5. Implicit Scheme for the 1-D Diffusion equation

As we discussed last time, the explicit scheme (3.5) has effective characteristics with gradients $dx/dt = \pm h/k$. Arguably, this gradient should approach infinity if it is to model the physics, which might explain why the explicit scheme requires $k = O(h^2)$ for stability. Mathematically, the solution at time level $j+1$ should depend on all the values at time level j. Today we consider **implicit** methods for this problem, which do have this property. Let us choose a parameter θ and approximate $u_t = u_{xx}$ by

$$U_n^{j+1} - U_n^j = r\left[\theta\delta^2 U_n^{j+1} + (1-\theta)\delta^2 U_n^j\right]. \tag{5.1}$$

Note that the RHS now involves the unknown variables U_n^{j+1}. To find them we will have to solve a set of simultaneous linear equations (unless $\theta = 0$). Before considering how to do this numerically, we observe that for this linear problem, both the PDE and our Finite Difference approximation may be solved exactly by the method of separation of variables.

We seek solutions of the form $U_n^j = X_n T^j$. Substituting into (5.1) and separating the terms depending on j and n leads to

$$\frac{T^{j+1} - T^j}{r\left[\theta T^{j+1} + (1-\theta)T^j\right]} = \frac{X_{n+1} - 2X_n + X_{n-1}}{X_n} = \sigma, \quad \text{say.} \tag{5.2}$$

As the LHS varies with j but not n, and the RHS the other way round, σ must be a constant. X_n therefore obeys the second order difference equation

$$X_{n+1} - (\sigma+2)X_n + X_{n-1} = 0, \tag{5.3}$$

which has solutions of Fourier form $X_n = e^{\pm in\xi}$ provided $\sigma + 2 = 2\cos\xi$. In general ξ is arbitrary, but if we impose the boundary conditions $X_0 = 0 = X_N$, then we can show that

$$\xi = \frac{m\pi}{N} = m\pi h \qquad \text{for} \quad m = 1, 2\ldots \qquad \text{and} \quad \sigma = -4\sin^2\left(\frac{m\pi h}{2}\right) = \sigma_m, \tag{5.4}$$

say. Then X_n is given by

$$X_n = A\sin(nm\pi h). \tag{5.5}$$

For each such permissible value $\sigma = \sigma_m$, T^j can be found from (5.2).

$$T^{j+1} = \lambda_m T^j, \qquad \text{so that} \quad T^j = C(\lambda_m)^j \qquad \text{where} \quad \lambda_m = \left[\frac{1+\sigma_m r(1-\theta)}{1-\sigma_m r\theta}\right]. \tag{5.6}$$

Now we have

$$\lambda_m = \frac{1 - 4r(1-\theta)\sin^2 \frac{1}{2}\xi}{1 + 4r\theta \sin^2 \frac{1}{2}\xi} = 1 - \frac{4r \sin^2 \frac{1}{2}\xi}{1 + 4r\theta \sin^2 \frac{1}{2}\xi}.$$

The stability requirement is that $|\lambda_m| \leqslant 1$ for all m. Clearly $\lambda_m \leqslant 1$, while $\lambda_m \geqslant -1$ if

$$1 + 4r\theta \sin^2 \tfrac{1}{2}\xi \geqslant 2r \sin^2 \tfrac{1}{2}\xi,$$

or

$$(1-2\theta)2r \sin^2 \tfrac{1}{2}\xi \leqslant 1. \tag{5.7}$$

If $\theta \geqslant \frac{1}{2}$, therefore, the FDM (5.1) is **unconditionally stable**. If on the other hand $0 \leqslant \theta < \frac{1}{2}$, stability for all m and h can be guaranteed only if

$$r \leqslant \frac{1}{2(1-2\theta)}. \tag{5.8}$$

Note that when $\theta = 0$ we recover the relation $r \leqslant \frac{1}{2}$ and (5.1) is then **conditionally stable**.

We can obtain the general solution by taking an arbitrary linear sum,

$$U_n^j = \sum_{m=1}^{\infty} B_m \sin(nm\pi h)(\lambda_m)^j, \tag{5.9}$$

where the coefficients B_m can be found from the Fourier expansion of $u_0(x)$ in (3.1). We can now compare (5.9) with the exact solution (4.4) evaluated at the grid points

$$u_n^j \equiv u(nh,\, jk) = \sum_{m=1}^{\infty} B_m \sin(nm\pi h)(e^{-m^2\pi^2 k})^j \tag{5.10}$$

The accuracy of the F.D. approximation may thus be determined by comparing λ_m and $\exp(-m^2\pi^2 k)$.

We see that it is the high modes, the large values of m for which $m \sim N$ and $\xi \sim \pi$ which are most likely to be unstable. This is typical behaviour for FDMs; instabilities tend to manifest themselves on the scale of the grid. The low modes, for which $m = O(1)$, and ξ is small are modelled well. For them we may approximate $\sin\beta \approx \beta - \frac{1}{6}\beta^3$ to give

$$\begin{aligned} \lambda_m &\approx 1 - \frac{r(\xi - \frac{1}{24}\xi^3 + O(\xi^5))^2}{1 + r\theta\xi^2 + O(r\xi^4)}, \\ &= 1 - r\xi^2(1 - \tfrac{1}{12}\xi^2)(1 - r\theta\xi^2) + O(r^3\xi^6). \\ &= 1 - m^2\pi^2 k + \left(\theta + \frac{1}{12r}\right) m^4\pi^4 k^2 + O(m^6 k^3) \end{aligned} \tag{5.11}$$

$$\text{while} \quad e^{-m^2\pi^2 k} = 1 - m^2\pi^2 k + \tfrac{1}{2}m^4\pi^4 k^2 + O(m^6 k^3).$$

The agreement between λ_m and $\exp(-m^2\pi^2 k)$ is quite good, while the greatest accuracy is achieved if

$$\theta = \frac{1}{2} - \frac{1}{12r} \qquad \text{or} \quad r(1-2\theta) = \tfrac{1}{6}.$$

We see from (5.7) that such a scheme would be stable. It is the generalisation of Milne's method ($\theta = 0,\ r = \frac{1}{6}$). If r is large, then the Crank-Nicolson scheme ($\theta = \frac{1}{2}$) is close to optimal.

Most importantly, using implicit methods we can produce stable schemes with $k \sim h$, and hence large values of $r = k/h^2$. Compared with the explicit schemes, we may choose relatively large time-steps.

The Crank-Nicolson method and (nearly) Tridiagonal systems

If we choose an unconditionally stable scheme with $\theta > 1/2$, then we may choose the timestep as large as we like, and our only concern is the accuracy of our approximation of the derivatives. A popular (and sensible) choice is the **Crank-Nicolson** scheme, $\theta = 1/2$. This scheme is centred about the time-level $(j + 1/2)$, and so is second-order accurate in both space and time, $R_n^j = O(k2, h2)$. The price we pay for an implicit scheme is that each timestep we have to solve some simultaneous linear equations. However, as the system is tri-diagonal this is not so hard. If we represent a list of all the U-values at time j by a vector U^j , then the system we need to solve is

$$AU^{j+1} = BU^j, \quad \text{where} \quad U^j = (U_1^j, U_2^j, \dots, U_N^j)^T$$

for suitable matrices A and B. In this problem, A and B are tridiagonal, which permits efficient solution. The Crank-Nicolson ($\theta = 1/2$) method for the equation $u_t = u_{xx}$ has

$$\mathcal{A} = \begin{pmatrix} 1+r & -r/2 & 0 & \ddots & 0 \\ -r/2 & 1+r & -r/2 & \ddots & \ddots \\ 0 & -r/2 & 1+r & -r/2 & 0 \\ \ddots & \ddots & \ddots & \ddots & -r/2 \\ 0 & \ddots & 0 & -r/2 & 1+r \end{pmatrix},$$

To solve $Ax = b$, where A is an $M \times M$ matrix. usually requires $O(M^3)$ operations. Here, however, the sparseness and structure of A renders the process much more efficient. Using Gaussian elimination, subtracting $(-r/2)/(1+r)$ times the first row from the second transforms A_{21} to zero and alters A_{22} and

b_2. Then subtracting a suitable multiple of the 2nd row from the 3rd and continuing, leaves us with

$$\mathcal{A} = \begin{pmatrix} 1+r & -r/2 & 0 & \ddots & 0 \\ 0 & a_2 & -r/2 & \ddots & \ddots \\ 0 & 0 & 1+r & a_3 & 0 \\ \ddots & \ddots & \ddots & \ddots & -r/2 \\ 0 & \ddots & 0 & 0 & a_n \end{pmatrix} x = b^*,$$

where a_i and b^* are known values. The last equation is now trivial, $a_n x_n = b^*_n$, while the penultimate $a_{n-1}x_{n-1} = b^*_{n-1} + (r/2)x_n$ which we now know, giving us x_{n-1}. Systematically back-substituting determines all the unknowns x_i in $O(M)$ operations. See the routine tridiag.m.

Aside – Separation of variables solution to Diffusion Eqn.

We consider a slight modification of (3.1), namely:

$$\begin{aligned} u_t &= u_{xx} \qquad \text{in} \quad 0 < x < 1, \;\; t > 0, \\ \text{with} \quad & u(0,t) = u(1,t) = 0, \quad u(x,0) = u_0(x). \end{aligned} \tag{5.12}$$

The PDE has separable solutions of the form $u(x,t) = X(x)T(t)$ provided

$$XT' = X''T \qquad \text{or} \quad \frac{T'}{T} = \frac{X''}{X} = -\omega^2, \qquad \text{say.}$$

As T'/T is a function of t only, while X''/X is a function of x only, both functions must be a constant, which we take to be negative. Then the functions $X(x)$ and $T(t)$ take the forms

$$X = A\cos\omega x + B\sin\omega x, \qquad \text{and} \quad T = Ce^{-\omega^2 t}.$$

If we require X to obey the boundary conditions in (4.3), namely $X(0) = X(1) = 0$, we obtain non-zero solutions only if $A = 0$ and $\omega = m\pi$, for some integer m, so that

$$u = B_m \sin m\pi x \, e^{-m^2\pi^2 t},$$

for some constant B_m. As (5.12) is a linear problem, we may combine solutions to obtain a more general solution in the form

$$u(x,\, t) = \sum_{m=1}^{\infty} B_m \sin m\pi x \, e^{-m^2\pi^2 t}. \tag{5.13}$$

The initial condition will be satisfied if

$$u(x,\,0)=\sum_{m=1}^{\infty}B_m\sin m\pi x=u_0(x). \tag{5.14}$$

Thus all we need do to obtain the solution of (5.12) is to expand the initial condition $u=u_0(x)$ in a Fourier series, and substitute the appropriate values of the constants B_m into (5.13).