17 Finite differences for the heat equation

In the example considered last time we used the forward difference for u_t and the centered difference for u_{xx} in the heat equation to arrive at the following difference equation.

$$\frac{u_j^{n+1}-u_j^n}{\Delta t}=\frac{u_{j+1}^n-2u_j^n+u_{j-1}^n}{(\Delta x)^2}. \tag{1}$$

Denoting $s=\Delta t/(\Delta x)^2$, this lead to the FTCS scheme,

$$u_j^{n+1}=s(u_{j+1}^n+u_{j-1}^n)+(1-2s)u_j^n, \tag{2}$$

which is an explicit numerical scheme. We also saw last time that choosing $s=1$ lead to a wild numerical solution, while the choice $s=\frac{1}{2}$ resulted in a somewhat accurate approximation to the solution.

To better understand the effect of s on the outcome of the scheme, let us consider a separated solution of the difference equation (1), $u_j^n=X_jT_n$. Substituting this solution gives

$$X_jT_{n+1}=s(X_{j+1}T_n+X_{j-1}T_n)+(1-2s)X_jT_n.$$

Dividing both sides of the above equation by u_j^n, we obtain

$$\frac{T_{n+1}}{T_n}=(1-2s)+s\frac{X_{j+1}+X_{j-1}}{X_j}=\xi, \tag{3}$$

where ξ is independent of both j and n. We then have for the T component

$$T_{n+1}=\xi T_n,\quad \text{and hence,}\quad T_n=\xi^nT_0.$$

For the scheme (2) to be *stable*, we need to have $|\xi|\le 1$, since otherwise it would lead to solutions that grow exponentially in time.

The equation for the X component is

$$1-2s+s\frac{X_{j+1}+X_{j-1}}{X_j}=\xi.$$

Plugging in a discretized Fourier mode $X_j=e^{ikj\Delta x}$ into this equation, we obtain

$$\xi=1-2s+s\frac{e^{ik(j+1)\Delta x}+e^{ik(j-1)\Delta x}}{e^{ikj\Delta x}},\quad \text{or}\quad \xi=1-2s+s(e^{ik\Delta x}+e^{-ik\Delta x}).$$

Using Euler's formula, we can rewrite the last identity as

$$\xi=1-2s+2s\cos k\Delta x=1-2s(1-\cos k\Delta x).$$

Since $\cos k\Delta x\le 1$, we see that $\xi\le 1$ for all values of s. On the other hand, the condition $-1\le\xi$ is equivalent to

$$-1\le 1-2s(1-\cos k\Delta x).$$

The worst case happens when the frequency k is such that $\cos k\Delta x\approx 1$, then the above inequality would give $-1\le 1-4s$, which implies the following necessary condition for stability

$$\frac{\Delta t}{(\Delta x)^2}=s\le\frac{1}{2}. \tag{4}$$

17.1 Boundary conditions

Last time we saw how one uses the discretized initial data to march forward in time using the explicit scheme (2). In the presence of Dirichlet boundary conditions, the discretized boundary data is also used when computing the numerical solution.

For the Neumann boundary conditions,

$$u_x(0,t) = g(t), \quad u_x(l,t) = h(t),$$

the values of the solution on the boundary points of the grid are not immediately available, so one needs to use a difference approximation for the conditions in order to march forward in time. Using the forward or backward differences to approximate u_x on the boundary would introduce a local truncation error of order $\mathcal{O}(\Delta x)$, which is bigger than the truncation error $\mathcal{O}(\Delta x)^2$, thus contaminating the numerical solution. To have errors of the same order as the difference equation, we must use the centered differences. For these we introduce "ghost points" u^n_{-1} and u^n_{J+1} in addition to the grid points $u^n_0, u^n_1, \ldots, u^n_J$. Then the centered difference approximation for the Neumann conditions will be

$$g(n\Delta t) = g^n = \frac{u^n_1 - u^n_{-1}}{2\Delta x}, \quad \text{and} \quad h(n\Delta t) = h^n = \frac{u^n_{J+1} - u^n_{J-1}}{2\Delta x}. \tag{5}$$

From the above identities we can compute the values u^n_{-1} and u^n_{J+1} at the ghost points at time level n, which will then be used to compute u^{n+1}_0 and u^{n+1}_J, thus arriving at numerical values at the boundaries.

17.2 The implicit BTCS scheme

Instead of approximating u_t in the heat equation $u_t = u_{xx}$ by the forward difference, which resulted in the difference equation (1), we can use the backward difference, which at time level $n+1$ will give the difference equation

$$\frac{u^{n+1}_j - u^n_j}{\Delta t} = \frac{u^{n+1}_{j+1} - 2u^{n+1}_j + u^{n+1}_{j-1}}{(\Delta x)^2}. \tag{6}$$

Denoting as before $s = \Delta t/(\Delta x)^2$, and separating values at different time levels, we can rewrite the above equation as

$$-su^{n+1}_{j-1} + (1+2s)u^{n+1}_j - su^{n+1}_{j-1} = u^n_j. \tag{7}$$

For a fixed n, the last equation can be thought of as a system for $u^{n+1}_1, u^{n+1}_2, \ldots, u^{n+1}_J$. Thus, (7) gives an implicit scheme, which corresponds to the template

$$\begin{array}{ccc} \frac{s}{1+2s}\bullet & & \frac{s}{1+2s}\bullet \\ & \frac{1}{1+2s}\bullet & \\ & * & \end{array}$$

To find the numerical solution using this scheme, one needs to solve the system of linear equations (7). In the presence of Dirichlet boundary conditions, this system can be written in the following vector form

$$\begin{pmatrix} 1+2s & -s & 0 & . & . & 0 \\ -s & 1+2s & -s & 0 & . & . \\ 0 & -s & 1+2s & . & & . \\ . & . & & . & . & 0 \\ . & & . & & . & -s \\ 0 & . & . & 0 & -s & 1+2s \end{pmatrix} \begin{bmatrix} u^{n+1}_1 \\ u^{n+1}_2 \\ . \\ . \\ u^{n+1}_{J-1} \\ u^{n+1}_J \end{bmatrix} = \begin{bmatrix} u^n_1 + su^{n+1}_0 \\ u^n_2 \\ . \\ . \\ u^n_{J-1} \\ u^n_J + su^{n+1}_J \end{bmatrix}.$$

We can see that the matrix of the coefficients is tridiagonal, which can be inverted by efficient algorithms. Although (7) is implicit, it requires about twice as many arithmetic operations as the scheme (2).

Let us now check for what values of s the BTCS scheme (7) will be stable. Plugging the separated

solution $u_j^n = \xi^n e^{ikj\Delta x}$ into the scheme and dividing both sides by u_j^n, we have

$$-s\xi e^{ik\Delta x} + (1+2s)\xi - s\xi e^{-ik\Delta x} = 1.$$

Factoring ξ on the left hand side, and solving for it gives

$$\xi = \frac{1}{1+2s-s(e^{ik\Delta x}+e^{-ik\Delta x})} = \frac{1}{1+2s(1-\cos k\Delta x)},$$

so $|\xi| \leq 1$, since $1-\cos k\Delta x \geq 0$. Thus, the implicit scheme (7) is stable for all values of s, i.e. unconditionally stable.

17.3 The θ-scheme

The two schemes for the heat equation considered so far have their advantages and disadvantages. On the one hand we have the FTCS scheme (2), which is explicit, hence easier to implement, but it has the stability condition $\Delta t \leq \frac{1}{2}(\Delta x)^2$. The last fact requires very small mesh size for the time variable, which leads one to consider more time steps to reach the values at a certain time.

On the other hand, the implicit scheme BTCS (7) requires more arithmetic operations to find the values at a certain time step, but it is unconditionally stable, allowing one to chose a larger mesh size for the time variable. Notice also that the two schemes use different sets of points in the computation of u_j^{n+1}

Combining the two schemes with different weights into a single scheme may emphasize some of the advantages of these schemes, and also be more accurate, since it would use a larger set of points to compute the same values.

Taking a parameter $0 \leq \theta \leq 1$, we form the scheme

$$\frac{u_j^{n+1}-u_j^n}{\Delta t} = (1-\theta)\frac{u_{j+1}^n - 2u_j^n + u_{j-1}^n}{(\Delta x)^2} + \theta\frac{u_{j+1}^{n+1} - 2u_j^{n+1} + u_{j-1}^{n+1}}{(\Delta x)^2}. \tag{8}$$

Observe that when $\theta = 0$ the above scheme is exactly the BTCS scheme (7), while when $\theta = 1$ the scheme become the FTCS scheme (2). Clearly for $0 < \theta \leq 1$ this scheme is implicit.

Denoting, as before, $s = \Delta t/(\Delta x)^2$, we analyze the stability of the scheme (8) by substituting in the separated solution $u_j^n = \xi^n e^{ikj\Delta x}$. This leads to

$$\xi - 1 = (1-\theta)s(e^{ik\Delta x} - 2 + e^{-ik\Delta x}) + \theta s\xi(e^{ik\Delta x} - 2 + e^{-ik\Delta x}).$$

Solving for ξ, we obtain

$$\xi = \frac{1-2(1-\theta)s(1-\cos k\Delta x)}{1+2\theta s(1-\cos k\Delta x)}.$$

Then the stability condition $|\xi| \leq 1$ is equivalent to

$$-1 \leq \frac{1-2(1-\theta)s(1-\cos k\Delta x)}{1+2\theta s(1-\cos k\Delta x)} \leq 1. \tag{9}$$

The inequality on the right is always satisfied, since the numerator of the fraction is always less than or equal to one, while the denominator is always more than or equal to one. Thus we need to only check the inequality on the left, which is equivalent to

$$-1 - 2\theta s(1-\cos k\Delta x) \leq 1 - 2s(1-\cos k\Delta x) + 2\theta s(1-\cos k\Delta x),$$

or, after combining like terms,

$$-2 \leq s(4\theta - 2)(1-\cos k\Delta x).$$

The worst case again corresponds to the frequency, for which $\cos k\Delta x \approx 1$, resulting in

$$-2 \le 2s(4\theta - 2).$$

The last inequality will be always satisfied, if $\theta \ge \frac{1}{2}$, since then the right hand side would be nonnegative. Thus, the θ-scheme (8) is unconditionally stable for $\frac{1}{2} \le \theta \le 1$. On the other hand, when $0 \le \theta < \frac{1}{2}$, the stability condition is

$$\frac{\Delta t}{(\Delta x)^2} = s \le \frac{1}{2 - 4\theta}.$$

Notice that for $\theta = 0$ this condition exactly coincides with the stability condition (4) for the FTCS scheme, which was expected, since the θ-scheme reduces to the FTCS scheme for $\theta = 0$.

In the special case $\theta = \frac{1}{2}$ the scheme (8) is called the Crank-Nicholson scheme. It can be written as the average of (2) and (7),

$$-\frac{s}{2}u_{j-1}^{n+1} + (1+s)u_j^{n+1} - \frac{s}{2}u_{j+1}^{n+1} = \frac{s}{2}u_{j-1}^{n} + (1-s)u_j^{n} + \frac{s}{2}u_{j-1}^{n}, \tag{10}$$

and will have the template

$$\begin{array}{ccc} \frac{1}{2}\frac{s}{1+s}\bullet & & \frac{1}{2}\frac{s}{1+s}\bullet \\ & * & \\ \frac{1}{2}\frac{s}{1+s}\bullet & \frac{1-s}{1+s}\bullet & \frac{1}{2}\frac{s}{1+s}\bullet \end{array}$$

The Crank-Nicholson scheme (10) is more accurate than (2) and (7) for small values of Δt, however, it is the most computationally involved.

17.4 Conclusion

Using either the forward difference approximation or the backward difference approximation for the time derivative in the heat equation, we obtained the explicit scheme (2) and the implicit scheme (7) respectively. The stability analysis of these schemes showed that for stability of the first scheme we need $\Delta t \le \frac{1}{2}(\Delta x)^2$, while the second scheme is unconditionally stable.

Combining these two schemes with different weights into one gave as the implicit θ-scheme (8), which is unconditionally stable for $\theta \ge \frac{1}{2}$. In the special case of $\theta = \frac{1}{2}$ it is called the Crank-Nicholson scheme, which is given by (10).