Department of Mathematics, University of Utah

Analysis of Numerical Methods, II

MATH 6620 – Section 001 – Spring 2024

Homework 4 Solutions

Finite difference methods for time-dependent problems

Due Friday, March 22, 2024

Submit your solutions online through Gradescope.

1. (Stability)

a. Compute the region of the stability for the leapfrog method:

$$D^0 u^n = f^n,$$

for the ODE $u' = f(t, u)$, with $u^n \approx u(t^n)$ and $f^n \approx f(t^n, u(t^n))$.

b. Investigate and discuss the consistency, stability, and practicality of the scheme,

$$D^0 u_j^n = D_+ D_- u_j^n,$$

for the PDE $u_t = u_{xx}$.

c. Use von Neumann analysis to determine stability requirements for the Lax-Friedrichs and Lax-Wendroff schemes for $u_t + a u_x = 0$, which are, respectively,

$$D^+ u_j^n = -a D_0 u_j^n + \frac{h^2}{2k} D_+ D_- u_j^n$$
$$D^+ u_j^n = -a D_0 u_j^n + \frac{a^2 k}{2} D_+ D_- u_j^n$$

<u>**Solution**</u>:

a. The leapfrog method is a multi-step scheme of the form,

$$u^{n+1} - u^{n-1} = 2k f^n.$$

Hence, the two characteristic polynomials of this scheme are,

$$\rho(w) = w^2 - 1, \qquad \sigma(w) = 2w.$$

The region of stability for this scheme is the collection of values $z \in \mathbb{C}$ such that the polynomial,

$$\rho(w) - z\sigma(w)$$

satisifes the root condition, i.e., has roots of magnitude at most 1, and also that any roots of magnitude exactly 1 must be simple. Given the characteristic polynomials above, we seek the roots of,

$$w^2 - 2zw - 1.$$

The roots of this polynomial are,

$$w = z \pm \sqrt{1+z^2}.$$

We will compute the boundary of the region of stability; i.e., these are points z such that the roots w above have magnitude 1, i.e., $w = e^{i\theta}$ for some $\theta \in [0, 2\pi)$. Using $w = e^{i\theta}$ above, we have,

$$e^{i\theta} = z \pm \sqrt{1+z^2},$$

and rearranging, this gives,

$$e^{2i\theta} - 2ze^{i\theta} - 1 = 0.$$

Dividing by $e^{i\theta}$ and simplifying yields,

$$2i \sin\theta = 2z.$$

I.e., z is purely imaginary, and in particular $z = iy$ with $y \in \mathbb{R}$ and $|y| \leq 1$. However, the points $z = \pm i$ are not in the region of stability because they are repeated (non-simple) and satisfy $|z| = |\pm i| = 1$, hence the value boundary points for the region of stability are the line segment $i(-1, 1)$ on the imaginary axis. Hence the boundary of the region of stability is a connected line segement in $\mathbb{C}$: Since roots of polynomials are continuous with respect to the coefficients of the polynomial, the region of stability is connected, and since it's boundary is a line segment, the region is either all of $\mathbb{C}$, or just the line segment itself. A direct computation with, say, $z = 2i$, shows that $z \notin \mathbb{C}$, and therefore the region of stability is just the line segment $i(-1, 1) \subset \mathbb{C}$, which is a subset of the imaginary axis.

b. This scheme is consistent with respect to the heat equation:

$$u_t = u_{xx}.$$

Establishing this consistency is a direct computation with local truncation errors:

$$D_+D_-u(x,t) = u_{xx}(x,t) + \mathcal{O}(h^2),$$
$$D^0u(x,t) = u_t(x,t) + \mathcal{O}(k^2).$$

Hence, this method is second-order in both space and time discretizations. However, the semi-discrete version of this problem corresponds to the ODE system,

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{u}(t) = \boldsymbol{A}\boldsymbol{u},$$

where $\boldsymbol{u}^T \approx (u(x_1,t), \ldots, u(x_M,t))^T$, with $\{x_j\}_{j\in[M]}$ an equispaced mesh on the spatial domain, and with $\boldsymbol{A}$ a symmetric (negative semi-definite) matrix, and hence has spectrum lying along the negative real axis. We have seen from the previous part that the region of stability for the leap-frog method is a subset of the imaginary axis, and in particular, there is no $k > 0$ such that $k\lambda(\boldsymbol{A})$ is a subset of the imaginary axis. Therefore, this scheme will not be stable since (all) modes will grow in time, whereas the exact solution has modes decaying in time. Because of this lack of stability, this method is also impractical as it will not produce a convergent scheme. (One can also come to the same conclusion through, say, von Neumann stability analysis.)

c. For von Neumann stability, we use the ansätze,

$$u_j^n = e^{ijh\omega}, \qquad\qquad u_j^{n+1} = g(\omega)e^{ijh\omega}$$

where ω is a (real-valued) frequency. By plugging these into the scheme, we will derive an expression for $g(\omega)$, and we seek to make this expression no larger than one in magnitude for arbitrary ω. To set up the computations, we note that our assumptions on u_j^n imply,

$$\begin{aligned} D_0 u_j^n &= e^{ijh\omega}\frac{1}{2h}\left(e^{ih\omega} - e^{-ih\omega}\right) = e^{ijh\omega}\frac{i}{h}\sin h\omega, \\ D_+D_- u_j^n &= e^{ijh\omega}\frac{1}{h^2}\left(-2 + e^{-ih\omega} + e^{ih\omega}\right) = e^{ijh\omega}\frac{2}{h^2}(\cos h\omega - 1) = -e^{ijh\omega}\frac{4}{h^2}\sin^2\left(\frac{h\omega}{2}\right). \\ D^+ u_j^n &= e^{ijh\omega}\frac{g(\omega)-1}{k}. \end{aligned}$$

Putting these all together, then the Lax-Friedrichs scheme reads,

$$\frac{1}{k}(g(\omega) - 1) = -\frac{ai}{h}\sin h\omega - \frac{2}{k}\sin^2\left(\frac{h\omega}{2}\right),$$

i.e.,

$$g(\omega) = \cos h\omega - i\frac{ak}{h}\sin h\omega.$$

Hence, the amplification factor is,

$$|g(\omega)| = \sqrt{\cos^2(h\omega) + \nu^2\sin^2(h\omega)}, \qquad\qquad n := \frac{ak}{h}.$$

We have that $|\nu| \leq 1$ implies $|g(\omega)| \leq 1$, and $|\nu| > 1$ implies $|g(\omega)| > 1$. Hence, the von Neumann stability condition for the Lax-Friedrichs scheme is,

$$|\nu| = \left|\frac{ak}{h}\right| \leq 1.$$

Similarly, with our set up computations, the Lax-Wendroff scheme reads,

$$\frac{1}{k}(g(\omega) - 1) = -\frac{ai}{h}\sin h\omega - \frac{2a^2k}{h^2}\sin^2\left(\frac{h\omega}{2}\right),$$

I.e.,

$$g(\omega) = 1 - 2\nu^2\sin^2\left(\frac{h\omega}{2}\right) - i\nu\sin h\omega$$

Note that $|\nu| > 1$ implies that,

$$|g(\omega)| \geq |\mathrm{Im} g(\omega)| = |\nu\sin h\omega| > 1,$$

for say $h\omega = \pi/2$. Hence, this scheme could only be von Neumann stable for $|\nu| \leq 1$. In the case when $|\nu| \leq 1$, then we have,

$$\begin{aligned} |g(\omega)|^2 &= \left(1 - 2\nu^2\sin^2\left(\frac{h\omega}{2}\right)\right)^2 + \nu^2\sin^2 h\omega \\ &= \left(1 - \nu^2 + \nu^2\cos h\omega\right)^2 + \nu^2\sin^2 h\omega. \end{aligned}$$

Now we make the substitution,

$$t = \cos h\omega, \qquad 1 - t^2 = \sin^2 h\omega,$$

and so the amplification factor (squared) is given by,

$$|g(\omega)|^2 = f(t) := (1 - \nu^2 + \nu^2 t)^2 + \nu^2(1 - t^2), \qquad t \in [0, 1].$$

Through direct calculus, we have,

$$f'(t) = 2\nu^2(1 - \nu^2)(1 - t) \geq 0, \qquad t \in [0, 1] \text{ and } |\nu| \leq 1,$$

and so f achieves its maximum at $t = 1$, which is $f(1) = 1$. Therefore, $\max_{h\omega} |g(\omega)|^2 = \max_{t\in[0,1]} f(t) = 1$ for $|\nu| \leq 1$. Hence, Lax-Wendroff is also von Neumann stable for $|\nu| = \left|\frac{ak}{h}\right| \leq 1$.

2. (Semi-Lagrangian methods) Consider the PDE,

$$u_t + a u_x = 0, \qquad u(x, 0) = u_0(x), \qquad a > 0,$$

with periodic boundary conditions for $x \in [0, 2\pi)$. Use our standard notation, $u_j^n \approx u(x_j, t^n)$ with $x_j = jh$ and $t^n = nk$ with $h = 2\pi/M$ and $k > 0$. In this problem, we will investgiate particular *semi-Lagrangian* schemes. Of particular use will be characteristic curves. Let $X(t; x_j, t^n)$ denote the characteristic curve passing through at $x = x_j$ at time $t = t^n$.

a. Consider the following scheme: Let the values of u_j^n for all j be given. We compute u_j^{n+1} as,

$$u_j^{n+1} = p_j(y), \qquad y := X(t^n; x_j, t^{n+1})$$

where p_j is the linear interpolant formed by the data (x_{j-1}, u_{j-1}^n) and (x_j, u_j^n). Show that this scheme is equivalent to the upwind scheme, and give conditions on a, k, h that ensure that $p_j(y)$ is an interpolatory evaluation instead of an extrapolatory evaluation.

b. Consider the scheme,

$$u_j^{n+1} = q_j(y), \qquad y := X(t^n; x_j, t^{n+1}),$$

where q is the quadratic interpolant formed from (x_{j-1}, u_{j-1}^n), (x_j, u_j^n), and (x_{j+1}, u_{j+1}^n). Explicitly write the scheme as a function of $\{u_j^n\}_j$ for this method, and compare against other schemes for this PDE that we've discussed in class.

Solution:

a. We first compute characteristics. The characteristics $X(t; x_0, t_0)$ for this problem are given by the ODE,

$$X'(t) = a, \qquad X(t_0) = x_0,$$

for all $t \in \mathbb{R}$, which has exact solution,

$$X(t; x_0, t_0) = x_0 + a(t - t_0)$$

Hence, we have,

$$y = X(t^n; x_j, t^{n+1}) = x_j + a(t^n - t^{n+1}) = x_j - ak$$

The linear interpolant p_j of the data in question is given by,

$$p_j(x) = u_{j-1}^n \frac{x - x_j}{x_{j-1} - x_j} + u_j^n \frac{x - x_{j-1}}{x_j - x_{j-1}},$$

and evaluating this at $y = x_j - ak = x_{j-1} + h - ak$ yields,

$$u_j^{n+1} = p_j(y) = u_{j-1}^n \frac{ak}{h} + u_j^n \frac{h - ak}{h} = u_j^n + \frac{ak}{h}\left(u_{j-1}^n - u_j^n\right) = u_j^n - akD_- u_j^n,$$

which coincides with the upwind scheme ($a > 0$). Interpolation occurs when,

$$x_{j-1} \le y \le x_j \quad \Longrightarrow \quad x_{j-1} \le x_j - ak \le x_j.$$

Since $a > 0$, the upper inequality is always satisfied. Using $h = x_j - x_{j-1}$, the lower inequality is satisfied, i.e., interpolation occurs, when,

$$\frac{ak}{h} \le 1.$$

If this condition is violated, the evaluation is extrapolation.

b. The characteristics haven't changed, so we have the same value of y as before:

$$y = x_j - ak = x_{j-1} + (h - ak) = x_{j+1} - (h + ak)$$

The new quadratic interpolant is given by,

$$q(x) = u_{j-1}^n \frac{(x - x_j)(x - x_{j+1})}{(x_{j-1} - x_j)(x_{j-1} - x_{j+1})} + u_j^n \frac{(x - x_{j-1})(x - x_{j+1})}{(x_j - x_{j-1})(x_j - x_{j+1})} + u_{j+1}^n \frac{(x - x_{j-1})(x - x_j)}{(x_{j+1} - x_j)(x_{j+1} - x_{j-1})}$$

Evaluating this at y gives,

$$\begin{aligned} q(y) &= u_{j-1}^n \frac{(-ak)(-h - ak)}{(-h)(-2h)} + u_j^n \frac{(h - ak)(-h - ak)}{h(-h)} + u_{j+1}^n \frac{(h - ak)(-ak)}{h(2h)} \\ &\overset{\nu := ak/h}{=} \frac{1}{2} u_{j-1}^n \left(\nu + \nu^2\right) + u_j^n \left(1 - \nu^2\right) + \frac{1}{2} u_{j+1}^n \left(-\nu + \nu^2\right). \end{aligned}$$

Hence, the full scheme is,

$$u_j^{n+1} = \frac{1}{2} u_{j-1}^n \left(\nu + \nu^2\right) + u_j^n \left(1 - \nu^2\right) + \frac{1}{2} u_{j+1}^n \left(-\nu + \nu^2\right), \qquad (\nu = ak/h).$$

To compare against other schemes, we rearrange and use D^+ on the left-hand side to obtain:

$$\begin{aligned} D^+ u_j^n &= \frac{\nu}{2k}\left(u_{j-1}^n - u_{j+1}^n\right) + \frac{\nu^2}{2k}\left(u_{j-1}^n - 2u_j^n + u_{j+1}^n\right) \\ &= \frac{a}{2h}\left(u_{j-1}^n - u_{j+1}^n\right) + \frac{a^2 k}{2h^2}\left(u_{j-1}^n - 2u_j^n + u_{j+1}^n\right) \\ &= -aD_0 u_j^n + \frac{a^2 k}{2} D_+ D_- u_j^n. \end{aligned}$$

We recognize this scheme as the Lax-Wendroff scheme, e.g., from the previous problem. Note that this evaluation is interpolatory under precisely the same condition as the previous part, $ak/h \le 1$.

3. (Orders of convergence)
This problem provides exercises to numerically confirm expected orders of convergence.

a. For the PDE,

$$u_t = u_{xx}, \qquad u(x,0) = \exp(\sin x), \qquad (x,t) \in [0, 2\pi) \times (0, T],$$

with periodic boundary conditions in x, numerically confirm that the scheme,

$$D^+ u_j^n = \frac{1}{2} D_+ D_- u_j^n + \frac{1}{2} D_+ D_- u_j^{n+1},$$

is second-order in space and time, say for $T = 1$.

b. For the PDE,

$$u_t = -u_x, \qquad u(x,0) = \exp(\sin x), \qquad (x,t) \in [0, 2\pi) \times (0, T],$$

with periodic boundary conditions in x, numerically confirm that the scheme,

$$D^0 u_j^n = -D_0 u_n^j,$$

is second-order in space and time, say for $T = 1$.

__Solution__:

a. The time-differencing scheme for this PDE is A-stable, and since the spectrum of $D_+ D_-$ is in the left half-plane, then this scheme is unconditionally stable. We expect the scheme is convergent to the order of the local truncation error, which is

$$\mathcal{O}(h^2 + k^2).$$

To numerically test this, we will fix some $k > 0$ very small, and decrease h, keeping $h \gg k$, so that we expect that the above error will behave like,

$$\mathcal{O}(h^2 + k^2) \overset{k \lll h}{=} \mathcal{O}(h^2).$$

To compute errors, we numerically solve the PDE using the scheme up to time $T = 1$, and compute errors in the scaled ℓ^2 norm:

$$\text{Scaled } \ell^2 \text{ error} = \|\boldsymbol{e}\|_{2,h} := \sqrt{h} \|\boldsymbol{e}\|_2,$$

where $\|\cdot\|_2$ is the standard Euclidean ℓ^2 norm, and $\boldsymbol{e}$ is the vector of spatial pointwise errors at time $T = 1$. The results are shown in Figure 1, showing that we numerically observe second-order convergent behavior in both h and k.

b. This scheme is leapfrog in time, whose region of stability is a subset of the imaginary axis. One can show that the spectrum of D_0 is also along the imaginary axis, with maximum spread $\pm i/h$ (see slide `D12-S08`). Therefore, we have a stability restriction of the form,

$$k \left| \frac{i}{h} \right| \leq 1 \quad \Longrightarrow \quad k \leq h.$$

The LTE of this scheme is second-order in both h and k, and so like the previous exercise we expect error at $T = 1$ to behave like,

$$\mathcal{O}(h^2 + k^2).$$

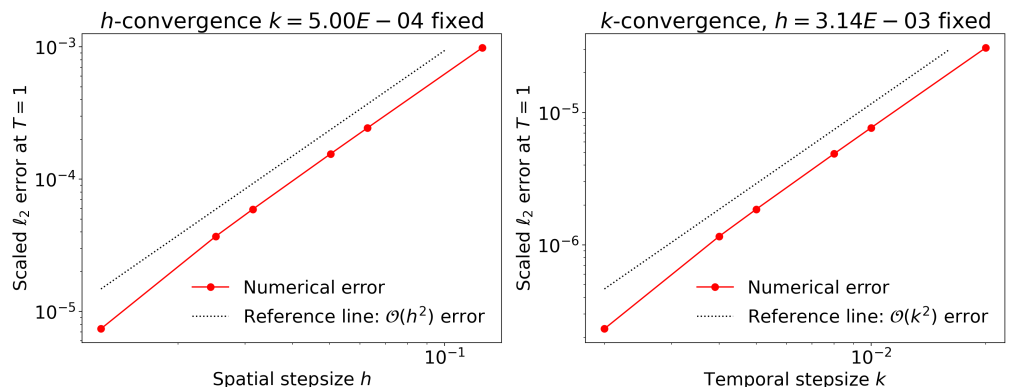

Figure 1: Numerical results for problem 3a. Left: Spatial convergence, where k is fixed at $k = 5 \times 10^{-4}$. Right: Temporal convergence, where h is fixed at $h = 3.14 \times 10^{-3}$. In both cases, we observe relatively clean second-order convergence, providing supporting evidence that this scheme is second-order in space and time.

With the restriction $k \leq h$, then we can test spatial accuracy by fixing $k \ll h$, and varying h. In this case the strategy from the previous part can be employed (using the same measure of scaled ℓ^2 error at $T = 1$), in order to measure h-convergence. The results are shown in Figure 2 (left). In order to measure convergence in the parameter k, we cannot take $h \ll k$ since this would violate stability (i.e., we wouldn't see convergence at all). Therefore, we amend the experiment as follows: If we choose $k = ch$, for some $c \leq 1$, then

$$\mathcal{O}(h^2 + k^2) \overset{h=k/c}{=} \mathcal{O}(k^2),$$

and using this strategy, we can observe second-order convergence in k. (Note that if actual k-order of convergence is larger than 2, then this experiment will not reveal this behavior; i.e., we will show that the order of convergence is *at least* 2.) With some fixed c, the actual implementation of this strategy is more subtle: in order to maintain uniform steps in space and time, taking $k = ch$ might not be possible. (E.g., if c is rational, then if h uniformly partitions $[0, 2\pi)$, it is impossible for it to uniformly partition $[0, 1]$ since 2π is irrational but 1 is rational.) Therefore, our actual approach is to fix an integer $N > 1$, define $k = T/N$, and then to define another integer M as,

$$M = \left\lceil c \frac{2\pi}{k} \right\rceil, \qquad\qquad h := \frac{2\pi}{M},$$

for any $c \leq 1$. The above prescription ensures that $h \leq ck$. In practice we choose $c = 0.1$. We show in Figure 2 the results for convergence in k (right plot), which verifies $\mathcal{O}(k^2)$ convergence.

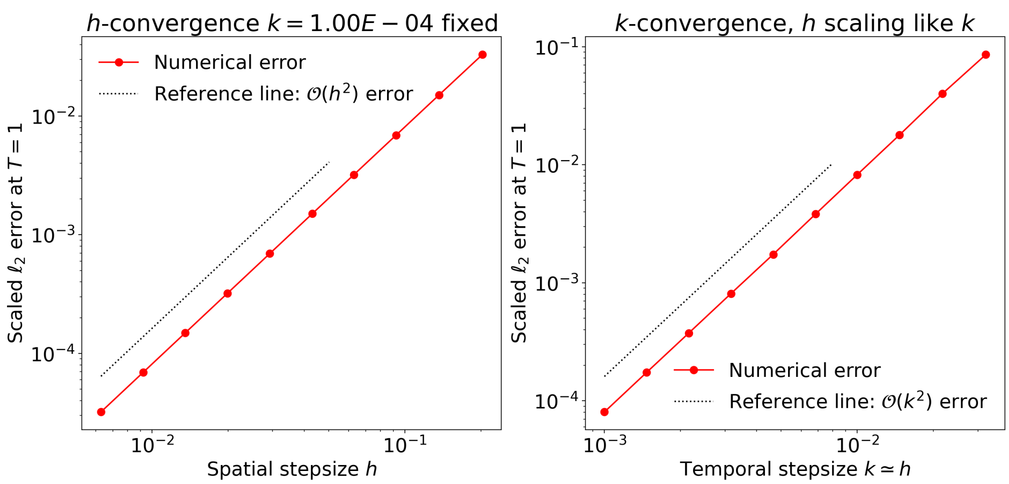

Figure 2: Numerical results for problem 3a. Left: Spatial convergence, where k is fixed at $k = 5 \times 10^{-4}$. Right: Temporal convergence, where k is varied to test for convergence, but $h = ck$, $c \approx 0.1 < 1$, is enforced to maintain stability. In both cases, we observe relatively clean second-order convergence, providing supporting evidence that this scheme is second-order in space and time.