Numerical Analysis - Part II

Anders C. Hansen

Lecture 7

Partial differential equations of evolution

Solving the diffusion equation

We consider the solution of the *diffusion equation*

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}\,, \qquad 0 \le x \le 1, \quad t \ge 0,$$

with *initial conditions* $u(x, 0) = u_0(x)$ for $t = 0$ and Dirichlet *boundary conditions* $u(0, t) = \phi_0(t)$ at $x = 0$ and $u(1, t) = \phi_1(t)$ at $x = 1$.

Semidiscretization

Let $u_m(t) = u(mh, t)$, $m = 1...M$, $t \geq 0$. Approximating $\partial^2/\partial x^2$ as before, we deduce from the PDE that the *semidiscretization*

$$\frac{du_m}{dt} = \frac{1}{h^2}(u_{m-1} - 2u_m + u_{m+1}), \qquad m = 1...M \tag{1}$$

carries an error of $\mathcal{O}(h^2)$. This is an ODE system, and we can solve it by any ODE solver.

Recall the trapezoidal rule

Suppose that we want to solve the differential equation

$$y' = f(t, y), \qquad y(t_0) = y_0.$$

The trapezoidal rule is given by the formula

$$y_{n+1} = y_n + \tfrac{1}{2}k\Big(f(t_n, y_n) + f(t_{n+1}, y_{n+1})\Big),$$

where $k = t_{n+1} - t_n$ is the step size.

The Crank–Nicolson scheme

Discretizing the ODE (1) with the trapezoidal rule, we obtain

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

where $m = 1...M$. Thus, each step requires the solution of an $M \times M$ TST system. The error of the scheme is $\mathcal{O}(k^3 + kh^2)$, so basically the same as with Euler's method. However, as we will see, Crank–Nicolson enjoys superior stability features.

Stability analysis via eigenvalues

Suppose that a numerical method (with zero boundary conditions) can be written in the form

$$\mathbf{u}_h^{n+1} = A_h \mathbf{u}_h^n,$$

where $\mathbf{u}_h^n \in \mathbb{R}^M$ are vectors, $A_h \in \mathbb{R}^{M\times M}$ is a matrix, and $h = \frac{1}{M+1}$. Then $\mathbf{u}_h^n = (A_h)^n \mathbf{u}_h^0$, and

$$\|\mathbf{u}_h^n\| = \|(A_h)^n \mathbf{u}_h^0\| \le \|(A_h)^n\| \cdot \|\mathbf{u}_h^0\| \le \|A_h\|^n \cdot \|\mathbf{u}_h^0\|,$$

for any vector norm $\|\cdot\|$ and the induced matrix norm $\|A\| = \sup \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}$. If we define stability as preserving the boundedness of $\mathbf{u}_h^n$ with respect to the norm $\|\cdot\|$, then, from the inequality above,

$$\|A_h\| \le 1 \text{ as } h \to 0 \quad \Rightarrow \quad \text{the method is stable.}$$

Crank–Nicolson method for diffusion equation

Let

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

where $m = 1 \ldots M$. Then $B\mathbf{u}^{n+1} = C\mathbf{u}^n$, where the matrices B and C are Toeplitz symmetric tridiagonal (TST),

$$\mathbf{u}^{n+1} = B^{-1}C\mathbf{u}^n, \qquad \begin{aligned} B &= I - \tfrac{1}{2}\mu A_*\,, \\ C &= I + \tfrac{1}{2}\mu A_*\,, \end{aligned} \qquad A_* = \begin{bmatrix} -2 & 1 & & \\ 1 & \ddots & \ddots & \\ & \ddots & \ddots & 1 \\ & & 1 & -2 \end{bmatrix}_{M\times M}.$$

All $M \times M$ TST matrices share the same eigenvectors, hence so does $B^{-1}C$. Moreover, these eigenvectors are orthogonal. Therefore, also $A = B^{-1}C$ is normal and its eigenvalues are

$$\lambda_k(A) = \frac{\lambda_k(C)}{\lambda_k(B)} = \frac{1 - 2\mu \sin^2 \frac{1}{2}\pi kh}{1 + 2\mu \sin^2 \frac{1}{2}\pi kh} \quad \Rightarrow \quad |\lambda_k(A)| \leq 1, \qquad k = 1 \ldots M.$$

Consequently Crank–Nicolson is stable for all $\mu > 0$.

Note: Similarly to the situation with stiff ODEs, this *does not* mean that $k = \Delta t$ may be arbitrarily large, but that the only valid consideration in the choice of $k = \Delta t$ vs $h = \Delta x$ is accuracy.

Convergence of the Crank-Nicolson method for diffusion equation

It is not difficult to verify that the local error of the Crank-Nicolson scheme is $\eta_m^n = \mathcal{O}(k^3 + kh^2)$, where $\mathcal{O}(k^3)$ is inherited from the trapezoidal rule (compared to $\mathcal{O}(k^2)$ for the Euler method). We also have

$$\|\boldsymbol{\eta}^n\| = \{h \textstyle\sum_{m=1}^{M} |\eta_m^n|^2\}^{1/2} = \mathcal{O}(k^3 + kh^2)\,.$$

Hence, for the error vectors $\mathbf{e}^n$ we have

$$B\mathbf{e}^{n+1} = C\mathbf{e}^n + \boldsymbol{\eta}^n \quad \Rightarrow \quad \|\mathbf{e}^{n+1}\| \le \|B^{-1}C\| \cdot \|\mathbf{e}^n\| + \|B^{-1}\| \cdot \|\boldsymbol{\eta}^n\|\,.$$

We have just proved that $\|B^{-1}C\| \le 1$, and we also have $\|B^{-1}\| \le 1$, because all the eigenvalues of B are greater than 1 (by Gershgorin's theorem). Therefore, $\|\mathbf{e}^{n+1}\| \le \|\mathbf{e}^n\| + \|\boldsymbol{\eta}^n\|$, and

$$\|\mathbf{e}^n\| \le \|\mathbf{e}^0\| + n\|\boldsymbol{\eta}\| = n\|\boldsymbol{\eta}\| \le \tfrac{cT}{k}(k^3 + kh^2) = cT(k^2 + h^2).$$

Thus, taking $k = \alpha h$ will result in $\mathcal{O}(h^2)$ error of approximation.

The advection equation

We consider the solution of the *advection equation*

$$\frac{\partial u}{\partial t} = \frac{\partial u}{\partial x}, \qquad 0 \leq x \leq 1, \quad t \geq 0,$$

with *initial conditions* $u(x,0) = u_0(x)$ for $t = 0$ and Dirichlet *boundary conditions* $u(0,t) = \phi_0(t)$ at $x = 0$ and $u(1,t) = \phi_1(t)$ at $x = 1$.

Crank–Nicolson for advection equation

Let

$$u_m^{n+1} - u_m^n = \tfrac{1}{4}\mu(u_{m+1}^{n+1} - u_{m-1}^{n+1}) + \tfrac{1}{4}\mu(u_{m+1}^n - u_{m-1}^n), \qquad m = 1 \ldots M\,.$$

(This is the trapezoidal rule applied to the semidiscretization of advection equation $\frac{\partial u}{\partial t} = \frac{\partial u}{\partial x}$). In this case, $\mathbf{u}^{n+1} = B^{-1}C\mathbf{u}^n$, where the matrices B and C are Toeplitz antisymmetric tridiagonal,

$$B = \begin{bmatrix} 1 & -\frac{1}{4}\mu & & \\ \frac{1}{4}\mu & 1 & \ddots & \\ & \ddots & \ddots & -\frac{1}{4}\mu \\ & & \frac{1}{4}\mu & 1 \end{bmatrix}, \qquad C = \begin{bmatrix} 1 & \frac{1}{4}\mu & & \\ -\frac{1}{4}\mu & 1 & \ddots & \\ & \ddots & \ddots & \frac{1}{4}\mu \\ & & -\frac{1}{4}\mu & 1 \end{bmatrix}.$$

Crank–Nicolson for advection equation

Similarly to Exercise 4, the eigenvalues and eigenvectors of the matrix

$$S = \begin{bmatrix} \alpha & \beta & & \\ -\beta & \alpha & \ddots & \\ & \ddots & \ddots & \beta \\ & & -\beta & \alpha \end{bmatrix},$$

are given by $\lambda_k = \alpha + 2\,\mathrm{i}\,\beta\cos kx$, and $\mathbf{w}_k = (\mathrm{i}^m \sin kmx)_{m=1}^{M}$, where $x = \pi h = \frac{\pi}{M+1}$. So, all such S are normal and share the same eigenvectors, hence so does $A = B^{-1}C$, hence A is normal and

$$\lambda_k(A) = \frac{\lambda_k(C)}{\lambda_k(B)} = \frac{1 + \frac{1}{2}\,\mathrm{i}\,\mu\cos kx}{1 - \frac{1}{2}\,\mathrm{i}\,\mu\cos kx} \quad \Rightarrow \quad |\lambda_k(A)| = 1, \qquad k = 1...M.$$

So, Crank–Nicolson is again stable for all $\mu > 0$.

Finally, consider the Euler method for advection equation

$$u_m^{n+1} - u_m^n = \mu(u_{m+1}^n - u_m^n), \qquad m = 1...M\,.$$

We have $\mathbf{u}^{n+1} = A\mathbf{u}^n$, where

$$A = \begin{bmatrix} 1-\mu & \mu & & \\ & 1-\mu & \ddots & \\ & & \ddots & \mu \\ & & & 1-\mu \end{bmatrix},$$

but A is *not* normal, and although its eigenvalues are bounded by 1 for $\mu \le 2$, it is the spectral radius of AA^T that matters, and we have $\rho(AA^T) \approx (|1-\mu| + |\mu|)^2$, so that the method is stable only if $\mu \le 1$.

Solving the diffusion equation

We consider the solution of the *diffusion equation*

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \qquad 0 \leq x \leq 1, \quad t \geq 0,$$

with *initial conditions* $u(x, 0) = u_0(x)$ for $t = 0$ and Dirichlet *boundary conditions* $u(0, t) = \phi_0(t)$ at $x = 0$ and $u(1, t) = \phi_1(t)$ at $x = 1$.

What if $-\infty < x < \infty$?

Let us now assume a recurrence of the form

$$\sum_{k=r}^{s} a_k u_{m+k}^{n+1} = \sum_{k=r}^{s} b_k u_{m+k}^{n}, \qquad n \in \mathbb{Z}^+, \tag{3}$$

where m ranges over $\mathbb{Z}$. (Within our framework of discretizing PDEs of evolution, this corresponds to $-\infty < x < \infty$ in the undelying PDE and so there are no explicit boundary conditions, but the initial condition must be square-integrable in $(-\infty, \infty)$: this is known as a *Cauchy problem.*)

The coefficients a_k and b_k are independent of m, n, but typically depend upon μ. We investigate stability by *Fourier analysis.* [Note that it doesn't matter what is the underlying PDE: numerical stability is a feature of algebraic recurrences, not of PDEs!]

Fourier analysis of stability

Let $\mathbf{v} = (v_m)_{m\in\mathbb{Z}} \in \ell_2[\mathbb{Z}]$. Its *Fourier transform* is the function

$$\widehat{v}(\theta) = \textstyle\sum_{m\in\mathbb{Z}} \mathrm{e}^{-\mathrm{i}m\theta} v_m, \qquad -\pi \le \theta \le \pi.$$

We equip sequences and functions with the norms

$$\|\mathbf{v}\| = \Big\{ \sum_{m\in\mathbb{Z}} |v_m|^2 \Big\}^{\frac{1}{2}} \qquad \text{and} \qquad \|\widehat{v}\|_* = \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} |\widehat{v}(\theta)|^2 d\theta \right\}^{\frac{1}{2}}.$$

Lemma 1 (Parseval's identity)

For any $\mathbf{v} \in \ell_2[\mathbb{Z}]$*, we have* $\|\mathbf{v}\| = \|\widehat{v}\|_*$*.*

Proof. By definition,

$$\begin{aligned}\|\widehat{v}\|_*^2 &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\Big|\sum_{m\in\mathbb{Z}} \mathrm{e}^{-\mathrm{i}m\theta} v_m\Big|^2 d\theta = \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{m\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} v_m \bar{v}_k \mathrm{e}^{-\mathrm{i}(m-k)\theta} d\theta \\ &= \frac{1}{2\pi}\sum_{m\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} v_m\bar{v}_k \int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}(m-k)\theta} d\theta \overset{(*)}{=} \sum_{m\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} v_m \bar{v}_k \delta_{m-k} = \|\mathbf{v}\|^2,\end{aligned}$$

where equality $(*)$ is due to the fact that

$$\int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}\ell\theta} d\theta = \begin{cases} 2\pi, & \ell = 0, \\ 0, & \ell \in \mathbb{Z}\setminus\{0\}, \end{cases}$$

□

The implication of the lemma is that the Fourier transform is an *isometry* of the Euclidean norm. This is an important reason underlying its many applications in mathematics and beyond.

Amplification factor

For $\theta \in [-\pi, \pi]$, let $\widehat{u}^n(\theta) = \sum_{m \in \mathbb{Z}} \mathrm{e}^{-\mathrm{i}m\theta} u_m^n$ be the Fourier transform of the sequence $\mathbf{u}^n \in \ell_2[\mathbb{Z}]$. We multiply the discretized equations (3) by $\mathrm{e}^{-\mathrm{i}m\theta}$ and sum up for $m \in \mathbb{Z}$. Thus, the left-hand side yields

$$\begin{aligned}\sum_{m=-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}m\theta} \sum_{k=r}^{s} a_k u_{m+k}^{n+1} &= \sum_{k=r}^{s} a_k \sum_{m=-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}m\theta} u_{m+k}^{n+1} \\ &= \sum_{k=r}^{s} a_k \sum_{m=-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}(m-k)\theta} u_m^{n+1} = \Big(\sum_{k=r}^{s} a_k \mathrm{e}^{\mathrm{i}k\theta} \Big) \widehat{u}^{n+1}(\theta).\end{aligned} \tag{4}$$

Similarly manipulating the right-hand side, we deduce that

$$\widehat{u}^{n+1}(\theta) = H(\theta) \widehat{u}^n(\theta), \quad \text{where} \quad H(\theta) = \frac{\sum_{k=r}^{s} b_k \mathrm{e}^{\mathrm{i}k\theta}}{\sum_{k=r}^{s} a_k \mathrm{e}^{\mathrm{i}k\theta}}. \tag{5}$$

The function H is sometimes called the *amplification factor* of the recurrence (3)

Fourier analysis of stability

Theorem 2

The method (3) is stable $\quad \Leftrightarrow \quad |H(\theta)| \leq 1$ *for all* $\theta \in [-\pi, \pi]$.

Fourier analysis of stability (proof)

Proof. The definition of stability is equivalent to the statement that there exists $c > 0$ such that $\|\mathbf{u}^n\| \le c$ for all $n \in \mathbb{Z}^+$. [Because we are solving a Cauchy problem, equations are identical for all $h = \Delta x$, and this simplifies our analysis and eliminates a major difficulty: there is no need to insist explicitly that $\|\mathbf{u}^n\|$ remains uniformly bounded when $h \to 0$]. The Fourier transform being an isometry, stability is thus equivalent to $\|\widehat{u}^n\|_* \le c$ for all $n \in \mathbb{Z}^+$. Iterating (5), we obtain

$$\widehat{u}^n(\theta) = [H(\theta)]^n \widehat{u}^0(\theta), \qquad |\theta| \le \pi, \quad n \in \mathbb{Z}^+. \tag{6}$$

Proof. (Continuing)
1) Assume first that $|H(\theta)| \leq 1$ for all $|\theta| \leq \pi$. Then, by (6),

$$|\widehat{u}^n(\theta)| \leq |\widehat{u}^0(\theta)|$$
$$\Rightarrow \quad \|\widehat{u}^n\|_*^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |\widehat{u}^n(\theta)|^2 \mathrm{d}\theta \leq \frac{1}{2\pi}\int_{-\pi}^{\pi} |\widehat{u}^0(\theta)|^2 \mathrm{d}\theta = \|\widehat{u}^0\|_*^2. \tag{7}$$

Hence stability.

Proof. (Continuing) 2) Suppose, on the other hand, that there exists $\theta_0 \in [-\pi, \pi]$ such that $|H(\theta_0)| = 1 + 2\epsilon > 1$, say. Since H is continuous, there exist $-\pi \le \theta_1 < \theta_2 \le \pi$ such that $|H(\theta)| \ge 1 + \epsilon$ for all $\theta \in [\theta_1, \theta_2]$. We set $\eta = \theta_2 - \theta_1$ and choose as our initial condition the function (or the $\ell_2[\mathbb{Z}]$-sequence)

$$\widehat{u}^0(\theta) = \begin{cases} \sqrt{\frac{2\pi}{\eta}}, & \theta_1 \le \theta \le \theta_2, \\ 0, & \text{otherwise}, \end{cases}$$

Then

$$\begin{aligned} \|\widehat{u}^n\|_*^2 &= \tfrac{1}{2\pi} \int_{-\pi}^{\pi} |H(\theta)|^{2n} |\widehat{u}^0(\theta)|^2 \mathrm{d}\theta = \tfrac{1}{2\pi} \int_{\theta_1}^{\theta_2} |H(\theta)|^{2n} |\widehat{u}^0(\theta)|^2 \mathrm{d}\theta \\ &\ge \tfrac{1}{2\pi} (1+\epsilon)^{2n} \int_{\theta_1}^{\theta_2} \tfrac{2\pi}{\eta} \mathrm{d}\theta = (1+\epsilon)^{2n} \to \infty \quad (n \to \infty). \end{aligned}$$

We deduce that the method is unstable. □

Stability: Euler and the diffusion equation

Consider the Cauchy problem for the diffusion equation.

1) For the Euler method

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

we obtain

$$H(\theta) = 1 + \mu\left(\mathrm{e}^{-\mathrm{i}\theta} - 2 + \mathrm{e}^{\mathrm{i}\theta}\right) = 1 - 4\mu\sin^2\tfrac{\theta}{2} \;\in\; [1-4\mu, 1]\,,$$

thus the method is stable iff $\mu \le \frac{1}{2}$.

2) For the backward Euler method

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

we have

$$H(\theta) = \left[1 - \mu\left(\mathrm{e}^{-\mathrm{i}\theta} - 2 + \mathrm{e}^{\mathrm{i}\theta}\right)\right]^{-1} = \left[1 + 4\mu\sin^2\tfrac{\theta}{2}\right]^{-1} \in (0,1]\,.$$

thus stability for all μ.

3) The Crank–Nicolson scheme

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

results in

$$H(\theta) = \frac{1 + \frac{1}{2}\mu(\mathrm{e}^{-\mathrm{i}\theta} - 2 + \mathrm{e}^{\mathrm{i}\theta})}{1 - \frac{1}{2}\mu(\mathrm{e}^{-\mathrm{i}\theta} - 2 + \mathrm{e}^{\mathrm{i}\theta})} = \frac{1 - 2\mu \sin^2 \frac{\theta}{2}}{1 + 2\mu \sin^2 \frac{\theta}{2}} \in (-1, 1]$$

Hence stability for all $\mu > 0$.