Finite Difference Methods for Conservation Laws

The canonical form for the 1D conservation law is

$$\mathbf{u}_t + \mathbf{f}(u)_x = 0\,, \tag{5.34}$$

and one famous benchmark problem is the Burger's equation

$$u_t + \left(\frac{u^2}{2}\right)_x = 0, \tag{5.35}$$

in which $f(u) = u^2/2$. The term $f(u)$ is often called the flux. This equation can be written in the nonconservative form

$$u_t + uu_x = 0, \tag{5.36}$$

and the solution likely develops shock(s) where the solution is discontinuous,[1] even if the initial condition is arbitrarily differentiable, i.e., $u_0(x) = \sin x$.

We can use the upwind scheme to solve the Burger's equation. From the nonconservative form, we obtain

$$\frac{U_j^{k+1} - U_j^k}{\Delta t} + U_j^k \frac{U_j^k - U_{j-1}^k}{h} = 0, \quad \text{if } U_j^k \geq 0,$$

$$\frac{U_j^{k+1} - U_j^k}{\Delta t} + U_j^k \frac{U_{j+1}^k - U_j^k}{h} = 0, \quad \text{if } U_j^k < 0,$$

or from the conservative form

$$\frac{U_j^{k+1} - U_j^k}{\Delta t} + \frac{(U_j^k)^2 - (U_{j-1}^k)^2}{2h} = 0, \quad \text{if } U_j^k \geq 0,$$

$$\frac{U_j^{k+1} - U_j^k}{\Delta t} + \frac{(U_{j+1}^k)^2 - (U_j^k)^2}{2h} = 0, \quad \text{if } U_j^k < 0.$$

If the solution is smooth, both methods work well (first-order accurate). However, if shocks develop the conservative form gives much better results than that of the nonconservative form.

We can derive the Lax–Wendroff scheme using the modified equation of the nonconservative form. Since $u_t = -uu_x$,

$$\begin{aligned} u_{tt} &= -u_t u_x - uu_{tx} \\ &= uu_x^2 + u(uu_x)_x \\ &= uu_x^2 + u\left(u_x^2 + uu_{xx}\right) \\ &= 2uu_x^2 + u^2 u_{xx}, \end{aligned}$$

so the leading term of the modified equation for the first-order method is

$$u_t + uu_x = \frac{\Delta t}{2}\left(2uu_x^2 + u^2 u_{xx}\right), \tag{5.37}$$

[1] There is no classical solution to the PDE when shocks develop because u_x is not well defined. We need to look for weak solutions.

and the nonconservative Lax–Wendroff scheme for Burger's equation is

$$U_j^{k+1} = U_j^k - \Delta t U_j^k \frac{U_{j+1}^k - U_{j-1}^k}{2h}$$

$$= + \frac{(\Delta t)^2}{2}\left(2U_j^k\left(\frac{U_{j+1}^k - U_{j-1}^k}{2h}\right)^2 + (U_j^k)^2\frac{U_{j-1}^k - 2U_j^k + U_{j+1}^k}{h^2}\right).$$

Conservative FD Methods for Conservation Laws

Consider the conservation law

$$\mathbf{u}_t + \mathbf{f}(u)_x = 0\,,$$

and let us seek a numerical scheme of the form

$$\mathbf{u}_j^{k+1} = \mathbf{u}_j^k - \frac{\Delta t}{h}\left(\mathbf{g}_{j+\frac{1}{2}}^k - \mathbf{g}_{j-\frac{1}{2}}^k\right), \tag{5.38}$$

where

$$\mathbf{g}_{j+\frac{1}{2}} = \mathbf{g}\left(\mathbf{u}_{j-p+1}^k, \mathbf{u}_{j-p+2}^k, \dots, \mathbf{u}_{j+q+1}^k\right)$$

is called the numerical flux, satisfying

$$g(u, u, \dots, u) = f(u)\,. \tag{5.39}$$

Such a scheme is called conservative. For example, we have $g(u) = u^2/2$ for the Burger's equation.

We can derive general criteria that g should satisfy, as follows.

1. Integrate the equation with respect to x from $x_{j-\frac{1}{2}}$ to $x_{j+\frac{1}{2}}$, to get

$$\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} u_t dx = -\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} f(u)_x dx$$

$$= -\left(f(u(x_{j+\frac{1}{2}}, t)) - f(u(x_{j-\frac{1}{2}}, t))\right).$$

2. Integrate the equation above with respect to t from t^k to t^{k+1}, to get

$$\int_{t^k}^{t^{k+1}} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} u_t\, dx\, dt = -\int_{t^k}^{t^{k+1}} \left(f(u(x_{j+\frac{1}{2}}, t)) - f(u(x_{j-\frac{1}{2}}, t))\right) dt\,,$$

$$\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \left(u(x, t^{k+1}) - u(x, t^k)\right) dx = -\int_{t^k}^{t^{k+1}} \left(f(u(x_{j+\frac{1}{2}}, t)) - f(u(x_{j-\frac{1}{2}}, t))\right) dt\,.$$

Define the average of $u(x,t)$ as

$$\bar{u}_j^k = \frac{1}{h}\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} u(x,t^k)dx\,, \tag{5.40}$$

which is the cell average of $u(x,t)$ over the cell $(x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}})$ at the time level k. The expression that we derived earlier can therefore be rewritten as

$$\begin{aligned}
\bar{u}_j^{k+1} &= \bar{u}_j^k - \frac{1}{h}\left(\int_{t^k}^{t^{k+1}} f(u(x_{j+\frac{1}{2}},t))dt - \int_{t^k}^{t^{k+1}} f(u(x_{j-\frac{1}{2}},t))dt\right) \\
&= \bar{u}_j^k - \frac{\Delta t}{h}\left(\frac{1}{\Delta t}\int_{t^k}^{t^{k+1}} f(u(x_{j+\frac{1}{2}},t))dt - \frac{1}{\Delta t}\int_{t^k}^{t^{k+1}} f(u(x_{j-\frac{1}{2}},t))dt\right) \\
&= \bar{u}_j^k - \frac{\Delta t}{h}\left(g_{j+\frac{1}{2}} - g_{j+\frac{1}{2}}\right),
\end{aligned}$$

where

$$g_{j+\frac{1}{2}} = \frac{1}{\Delta t}\int_{t^k}^{t^{k+1}} f(u(x_{j+\frac{1}{2}},t))dt\,.$$

Different conservative schemes can be obtained, if different approximations are used to evaluate the integral.

Some Commonly Used Numerical Scheme for Conservation Laws

Some commonly used schemes are:

- Lax–Friedrichs scheme

$$U_j^{k+1} = \frac{1}{2}\left(U_{j+1}^k + U_{j-1}^k\right) - \frac{\Delta t}{2h}\left(f(U_{j+1}^k) - f(U_{j-1}^k)\right); \tag{5.41}$$

- Lax–Wendroff scheme

$$\begin{aligned}
U_j^{k+1} &= U_j^k - \frac{\Delta t}{2h}\left(f(U_{j+1}^k) - f(U_{j-1}^k)\right) \\
&\quad + \frac{(\Delta t)^2}{2h^2}\left\{A_{j+\frac{1}{2}}\left(f(U_{j+1}^k) - f(U_j^k)\right) - A_{j-\frac{1}{2}}\left(f(U_j^k) - f(U_{j-1}^k)\right)\right\},
\end{aligned} \tag{5.42}$$

where $A_{j+\frac{1}{2}} = Df(u(x_{j+\frac{1}{2}},t))$ is the Jacobian matrix of $f(u)$ at $u(x_{j+\frac{1}{2}},t)$.

Numerical Algorithms for the Wave Equation

Let us now develop some basic numerical solution techniques for the second-order wave equation. As above, although we are in possession of the explicit d'Alembert solution formula (2.82), the lessons learned in designing viable schemes here will carry over to more complicated situations, including inhomogeneous media and higher-dimensional problems, for which analytic solution formulas may no longer be readily available.

Consider the wave equation

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < \ell, \qquad t \geq 0, \tag{5.52}$$

on a bounded interval of length ℓ with constant wave speed $c > 0$. For specificity, we impose (possibly time-dependent) Dirichlet boundary conditions

$$u(t,0) = \alpha(t), \qquad u(t,\ell) = \beta(t), \qquad t \geq 0, \tag{5.53}$$

along with the usual initial conditions

$$u(0,x) = f(x), \qquad \frac{\partial u}{\partial t}(0,x) = g(x), \qquad 0 \le x \le \ell. \tag{5.54}$$

As usual, we adopt a uniformly spaced mesh

$$t_j = j\,\Delta t, \qquad x_m = m\,\Delta x, \qquad \text{where} \qquad \Delta x = \frac{\ell}{n}.$$

Discretization is implemented by replacing the second-order derivatives in the wave equation by their standard finite difference approximations (5.5):

$$\begin{aligned} \frac{\partial^2 u}{\partial t^2}(t_j, x_m) &\approx \frac{u(t_{j+1}, x_m) - 2\,u(t_j, x_m) + u(t_{j-1}, x_m)}{(\Delta t)^2} + \mathrm{O}\big((\Delta t)^2\big), \\ \frac{\partial^2 u}{\partial x^2}(t_j, x_m) &\approx \frac{u(t_j, x_{m+1}) - 2\,u(t_j, x_m) + u(t_j, x_{m-1})}{(\Delta x)^2} + \mathrm{O}\big((\Delta x)^2\big). \end{aligned} \tag{5.55}$$

Since the error terms are both of second order, we anticipate being able to choose the space and time step sizes to have comparable magnitudes: $\Delta t \approx \Delta x$. Substituting the finite difference formulas (5.55) into the partial differential equation (5.52) and rearranging terms, we are led to the iterative system

$$u_{j+1,m} = \sigma^2\,u_{j,m+1} + 2\,(1-\sigma^2)\,u_{j,m} + \sigma^2\,u_{j,m-1} - u_{j-1,m}, \qquad \begin{aligned} j &= 1,2,\ldots\,, \\ m &= 1,\ldots,n-1, \end{aligned} \tag{5.56}$$

for the numerical approximations $u_{j,m} \approx u(t_j, x_m)$ to the solution values at the nodes. The parameter

$$\sigma = \frac{c\,\Delta t}{\Delta x} > 0 \tag{5.57}$$

depends on the wave speed and the ratio of space and time step sizes. The boundary conditions (5.53) require that

$$u_{j,0} = \alpha_j = \alpha(t_j), \qquad u_{j,n} = \beta_j = \beta(t_j), \qquad j = 0,1,2,\ldots\,. \tag{5.58}$$

This allows us to rewrite the iterative system in vectorial form

$$\mathbf{u}^{(j+1)} = B\,\mathbf{u}^{(j)} - \mathbf{u}^{(j-1)} + \mathbf{b}^{(j)}, \tag{5.59}$$

where

$$B = \begin{pmatrix} 2\,(1-\sigma^2) & \sigma^2 & & & \\ \sigma^2 & 2\,(1-\sigma^2) & \sigma^2 & & \\ & \sigma^2 & \ddots & \ddots & \\ & & \ddots & \ddots & \sigma^2 \\ & & & \sigma^2 & 2\,(1-\sigma^2) \end{pmatrix}, \quad \mathbf{u}^{(j)} = \begin{pmatrix} u_{j,1} \\ u_{j,2} \\ \vdots \\ u_{j,n-2} \\ u_{j,n-1} \end{pmatrix}, \quad \mathbf{b}^{(j)} = \begin{pmatrix} \sigma^2\alpha_j \\ 0 \\ \vdots \\ 0 \\ \sigma^2\beta_j \end{pmatrix}. \tag{5.60}$$

The entries of $\mathbf{u}^{(j)} \in \mathbb{R}^{n-1}$ are, as in (5.18), the numerical approximations to the solution values at the *interior* nodes. Note that (5.59) describes a *second-order iterative scheme*, since computing the subsequent iterate $\mathbf{u}^{(j+1)}$ requires knowing the values of the preceding two: $\mathbf{u}^{(j)}$ and $\mathbf{u}^{(j-1)}$.

The one subtlety is how to get the method started. We know $\mathbf{u}^{(0)}$, since its entries $u_{0,m} = f_m = f(x_m)$ are determined by the initial position. However, we also need $\mathbf{u}^{(1)}$

in order to launch the iteration and compute $\mathbf{u}^{(2)}, \mathbf{u}^{(3)}, \ldots$. Its entries $u_{1,m} \approx u(\Delta t, x_m)$ approximate the solution at time $t_1 = \Delta t$, whereas the initial velocity $u_t(0,x) = g(x)$ prescribes the derivatives $u_t(0,x_m) = g_m = g(x_m)$ at the initial time $t_0 = 0$. To resolve this difficulty, a first thought might be to use the finite difference approximation

$$g_m = \frac{\partial u}{\partial t}(0,x_m) \approx \frac{u(\Delta t, x_m) - u(0,x_m)}{\Delta t} \approx \frac{u_{1,m} - f_m}{\Delta t} \tag{5.61}$$

to compute the required values $u_{1,m} = f_m + g_m \Delta t$. However, the approximation (5.61) is accurate only to order Δt, whereas the rest of the scheme has errors proportional to $(\Delta t)^2$. The effect would be to introduce an unacceptably large error at the initial step, and the resulting solution would fail to conform to the desired order of accuracy.

To construct an initial approximation to $\mathbf{u}^{(1)}$ with error on the order of $(\Delta t)^2$, we need to analyze the error in the approximation (5.61) in more depth. Note that, by Taylor's Theorem,

$$\begin{aligned}\frac{u(\Delta t, x_m) - u(0,x_m)}{\Delta t} &= \frac{\partial u}{\partial t}(0,x_m) + \frac{1}{2}\frac{\partial^2 u}{\partial t^2}(0,x_m)\Delta t + \mathrm{O}(\,(\Delta t)^2\,) \\ &= \frac{\partial u}{\partial t}(0,x_m) + \frac{c^2}{2}\frac{\partial^2 u}{\partial x^2}(0,x_m)\,\Delta t + \mathrm{O}(\,(\Delta t)^2\,),\end{aligned}$$

since $u(t,x)$ solves the wave equation. Therefore,

$$\begin{aligned}u_{1,m} = u(\Delta t, x_m) &\approx u(0,x_m) + \frac{\partial u}{\partial t}(0,x_m)\Delta t + \frac{c^2}{2}\frac{\partial^2 u}{\partial x^2}(0,x_m)(\Delta t)^2 \\ &= f(x_m) + g(x_m)\,\Delta t + \frac{c^2}{2}f''(x_m)(\Delta t)^2 \\ &\approx f_m + g_m\,\Delta t + \frac{c^2(f_{m+1} - 2f_m + f_{m-1})(\Delta t)^2}{2(\Delta x)^2},\end{aligned}$$

where the last line, which employs the finite difference approximation (5.5) to the second derivative, can be used if the explicit formula for $f''(x)$ is either not known or too complicated to evaluate directly. Therefore, we initiate the scheme by setting

$$u_{1,m} = \tfrac{1}{2}\sigma^2 f_{m+1} + (1-\sigma^2)f_m + \tfrac{1}{2}\sigma^2 f_{m-1} + g_m\,\Delta t, \tag{5.62}$$

or, in vectorial form,

$$\mathbf{u}^{(0)} = \mathbf{f}, \qquad \mathbf{u}^{(1)} = \tfrac{1}{2}B\,\mathbf{u}^{(0)} + \mathbf{g}\,\Delta t + \tfrac{1}{2}\mathbf{b}^{(0)}, \tag{5.63}$$

where $\mathbf{f} = \left(f_1, f_2, \ldots, f_{n-1}\right)^T$, $\mathbf{g} = \left(g_1, g_2, \ldots, g_{n-1}\right)^T$, are the sampled values of the initial data. This serves to maintain the desired second-order accuracy of the scheme.

Example 5.6. Consider the particular initial value problem

$$u_{tt} = u_{xx}, \qquad \begin{aligned} u(0,x) = e^{-400\,(x-.3)^2}, \qquad u_t(0,x) = 0, \qquad & 0 \le x \le 1, \\ u(t,0) = u(t,1) = 0, \qquad & t \ge 0,\end{aligned}$$

subject to homogeneous Dirichlet boundary conditions on the interval $[0,1]$. The initial data is a fairly concentrated hump centered at $x = .3$. As time progresses, we expect the initial hump to split into two half-sized humps, which then collide with the ends of the interval, reversing direction and orientation.

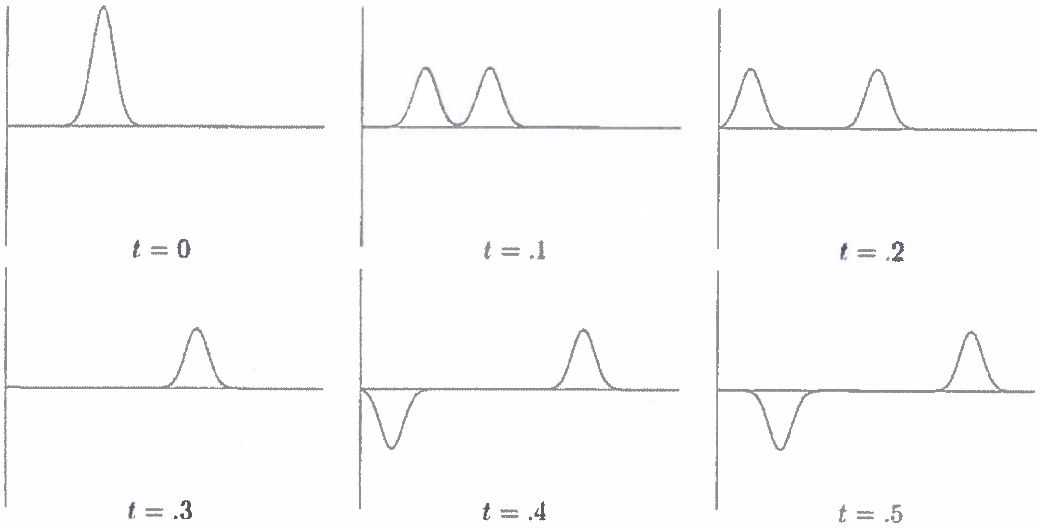

Figure 5.9. Numerically stable waves.

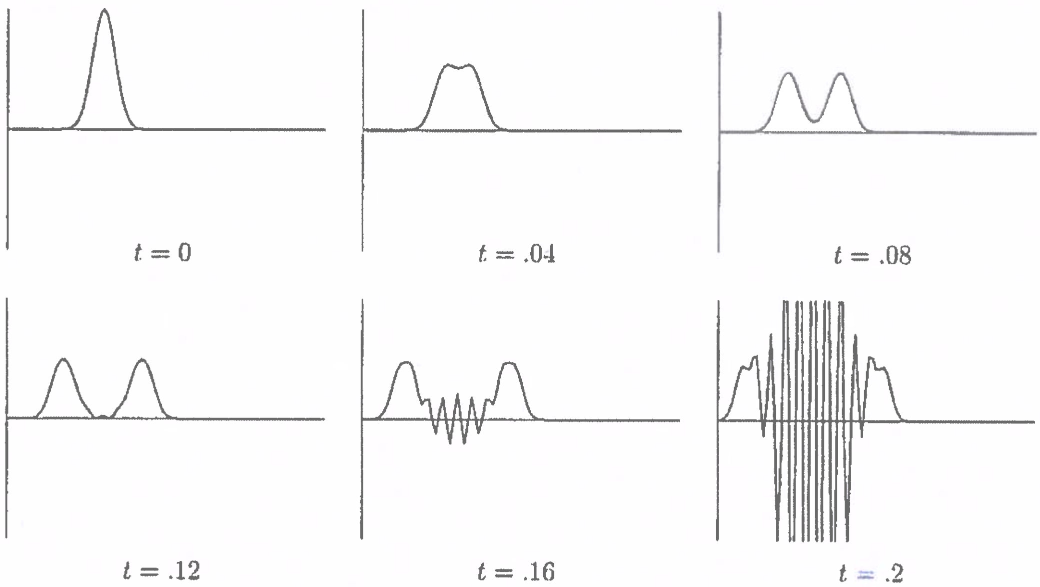

Figure 5.10. Numerically unstable waves.

For our numerical approximation, let us use a space discretization consisting of 90 equally spaced points, and so $\Delta x = \frac{1}{90} = .0111\ldots$. If we choose a time step of $\Delta t = .01$, whereby $\sigma = .9$, then we obtain a reasonably accurate solution over a fairly long time range, as plotted in Figure 5.9. On the other hand, if we double the time step, setting $\Delta t = .02$, so $\sigma = 1.8$, then, as shown in Figure 5.10, we induce an instability that eventually

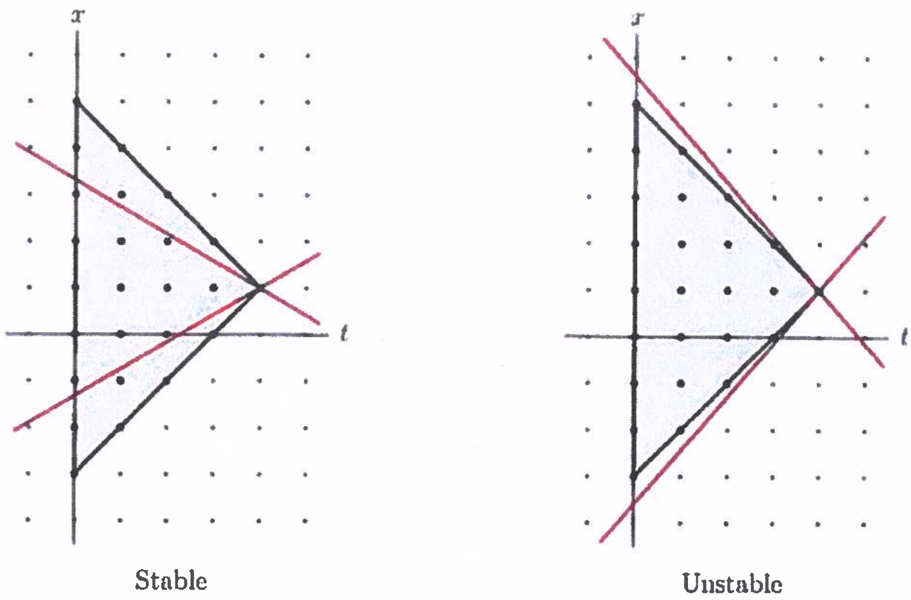

Figure 5.11. The CFL condition for the wave equation.

overwhelms the numerical solution. Thus, the preceding numerical scheme appears to be only conditionally stable.

Stability analysis proceeds along the same lines as in the first-order case. The CFL condition requires that the characteristics emanating from a node (t_j, x_m) remain, for times $0 \le t \le t_j$, in its numerical domain of dependence, which, for our particular numerical scheme, is the same triangle

$$\widetilde{T}_{(t_j,x_m)} = \{\, (t,x) \mid 0 \le t \le t_j,\ x_m - t_j + t \le x \le x_m + t_j - t \,\},$$

now plotted in Figure 5.11. Since the characteristics are the lines of slope $\pm c$, the CFL condition is the same as in (5.48):

$$\sigma = \frac{c\,\Delta t}{\Delta x} \le 1, \qquad \text{or, equivalently,} \qquad 0 < c \le \frac{\Delta x}{\Delta t}. \tag{5.64}$$

The resulting stability criterion explains the observed difference between the numerically stable and unstable cases.

However, as we noted above, the CFL condition is, in general, only necessary for stability of the numerical scheme; sufficiency requires that we perform a von Neumann stability analysis. To this end, we specialize the calculation to a single complex exponential $e^{\,i\,k\,x}$. After one time step, the scheme will have the effect of multiplying it by the *magnification factor* $\lambda = \lambda(k)$, after another time step by λ^2, and so on. To determine λ, we substitute the relevant sampled exponential values

$$u_{j-1,m} = e^{\,i\,k\,x_m}, \qquad u_{j,m} = \lambda\, e^{\,i\,k\,x_m}, \qquad u_{j+1,m} = \lambda^2\, e^{\,i\,k\,x_m}, \tag{5.65}$$

into the scheme (5.56). After canceling the common exponential, we find that the magnification factor satisfies the following quadratic equation:

$$\lambda^2 = \left(2 - 4\sigma^2 \sin^2\left(\tfrac{1}{2}k\,\Delta x\right)\right)\lambda - 1,$$

whence

$$\lambda = \alpha \pm \sqrt{\alpha^2 - 1}\,, \qquad \text{where} \qquad \alpha = 1 - 2\sigma^2 \sin^2\left(\tfrac{1}{2} k \Delta x\right). \tag{5.66}$$

Thus, there are *two* different magnification factors associated with each complex exponential — which is, in fact, a consequence of the scheme being of second order. Stability requires that *both* be ≤ 1 in modulus. Now, if the CFL condition (5.64) holds, then $|\,\alpha\,| \leq 1$, which implies that both magnification factors (5.66) are complex numbers of modulus $|\,\lambda\,| = 1$, and thus the numerical scheme satisfies the stability criterion (5.26). On the other hand, if $\sigma > 1$, then $\alpha < -1$ over a range of values of k, which implies that the two magnification factors (5.66) are both real and one of them is < -1, thus violating the stability criterion. Consequently, the CFL condition (5.64) does indeed distinguish between the stable and unstable finite difference schemes for the wave equation.