Numerical Analysis – Lecture 23[1]

Problem 5.21 *(Stability of leapfrog)* We have

$$\hat{u}^{n+1}(\theta) - 2\mathrm{i}\mu(\sin\theta)\hat{u}^n(\theta) - \hat{u}^{n-1}(\theta) = 0, \qquad n = 1, 2, \ldots,$$

and our goal is to determine values of μ such that $|\hat{u}^n(\theta)|$ is uniformly bounded for $\theta \in [-\pi, \pi]$ and $n \geq 1$. This is an example of a *difference equation*. Specifically, given $aw_{n+1} + bw_n + cw_{n-1} = 0$, $n = 1, 2, \ldots$, where $a \neq 0$, we let $\omega_\pm$ be the zeros of $a\omega^2 + b\omega + c = 0$. Provided that $\omega_- \neq \omega_+$, the general solution is $w_n = \alpha\omega_+^n + \beta\omega_-^n$, where α, β are constants, dependent on the initial values w_0 and w_1 (if $\omega_- = \omega_+$ then the solution is $w_n = (\alpha + \beta n)\omega_+^n$). In our case, we obtain

$$\omega_\pm(\theta) = \mathrm{i}\mu\sin\theta \pm \sqrt{1 - \mu^2\sin^2\theta}, \qquad |\theta| \leq \pi.$$

However, stability $\Leftrightarrow$ $|\omega_\pm(\theta)| \leq 1$ (for all $|\theta| \leq \pi$) and this, as can be easily verified, is true when $|\mu| \leq 1$.

Problem 5.22 *(Stability in the presence of boundaries)* It is easy to extend Fourier analysis to cater for the Euler method $u_m^{n+1} = u_m^n + \mu(u_{m+1}^n - u_m^n)$, $m = 0, 1, \ldots, M+1$ (where $M\Delta x = 1$), with the initial condition $u(x, 0) = \phi(x)$, $x \in [0, 1]$, and zero boundary condition along $x = 1$. Thus, we consider the Cauchy problem for the advection equation with the initial condition $u(x, 0) = \phi(x)$ for $x \in [0, 1]$, $u(x, 0) = 0$ otherwise (it isn't differentiable, but this is not much of a problem). Solving the Cauchy problem with Euler, we recover identical u_m^n for $m = 0, 1, \ldots, M$. This justifies using Fourier analysis for the problem with a boundary.

Unfortunately, this is no longer true for leapfrog. Closer examination reveals that we cannot use leapfrog at $m = 0$, since u_{-1}^n is unknown. The naive remedy, setting $u_{-1}^n = 0$, leads to instability, which propagates from the boundary inwards. We can recover stability letting, for example, $u_0^{n+1} = u_1^n$ (the proof is *very* difficult).

Problem 5.23 *(The wave equation)* Once we know how to solve the advection equation, it is easy to derive methods for the wave equation

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

given in an appropriate domain of $\mathbb{R} \times \mathbb{R}_+$ with appropriate initial (for u and $\partial u/\partial t$) and boundary conditions. Specifically, suppose that (v, w) are the solution of the system

$$\frac{\partial v}{\partial t} = \frac{\partial w}{\partial x}, \qquad \frac{\partial w}{\partial t} = \frac{\partial v}{\partial x} \qquad \text{of advection equations. Then} \qquad \frac{\partial^2 v}{\partial t^2} = \frac{\partial^2 w}{\partial t \partial x} = \frac{\partial^2 w}{\partial x \partial t} = \frac{\partial^2 v}{\partial x^2}.$$

Therefore (imposing correctly initial and boundary conditions) $u = v$.
Once we have a method for the advection equation, we may generalize it easily to the system $\partial \boldsymbol{u}/\partial t = A\partial \boldsymbol{u}/\partial x$, where all the eigenvalues of A are real and nonzero (to ensure hyperbolicity). For the wave equation we have $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. If the original method is stable for all $a \leq \mu \leq b$, say, where $a < 0 < b$, then the more general method is stable, provided that $a \leq \lambda\mu \leq b$ for all $\lambda \in \sigma(A)$.
For the wave equation solved with the method (5.9), (i.e., in one dimension, $u_m^{n+1} = u_m^n + \mu(u_{m+1}^n - u_m^n)$) we eliminate the w_m^ns from

$$v_m^{n+1} = v_m^n + \mu(w_{m+1}^n - w_m^n), \qquad w_m^{n+1} = w_m^n + \mu(v_m^n - v_{m-1}^n).$$

[1]Please email all corrections and suggestions to these notes to `A.Iserles@damtp.cam.ac.uk`. All handouts are available on the WWW at the URL `http://www.damtp.cam.ac.uk/user/na/PartII/Handouts.html`.

This results (letting $u_m^n = v_m^n$) in the *leapfrog scheme*

$$u_m^{n+1} - 2u_m^n + u_m^{n-1} = \mu^2(u_{m+1}^n - 2u_m^n + u_{m-1}^n)$$

(note that the Courant number is now μ^2).

Implementation 5.24 It is quite usual to solve *hyperbolic* PDEs (advection equation, wave equation, Schrödinger equation, Euler equations of invicid compressible flow...) by explicit methods, because stability conditions can be typically satisfied with $\Delta t \sim \Delta x$. However, solving parabolic equations with explicit methods leads typically to restrictions of the form $\Delta t \sim (\Delta x)^2$, and this is unacceptable. Instead, we use implicit methods, e.g. Crank–Nicolson. This means that in each time step we need to solve a system of linear algebraic equations, and this can be costly if there are several space dimensions. Thus, consider the diffusion equation

$$\frac{\partial u}{\partial t} = \nabla^\top(a(x,y)\nabla u) + f(x,y) = \frac{\partial}{\partial x}\left(a(x,y)\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left(a(x,y)\frac{\partial u}{\partial y}\right) + f(x,y), \tag{5.12}$$

where $a(x,y) > 0, f(x,y)$ are given, together with initial conditions on $[0,1]^2$ and Dirichlet boundary conditions along $\partial[0,1]^2 \times [0,\infty)$. Replace each space derivative by *central differences* at midpoints,

$$\frac{\mathrm{d}g(\xi)}{\mathrm{d}\xi} \approx \frac{g(\xi + \frac{1}{2}\Delta x) - g(\xi - \frac{1}{2}\Delta x)}{\Delta x},$$

resulting in the ODE system

$$\begin{aligned} u'_{l,m} = \frac{1}{(\Delta x)^2}[&a_{l-\frac{1}{2},m}u_{l-1,m} + a_{l+\frac{1}{2},m}u_{l+1,m} + a_{l,m-\frac{1}{2}}u_{l,m-1} + a_{l,m+\frac{1}{2}}u_{l,m+1} \\ &- (a_{l-\frac{1}{2},m} + a_{l+\frac{1}{2},m} + a_{l,m-\frac{1}{2}} + a_{l,m+\frac{1}{2}})u_{l,m}] + f_{l,m}. \end{aligned} \tag{5.13}$$

The system (5.13) can be in turn solved by an implicit ODE method, e.g. Crank–Nicolson, except that this requires a (costly) solution of a large algebraic system in each time step.

Intermezzo 5.25 *(Linear systems of ODEs)* The system (5.13) is linear and (assuming for the time being zero boundary conditions and $f \equiv 0$) homogeneous. With greater generality, let us consider the ODE system

$$\boldsymbol{y}' = A\boldsymbol{y}, \qquad \boldsymbol{y}(0) = \boldsymbol{y}_0. \tag{5.14}$$

We define formally a *matrix exponential* by Taylor series, $\mathrm{e}^B = \sum_{k=0}^{\infty} \frac{1}{k!}B^k$, and easily verify by formal differentiation that $\mathrm{d}\,\mathrm{e}^{tA}/\mathrm{d}t = A\mathrm{e}^{tA}$, therefore $\boldsymbol{y}(t) = \mathrm{e}^{tA}\boldsymbol{y}_0$.
Much of our theory of one-step methods for ODEs is illuminated by the observation that, in a linear case, we are approximating a matrix exponential. Thus,

Euler: $\boldsymbol{y}_n = (I + hA)^n\boldsymbol{y}_0$ and $1 + z = \mathrm{e}^z + \mathcal{O}(z^2)$;

TR: $\boldsymbol{y}_n = \left[\left(I - \frac{1}{2}hA\right)^{-1}\left(I + \frac{1}{2}hA\right)\right]^n \boldsymbol{y}_0$ and $\frac{1+\frac{1}{2}z}{1-\frac{1}{2}z} = \mathrm{e}^z + \mathcal{O}(z^3)$.

Technique 5.26 *(Splitting methods)* Going back to (5.13), we *split* $A = B + C$, so that B and C are constructed from the contribution of discretizations in the x and y directions respectively. In other words, B includes all the $a_{\ell\pm\frac{1}{2},m}$ terms and C consists of the remaining $a_{\ell,m\pm\frac{1}{2}}$ components. Note that, if the grid is ordered by columns, B is tridiagonal. However, if the grid is ordered by rows, C is tridiagonal. Recall that, for $z_1, z_2 \in \mathbb{C}$, $\mathrm{e}^{z_1+z_2} = \mathrm{e}^{z_1}\mathrm{e}^{z_2}$ and suppose for a moment that this property extends to matrices, i.e. that $\mathrm{e}^{tA} = \mathrm{e}^{t(B+C)} = \mathrm{e}^{tB}\mathrm{e}^{tC}$. Had this been true, we could have approximated each component with the trapezoidal rule, say, to produce

$$\boldsymbol{u}^{n+1} = \left(I - \tfrac{\Delta t}{2}B\right)^{-1}\left(I + \tfrac{\Delta t}{2}B\right)\left(I - \tfrac{\Delta t}{2}C\right)^{-1}\left(I + \tfrac{\Delta t}{2}C\right)\boldsymbol{u}^n. \tag{5.15}$$

The advantage of (5.15) lies in the fact that (up to a known permutation) both $I - \frac{\Delta t}{2}B$ and $I - \frac{\Delta t}{2}C$ are tridiagonal, hence can be solved very cheaply.
Unfortunately, the assumption that $\mathrm{e}^{t(B+C)} = \mathrm{e}^{tB}\mathrm{e}^{tC}$ is, in general, false. *[Note: It is true, however, for $a(x,y) \equiv$ const.]* Not all hope is lost, though, and we will demonstrate that, suitably implemented, splitting is a powerful technique to reduce drastically the expense of numerical solution.