability condition is that both roots satisfy $|\xi| \leqslant 1$.

4.2.2 Other Varieties of Error

Finite-difference schemes for hyperbolic equations can exhibit dispersion, or phase errors.

$$\xi = e^{-ik\Delta x} + i\left(1 - \frac{v\Delta t}{\Delta x}\right)\sin k\Delta x \tag{95}$$

An arbitrary initial wave packet is a superposition of modes with different k's. At each timestep the modes get multiplied by different phase factors, depending on their value of k.

If $\Delta t = \Delta x/v$, then the exact solution for each mode of a wave packet $f(xvt)$ is obtained if each mode gets multiplied by $\exp(-ik\Delta x)$. For this value of Δt, equation shows that the finite-difference solution gives the exact analytic result. However, if $v\Delta t/\Delta x$ is not exactly 1, the phase relations of the modes can become hopelessly garbled and the wave packet disperses. The dispersion becomes large as soon as the wavelength becomes comparable to the grid spacing Δx.

nonlinear instability :

$$\frac{\partial v}{\partial t} = -v\frac{\partial v}{\partial x} + \cdots \tag{96}$$

The nonlinear term in v can cause a transfer of energy in Fourier space from long wavelengths to short wavelengths. This results in a wave profile steepening until a vertical profile or "shock" develops. Since the von Neumann analysis suggests that the stability can depend on $k\Delta x$, a scheme that was stable for shallow profiles can become unstable for steep profiles. This kind of difficulty arises in a differencing scheme where the cascade in Fourier space is halted at the shortest wavelength representable on the grid, that is, at $k \sim 1/\Delta x$. If energy simply accumulates in these modes, it eventually swamps the energy in the long wavelength modes of interest.

For wave equations, propagation errors (amplitude or phase) are usually most worrisome. For advective equations, transport errors are usually of greater concern.

The simplest way to model the transport properties "better" is to use upwind differencing

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} = -v_j^n \begin{cases} \dfrac{u_j^n - u_{j-1}^n}{\Delta x} \;, & u_j^n > 0 \\ \dfrac{u_{j+1}^n - u_j^n}{\Delta x} \;, & u_j^n < 0 \end{cases} \tag{97}$$

This scheme is only first-order, not second-order, accurate in the calculation of the spatial derivatives. Upwind differencing generally adds fidelity to problems where the advected variables are liable to undergo sudden changes of state, e.g., as they pass through shocks or other discontinuities. The amplification factor (for constant v) is

4.2.3 Second-Order Accuracy in Time

When using a method that is first-order accurate in time but second-order accurate in space, one generally has to take $v\Delta t$ significantly smaller than Δx to achieve desired accuracy

The staggered leapfrog method for the conservation equation is to use the values of u^n at time t^n, compute the fluxes F^n_j. Then compute new values u^{n+1} using the time-centered values of the fluxes:

$$u_j^{n+1} - u_j^{n-1} = -\frac{\Delta t}{\Delta x}(F^n_{j+1} - F^n_{j-1}) \tag{98}$$

The time levels in the time derivative term "leapfrog" over the time levels in the space derivative term.

For equation

$$\frac{\partial u}{\partial t} = -v\frac{\partial u}{\partial x} \ , \tag{99}$$

staggered leapfrog takes the form

$$u_j^{n+1} - u_j^{n-1} = -\frac{v\Delta t}{\Delta x}(u^n_{j+1} - u^n_{j-1}) \tag{100}$$

The von Neumann stability analysis now gives a quadratic equation for ξ,

$$\xi^2 - 1 = -2i\xi\frac{v\Delta t}{\Delta x}\sin k\Delta x \tag{101}$$

$$\xi = -i\frac{v\Delta t}{\Delta x}\sin k\Delta x \pm \sqrt{1 - \left(\frac{v\Delta t}{\Delta x}\sin k\Delta x\right)^2} \tag{102}$$

$|\xi|^2 = 1$ for any $v\Delta t \leqslant \Delta x$. The great advantage of the staggered leapfrog method is that there is no amplitude dissipation.

For equation (**??**), if the variables are centered on appropriate half-mesh points

$$r^n_{j+1/2} \equiv v \frac{\partial u}{\partial x}\bigg|^n_{j+1/2} = v \frac{u^n_{j+1} - u^n_j}{\Delta x} \tag{103}$$

$$s^{n+1/2}_j \equiv \frac{\partial u}{\partial t}\bigg|^{n+1/2}_j = \frac{u^{n+1}_j - u^n_j}{\Delta t} \tag{104}$$

the leapfrog differencing is

$$\frac{r^{n+1}_{j+1/2} - r^n_{j+1/2}}{\Delta t} = \frac{s^{n+1/2}_{j+1} - s^{n+1/2}_j}{\Delta x} \tag{105}$$

$$\frac{s^{n+1/2}_j - s^{n-1/2}_j}{\Delta t} = v \frac{r^n_{j+1/2} - r^n_{j-1/2}}{\Delta x} \ , \tag{106}$$

there is no amplitude dissipation when Courant condition required for stability is satisfied.

$$\frac{u^{n+1}_j - 2u^n_j + u^{n-1}_j}{(\Delta t)^2} = v^2 \frac{u^n_{j+1} - 2u^n_j + u^n_{j-1}}{(\Delta x)^2} \tag{107}$$

For nonlinear equations, the leapfrog method usually becomes unstable when the gradients get large. The instability is related to the fact that odd and even mesh points are completely decoupled.

The Two-Step Lax-Wendroff scheme is a second-order in time method which avoids large numerical dissipation and mesh drifting.

4.3 Fluid Dynamics with Shocks

There are basically three important general methods for handling shocks. The oldest and simplest method is to add artificial viscosity to the equations, modeling the way

Nature uses real viscosity to smooth discontinuities. This scheme is excellent for nearly all problems in one spatial dimension.

The second method combines a high-order differencing scheme that is accurate for smooth flows with a low order scheme that is very dissipative and can smooth the shocks. Various upwind differencing schemes are combined using weights chosen to zero the low order scheme unless steep gradients are present, and also chosen to enforce various "monotonicity" constraints that prevent nonphysical oscillations from appearing in the numerical solution.

The third is Godunov's approach. Here one gives up the simple linearization inherent in finite differencing based on Taylor series and includes the nonlinearity of the equations explicitly. There is an analytic solution for the evolution of two uniform states of a fluid separated by a discontinuity, the Riemann shock problem. Godunov's idea was to approximate the fluid by a large number of cells of uniform states, and piece them together using the Riemann solution.

4.4 Diffusive Initial Value Problems

Assume D is a constant. The equation becomes

$$\frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2} \,, \tag{108}$$

which is

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} = D\left[\frac{u_{j+1}^n - 2u_j^n + u_{j-1}^n}{(\Delta x)^2}\right] \tag{109}$$

The FTCS scheme was unstable for the hyperbolic equation. The amplification factor is

$$\xi = 1 - \frac{4D\Delta t}{(\Delta x)^2}\sin^2\left(\frac{k\Delta x}{2}\right) \tag{110}$$

The requirement $|\xi| \leqslant 1$ leads to the stability criterion

$$\frac{2D\Delta t}{(\Delta x)^2} \leqslant 1 \tag{111}$$

i.e. the maximum allowed timestep is, up to a numerical factor, the diffusion time across a cell of width Δx.

The diffusion time τ across a spatial scale of size λ is of order

$$\tau \sim \frac{\lambda^2}{D} \tag{112}$$

If we are limited to timesteps satisfying (**??**), we will need to evolve through of order $\left(\frac{\lambda^2}{(\Delta x)^2}\right)$ steps before things start to happen on the scale of interest.

There are two different answers, each of which has its pros and cons. The first answer is to seek a differencing scheme that drives small-scale features to their equilibrium forms, e.g., satisfying equation (**??**) with the left-hand side set to zero. This answer generally makes the best physical sense. But it leads to a differencing scheme (fully implicit) that is only first-order accurate in time for the scales that we are interested in. The second answer is to let small-scale features maintain their initial amplitudes, so that the evolution of the larger-scale features of interest takes place superposed with a kind of "frozen in" (though fluctuating) background of small-scale stuff. This gives a differencing scheme ("Crank- Nicholson") that is second-order accurate in time. Toward the end of an evolution calculation, however, one might want to switch over to some steps of the other kind, to drive the small-scale stuff into equilibrium.

Consider

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} = D\left[\frac{u_{j+1}^{n+1} - 2u_j^{n+1} + u_{j-1}^{n+1}}{(\Delta x)^2}\right] \tag{113}$$

Schemes with this character are called fully implicit or backward time, by contrast with FTCS (which is called fully explicit).

$$-\alpha u_{j-1}^{n+1} + (1+2\alpha)u_j^{n+1} - \alpha u_{j+1}^{n+1} = u_j^n \ , \quad j = 1, 2, \cdots, J-1 \tag{114}$$

where

$$\alpha \equiv \frac{D\Delta t}{(\Delta x)^2} \tag{115}$$

In the limit $\alpha \to \infty(\Delta t \to \infty)$. Dividing by α, the difference equations are just the finite-difference form of the equilibrium equation

$$\frac{\partial^2 u}{\partial x^2} = 0 \tag{116}$$

The amplification factor for equation (**??**) is

$$\xi = \frac{1}{1 + 4\alpha \sin^2\left(\dfrac{k\Delta x}{2}\right)} \ , \tag{117}$$

$|\xi| < 1$ for any stepsize Δt. The scheme is unconditionally stable. The details of the small-scale evolution from the initial conditions are obviously inaccurate for large Δt. But, as advertised, the correct equilibrium solution is obtained. This is the characteristic feature of implicit methods.

Form the average of the explicit and implicit FTCS schemes:

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} = \frac{D}{2}\left[\frac{(u_{j+1}^{n+1} - 2u_j^{n+1} + u_{j-1}^{n+1}) + (u_{j+1}^n - 2u_j^n + u_{j-1}^n)}{(\Delta x)^2}\right] \ , \tag{118}$$

both the left- and right-hand sides are centered at timestep $n + \dfrac{1}{2}$, so the method is second-order accurate in time as claimed. The amplification factor is

$$\xi = \frac{1 - 2\alpha \sin^2\left(\dfrac{k\Delta x}{2}\right)}{1 + 2\alpha \sin^2\left(\dfrac{k\Delta x}{2}\right)} \ , \tag{119}$$

the method is stable for any size Δt. This scheme is called the Crank-Nicholson scheme.

If $D = D(x)$, make an analytic change of variable

$$y = \int \frac{\mathrm{d}x}{D(x)} \ , \tag{120}$$

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(D(x)\frac{\partial u}{\partial x}\right) \tag{121}$$

$$\Longrightarrow \frac{\partial u}{\partial t} = \frac{1}{D(y)}\frac{\partial^2 u}{\partial y^2} \tag{122}$$

D is evaluated at the appropriate y_j. The stability criterion in an explicit scheme becomes

$$\Delta t \leqslant \min_j \left[\frac{(\Delta y)^2}{2D_j^{-1}}\right] \tag{123}$$

The constant spacing Δy in y does not imply constant spacing in x.

Alternative method is

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} = \left[\frac{D_{j+1/2}(u_{j+1}^n - u_j^n) - D_{j-1/2}(u_j^n - u_{j-1}^n)}{(\Delta x)^2}\right] \tag{124}$$

where

$$D_{j+1/2} \equiv D(x_{j+1/2}) \tag{125}$$

and the heuristic stability criterion is

$$\Delta t \leqslant \min_j \left[\frac{(\Delta x)^2}{2D_{j+1/2}}\right] \tag{126}$$

For nonlinear diffusion, $D = D(u)$,

4.4.1 Schrödinger Equation

The physical problem imposes constraints on the differencing scheme. Consider the time-dependent Schrödinger equation of quantum mechanics. For the scattering of a wavepacket by a one-dimensional potential $V(x)$,

$$i\frac{\partial \psi}{\partial t} = -\frac{\partial^2 \psi}{\partial x^2} + V\psi \tag{127}$$

One is given the initial wavepacket, $\psi(x, t = 0)$, together with boundary conditions that $\psi \to 0$ at $x \to \pm\infty$. An first- order accuracy in time and implicit scheme for stability is

$$i\left[\frac{\psi_j^{n+1} - \psi_j^n}{\Delta t}\right] = -\left[\frac{u_{j+1}^{n+1} - 2u_j^{n+1} + u_{j-1}^{n+1}}{(\Delta x)^2}\right] + V_j\psi_j^{n+1} \tag{128}$$

for which

$$\xi = \frac{1}{1 + i\left[\frac{4\Delta t}{(\Delta x)^2}\sin^2\left(\frac{k\Delta x}{2}\right) + V_j\Delta t\right]} \tag{129}$$

It is unconditionally stable, but unfortunately is not unitary. It requires that the total probability of finding the particle somewhere remains unity, i.e.

$$\int_{-\infty}^{\infty} |\psi|^2 \mathrm{d}x = 1 \tag{130}$$

The initial wave function $\psi(x, 0)$ is normalized to satisfy.

$$i\frac{\partial \psi}{\partial t} = H\psi \tag{131}$$

where the operator H is

$$H = -\frac{\partial^2}{\partial x^2} + V(x) \tag{132}$$

$$\psi(x, t) = e^{-iHt}\psi(x, 0) \tag{133}$$

The unstable explicit FTCS scheme is

$$\psi_j^{n+1} = (1 - iH\Delta t)\psi_j^n \tag{134}$$

where H is represented by a centered finite-difference approximation in x. The stable implicit scheme is

$$\psi_j^{n+1} = (1 + iH\Delta t)^{-1}\psi_j^n \tag{135}$$

These are both first-order accurate in time. However, neither operator is unitary.

Cayley's form for the finite-difference representation of e^{-iHt}, which is second-order accurate and unitary, is

$$e^{-iHt} \simeq \frac{1 - \frac{1}{2}iH\Delta t}{1 + \frac{1}{2}iH\Delta t} \tag{136}$$

i.e.

$$(1 + \frac{1}{2}iH\Delta t)\psi_j^{n+1} = (1 - \frac{1}{2}iH\Delta t)\psi_j^n \tag{137}$$

It is stable, unitary, and second-order accurate in space and time.

4.5 Initial Value Problems in Multidimensions

First run your programs on very small grids, e.g., 8 × 8, even though the resulting accuracy is so poor as to be useless. When your program is all debugged and demonstrably stable, then you can increase the grid size to a reasonable one and start looking at the results. New instabilities sometimes do show up on larger grids; but old instabilities never (in our experience) just go away.

4.5.1 Lax Method for a Flux-Conservative Equation

Consider

$$\frac{\partial u}{\partial t} = -\nabla \cdot \boldsymbol{F} = -\left(\frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y}\right) \tag{138}$$

$$x_j = x_0 + j\Delta \ , y_l = y_0 + l\Delta \ , \tag{139}$$

where $\Delta x = \Delta y \equiv \Delta$. The Lax scheme is

$$u_{j,l}^{n+1} = \frac{1}{4}(u_{j+1,l}^n + u_{j-1,l}^n + u_{j,l+1}^n + u_{j,l-1}^n) - \frac{\Delta t}{2\Delta}(F_{j+1,l}^n - F_{j-1,l}^n + F_{j,l+1}^n - F_{j,l-1}^n) \quad (140)$$

For

$$F_x = v_x u, \quad F_y = v_y u \ , \quad (141)$$

this requires an eigenmode with two dimensions in space, only a dependence on powers of ξ in time,

$$u_{j,l}^{n+1} = \xi^n e^{ik_x j\Delta} e^{ik_y l\Delta} \quad (142)$$

$$\xi = \frac{1}{2}(\cos k_x\Delta + \cos k_y\Delta) - i\alpha_x \sin k_x\Delta - i\alpha_y \sin k_y\Delta \quad (143)$$

where

$$\alpha_x = \frac{v_x \Delta t}{\Delta} \ , \quad \alpha_y = \frac{v_y \Delta t}{\Delta} \ , \quad (144)$$

$$\begin{aligned} |\xi|^2 = 1 &- (\sin^2 k_x\Delta + \sin^2 k_y\Delta)\left[\frac{1}{2} - (\alpha_x^2 + \alpha_y^2)\right] \\ &- \frac{1}{4}(\cos k_x\Delta - \cos k_y\Delta)^2 - (\alpha_y \sin k_x\Delta - \alpha_x \sin k_y\Delta)^2 \end{aligned} \quad (145)$$

The stability requirement $|\xi|^2 \leqslant 1$ becomes

$$\frac{1}{2} - (\alpha_x^2 + \alpha_y^2) \geqslant 0 \quad (146)$$

$$\text{or } \Delta t \leqslant \frac{\Delta}{\sqrt{2}(v_x^2 + v_y^2)^{1/2}} \quad (147)$$

The N-dimensional Courant condition: If $|v|$ is the maximum propagation velocity in the problem, then

$$\Delta t \leqslant \frac{\Delta}{\sqrt{N}|v|} \quad (148)$$

4.5.2 Diffusion Equation in Multidimensions

Consider

$$\frac{\partial u}{\partial t} = D\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) \tag{149}$$

The Crank-Nicholson scheme in two dimensions is

$$u_{j,l}^{n+1} = u_{j,l}^{n} + \frac{\alpha}{2}\left(\delta_x^2 u_{j,l}^{n+1} + \delta_x^2 u_{j,l}^{n} + \delta_y^2 u_{j,l}^{n+1} + \delta_y^2 u_{j,l}^{n}\right) \tag{150}$$

where

$$\alpha \equiv \frac{D\Delta t}{\Delta^2}\ , \quad \Delta \equiv \Delta x = \Delta y\ , \tag{151}$$

$$\delta_x^2 u_{j,l}^{n} \equiv u_{j+1,l}^{n} - 2u_{j,l}^{n} + u_{j-1,l}^{n} \tag{152}$$

A generalizing the Crank-Nicholson algorithm. It is still second-order accurate in time and space, and unconditionally stable, called the alternating-direction implicit method (ADI), which embodies the concept of operator splitting or time splitting. Divide each timestep into two steps of size $\Delta t/2$. In each substep, a different dimension is treated implicitly:

$$u_{j,l}^{n+1/2} = u_{j,l}^{n} + \frac{\alpha}{2}\left(\delta_x^2 u_{j,l}^{n+1/2} + \delta_y^2 u_{j,l}^{n}\right) \tag{153}$$

$$u_{j,l}^{n+1} = u_{j,l}^{n+1/2} + \frac{\alpha}{2}\left(\delta_x^2 u_{j,l}^{n+1/2} + \delta_y^2 u_{j,l}^{n+1}\right) \tag{154}$$

in which each substep requires only the solution of a simple tridiagonal system.

4.5.3 Operator Splitting Methods Generally

The basic idea of operator splitting, which is also called time splitting or the method of fractional steps, is that suppose an initial value equation of the form

$$\frac{\partial u}{\partial t} = \mathscr{L} u\ , \tag{155}$$

where $\mathscr{L}$ is some operator. While $\mathscr{L}$ is not necessarily linear, suppose that it can at least be written as a linear sum of m pieces, which act additively on u,

$$\mathscr{L}u = \mathscr{L}_1 u + \mathscr{L}_2 u + \cdots + \mathscr{L}_m u \ . \tag{156}$$

Finally, suppose that for each of the pieces, a differencing scheme for updating the variable u from timestep n to timestep $n+1$ is already known, valid if that piece of the operator were the only one on the right-hand side. Write these updatings symbolically as

$$\begin{aligned}
u^{n+1} &= \mathscr{U}_1(u^n, \Delta t) \ , \\
u^{n+1} &= \mathscr{U}_2(u^n, \Delta t) \ , \\
&\cdots \\
u^{n+1} &= \mathscr{U}_m(u^n, \Delta t) \ .
\end{aligned} \tag{157}$$

One form of operator splitting would be to get from n to $n+1$ by the following sequence of updatings:

$$\begin{aligned}
u^{n+1/m} &= \mathscr{U}_1(u^n, \Delta t) \ , \\
u^{n+2/m} &= \mathscr{U}_2(u^{n+1/m}, \Delta t) \ , \\
&\cdots \\
u^{n+1} &= \mathscr{U}_m(u^{n+(m-1)/m}, \Delta t) \ .
\end{aligned} \tag{158}$$

Let $\mathscr{U}_1$ now denote an updating method that includes algebraically all the pieces of the total operator $\mathscr{L}$, but which is desirably stable only for the $\mathscr{L}_1$ piece; likewise $\mathscr{U}_2, \cdots, \mathscr{U}_m$.

Then a method of getting from u^n to u^{n+1} is

$$\begin{aligned} u^{n+1/m} &= \mathscr{U}_1(u^n, \Delta t/m) \ , \\ u^{n+2/m} &= \mathscr{U}_2(u^{n+1/m}, \Delta t/m) \ , \\ &\cdots \\ u^{n+1} &= \mathscr{U}_m(u^{n+(m-1)/m}, \Delta t/m) \ . \end{aligned} \tag{159}$$

The timestep for each fractional step is now only $1/m$ of the full timestep, because each partial operation acts with all the terms of the original operator. Equation is usually, though not always, stable as a differencing scheme for the operator $\mathscr{L}$. In fact, as a rule of thumb, it is often sufficient to have stable $\mathscr{U}_i$'s only for the operator pieces having the highest number of spatial derivatives - the other $\mathscr{U}_i$'s can be unstable - to make the overall scheme stable!

5 Boundary Value Problems

In contrast to initial value problems, stability is relatively easy to achieve for boundary value problems. The efficiency of the algorithms, both in computational load and storage requirements, becomes the principal concern.

Represent the function $u(x, y)$ by its values at the discrete set of points

$$x_j = x_0 + j\Delta \ , \quad j = 0, 1, \cdots, J \tag{160}$$

$$y_l = y_0 + l\Delta \ , \quad l = 0, 1, \cdots, L \tag{161}$$

where Δ is the grid spacing.

$$\frac{u_{j+1,l} - 2u_{j,l} + u_{j-1,l}}{\Delta^2} + \frac{u_{j,l+1} - 2u_{j,l} + u_{j,l-1}}{\Delta^2} = \rho_{j,l} \tag{162}$$

$$\text{or } u_{j+1,l} + u_{j-1,l} + u_{j,l+1} + u_{j,l-1} - 4u_{j,l} = \rho_{j,l}\Delta^2 \tag{163}$$

$$\boldsymbol{A} \cdot \boldsymbol{u} = \boldsymbol{b} \tag{164}$$

There are three different approaches to the solution, not all applicable in all cases: relaxation methods, "rapid" methods (e.g., Fourier methods), and direct matrix methods.