2.7 Stability

We shall first introduce the problem of stability of a finite difference calculation used to solve equation (1) as consisting of finding a condition under which

$$U_m^n - \tilde{U}_m^n \ (\equiv Z_m^n),$$

the difference between the theoretical and numerical solutions of the difference equation, remains bounded as *n tends to infinity*. The latter occurs in a calculation where either

(i) k remains fixed for all n and $t \to \infty$, or
(ii) $h, k \to 0$ (k/h^2 fixed) for a fixed value of $t = nk$.

The following methods are used for examining this notion of stability of a finite difference calculation.

The von Neumann method

Here, a harmonic decomposition is made of the error Z at grid points at a given time level, leading to the error function

$$E(x) = \sum_j A_j e^{i\beta_j x},$$

where in general the frequencies $|\beta_j|$ and j are arbitrary. It is necessary to consider only the single term $e^{i\beta x}$ where β is any *real* number. For convenience, suppose that the time level being considered corresponds to $t = 0$. To investigate the error propagation as t increases, it is necessary to find a solution of the finite difference equation which reduces to $e^{i\beta x}$ when $t = 0$. Let such a solution be

$$e^{\alpha t} e^{i\beta x}$$

where $\alpha = \alpha(\beta)$ is, in general, complex. The original error component $e^{i\beta x}$ will not grow with time if

$$|e^{\alpha k}| \leqslant 1$$

for all α. This is von Neumann's criterion for stability.*

* In order to allow for exponentially growing solutions of the partial differential equation itself, a more general form is

$$|e^{\alpha k}| \leqslant 1 + 0(k).$$

Example 4 Determine the stability condition for the formula

$$U_m^{n+1} = (1 - 2r)U_m^n + r(U_{m+1}^n + U_{m-1}^n)$$

which is used to solve the equation

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

in region (c) of figure 1.

In this case, let $Mh = 1$. Denote the errors at the grid points in the range $0 \leqslant x \leqslant 1$ at $t = 0$ by $Z(mh) = Z_m^0 (m = 0, 1, \ldots, M)$. Then

$$Z_m^0 = \sum_{j=0}^{M} A_j e^{i\beta_j mh} \quad (m = 0, 1, \ldots, M)$$

determines the unknowns $A_j (j = 0, 1, \ldots, M)$. In this problem, the number of harmonics corresponds to the number of grid points at any time level, and so the number of harmonics increases as the mesh size is reduced. Now, since Z_m^n satisfies the original difference equation, we get

$$Z_m^{n+1} = (1 - 2r)Z_m^n + r(Z_{m+1}^n + Z_{m-1}^n). \tag{44}$$

Substitute

$$Z_m^n = e^{\alpha nk} e^{i\beta mh} = \xi^n e^{i\beta mh}$$

where

$$\xi = e^{\alpha k},$$

and equation (44) gives

$$\xi^{n+1} e^{i\beta mh} = \xi^n [(1 - 2r) e^{i\beta mh} + r(e^{i\beta(m+1)h} + e^{i\beta(m-1)h})].$$

Cancellation of $\xi^n e^{i\beta mh}$ leads to

$$\xi = (1 - 2r) + r(e^{i\beta h} + e^{-i\beta h})$$
$$= 1 - 2r(1 - \cos \beta h)$$
$$= 1 - 4r \sin^2 \frac{\beta h}{2}.$$

The quantity ξ is called the *amplification factor*. The condition for stability, $|\xi| \leqslant 1$, for *all values of* βh, leads to

$$-1 \leqslant 1 - 4r \sin^2 \frac{\beta h}{2} \leqslant +1 \quad (\text{all } \beta h).$$

The right-hand side of the inequality is trivially satisfied if $r > 0$, and the left-hand side gives

$$r \leqslant \frac{1}{2\sin^2 \frac{\beta h}{2}},$$

leading to the stability condition $0 < r \leqslant \frac{1}{2}$.

The following important points should be noted concerning the von Neumann method of examining stability.

(i) The method which is based on Fourier series applies only if the coefficients of the linear difference equation are constant. If the difference equation has variable coefficients, the method can still be applied locally and it might be expected that a method will be stable if the von Neumann condition, derived as though the coefficients were constant, is satisfied at every point of the field. There is much numerical evidence to support this contention.

(ii) For two level difference schemes with *one dependent variable* and any number of independent variables, the von Neumann condition is sufficient as well as necessary for stability. Otherwise, the condition is *necessary* only.

(iii) Boundary conditions are neglected by the von Neumann method which applies in theory only to pure initial value problems with periodic initial data. It does however provide necessary conditions for stability of constant coefficient problems regardless of the type of boundary condition.

Exercise

6. Find the stability condition for formula (10).

7. Show that the implicit formulae (31) and (32) are unconditionally stable [i.e. stable for all values of $r > 0$].

The matrix method

This is applicable to the initial boundary value problem illustrated in figure 1 (c). If $Mh = 1$, the totality of difference equations connecting values of U at two neighbouring time levels can be written in the matrix form

$$A_n \mathbf{U}^{n+1} = B_n \mathbf{U}^n \tag{45}$$

where the column vector $\mathbf{U}^s$ ($s = n, n+1$) contains the grid values at the time level s, and A_n, B_n are square matrices of order $(M+1)$ which may vary with n. If the difference equation is explicit, $A_n = I$ for all n. Also if n tends to infinity while $h, k \to 0$ the order of A_n and B_n tends to infinity.

Equation (45) may be written in the explicit form

$$\mathbf{U}^{n+1} = C_n \mathbf{U}^n$$

where $C_n = A_n^{-1} B_n$, provided $|A_n| \neq 0$. The error vector $\mathbf{Z}^n$ ($\equiv \mathbf{U}^n - \tilde{\mathbf{U}}^n$) therefore satisfies

$$\mathbf{Z}^{n+1} = C_n \mathbf{Z}^n,$$

which leads to

$$\mathbf{Z}^{n+1} = \left(\prod_{i=0}^{n} C_i\right)\mathbf{Z}^0,$$

The method defined by (45) will be stable provided the norm $\|\mathbf{Z}^{n+1}\|$ is bounded for all $n \geqslant 0$. Since

$$\|\mathbf{Z}^{n+1}\| \leqslant \left\|\prod_{i=0}^{n} C_i\right\| \, \|\mathbf{Z}^0\|,$$

this occurs if and only if a constant K (independent of h and k) can be found such that

$$\left\|\prod_{i=0}^{n} C_i\right\| \leqslant K.$$

Now for any matrix C with elements independent of n, the spectral radius (maximum modulus eigenvalue), $\rho(C)$, is related to the norm, $\|C\|$, by the inequality (see Chapter 1)

$$\rho^{n+1}(C) \leqslant \|C^{n+1}\| \leqslant \|C\|^{n+1}$$

for all $n \geqslant 0$. Bearing these inequalities in mind, we consider two simple criteria for regulating the error growth in a calculation based on (45). They are

(i) The spectral radius condition

$$\rho(C) \leqslant 1, \tag{46a}$$

which is *necessary* for stability and for $\rho(C) < 1$ guarantees that the error vector $\mathbf{Z}^n \to 0$ as $n \to \infty$ (see theorem II, Chapter 1), but gives no indication of the magnitude of $\mathbf{Z}^n$ for finite n, and

(ii) The norm condition

$$\|C\| \leqslant 1, \tag{46b}$$

which is *sufficient* for stability and guarantees an ever-diminishing error as n increases. (see theorem I, Chapter 1)

As with the von Neumann criterion for stability, if the right hand sides of (46a) and (46b) are replaced by $1 + O(k)$, exponentially growing solutions of the differential system are allowed. The gap between the conditions (46a) and (46b) depends on the matrix C and we now look at particular types of matrix C.

(i) *Symmetric* Here if we consider the norm $\| \ \|_2$, then

$$\rho(C) = \|C\|_2,$$

and (46a) is necessary and sufficient for stability.

(ii) *Similar to a symmetric matrix* This time, we have

$$P^{-1} CP = \widetilde{C}$$

where P is some non-singular matrix, and $\widetilde{C}$ is symmetric. Although (46a) is again necessary and sufficient for stability, (proved in example 5) it does not imply (46b) and so the error $\mathbf{Z}^n$ can be large for finite n.

(iii) *Otherwise* If the matrix C is other than (i) or (ii) above we, are unable to comment on the gap between conditions (46a) and (46b).

The reader interested in this material will find numerical experiments in section 6.7, which demonstrate that substantial error growth can arise when C is of type (ii) or (iii) above, even if condition (46a) is satisfied [see also Morton (1979)].

Example 5. Show that a calculation carried out according to (45), with C_n the constant matrix C, is stable if condition (46a) holds and C is similar to a symmetric matrix.

From

$$\widetilde{C} = P^{-1} CP \qquad (P \text{ non-singular}),$$

we get

$$\widetilde{C}^n = P^{-1} C^n P \qquad (n = 1, 2, \ldots),$$

and

$$C^n = P\widetilde{C}^n P^{-1}.$$

These results lead to

$$\|\widetilde{C}^n\|_2 \leqslant \|P^{-1}\|_2 \, \|P\|_2 \, \|C^n\|_2,$$

and

$$\|C^n\|_2 \leqslant \|P\|_2 \, \|P^{-1}\|_2 \, \|\widetilde{C}^n\|_2$$

respectively, and so

$$\frac{\|\widetilde{C}^n\|_2}{\|P^{-1}\|_2 \, \|P\|_2} \leqslant \|C^n\|_2 \leqslant \|P\|_2 \, \|P^{-1}\|_2 \, \|\widetilde{C}^n\|_2.$$

But

$$\|\widetilde{C}^n\|_2 = \rho(\widetilde{C}^n) = \rho^n(\widetilde{C}) = \rho^n(C) \leqslant 1,$$

and so provided $\|P\|_2$ and $\|P^{-1}\|_2$ are finite, it follows that $\|C^n\|_2$ is bounded, and so the procedure is stable.

As an example of the matrix method for examining stability, we consider the initial boundary value problem consisting of the heat conduction equation

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

subject to the boundary conditions $u = 0$ at $x = 0,1$ $(t \geqslant 0)$ and the initial condition $u = f(x)$ at $t = 0$ $(0 \leqslant x \leqslant 1)$, being solved by the explicit difference equation

$$U_m^{n+1} = (1 - 2r)U_m^n + r(U_{m+1}^n + U_{m-1}^n).$$

In this case, $A_n = I$ for all n, and

$$B_n = C_n = \begin{bmatrix} (1-2r) & r & & & \bigcirc \\ r & (1-2r) & r & & \\ & \cdot & \cdot & \cdot & \\ & & \cdot & \cdot & r \\ \bigcirc & & & r & (1-2r) \end{bmatrix} \quad \text{for all } n.$$

The eigenvalues of this matrix are

$$\lambda_s = 1 - 4r \sin^2 \frac{s\pi}{2M} \; (s = 1, 2, \ldots, M-1),$$

and from equation (46a), it follows that the method is stable if

$$-1 \leqslant 1 - 4r \sin^2 \frac{s\pi}{2M} \leqslant 1 \; (s = 1, 2, \ldots, M-1),$$

which leads to

$$0 < r \leqslant \tfrac{1}{2},$$

an identical result to that obtained by the method of von Neumann.

The energy method

This method of analysing stability may be applied in principle to problems with variable coefficients and to non-linear equations. Each problem requires a different treatment but we can illustrate the general philosophy by means of an example (see Richtmyer and Morton, (1967) p. 13). Consider the problem of solving the heat conduction equation

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} \quad \text{in } (0 \leqslant x \leqslant 1) \times (t \geqslant 0)$$

with homogeneous Dirichlet boundary conditions

$$u(0, t) = u(1, t) = 0 \qquad (t \geqslant 0)$$

and initial data

$$u(x, 0) = f(x) \qquad (0 \leqslant x \leqslant 1).$$

Multiplying the equation by u and integrating with respect to x, we obtain

$$\int_0^1 u \frac{\partial u}{\partial t} \mathrm{d}x = \int_0^1 u \frac{\partial^2 u}{\partial x^2} \mathrm{d}x,$$

which on integration by parts of the right hand side gives

$$\frac{\partial}{\partial t} \int_0^1 u^2 \mathrm{d}x = -\int_0^1 \left(\frac{\partial u}{\partial x}\right)^2 \mathrm{d}x \leqslant 0.$$

We deduce from this, that

$$\int_0^1 u^2(x, t)\mathrm{d}x \leqslant \int_0^1 u^2(x, 0)\mathrm{d}x = \int_0^1 f^2(x)\mathrm{d}x$$

and therefore the quantity $\int_0^1 u^2(x, t)\mathrm{d}x$ remains bounded as $t \to \infty$.

When the explicit scheme (9) is used, the energy method may be employed to show that the analogous quantity $h \sum_{m+1}^{M-1} (Z_m^n)^2$ remains bounded as $n \to \infty$ as follows:

The nodal errors Z_m^n satisfy the difference equation (see (44))

$$Z_m^{n+1} - Z_m^n = r(Z_{m+1}^n + Z_{m-1}^n - 2Z_m^n) \qquad (m = 1, 2, \ldots, M-1),$$

and the boundary conditions give

$$Z_0^n = Z_M^n = 0 \qquad \text{(for all } n).$$

The difference equation is multiplied by $(Z_m^{n+1} + Z_m^n)$ and the result summed over $m = 1, 2, \ldots, M-1$ to give

$$\|\mathbf{Z}^{n+1}\|^2 - \|\mathbf{Z}^n\|^2 = r \sum_{m=1}^{M-1} (Z_m^{n+1} + Z_m^n)(Z_{m+1}^n + Z_{m-1}^n - 2Z_m^n) \qquad (47)$$

where $\|\mathbf{Z}^n\|^2 = \sum_{m=1}^{M-1} (Z_m^n)^2$. In order to proceed further we require the identities given in the following exercise.

Exercise

8. Show that

$$\sum_{m=1}^{M-1} V_m(Z_{m-1} - Z_m) = \sum_{m=1}^{M-1} Z_m(V_{m+1} - V_m) \qquad (48a)$$

if $Z_0 = V_M = 0$ and

$$\sum_{m=1}^{M-1} V_m(Z_{m+1} - Z_m) = \sum_{m=1}^{M-1} Z_{m+1}(V_m - V_{m+1}) - V_1 Z_1 \qquad (48b)$$

if $Z_M = 0$.

The right hand side of (47) can be rearranged to read

$$r \sum_{m=1}^{M-1} V_m [(Z_{m+1}^n - Z_m^n) - (Z_m^n - Z_{m-1}^n)]$$

with $V_m = Z_m^{n+1} + Z_m^n$. Substituting (48a) and (48b) into this expression and bearing in mind that

$$Z_0^n = Z_0^{n+1} = Z_M^n = Z_M^{n+1} = 0$$

we find that (47) becomes

$$\| Z^{n+1} \|^2 - \| Z^n \|^2 = -r\left[\sum_{m=0}^{M-1} \left\{(Z_{m+1}^n - Z_m^n)^2 + (Z_{m+1}^{n+1} - Z_m^{n+1})(Z_{m+1}^n - Z_m^n)\right\}\right] \qquad (49)$$

If we define

$$E_n = \| \mathbf{Z}^n \|^2 - ½ r \sum_{m=0}^{M-1} (Z_{m+1}^n - Z_m^n)^2 \quad (n = 0, 1, 2, \ldots),$$

it follows that

$$E_n \leqslant \| \mathbf{Z}^n \|^2 \quad ,$$

and

$$E_{n+1} - E_n = \| \mathbf{Z}^{n+1} \|^2 - \| \mathbf{Z}^n \|^2 - \tfrac{1}{2} r \sum_{m=0}^{M-1} \{(Z_{m+1}^{n+1} - Z_m^{n+1})^2 - (Z_{m+1}^n - Z_m^n)^2\}$$

Substituting for $\| \mathbf{Z}^{n+1} \|^2 - \| \mathbf{Z}^n \|^2$ from (49), we get

$$E_{n+1} - E_n = -\tfrac{1}{2} r \sum_{m=0}^{M-1} (Z_{m+1}^{n+1} - Z_m^{n+1} + Z_{m+1}^n - Z_m^n)^2 \leqslant 0.$$

We conclude therefore that E_n is a monotonic decreasing function of n and it remains to show that $\| \mathbf{Z}^n \|^2$ is bounded as $n \to \infty$.

Summing the inequalities

$$(Z_{m+1}^n - Z_m^n)^2 \leqslant (Z_{m+1}^n - Z_m^n)^2 + (Z_{m+1}^n + Z_m^n)^2 = 2(Z_{m+1}^n)^2 + 2(Z_m^n)^2$$

for $m = 0, 1, \ldots, M-1$, leads to

$$\sum_{m=0}^{M-1} (Z_{m+1}^n - Z_m^n)^2 \leqslant 2 \sum_{m=0}^{M-1} [\![(Z_{m+1}^n)^2 + (Z_m^n)^2]\!] = 4 \sum_{m=0}^{M-1} (Z_m^n)^2 = 4 \| \mathbf{Z}^n \|^2,$$

and so from the definition of E_n we get

$$E_n \geqslant (1 - 2r) \| \mathbf{Z}^n \|^2.$$

If $0 < r < \tfrac{1}{2}$,

$$\| \mathbf{Z}^n \|^2 \leqslant \frac{1}{1-2r} E_n,$$

and in view of the inequalities

$$E_n \leqslant E_{n-1} \leqslant \ldots \leqslant E_0 \leqslant \| \mathbf{Z}^0 \|^2,$$

we have

$$\| \mathbf{Z}^n \|^2 \leqslant \frac{1}{1-2r} \| \mathbf{Z}^0 \|^2.$$

The vectors $\mathbf{Z}^n$, $n = 0, 1, 2, \ldots$ are therefore bounded provided $0 < r < \tfrac{1}{2}$ and stability results.

The quantity $\| \mathbf{Z}^n \|^2$ is called the *energy* from which the method gets its name, but it is in no way related to the physical energy of the system. The successful application of this method to more general problems relies heavily on the ingenuity of the user to identify a suitable quantity E_n. For this reason, only *sufficient* conditions for stability can be derived. A comprehensive discussion, along with several examples, may be found in Richtmyer and Morton (1967) (Chapter 6).

Exercise

9. If the Crank–Nicolson formula (31) is used to solve the heat conduction equation, use the method outlined above to show that

$$\| \mathbf{Z}^{n+1} \|^2 - \| \mathbf{Z}^n \|^2 = -r \sum_{m=0}^{M-1} (V_{m+1} - V_m)^2$$

where $V_m = Z_m^{n+1} + Z_m^n$. Deduce directly from this result that the method is stable for $r > 0$.

So far, no mention has been made of the *consistency* of a finite difference approximation to a differential equation. Consistency ensures that the difference equation converges to the correct differential equation as the grid spacing tends to zero. A difference approximation to a parabolic equation is *consistent* if

$$\frac{\text{Truncation error}}{k} \to 0 \text{ as } h, k \to 0.$$

For example, the standard four point explicit formula (9) is consistent since

$$\frac{\text{Truncation error}}{k} = (\tfrac{1}{2} k \frac{\partial^2 u}{\partial t^2} - \tfrac{1}{12} h^2 \frac{\partial^4 u}{\partial x^4} + \ldots)$$

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

In fact, all difference replacements so far mentioned in this book are consistent.

A full discussion of the concepts of stability, convergence and consistency is well beyond the scope of this book and readers who wish to study these topics further are referred to Isaacson and Keller (1967) and Richtmyer and Morton (1967), where there is included a discussion of Lax's equivalence theorem. This theorem states that: 'Given a properly posed linear initial value problem and a finite difference approximation to it that satisfies the consistency condition, stability (as $h, k \to 0$) is the necessary and sufficient condition for convergence'.

Exercise

10. Show that the difference method given by (31) is consistent.

2.8 Derivative boundary conditions

Consider now the equation of heat conduction