Mat1062: Computational Methods for PDE

Mary Pugh

February 26, 2008

1 Modified equations

The *modified equation* or *effective equation* is a way to understand qualitatively the effect of discretization errors on the behaviour of the solution. The idea is that we look at the leading-order truncation errors, write them in terms of derivatives of the solution, and then write the PDE that the scheme more closely approximates than the original one. This modified equation will have coefficients containing factors of h or h^2, so as the mesh is refined they go to zero. But for any finite h, the extra terms give us insight into the behavior of the discrete solutions.

As an example, let's consider the explicit upwind scheme. We already know that if $u(x,t)$ is the solution of $u_t + au_x = 0$ and $u_j^n = u(x_j, t_n)$ then the truncation error

$$\epsilon_j^n = u_j^{n+1} - u_j^n + \mu(u_j^n - u_{j-1}^n)$$

satisfies

$$\left|\epsilon_j^n - k\left[u_t(x_j,t_n) + au_x(x_j,t_n)\right]\right| = \mathcal{O}(k^2, hk).$$

On the other hand, if $v(x,t)$ is the solution of $v_t + av_x + k/2\, v_{tt} - ah/2\; v_{xx} = 0$ and $v_j^n = v(x_j, t_n)$ then the truncation error

$$\epsilon_j^n = v_j^{n+1} - v_j^n + \mu(v_j^n - v_{j-1}^n)$$

satisfies

$$\left|\epsilon_j^n - k\left(v_t(x_j,t_n) + av_x(x_j,t_n) + \frac{k}{2}\, v_{tt}(x_j,t_n) - \frac{ah}{2}\, v_{xx}(x_j,t_n)\right)\right| = \mathcal{O}(k^3, h^2 k).$$

This means that the discrete solution generated by explicit upwinding is actually closer to the solution of $v_t + av_x + k/2\, v_{tt} - ah/2\; v_{xx} = 0$ than it is to the solution of $u_t + au_x = 0$. And so to understand the properties of the discrete solution we would be better off studying properties of the solution of the v equation. In fact, we're going to find an even better equation to study. The v equation is

$$v_t = -av_x - \frac{k}{2} v_{tt} + \frac{ah}{2} v_{xx} \tag{1}$$

This means that $v_t = -av_x + \mathcal{O}(h,k)$ and hence $v_{tt} = a^2 v_{xx} + \mathcal{O}(h,k)$. As a result, if we replace the v_{tt} in equation (1) with $a^2 v_{xx}$ we will get an equation which is close to to it. And so we introduce

$$w_t + aw_x = -\frac{k}{2} a^2 w_{xx} + \frac{ah}{2} w_{xx} = \frac{ah}{2}\left(1 - \frac{ak}{h}\right) w_{xx} \tag{2}$$

If $w(x,t)$ is the solution of (2) and $w_j^n = w(x_j, t_n)$ then the truncation error

$$\epsilon_j^n = w_j^{n+1} - w_j^n + \mu(w_j^n - w_{j-1}^n)$$

satisfies

$$\left| \epsilon_j^n - k \left(w_t(x_j, t_n) + wv_x(x_j, t_n) - \frac{ah}{2}\left(1 - \frac{ak}{h}\right) w_{xx}(x_j, t_n) \right) \right| = \mathcal{O}(k^3, h^2 k).$$

In this way, we find that the discrete solution generated by explicit upwinding is closer to the solution of equation (2) than to the solution of $u_t + au_x = 0$: see Figure 1. We call equation (2) the *modified equation* for the explicit upwinding scheme.

In a similar manner, we find that the modified equation for the Lax-Friedrichs scheme is

$$u_t + au_x = \frac{h^2}{2k}\left(1 - \left(\frac{ak}{h}\right)^2\right) u_{xx}, \tag{3}$$

the modified equation for the Lax-Wendroff scheme is

$$u_t + au_x = \frac{h^2 a}{6}\left(\frac{a^2k^2}{h^2} - 1\right) u_{xxx}, \tag{4}$$

and the modified equation for the Beam-Warming scheme is

$$u_t + au_x = \frac{h^2 a}{6}\left(\frac{ak}{h} - 1\right)\left(\frac{ak}{h} - 2\right) u_{xxx} \tag{5}$$

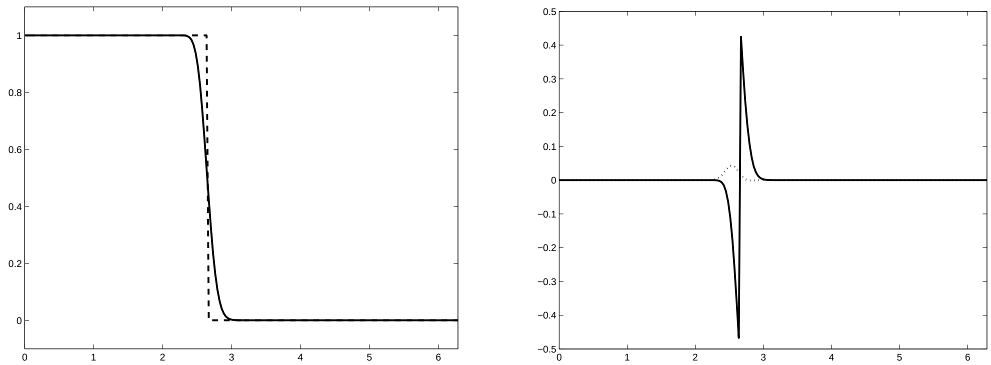

Figure 1: The parameters are as in Figures 1 and 2 of the February 14 notes. Left: The solution of the explicit upwind scheme is shown with a solid line. The exact solution of $u_t + au_x = 0$ is shown with a dashed line. The exact solution of the modified equation (2) for the explicit upwind scheme is shown with (barely visible) open circles. Right: the discrete solution minus the exact travelling wave solution is shown with a solid line. The discrete solution minus the exact solution of the modified equation is shown with a dashed line. Note that the discrete solution is much closer to the solution of the modified equation.

1.1 First-order schemes and diffusion

Both the explicit upwind scheme and the Lax-Friedrichs scheme have modified equations that are advection diffusion equations:

$$u_t + au_x = Du_{xx}.$$

One can explicitly solve this equation by going into moving coordinates $v(y,t) = u(x+at,t)$ and then solving the resulting heat equation $v_t = Dv_{yy}$. For example, for initial data

$$u_0(x) = \begin{cases} 1 & \text{for } x < 0 \\ 0 & \text{otherwise} \end{cases} \quad \Longrightarrow \quad u(x,t) = 1 - \operatorname{erf}\left(\frac{x-at}{\sqrt{4Dt}}\right) \tag{6}$$

where

$$\operatorname{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{x} e^{-t^2}\, dt.$$

This is a different error function than what's built in to matlab and maple. I chose this one for analytical convenience: it tends to zero as x tends to

$-\infty$ and it tends to one as x tends to $+\infty$. The error function built in to maple and matlab has limit -1 at $x = -\infty$ and $+1$ at $x = +\infty$.

The diffusion constants depend on h and k:

$$D_{\text{EU}} = h\frac{a}{2}\left(1 - \frac{ak}{h}\right) = h\frac{a}{2}\left(1 - \mu\right) \qquad \text{(explicit upwind)} \tag{7}$$

$$D_{\text{LF}} = \frac{h^2}{2k}\left(1 - \left(\frac{ak}{h}\right)^2\right) = h\frac{a}{2\mu}\left(1 - \mu^2\right) \qquad \text{(Lax-Friedrichs)} \tag{8}$$

If one is considering the usual "refinement path" in which one chooses a fixed value of μ within the stability region and then refines h and k so that μ remains unchanged then one sees that as $h \to 0$ both diffusion coefficients D_{EU} and D_{LF} tend to zero. In this way, we see that for a fixed value of h and k the modified equation is a smoothed version of $u_t + au_x = 0$ and as $h \to 0$ this smoothing vanishes.

Those of you who have some familiarity with PDE see how this could lead to an existence result: 1) prove existence for the regularized equation and understand the properties of its (unique) solution and 2) take the regularization to zero and use a compactness argument to prove that there's a sequence of solutions of the regularized equation that converge to a solution of the hyperbolic equation. Of course, this is all irrelevant in this particular case because the hyperbolic equation in question $u_t + au_x = 0$ not only has explicit solutions but they're unique. However, in general proving existence and/or uniqueness for hyperbolic problems is much Much **MUCH** harder than for parabolic or elliptic problems. The very first proof for a fairly simple nonlinear hyperbolic problem was actually done by proving that a particular numerical method converged. (Glimm, Communications in Pure and Applied Mathematics 18(1965)697-715.)

Finally, note that for $-1 < \mu < 1$

$$D_{\text{EU}} + \frac{h^2}{2k}\left(1 - \mu\right) = D_{\text{LF}} \quad \Longrightarrow \quad D_{\text{EU}} < D_{\text{LF}}.$$

Thus the diffusivity is greater in the modified equation for Lax-Friedrichs than for the modified equation for the explicit upwinding scheme. This explains what you observed in Figure 1 of the February 14 notes: the shock was smoothed out more by the Lax-Friedrichs scheme than by the explicit upwind scheme.

1.2 Convergence rate for the explicit upwind and Lax-Friedrichs schemes

Here I will try to make a suggestive argument as to why the convergence rate for the shock wave solution is $\mathcal{O}(\sqrt{k})$. If we have the discrete solution computed up to a fixed time T and the exact solution $u_0(x - aT)$, we want to see that the L^1 norm of their difference goes to zero as h and k go to zero along a fixed refinement path (μ fixed).

We know that the explicit upwind scheme and the Lax-Friedrichs scheme result in discrete solutions which are very close to the solution (6) of the advection diffusion equation $u_t + au_x = Du_{xx}$. And so rather than trying to control the L^1 norm of the discrete solution minus the travelling wave solution of $u_t + au_x = 0$, I'll try to control the L^1 norm of the solution (6) minus the travelling wave solution. In short, I want to understand

$$||u(\cdot, T) - v(\cdot, T)||_{L^1(\mathbb{R})} = \int_{-\infty}^{\infty} |u(x,T) - v(x,T)| \, dx$$

where

$$u(x,T) = 1 - \mathrm{erf}\left(\frac{x - aT}{\sqrt{4DT}}\right), \qquad v(x,T) = \begin{cases} 1 & \text{if } x - aT < 0 \\ 0 & \text{otherwise} \end{cases}$$

The L^1 norm is translation invariant and both profiles are moving with the same speed a, I can go into a moving frame and take

$$u(x,T) = 1 - \mathrm{erf}\left(\frac{x}{\sqrt{4DT}}\right), \qquad v(x,T) = \begin{cases} 1 & \text{if } x < 0 \\ 0 & \text{otherwise} \end{cases}$$

Next, write $u(x,T)$ as $u(x,T) = 1/2 + \widetilde{u}(x,T)$ and $v(x,T) = 1/2 + \widetilde{u}(x,T)$ where $\widetilde{u}$ and $\widetilde{v}$ are odd about $x = 0$. Hence, $|u(x,T) - v(x,T)| = |\widetilde{u}(x,T) - \widetilde{v}(x,T)|$ is an even function and so it suffices to find

$$2\int_{-\infty}^{0} |\widetilde{u}(x,T) - \widetilde{v}(x,T)| \, dx = 2\int_{-\infty}^{0} \mathrm{erf}\left(\frac{x}{\sqrt{4DT}}\right) \, dx = C\sqrt{DT}$$

where C is a constant that depends on the error function.

We are now ready to look for the convergence rate. We have just shown that the L^1 norm of the error is a constant times $\sqrt{D}$. For both the explicit upwind scheme and the Lax-Friedrichs scheme, the diffusion constant (7–8) is h times a constant as long as you have taken a constant-μ refinement path. As a result, the L^1 norm of the error decreases like $\sqrt{h}$ (or like $\sqrt{k}$ because μ is fixed.)

1.3 Second-order schemes and dispersion

We now turn to the Law-Wendroff and Beam-Warming schemes. Both have modified equations of the form:

$$u_t + a u_x = D u_{xxx}$$

where

$$D_{\text{LW}} = \frac{h^2 a}{6}\left(\mu^2 - 1\right) \qquad \text{(Lax-Wendroff)} \qquad (9)$$

$$D_{\text{BW}} = \frac{h^2 a}{6}\left(\mu - 1\right)\left(\mu - 2\right) \qquad \text{(Beam-Warming)} \qquad (10)$$

The equation $u_t + a u_x = D u_{xxx}$ is the advection equation $u_t + a u_x = 0$ with dispersive smoothing. As with the first-order schemes, we see immediately that if one chooses the usual "refinement path" of fixed μ then as $h \to 0$ the dispersive smoothing gets smaller and smaller. Again, one could use this as a method to construct solutions of the underlying hyperbolic problem.

We find solutions of $u_t + a u_x = D u_{xxx}$ via Fourier methods. Specifically, we take $u(x,t)$ of the form

$$u(x,t) = \int_{-\infty}^{\infty} \widehat{u}(\xi,t) e^{i\xi x}\, d\xi.$$

To find $\widehat{u}(\xi,t)$ we look at a single frequency in isolation and seek $c(\xi)$ such that

$$e^{i(\xi x - c(\xi) t)}$$

solves $u_t + a u_x = D u_{xxx}$. This results in

$$c(\xi) = a\xi + D\xi^3.$$

And so we see that this particular oscillating wave moves with speed $a + D\xi^2$:

$$u(x,t) = e^{i\xi(x - (a + D\xi^2)t)}$$

is a travelling wave solution of $u_t + a u_x = D u_{xxx}$. Because the speed depends on ξ, different frequencies travel with different speeds: the equation is dispersive. And so the solution of the initial value problem is

$$u(x,t) = \int_{-\infty}^{\infty} \widehat{u_0}(\xi) e^{i(\xi x - (a\xi + D\xi^3)t)}\, d\xi \text{ where } u_0(x) = \int_{-\infty}^{\infty} \widehat{u_0}(\xi) e^{i\xi x}\, d\xi.$$

Because the equation is dispersive, each frequency travels with different speed and the only travelling wave solutions must be made of a single frequency. Any other type of initial data will lose its shape as the different frequencies that make it up travel at different speeds. For dispersive equations, one considers the *group velocity*

$$c_g(\xi) = c'(\xi) = a + 3D\xi^2.$$

For Lax-Wendroff, if $\mu \in (-1, 1)$ the dispersive coefficient $D_{\text{LW}} < 0$ and so $c_g(\xi) < a$. That is, the group velocity is smaller than the advection speed a. This is why the ripples in Figure 2 of the February 14 notes occur *behind* the shock. On the other hand, for Beam-Warming if $\mu \in (0, 1)$ then $D_{\text{BW}} > 0$ and so $c_g(\xi) > a$. This is why the ripples occur *ahead* of the shock in Figure 2.

2 Phase error

Another way to understand the error in the discrete computation is to consider solutions of the form $u(x, t) = U(t)\exp(i\xi x)$, for $\xi \in \mathbb{R}$. The true solution has $U(t) = \exp(-ia\xi t)$, so that $u(x, t) = \exp(i\xi(x - at))$. To relate these waves to the solutions of the discrete formula, we identify $x_j = jh$ and $t_n = nk$, and then see that in terms of the discrete dispersion relation $\eta(\omega)$

$$\omega \;=\; e^{i\xi h}, \qquad \eta = e^{-i\kappa k}$$

so

$$\eta^n\omega^j \;=\; e^{-i\kappa nk}\,e^{i\xi jh} \;=\; e^{i(\xi x_j - \kappa t_n)} \;=\; e^{i\xi\left(x_j - \frac{\operatorname{Re}\kappa}{\xi}t_n\right)}\,e^{\operatorname{Im}\kappa\, t_n}$$

The scheme would be perfect if $\kappa = a\xi$:

$$\kappa = \frac{a}{\xi} \quad\Longrightarrow\quad \eta^n\omega^j = e^{i\xi(x_j - at_n)}$$

If κ has an imaginary part then this will cause growth or decay of the Fourier modes due to the discretization. For a stable scheme we will always have $|\eta| \le 1$ so $\operatorname{Im}\kappa \le 0$. If $\operatorname{Re}\kappa > a\xi$ (for $\xi > 0$) then the mode with wavenumber ξ is traveling too *quickly* and if $\operatorname{Re}\kappa < a\xi$ then the mode is traveling too *slowly*.

For the explicit upwind scheme we know η from the von Neumann stability analysis

$$\eta = 1 - \mu(1 - \overline{\omega}) = 1 - \mu(1 - e^{-i\xi h})$$

and so

$$\kappa = \frac{i}{k}\ln(\eta) \quad \Longrightarrow \quad \kappa = a\xi - i\frac{a}{2}h(\mu-1)\xi^2 + \mathcal{O}(h^2) = a\xi - iD_{\text{EU}}\xi^2 + \mathcal{O}(h^2)$$

where the diffusion constant D_{EU} is from the modified equation (2). And so, the $\mathcal{O}(h)$ term is precisely the $-D\xi^2$ that we would expect for a diffusion.

For the Lax-Friedrichs scheme,

$$\eta = \cos(\xi h) - i\mu\sin(\xi h)$$

hence

$$\kappa = a\xi - i\frac{h^2}{2k}(1-\mu^2)\xi^2 + \mathcal{O}(h^2) = a\xi - iD_{\text{LF}}\xi^2 + \mathcal{O}(h^2).$$

Again, we see the diffusive effect.

For the Lax-Wendroff scheme,

$$\eta = 1 - \mu^2 + \mu^2\cos(\xi h) - i\mu\sin(\xi h)$$

hence

$$\kappa = a\xi + h^2\frac{a}{6}\left(\mu^2-1\right)\xi^3 + \mathcal{O}(h^3) = a\xi + D_{\text{LW}}\xi^3 + \mathcal{O}(h^3).$$

Bearing in mind that we take $|\mu|^2 \le 1$ in order to ensure stability, we see that modes with $\xi \neq 0$ travel slightly too *slowly*, by an amount that is second-order in h, but are not damped ($\operatorname{Im}\kappa = 0$). The errors are *dispersive.* In practice, this can be even worse than dissipation since the maximum principle is not maintained. Note that the errors depend on the product ξh, the wavelength relative to the grid size, and is exactly zero if $a = 0$ or $\mu = 1$.

Finally, for the Beam-Warming scheme,

$$\eta = 1 + \frac{\mu}{2}\left(-\overline{\omega}^2 + 4\overline{\omega} - 3\right) + \frac{\mu^2}{2}\left(\overline{\omega}^2 - 2\overline{\omega} + 1\right)$$

hence

$$\kappa = a\xi + \frac{a}{6}h^2(\mu-1)(\mu-2)\xi^3 + \mathcal{O}(h^3) = a\xi + D_{\text{BW}}\xi^3 + \mathcal{O}(h^3).$$

Bearing in mind that we take $0 \le \mu \le 2$ in order to ensure stability, we see that for $0 \le \mu < 1$ modes with $\xi \neq 0$ travel slightly too *quickly* by an amount which is second-order in h. And if $1 < \mu < 2$ then modes travel slightly too slowly.

In this way we see that phase analysis produces a number of the qualitative observations we made based on the modified equations.