Lecture 27: Consistency and Stability of Finite Difference Schemes

Rajen Kumar Sinha

Department of Mathematics
IIT Guwahati

Let

$$L(U) = 0, \tag{1}$$

represent the partial differential equation (PDE) in the independent variables x and t, with exact solution U. Let

$$F_{i,j}(u) = 0 \tag{2}$$

represent the finite difference equation approximation the PDE at the $(i,j)^{th}$ mesh point, with exact solution u.
The local truncation error $T_{i,j}(U)$ at the point (ih, jk) is defined by

$$T_{i,j}(U) = F_{i,j}(U) - L(U_{i,j}) = F_{i,j}(U).$$

$T_{i,j}$ gives an indication of the error resulting from the replacement of $L(U_{i,j})$ by $F_{i,j}(U)$.

Definition. If $T_{i,j}(U) \to 0$ as $h \to 0$, $k \to 0$, the difference equation (2) is said to be consistent or compartible with the PDE (1).

Example: Compute $T_{i,j}$ of the two-level explicit scheme approximating $U_t = U_{xx}$ at the point (ih, jk).

$$F_{i,j}(u) = \frac{u_{i,j+1} - u_{i,j}}{k} - \frac{u_{i-1,j} - 2u_{i,j} + u_{i+1,j}}{h^2} = 0.$$

$$T_{i,j}(U) = F_{i,j}(U) = \frac{U_{i,j+1} - U_{i,j}}{k} - \frac{U_{i-1,j} - 2U_{i,j} + U_{i+1,j}}{h^2} \tag{3}$$

By Taylor's expansion

$$\begin{aligned}
U_{i+1,j} &= U((i+1)h, jk) = U(x_i + h, t_j) \\
&= U_{i,j} + h\left(\frac{\partial U}{\partial x}\right)_{i,j} + \frac{1}{2}h^2\left(\frac{\partial^2 U}{\partial x^2}\right)_{i,j} + \frac{1}{6}h^3\left(\frac{\partial^3 U}{\partial x^3}\right)_{i,j} + \cdots \\
U_{i-1,j} &= U((i-1)h, jk) = U(x_i - h, t_j) \\
&= U_{i,j} - h\left(\frac{\partial U}{\partial x}\right)_{i,j} + \frac{1}{2}h^2\left(\frac{\partial^2 U}{\partial x^2}\right)_{i,j} - \frac{1}{6}h^3\left(\frac{\partial^3 U}{\partial x^3}\right)_{i,j} + \cdots \\
U_{i,j+1} &= U(ih, (j+1)k) = U(x_i, t_j + k) \\
&= U_{i,j} + k\left(\frac{\partial U}{\partial t}\right)_{i,j} + \frac{1}{2}k^2\left(\frac{\partial^2 U}{\partial t^2}\right)_{i,j} + \frac{1}{6}k^3\left(\frac{\partial^3 U}{\partial t^3}\right)_{i,j} + \cdots
\end{aligned}$$

Substituting the above in (3) to obtain

$$T_{i,j} = \left(\frac{\partial U}{\partial t} - \frac{\partial^2 U}{\partial x^2}\right)_{i,j} + \frac{1}{2}k\left(\frac{\partial^2 U}{\partial t^2}\right)_{i,j} - \frac{1}{12}h^2\left(\frac{\partial^4 U}{\partial x^4}\right)_{i,j} + \frac{1}{6}k^2\left(\frac{\partial^3 U}{\partial t^3}\right)_{i,j} - \frac{1}{360}h^4\left(\frac{\partial^6 U}{\partial x^6}\right)_{i,j} + \cdots$$

Since $\left(\frac{\partial U}{\partial t} - \frac{\partial^2 U}{\partial x^2}\right)_{i,j} = 0$, the principal part of the local truncation error is

$$\left(\frac{1}{2}k\frac{\partial^2 U}{\partial t^2} - \frac{1}{12}h^2\frac{\partial^4 U}{\partial x^4}\right)_{i,j}.$$

Therefore, $T_{i,j} = O(k) + O(h^2)$. Note that

$$T_{i,j} \to 0 \text{ as } h \to 0, k \to 0.$$

Thus, the explicit scheme approximating $U_t = U_{xx}$ is consistent with the differential equation.

Remark: This error may further be reduced by choosing special value for k/h^2. $T_{i,j}$ can be written as

$$T_{i,j} = \frac{1}{12}h^2 \left(6\frac{k}{h^2}\frac{\partial^2 U}{\partial t^2} - \frac{\partial^4 U}{\partial x^4}\right)_{i,j} + O(k^2) + O(h^4).$$

Note that

$$\frac{\partial}{\partial t} = \frac{\partial^2}{\partial x^2} \implies \frac{\partial}{\partial t}\left(\frac{\partial U}{\partial t}\right) = \frac{\partial^2}{\partial x^2}\left(\frac{\partial^2 U}{\partial x^2}\right).$$

If $6\frac{k}{h^2} = 1$ then $T_{i,j} = O(k^2) + O(h^4)$.

Note: This is of little use because $k = \frac{1}{6}h^2$ is very small for small h and the volume of arithmatic operation needed to advance the solution to a large-time level is enourmous).

Stability

Stability is related to the round-off error. We say that the scheme is stable if the round-off error in the numerical process is bounded.

Let $R = \{(x,t) \mid 0 \le x \le 1,\ 0 \le t \le T\}$ be a rectangle. Consider the PDE

$$L(U) = 0 \quad \text{in} \quad R$$

with prescribed initial and boundary conditions. Let h and k be the discretisation/mesh parameters such that

$$\begin{aligned} x_i &= ih, \quad i = 0(1)N \ \text{ with } Nh = 1, \\ t_j &= jk, \quad j = 0(1)J \ \text{ with } Jk = T. \end{aligned}$$

Assume that h is related to k (e.g., $k = O(h^2)$). That is, As $h \to 0$, $k \to 0$.
Consider the finite difference approximation of the form:

$$b_{i-1}u_{i-1,j+1} + b_i u_{i,j+1} + b_{i+1}u_{i+1,j+1} = c_{i-1}u_{i-1,j} + c_i u_{i,j} + c_{i+1}u_{i+1,j},$$

where b_i's and c_i's are constants.

Suppose the boundary values $u_{0,j}$ and $u_{N,j}$ for $j > 0$ are known. Then for $i = 1(1)N-1$, we have

$$\begin{bmatrix} b_1 & b_2 & & & \\ b_1 & b_2 & b_3 & & \\ \ddots & \ddots & \ddots & & \\ & & b_{N-3} & b_{N-2} & b_{N-3} \\ & & & b_{N-2} & b_{N-1} \end{bmatrix} \begin{bmatrix} u_{1,j+1} \\ u_{2,j+1} \\ \vdots \\ u_{N-2,j+1} \\ u_{N-1,j+1} \end{bmatrix}$$

$$= \begin{bmatrix} c_1 & c_2 & & & \\ c_1 & c_2 & c_3 & & \\ \ddots & \ddots & \ddots & & \\ & & c_{N-3} & c_{N-2} & c_{N-3} \\ & & & c_{N-2} & c_{N-1} \end{bmatrix} \begin{bmatrix} u_{1,j} \\ u_{2,j} \\ \vdots \\ u_{N-2,j} \\ u_{N-1,j} \end{bmatrix} + \begin{bmatrix} c_0 u_{0,j} - b_0 u_{0,j+1} \\ \\ \vdots \\ \\ c_N u_{N,j} - b_{N,j+1} \end{bmatrix}$$

In vector and matrix notation,

$$\begin{aligned} & \mathbf{B}\mathbf{u}_{j+1} = \mathbf{C}\mathbf{u}_j + \mathbf{d}_j \\ \Longrightarrow \quad & \mathbf{u}_{j+1} = \mathbf{B}^{-1}\mathbf{C}\mathbf{u}_j + \mathbf{B}^{-1}\mathbf{d}_j \\ \Longrightarrow \quad & \mathbf{u}_{j+1} = \mathbf{A}\mathbf{u}_j + \mathbf{f}_j, \text{ where } \mathbf{A} = \mathbf{B}^{-1}\mathbf{C}, \ \mathbf{f}_j = \mathbf{B}^{-1}\mathbf{d}_j \end{aligned}$$

Apply recursively to obtain

$$\begin{aligned}
\mathbf{u}_j &= \mathbf{A}\mathbf{u}_{j-1} + \mathbf{f}_{j-1} = \mathbf{A}(\mathbf{A}\mathbf{u}_{j-2} + \mathbf{f}_{j-2}) + \mathbf{f}_{j-1} \\
&= \mathbf{A}^2\mathbf{u}_{j-2} + \mathbf{A}\mathbf{f}_{j-2} + \mathbf{f}_{j-1} \\
&= \cdots \\
&= \mathbf{A}^j\mathbf{u}_0 + \mathbf{A}^{j-1}\mathbf{f}_0 + \mathbf{A}^{j-2}\mathbf{f}_1 + \cdots + \mathbf{f}_{j-1}, \qquad (4) \\
\text{where } \mathbf{u}_0 &\to \text{the vector of initial values} \\
\mathbf{f}_0, \mathbf{f}_1, \ldots, \mathbf{f}_{j-1} &\to \text{are the vectors of known boundary values}
\end{aligned}$$

Perturb the vector of initial value $\mathbf{u}_0$ to $\mathbf{u}_0^*$. The exact solution at the jth time-level will be

$$\mathbf{u}_j^* = \mathbf{A}^j\mathbf{u}_0^* + \mathbf{A}^{j-1}\mathbf{f}_0 + \mathbf{A}^{j-2}\mathbf{f}_1 + \cdots + \mathbf{f}_{j-1} \qquad (5)$$

Define the perturbation error by $\mathbf{e} = \mathbf{u}^* - \mathbf{u}$. Then, it follows from (4) and (5) that

$$\mathbf{e}_j = \mathbf{u}_j^* - \mathbf{u}_j = \mathbf{A}^j(\mathbf{u}_0^* - \mathbf{u}_0) = \mathbf{A}^j\mathbf{e}_0,$$

where $\mathbf{e}_0$ is the perturbation error of initial values.

$$\|\mathbf{e}_j\| = \|\mathbf{A}^j\mathbf{e}_0\| \le \|\mathbf{A}^j\|\|\mathbf{e}_0\|.$$

If there exists a positive number M, independent of j, h and k such that $\|\mathbf{A}^j\| \le M, \ \ j = 1(1)J$ (Due to Lax and Richtmyer), then

$$\|\mathbf{e}_j\| \le M\|\mathbf{e}_0\|,$$

which limits the amplification of initial error. Since

$$\|\mathbf{A}^j\| = \|\mathbf{A}\mathbf{A}^{j-1}\| \le \|\mathbf{A}\|\|\mathbf{A}^{j-1}\| \le \cdots \le \|\mathbf{A}\|^j,$$

we have

$$\|\mathbf{e}_j\| \le \|\mathbf{A}\|^j\|\mathbf{e}_0\|.$$

Thus, the Lax and Richtmyer definition of stability is satisfied if

$$\| \underbrace{A}_{\textbf{amplification matrix}} \| \le 1.$$

This is the necessary and sufficient condition for the difference equations to be stable when the solution of the PDE does not increase as t increases.

*** Ends ***