The 2-norm of matrix $\mathbf{A}$ is the square root of the spectral radius of $\mathbf{A}^H\mathbf{A}$, where $\mathbf{A}^H = (\bar{\mathbf{A}})^T$, the transpose of the conjugate complex of $\mathbf{A}$. For example, if

$$\mathbf{A} = \begin{bmatrix} -1 & 1 \\ 3 & -2 \end{bmatrix}, \text{ then } \mathbf{A}^H\mathbf{A} = \mathbf{A}^T\mathbf{A} = \begin{bmatrix} 10 & -7 \\ -7 & 5 \end{bmatrix}$$

with eigenvalues 14.93 and 0.067. Hence $\|\mathbf{A}\|_1 = 1+3=4$, $\|\mathbf{A}\|_\infty = 3+2=5$ and $\|\mathbf{A}\|_2 = \sqrt{14.93} = 3.86$.

When matrix $\mathbf{A}$ is real and symmetric, $\mathbf{A}^H = \mathbf{A}$, and

$$\|\mathbf{A}\|_2 = [\rho(\mathbf{A}^2)]^{\frac{1}{2}} = [\rho^2(\mathbf{A})]^{\frac{1}{2}} = \rho(\mathbf{A}) = \max_i |\lambda_i|.$$

A bound for the spectral radius

Let λ_i be an eigenvalue of the $n \times n$ matrix $\mathbf{A}$ and $\mathbf{x}_i$ the corresponding eigenvector. Hence

$$\mathbf{A}\mathbf{x}_i = \lambda_i \mathbf{x}_i$$

and

$$\|\mathbf{A}\mathbf{x}_i\| = \|\lambda_i \mathbf{x}_i\| = |\lambda_i|\,\|\mathbf{x}_i\|.$$

For all compatible matrix and vector norms it follows that

$$|\lambda_i|\,\|\mathbf{x}_i\| = \|\mathbf{A}\mathbf{x}_i\| \leq \|\mathbf{A}\|\,\|\mathbf{x}_i\|.$$

Therefore,

$$|\lambda_i| \leq \|\mathbf{A}\|, \quad i = 1(1)n.$$

Hence,

$$\rho(\mathbf{A}) \leq \|\mathbf{A}\|.$$

A necessary and sufficient condition for stability. *(Constants coefficients)*

Let the solution domain of the partial differential equation be the finite rectangle $0 \leq x \leq 1$, $0 \leq t \leq T$, and subdivide it into uniform rectangular meshes by the lines $x_i = ih$, $i = 0(1)N$, where $Nh = 1$, and the lines $t_j = jk$, $j = 0(1)J$ where $Jk = T$. It will be assumed that h is related to k by some relationship such as $k = rh$ or $k = rh^2$, $r > 0$ and finite, so that $h \to 0$ as $k \to 0$.

Assume that the finite-difference equation relating the mesh-point values along the $(j+1)$th and jth time-rows is

$$b_{i-1}u_{i-1,j+1}+b_iu_{i,j+1}+b_{i+1}u_{i+1,j+1}=c_{i-1}u_{i-1,j}+c_iu_{i,j}+c_{i+1}u_{i+1,j}$$

where the coefficients are constants.

If the boundary values at $i=0$ and N, $j>0$, are known, these $(N-1)$ equations for $i=1(1)N-1$ can be written in matrix form as

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

$$= \begin{bmatrix} c_1 & c_2 & & & \\ c_1 & c_2 & c_3 & & \\ \cdot & \cdot & \cdot & & \\ & \cdot & \cdot & \cdot & \\ & & c_{N-3} & c_{N-2} & c_{N-1} \\ & & & c_{N-2} & c_{N-1} \end{bmatrix} \begin{bmatrix} u_{1,j} \\ u_{2,j} \\ \cdot \\ \cdot \\ u_{N-2,j} \\ u_{N-1,j} \end{bmatrix} + \begin{bmatrix} c_0u_{0,j}-b_0u_{0,j+1} \\ 0 \\ \cdot \\ \cdot \\ 0 \\ c_Nu_{N,j}-b_Nu_{N,j+1} \end{bmatrix}$$

i.e. as $\mathbf{B}\mathbf{u}_{j+1}=\mathbf{C}\mathbf{u}_j+\mathbf{d}_j$, where the matrices $\mathbf{B}$ and $\mathbf{C}$ of order $(N-1)$ are as shown, $\mathbf{u}_{j+1}$ denotes the column vector with components $u_{1,j+1}, u_{2,j+1}, \ldots, u_{N-1,j+1}$ and $\mathbf{d}_j$ denotes the column vector of known boundary values and zeros.

Hence,

$$\mathbf{u}_{j+1}=\mathbf{B}^{-1}\mathbf{C}\mathbf{u}_j+\mathbf{B}^{-1}\mathbf{d}_j,$$

which may be expressed more conveniently as

$$\mathbf{u}_{j+1}=\mathbf{A}\mathbf{u}_j+\mathbf{f}_j,$$

where $\mathbf{A}=\mathbf{B}^{-1}\mathbf{C}$ and $\mathbf{f}_j=\mathbf{B}^{-1}\mathbf{d}_j$. Applied recursively, this leads to

$$\begin{aligned}\mathbf{u}_j &= \mathbf{A}\mathbf{u}_{j-1}+\mathbf{f}_{j-1}=\mathbf{A}(\mathbf{A}\mathbf{u}_{j-2}+\mathbf{f}_{j-2})+\mathbf{f}_{j-1}\\ &= \mathbf{A}^2\mathbf{u}_{j-2}+\mathbf{A}\mathbf{f}_{j-2}+\mathbf{f}_{j-1}\\ &= \ldots\\ &= \mathbf{A}^j\mathbf{u}_0+\mathbf{A}^{j-1}\mathbf{f}_0+\mathbf{A}^{j-2}\mathbf{f}_1+\ldots+\mathbf{f}_{j-1}, \qquad (2.30)\end{aligned}$$

where $\mathbf{u}_0$ is the vector of initial values and $\mathbf{f}_0, \mathbf{f}_1, \ldots, \mathbf{f}_{j-1}$ are vectors of known boundary-values. When we are concerned more with a property of the equations, such as stability, than with a numerical solution, the constant vectors can be eliminated by investigating the propagation of a perturbation.

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

$$\mathbf{u}_j^* = \mathbf{A}^j\mathbf{u}_0^* + \mathbf{A}^{j-1}\mathbf{f}_0 + \mathbf{A}^{j-2}\mathbf{f}_1 + \ldots + \mathbf{f}_{j-1}. \qquad (2.31)$$

If the perturbation or 'error' vector $\mathbf{e}$ is defined by

$$\mathbf{e} = \mathbf{u}^* - \mathbf{u},$$

it follows by eqns (2.30) and (2.31) that

$$\mathbf{e}_j = \mathbf{u}_j^* - \mathbf{u}_j = \mathbf{A}^j(\mathbf{u}_0^* - \mathbf{u}_0) = \mathbf{A}^j\mathbf{e}_0,\ j = 1(1)J. \qquad (2.32)$$

In other words, a perturbation $\mathbf{e}_0$ of the initial values will be propagated according to the equation

$$\mathbf{e}_j = \mathbf{A}\mathbf{e}_{j-1} = \mathbf{A}^2\mathbf{e}_{j-2} = \ldots = \mathbf{A}^j\mathbf{e}_0, \quad j = 1(1)J.$$

Hence, for compatible matrix and vector norms,

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

Lax and Richtmyer define the difference scheme to be stable when there exists a positive number M, independent of j, h, and k, such that

$$\|\mathbf{A}^j\| \leqslant M, \quad j = 1(1)J.$$

This clearly limits the amplification of any initial perturbation, and therefore of any arbitrary initial rounding errors, because

$$\|\mathbf{e}_j\| \leqslant M\,\|\mathbf{e}_0\|.$$

In reference 25, this definition of stability is related to convergence via Lax's equivalence theorem (p. 72).

Since

$$\|\mathbf{A}^j\| = \|\mathbf{A}\mathbf{A}^{j-1}\| \leqslant \|\mathbf{A}\|\,\|\mathbf{A}^{j-1}\| \leqslant \ldots \leqslant \|\mathbf{A}\|^j,$$

it follows that the Lax–Richtmyer definition of stability is satisfied by

$$\|\mathbf{A}\| \leqslant 1.$$

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

When this condition is satisfied it follows automatically that the spectral radius $\rho(\mathbf{A}) \leq 1$ since $\rho(\mathbf{A}) \leq \|\mathbf{A}\|$.

If, however, $\rho(\mathbf{A}) \leq 1$, it does not follow that $\|\mathbf{A}\| \leq 1$. This is demonstrated by the simple example

$$\mathbf{A} = \begin{bmatrix} -0.8 & 0 \\ 0.4 & 0.7 \end{bmatrix},$$

for which $\lambda_1 = -0.8$, $\lambda_2 = 0.7$, $\rho(\mathbf{A}) = 0.8$, $\|\mathbf{A}\|_1 = 1.2$ and $\|\mathbf{A}\|_\infty = 1.1$. If, however, $\mathbf{A}$ is real and symmetric, then

$$\|\mathbf{A}\|_2 = \sqrt{\rho(\mathbf{A}^T\mathbf{A})} = \sqrt{\rho(\mathbf{A}^2)} = \sqrt{\rho^2(\mathbf{A})} = \rho(\mathbf{A}).$$

Example 2.8

Consider the stability of the classical explicit equations

$$u_{i,j+1} = ru_{i-1,j} + (1-2r)u_{i,j} + ru_{i+1,j}, \quad i = 1(1)N-1,$$

for which the $(N-1)\times(N-1)$ matrix $\mathbf{A}$ is

$$\begin{bmatrix} (1-2r) & r & & \\ r & (1-2r)r & & \\ & r & (1-2r) & r \\ & & r & (1-2r), \end{bmatrix}$$

where $r = k/h^2 > 0$, and it is assumed that the boundary values $u_{0,j}$ and $u_{N,j}$ are known for $j = 1, 2, \ldots$.

When $1-2r \geq 0$, then $0 < r \leq \frac{1}{2}$ and

$$\|\mathbf{A}\|_\infty = r + (1-2r) + r = 1.$$

When $1-2r < 0$, $r > \frac{1}{2}$, $|1-2r| = 2r-1$ and

$$\|\mathbf{A}\|_\infty = r + 2r - 1 + r = 4r - 1 > 1.$$

Therefore the scheme is stable for $0 < r \leq \frac{1}{2}$.

Alternatively, since matrix $\mathbf{A}$ is real and symmetric,

$$\|\mathbf{A}\|_2 = \rho(\mathbf{A}) = \max_s |\mu_s|,$$

where μ_s is the sth eigenvalue of $\mathbf{A}$. Now $\mathbf{A}$ can be written as

$$\begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ & & & & 1 \end{bmatrix} + r\begin{bmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & & \ddots & & \\ & & 1 & -2 & 1 \\ & & & 1 & -2 \end{bmatrix} = \mathbf{I}_{N-1} + r\mathbf{T}_{N-1},$$

where $\mathbf{I}_{N-1}$ is the unit matrix of order $(N-1)$ and $\mathbf{T}_{N-1}$ an $(N-1)\times(N-1)$ matrix whose eigenvalues λ_s are given by

$$\lambda_s = -4\sin^2 s\pi/2N, \quad s = 1(1)N-1. \text{ (See p. 59.)}$$

Hence the eigenvalues of $\mathbf{A}$, as shown later in 'A note on eigenvalues and eigenvectors' are $\mu_s = 1-4r\sin^2 s\pi/2N$.

Therefore the equations will be stable when

$$\|\mathbf{A}\|_2 = \max_s |1-4r\sin^2 s\pi/2N| \leq 1,$$

i.e.,

$$-1 \leq 1-4r\sin^2 s\pi/2N \leq 1, \quad s = 1(1)N-1.$$

The left-hand inequality gives that

$$r \leq 1/2\sin^2 (N-1)\pi/2N.$$

As $h \to 0$, $N \to \infty$ and $\sin^2(N-1)\pi/2N \to 1$.
Hence $r \leq \frac{1}{2}$.

It has been shown that these equations are also consistent. Hence by Lax's equivalence theorem they are also convergent for $0 < r \leq \frac{1}{2}$.

Example 2.9

The Crank–Nicolson equations (2.10) are

$$-ru_{i-1,j+1} + (2+2r)u_{i,j+1} - ru_{i+1,j+1}$$
$$= ru_{i-1,j} + (2-2r)u_{i,j} + ru_{i+1,j}, \; i = 1(1)N-1.$$

In matrix form, for known boundary values, these give

$$\begin{bmatrix} (2+2r) & -r & & & \\ -r & (2+2r) & -r & & \\ & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & -r & (2+2r) & -r \\ & & & -r & (2+2r) \end{bmatrix} \begin{bmatrix} u_{1,j+1} \\ u_{2,j+1} \\ \cdot \\ \cdot \\ \cdot \\ u_{N-2,j+1} \\ u_{N-1,j+1} \end{bmatrix}$$

$$= \begin{bmatrix} (2-2r) & r & & & \\ r & (2-2r) & r & & \\ & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & r & (2-2r) & r \\ & & & r & (2-2r) \end{bmatrix} \begin{bmatrix} u_{1,j} \\ u_{2,j} \\ \cdot \\ \cdot \\ \cdot \\ u_{N-2,j} \\ u_{N-1,j} \end{bmatrix} + \mathbf{b}_j,$$

where $\mathbf{b}_j$ is a vector of known boundary values and zeros. This can be written as

$$(2\mathbf{I}_{N-1} - r\mathbf{T}_{N-1})u_{j+1} = (2\mathbf{I}_{N-1} + r\mathbf{T}_{N-1})\mathbf{u}_j + \mathbf{b}_j$$

from which it follows that matrix $\mathbf{A}$ of eqn (2.30) is

$$\mathbf{A} = (2\mathbf{I}_{N-1} - r\mathbf{T}_{N-1})^{-1}(2\mathbf{I}_{N-1} + r\mathbf{T}_{N-1}).$$

In Exercise 14 it is proved that if the $n \times n$ symmetric matrices $\mathbf{B}$ and $\mathbf{C}$ commute then $\mathbf{B}^{-1}\mathbf{C}$, $\mathbf{B}\mathbf{C}^{-1}$ and $\mathbf{B}^{-1}\mathbf{C}^{-1}$ are symmetric. Matrix $\mathbf{T}_{N-1}$ is symmetric so $2\mathbf{I}_{N-1} - r\mathbf{T}_{N-1}$ and $2\mathbf{I}_{N-1} + r\mathbf{T}_{N-1}$ are symmetric. They also commute as their multiplication immediately shows. Hence matrix $\mathbf{A}$ is symmetric. Since the eigenvalues of $\mathbf{T}_{N-1}$ are $\lambda_s = -4\sin^2 s\pi/2N$, $s = 1(1)N-1$, it follows that the eigenvalues of $\mathbf{A}$ are $(2+4r\sin^2 s\pi/2N)^{-1}(2-4r\sin^2 s\pi/2N)$. (See p. 59.)
Therefore

$$\|\mathbf{A}\|_2 = \rho(\mathbf{A}) = \max_s \left| \frac{1-2r\sin^2 s\pi/2N}{1+2r\sin^2 s\pi/2N} \right| < 1 \text{ for all } r > 0,$$

proving that the Crank–Nicolson equations are unconditionally stable. They are also consistent, reference Exercise 11, so they are also convergent, although it will be shown under A_0-stability that r must be restricted in order to avoid the possibility of finite oscillations near points of discontinuity.

Matrix method of analysis, fixed mesh lengths

The following method of analysis will establish the conditions necessary for the boundedness of the analytical solution of the finite-difference equations as $t_j = jk$ tends to infinity for fixed mesh lengths h and k. These conditions may not, however, be sufficient to ensure convergence when the equations are also consistent and large finite errors can occur near the end points of the range of values of some parameter, such as $r = k/h^2$, for which the equations are bounded.

Consider the classical explicit finite-difference equations incorporating known boundary values, namely,

$$\begin{bmatrix} u_{1,j+1} \\ u_{2,j+1} \\ \cdot \\ \cdot \\ \cdot \\ u_{N-2,j+1} \\ u_{N-1,j+1} \end{bmatrix} = \begin{bmatrix} (1-2r) & r & & & & \\ r & (1-2r) & r & & & \\ & \cdot & \cdot & \cdot & & \\ & & \cdot & \cdot & \cdot & \\ & & & \cdot & \cdot & \cdot \\ & & & r & (1-2r) & r \\ & & & & r & (1-2r) \end{bmatrix} \begin{bmatrix} u_{1,j} \\ u_{2,j} \\ \cdot \\ \cdot \\ \cdot \\ u_{N-2,j} \\ u_{N-1,j} \end{bmatrix} + \begin{bmatrix} ru_{0,j} \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ ru_{N,j} \end{bmatrix},$$

i.e.

$$\mathbf{u}_{j+1} = \mathbf{A}\mathbf{u}_j + \mathbf{b}_j.$$

As shown by eqn (2.32), if the vector of initial values $\mathbf{u}_0$ is perturbed to $\mathbf{u}_0^*$ and no further perturbations or errors are introduced into the subsequent calculations, the perturbation vector $\mathbf{e} = \mathbf{u}^* - \mathbf{u}$ will be propagated forward in time, according to the equation

$$\mathbf{e}_j = \mathbf{A}^j \mathbf{e}_0,$$

a procedure that eliminates the boundary values.

For fixed mesh lengths h and k the difference equations will be stable if $\mathbf{e}_j$ remains bounded as j increases indefinitely. This can always be investigated by expressing the initial perturbation vector in terms of the eigenvectors of $\mathbf{A}$, which remain fixed as j increases.

Assume that matrix $\mathbf{A}$ is non-deficient, i.e. has $(N-1)$ linearly independent eigenvectors $\mathbf{v}_s$, which will be so if the eigenvalues λ_s of $\mathbf{A}$ are all distinct or $\mathbf{A}$ is real and symmetric. Then these eigenvectors can be used as a basis for our $(N-1)$ dimensional vector space and the perturbation vector $\mathbf{e}_0$, with its $(N-1)$ components, can be expressed uniquely as a linear combination of them, namely,

$$\mathbf{e}_0 = \sum_{s=1}^{N} c_s \mathbf{v}_s,$$

where the c_s, $s = 1(1)N-1$, are known scalars.

The perturbations along time-level $t = k$, resulting from the initial perturbation $\mathbf{e}_0$ will be

$$\mathbf{e}_1 = \mathbf{A}\mathbf{e}_0 = \mathbf{A} \sum_{s=1}^{N-1} c_s \mathbf{v}_s = \sum c_s \mathbf{A}\mathbf{v}_s.$$

But $\mathbf{A}\mathbf{v}_s = \lambda_s \mathbf{v}_s$ by the definition of an eigenvalue. Therefore,

$$\mathbf{e}_1 = \sum c_s \lambda_s \mathbf{v}_s.$$

Similarly,

$$\mathbf{e}_j = \sum_{s=1}^{N-1} c_s \lambda_s^j \mathbf{v}_s.$$

This shows that the perturbations will not increase exponentially with j provided

$$\max_s |\lambda_s| \leq 1, \quad s = 1(1)N-1.$$

By p. 59, $\lambda_s = 1 - 4r \sin^2 s\pi/2N$.
Therefore any perturbation, rounding errors and $u_{i,j}$ will be bounded as j increases if

$$-1 \leq 1 - 4r \sin^2 s\pi/2N \leq 1, \quad \text{where } r > 0.$$

This is satisfied by $r \leq \frac{1}{2}$.

A note on eigenvalues and eigenvectors

Let $\mathbf{x}$ be an eigenvector of the matrix $\mathbf{A}$ corresponding to the eigenvalue λ. Then $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$. Hence $\mathbf{A}(\mathbf{A}\mathbf{x}) = \mathbf{A}^2\mathbf{x} = \lambda\, \mathbf{A}\mathbf{x} = \lambda^2\mathbf{x}$, showing that the matrix $\mathbf{A}^2$ has an eigenvalue λ^2 corresponding to the eigenvector $\mathbf{x}$. Similarly $\mathbf{A}^p\mathbf{x} = \lambda^p\mathbf{x}$, $p = 3, 4, \ldots$.

(i) If $f(\mathbf{A}) = a_p\mathbf{A}^p + a_{p-1}\mathbf{A}^{p-1} + \ldots + a_0\mathbf{I}$ is a polynomial in $\mathbf{A}$ with scalar coefficients $a_p, \ldots, a_0$, then $f(\mathbf{A})\mathbf{x} = (a_p\lambda^p + \ldots + a_0)\mathbf{x} = f(\lambda)\mathbf{x}$, showing that $f(\mathbf{A})$ has an eigenvalue $f(\lambda)$ corresponding to the eigenvector $\mathbf{x}$.

(ii) The eigenvalue of $[f_1(\mathbf{A})]^{-1}f_2(\mathbf{A})$ corresponding to the eigenvector $\mathbf{x}$ is $f_2(\lambda)/f_1(\lambda)$, where $f_1(\mathbf{A})$ and $f_2(\mathbf{A})$ are polynomials in $\mathbf{A}$. The proof is as follows. By (i),

$$f_1(\mathbf{A})\mathbf{x} = f_1(\lambda)\mathbf{x} \quad \text{and} \quad f_2(\mathbf{A})\mathbf{x} = f_2(\lambda)\mathbf{x}.$$

Premultiply both equations by $[f_1(\mathbf{A})]^{-1}$ and write as

$$[f_1(\mathbf{A})]^{-1}\mathbf{x} = \mathbf{x}/f_1(\lambda) \quad \text{and} \quad [f_1(\mathbf{A})]^{-1}f_2(\mathbf{A})\mathbf{x} = f_2(\lambda)[f_1(\mathbf{A})]^{-1}\mathbf{x}.$$

Then the elimination of $[f_1(\mathbf{A})]^{-1}\mathbf{x}$ between these two equations shows that

$$[f_1(\mathbf{A})]^{-1}f_2(\mathbf{A})\mathbf{x} = \{f_2(\lambda)/f_1(\lambda)\}\mathbf{x},$$

which states, by the definition of an eigenvalue, that $f_2(\lambda)/f_1(\lambda)$ is an eigenvalue of $[f_1(\mathbf{A})]^{-1}f_2(\mathbf{A})$ corresponding to the eigenvector $\mathbf{x}$. In a similar manner the eigenvalue of $f_2(\mathbf{A})[f_1(\mathbf{A})]^{-1}$ corresponding to the eigenvector $\mathbf{x}$ is $f_2(\lambda)/f_1(\lambda)$.

In particular, the eigenvalue of $[f_1(\mathbf{A})]^{-1}$ corresponding to the eigenvector $\mathbf{x}$ is $1/f_1(\lambda)$.

The eigenvalues of a common tridiagonal matrix

The eigenvalue of the $N \times N$ matrix

$$\begin{bmatrix} a & b & & & & \\ c & a & b & & & \\ & c & a & b & & \\ & & & \cdots & & \\ & & & c & a & b \\ & & & & c & a \end{bmatrix}$$

are

$$\lambda_s = a + 2\{\surd(bc)\}\cos\frac{s\pi}{N+1}, \quad s = 1(1)N,$$

where a, b, and c may be real or complex. A proof is given on p. 154.

Another useful result is the following. If a real tridiagonal matrix has either all its off-diagonal elements positive or all its off-diagonal elements negative, then all its eigenvalues are real. A proof is given in Exercise 16.

Useful theorems on bounds for eigenvalues

Gerschgorin's first theorem

The largest of the moduli of the eigenvalues of the square matrix $\mathbf{A}$ cannot exceed the largest sum of the moduli of the elements along any row or any column. In other words

$$\rho(A) \leq \|\mathbf{A}\|_1 \quad \text{or} \quad \|\mathbf{A}\|_\infty.$$

Proof

Let λ_i be an eigenvalue of the $N \times N$ matrix $\mathbf{A}$, and $\mathbf{x}_i$ the corresponding eigenvector with components $v_1, v_2 \ldots v_n$. Then the equation

$$\mathbf{A}\mathbf{x}_i = \lambda_i \mathbf{x}_i$$

in detail, is

$$\begin{aligned}
a_{1,1}v_1 + a_{1,2}v_2 + \ldots + a_{1,n}v_n &= \lambda_i v_1,\\
a_{2,1}v_1 + a_{2,2}v_2 + \ldots + a_{2,n}v_n &= \lambda_i v_2,\\
\vdots \qquad \vdots \qquad & \quad \vdots \qquad \vdots\\
a_{s,1}v_1 + a_{s,2}v_2 + \ldots + a_{s,n}\,v_n &= \lambda_i v_s,\\
\vdots \qquad \vdots \qquad & \quad \vdots \qquad \vdots
\end{aligned}$$

Let v_s be the largest in modulus of $v_1, v_2, \ldots, v_n$. Select the sth equation and divide by v_s, giving

$$\lambda_i = a_{s,1}\left(\frac{v_1}{v_s}\right) + a_{s,2}\left(\frac{v_2}{v_s}\right) + \ldots + a_{s,n}\left(\frac{v_n}{v_s}\right).$$

Therefore

$$|\lambda_i| \leq |a_{s,1}| + |a_{s,2}| + \ldots + |a_{s,n}|,$$

because

$$\left|\frac{v_i}{v_s}\right| \leq 1, \quad i = 1, 2, \ldots, n.$$

If this is not the largest row sum then $|\lambda_i| <$ the largest row sum. In particular this holds for $|\lambda_i| = \max|\lambda_s|$, $s = 1(1)N$. Since the eigenvalues of the transpose of $\mathbf{A}$ are the same as those of $\mathbf{A}$ the theorem is also true for columns.

Gerschgorin's circle theorem or Brauer's theorem

Let P_s be the sum of the moduli of the elements along the sth row excluding the diagonal element $a_{s,s}$. Then each eigenvalue of **A** lies inside or on the boundary of at least one of the circles $|\lambda - a_{s,s}| = P_s$.

Proof

By the previous proof,

$$\lambda_i = a_{s,1}\left(\frac{v_1}{v_s}\right) + a_{s,2}\left(\frac{v_2}{v_s}\right) + \ldots + a_{s,s} + \ldots + a_{s,n}\left(\frac{v_n}{v_s}\right).$$

Hence

$$\begin{aligned}|\lambda_i - a_{s,s}| &= \left|a_{s,1}\left(\frac{v_1}{v_s}\right) + \ldots + 0 + \ldots + a_{s,n}\left(\frac{v_s}{v_s}\right)\right| \\ &\leq |a_{s,1}| + |a_{s,2}| + \ldots + 0 + \ldots + |a_{s,n}| \\ &= P_s,\end{aligned}$$

which completes the proof.

As an illustrative example consider the Crank–Nicolson equations with known boundary-values, namely

$$\begin{aligned}&(2\mathbf{I}_{N-1} - r\mathbf{T}_{N-1})\mathbf{u}_{j+1} \\ &\quad = (2\mathbf{I}_{N-1} + r\mathbf{T}_{N-1})\mathbf{u}_j + \mathbf{b}_j = \{4\mathbf{I}_{N-1} - (2\mathbf{I}_{N-1} - r\mathbf{T}_{N-1})\}\mathbf{u}_j + \mathbf{b}_j,\end{aligned}$$

which can be written as

$$\mathbf{B}\mathbf{u}_{j+1} = (4\mathbf{I}_{N-1} - \mathbf{B})\mathbf{u}_j + \mathbf{b}_j,$$

giving

$$\mathbf{u}_{j+1} = (4\mathbf{B}^{-1} - \mathbf{I}_{N-1})\mathbf{u}_j + \mathbf{B}^{-1}\mathbf{b}_j,$$

where

$$\mathbf{B} = \begin{bmatrix} (2+2r) & -r & & & \\ -r & (2+2r) & -r & & \\ & \ddots & \ddots & \ddots & \\ & & -r & (2+2r) & -r \\ & & & -r & (2+2r) \end{bmatrix}$$

The equations will be stable in the Lax–Richtmyer sense when $\|4\mathbf{B}^{-1}-\mathbf{I}_{N-1}\|\leq 1$. Since $\mathbf{B}$ is real and symmetric it follows by Exercise 14 that $4\mathbf{B}^{-1}-\mathbf{I}_{N-1}$ is real and symmetric, so $\|4\mathbf{B}^{-1}-\mathbf{I}_{N-1}\|=\rho(4\mathbf{B}^{-1}-\mathbf{I}_{N-1})$. The stability condition will therefore be satisfied when the modulus of every eigenvalue of $4\mathbf{B}^{-1}-\mathbf{I}_{N-1}$ does not exceed one; that is, when

$$\left|\frac{4}{\lambda}-1\right|\leq 1,\quad \text{implying}\quad -1\leq\frac{4}{\lambda}-1\leq 1,$$

where λ is an eigenvalue of $\mathbf{B}$. This states that $\lambda\geq 2$.

For the matrix $\mathbf{B}$, $a_{s,s}=2+2r$, $\max P_s=2r$, so Gerschgorin's circle theorem leads to

$$|\lambda-2-2r|\leq 2r,$$

from which it follows that

$$-2r\leq\lambda-2-2r\leq 2r,$$

or

$$2\leq\lambda\leq 2+4r,$$

proving that the equations are unconditionally stable, since $\lambda\geq 2$ for all $r>0$.

Gerschgorin's circle theorem and the norm of matrix A

It should be noted that when the eigenvalues λ_i of matrix $\mathbf{A}$ are estimated by the circle theorem, the condition $|\lambda_i|\leq 1$ is equivalent to $\|\mathbf{A}\|_\infty\leq 1$ or $\|\mathbf{A}\|_1\leq 1$. The theorem states that

$$|\lambda-a_{s,s}|\leq P_s.$$

Hence,

$$-P_s\leq\lambda-a_{ss}\leq P_s,$$

so that

$$-P_s+a_{ss}\leq\lambda\leq P_s+a_{ss}.$$

The eigenvalue λ will therefore satisfy $-1\leq\lambda\leq 1$ if

$$-1\leq -P_s+a_{ss}\leq P_s+a_{ss}\leq 1,\quad s=1(1)N-1.$$

Remembering that P_s is the sum of the moduli of the elements of $\mathbf{A}$ along the sth row and that a_{ss} may be positive or negative, this