Pré-Publicações do Departamento de Matemática
Universidade de Coimbra
Preprint Number 07–10

ON THE EDGE OF STABILITY ANALYSIS

ERCÍLIA SOUSA

Abstract: The application of high order methods to solve problems with physical boundary conditions in many cases implies to consider a different numerical approximation on the discrete points near the boundary. The choice of these approximations, called the numerical boundary conditions, influence most of the times the stability of the numerical method.

Some theoretical analysis for stability, such as the von Neumann analysis, do not take into account the influence of the numerical representation of the boundary conditions on the overall stability of the scheme. The spectral analysis considers the eigenvalues of the matrix iteration of the scheme and although they reflect some of the influence of numerical boundary conditions on the stability, many times eigenvalues fail to capture the transient effects in time-dependent partial differential equations.

The Lax analysis does provide information on the influence of numerical boundary conditions although in practical situations it is generally not easy to derive the corresponding stability conditions. In this paper we present properties that relates the von Neumann analysis, the spectral analysis and the Lax analysis and show under which circumstances the von Neumann analysis together with the spectral analysis provides sufficient conditions to achieve Lax stability.

Keywords: Stability, high-order methods, von Neumann analysis, spectral.

1. Introduction

The analysis of numerical schemes involves the study of consistency, accuracy, stability and convergence. Clearly, the conditions of consistency, stability and convergence are related to each other and the precise relation is contained in the fundamental Equivalence Theorem of Lax a proof of which can be found in Ritchmyer and Morton [9]. The theorem says that for a well-posed initial value problem and a consistent discretisation scheme, stability is the necessary and sufficient condition for convergence.

The von Neumann analysis is the most well known classical method to determine stability conditions. However, when applicable to finite domain problems, this method only provides necessary and sufficient conditions if we assume periodic boundary conditions. The spectral analysis takes in consideration the spectral radius of the matrix iteration of the numerical method.

Received March 22, 2007.

If the matrix is normal, the spectral radius gives accurate information about the matrix norm and this guarantees sufficient and necessary conditions for stability. When the matrix iteration is not normal the spectral radius gives no indication of the magnitude of the error for a finite time. It guarantees eventual decay of the solution, but does not control the intermediate growth. Therefore, the spectral condition for a non normal matrix it is only a necessary condition for stability. Additionally, practical computations are performed for finite values which reinforces the importance of controlling the intermediate growth of the error.

The Lax stability guarantees that the norm of the powers of the matrix iteration is uniformly bounded. In this paper we show that under certain circumstances the von Neumann and spectral stability together ensure Lax stability even when we have a finite domain problem with non-periodic boundary conditions.

Nowadays there are many textbooks that describes the three types of stability analysis mentioned above, such as, the classical book by Ritchmyer and Morton [9], or some more recent books [4], [5], [8], [10].

In many works the von Neumann analysis has been used to study the stability even when we have finite domains without periodic boundary conditions. Therefore it is assumed in practice that the von Neumann analysis is a necessary condition for stability. Also the work of Godunov and Ryabenkii [2],[3] considers the von Neumann condition has still being necessary for stability.

The outline of the paper is as follows. In the next section we introduce the model problem and the general form of the finite difference schemes considered. In section 3, we give an overview of the stability theories we are discussing. In section 4, we present the main results. In section 5, we obtain stability conditions for some practical examples using the theories considered. The last section includes some conclusions.

2. The finite difference schemes

The initial boundary problem we consider is a linear equation and is defined on a half-real line:

$$\frac{\partial u}{\partial t} = \mathcal{L}u(x,t) \quad x \geq 0, \quad t \geq 0, \tag{1}$$

where $\mathcal{L}$ is a general differential operator in x. The initial condition and the boundary conditions are given by

$$u(x,0) = f(x), \tag{2}$$

$$u(0,t) = g(t), \qquad \lim_{x\to\infty} u(x,t) = 0. \tag{3}$$

Suppose we have approximations U_j^n to the values $u(x_j, t_n)$ at the mesh points $x_j, j = 0, 1, 2, \ldots$ and assume we approximate the problem (1)–(3) by the difference scheme

$$U_j^{n+1} = QU_j^n, \qquad j = r, r+1, \ldots \tag{4}$$

$$Q = \sum_{j=-r}^{p} a_j E^j, \quad E^a U_j^n = U_{j+a}^n, \tag{5}$$

where a_{-r} and a_p are non-zero. The a_j's also depend of parameters Δx and Δt, where Δt is the time step and Δx the space step.

Considering the finite difference scheme (4) we observe that as Q uses r points upstream, the basic approximation can not be used at x_0, x_1, x_2, $\ldots$, x_{r-1}, so there we need to apply numerical boundary conditions. In our particular case the boundary given by the physical problem is associated only with the point x_0. At the other points the boundary conditions, called numerical boundary conditions, affect the difference scheme. Let us assume that the boundary conditions can be written as

$$U_\beta^{n+1} = \sum_{j=0}^{q} b_{\beta j} U_j^n \quad \beta = 0, 1, \ldots, r-1 \tag{6}$$

and $b_{\beta j}$ depend also of parameters Δx and Δt.

3. Stability analysis

3.1. The von Neumann analysis. The von Neumann (Fourier) method is the most well known classical method to determine necessary and sufficient stability conditions. If we assume periodic boundary conditions the von Neumann analysis is based on the decomposition of the numerical solution into a Fourier sum as

$$U_j^n = \sum_{p=0}^{N-1} \kappa_p^n \mathrm{e}^{\mathrm{i}\xi_p (j\Delta x)}$$

where $\mathrm{i} = \sqrt{-1}$, κ_p^n is the amplification factor of the p-th harmonic and $\xi_p = \dfrac{p2\pi}{N\Delta x}$. The product $\xi_p \Delta x$ is often called the phase angle: $\theta = \xi_p \Delta x$ and covers the domain $[0, 2\pi)$ in steps of $2\pi/N$. The region around $\theta = 0$ corresponds to the low frequencies while the region $\theta = \pi$ is associated with the high-frequencies. In particular, the value $\theta = \pi$ corresponds to the highest frequency resolvable on the mesh, namely the frequency of wavelength $2\Delta x$.

Considering a single mode, $\kappa^n \mathrm{e}^{\mathrm{i}j\theta}$, its time evolution is determined by the same numerical scheme as the complete numerical solution U_j^n. Hence inserting a representation of this form into a numerical scheme we obtain a stability condition by imposing an upper bound to the amplification factor, κ.

The amplification factor is said to satisfy the **von Neumann condition** if there is a constant K such that

$$|\kappa(\xi)| \leq 1 + K\Delta t, \quad \forall \xi \in \mathbb{R}. \tag{7}$$

However, for some problems the presence of the arbitrary constant in (7) is too generous for practical purposes, although being adequate for eventual convergence in the limit $\Delta t \to 0$. In practice, the inequality (7) is substituted by the following stronger condition,

$$|\kappa(\xi)| \leq 1, \quad \forall \xi \in \mathbb{R}, \tag{8}$$

or in terms of the phase angle,

$$|\kappa(\theta)| \leq 1, \quad \forall \theta \in [0, 2\pi). \tag{9}$$

This has been called practical stability by Richtmyer and Morton [9] or strict stability by other authors. In some cases condition (7) allows numerical modes to grow exponentially in time for finite values of Δt. Therefore, the practical, or strict, stability condition (8) is recommended in order to prevent numerical modes from growing faster than the physical modes of the differential equation.

Von Neumann stability: The amplification factor satisfy the practical von Neumann condition if

$$|\kappa(\xi)| \leq 1, \quad \forall \xi \in \mathbb{R}.$$

3.2. The spectral analysis. Let us assume that we are considering an explicit method that can be written in the form of a matrix iteration, where the nodal points are $U_j^n, j = 0, \ldots, N-1$.

Introducing the vector $U^n = [U_0^n, \ldots, U_{N-1}^n]^T$, the scheme may be written as a matrix equation

$$U^{n+1} = AU^n + \mathbf{v}, \quad n = 0, 1, 2, \ldots. \tag{10}$$

where A is an $N \times N$ matrix and $\mathbf{v}$ is a vector that may contain some data from the physical boundary conditions.

Any errors E^n in a calculation based on (10) will grow according to

$$E^{n+1} = AE^n, \quad n = 0, 1, 2, \ldots. \tag{11}$$

where $E^n = u^n - U^n$ with u^n, U^n the exact and numerical solutions of (10), respectively, at $t = n\Delta t$.

Given $A \in \mathrm{I\!R}^{N\times N}$ denote the spectral radius of A by $\rho(A)$ and the 2-norm of the matrix A by $||A||$. We recall that

$$||A|| = \rho(A) \quad \text{if} \quad A \in \mathrm{I\!R}^{N\times N} \quad \text{is normal.}$$

It is well known that for any $A \in \mathrm{I\!R}^{N\times N}$

$$\rho(A) \leq ||A||,$$

and that

$$A^m \to 0 \quad \text{as} \quad m \to \infty \quad \text{if and only if} \quad \rho(A) \leq 1,$$

where the eigenvalues on the unit circle must be simple.

A simple criterion for regulating the error growth governed by (11) is given by

$$\rho(A) \leq 1. \tag{12}$$

When the matrix A is not normal the spectral radius gives no indication of the magnitude of E^n for finite n. In this case a condition of the form $\rho(A) \leq 1$ guarantees eventual decay of the solution, but does not control the intermediate growth of the solution. Then, it is easy to understand that the condition (12) is a necessary condition for stability but not always sufficient.

Spectral stability: Let A be the matrix iteration of a numerical method, then the spectral condition is given by

$$\rho(A) \leq 1.$$

3.3. The Lax analysis. The Lax stability is in some way connected with the spectral stability. The spectral stability ensures that $||E^n|| \to 0$ when $n \to \infty$. Practical computations are however, performed at finite values of n, and the spectral condition does not ensure that the norm $||A^n||$ does not become large at finite values of n before decaying as n goes to infinity. In order to guarantee that $||A^n||$ does not become too large we require $||A^n||$ to remain uniformly bounded for all values of n.

Lax stability: In order for all $\mathbf{U}^n$ to remain bounded and the scheme, defined by the operator A, to remain stable the infinite set of operators A^n has to be uniformly bounded. That is, a constant K exists, such that,

$$||A^n|| \leq K, \quad 0 < \Delta t < \tau, \quad 0 \leq n\Delta t \leq T,$$

for fixed values of τ and T and for all n. This condition implies the definition of some norm in the considered functional space.

4. Main results

We start this section by presenting the results for problems with periodic boundary conditions before showing the results for the case we have numerical boundary conditions additionally to the physical boundaries. We assume the norm $|| \cdot ||$ of vectors and matrices is the 2-norm.

4.1. Periodic boundary conditions. Consider the difference scheme

$$U_j^{n+1} = QU_j^n, \qquad j = r, r+1, \ldots \tag{13}$$

$$Q = \sum_{j=-r}^{p} a_j E^j, \quad E^a U_j^n = U_{j+a}^n, \tag{14}$$

and assume we have periodic boundary conditions. The matricial form of the numerical scheme is given by

$$\mathbf{U}^{n+1} = Q\mathbf{U}^n \tag{15}$$

for $\mathbf{U}^n = [U_0^n, U_1^n, \ldots, U_{N-1}^n]^T$ and where Q is a circulant $N \times N$ matrix given by

$$Q=\begin{bmatrix}
a_0 & & \dots & a_p & 0 & \dots & & 0 & a_{-r} & \dots & & a_{-1}\\
\vdots & & \ddots & & & \ddots & & & \ddots & \ddots & & \vdots\\
a_{-r+1} & \dots & a_0 & \dots & & a_p & & 0 & \dots & 0 & & a_{-r}\\
a_{-r} & \dots & & a_0 & & \dots & a_p & & 0 & \dots & & 0\\
\vdots & & \ddots & & \ddots & \ddots & & & \ddots & \ddots & & \vdots\\
0 & \dots & 0 & a_{-r} & & \dots & & a_0 & \dots & & a_p & 0\\
0 & & & & a_{-r} & & & & a_0 & & & a_p\\
a_p & & & & & & & & & & & a_{p-1}\\
\vdots & \ddots & & & \ddots & \ddots & & & \ddots & \ddots & & \vdots\\
a_1 & a_p & 0 & & \dots & 0 & & a_{-r} & & & & a_0
\end{bmatrix} \tag{16}$$

Theorem 1. *The finite difference scheme (13)-(14) is spectral stable if and only if is von Neumann stable.*

Proof: Let us assume that our methods have the matricial form (15)-(16). The eigenvalues of the matrix Q, in (16), are [1] given by

$$\lambda_m=\sum_{k=0}^{p}a_k e^{i2\pi mk/N}+\sum_{k=N-r}^{N-1}a_{k-N}e^{i2\pi mk/N},\quad m=0,\dots,N-1 \tag{17}$$

with associated eigenvectors

$$\psi_m = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 \\ e^{i2\pi m/N} \\ e^{i2\pi m2/N} \\ \vdots \\ e^{i2\pi m(N-1)/N} \end{bmatrix} \tag{18}$$

If now we consider the von Neumann analysis for the finite difference scheme represented by the matrix Q, we need to substitute in (13) the Fourier components

$$\kappa_m^n e^{ijm2\pi/N}, \quad m = 0, \ldots, N-1$$

usually represented by

$$\kappa_m^n e^{ij\theta_m}, \quad m = 0, \ldots, N-1,$$

where $\theta_m = m2\pi/N$ represents the phase angle and covers the domain $[0, 2\pi)$ in steps of $2\pi/N$.

For

$$\mathbf{U}^{n+1} = Q\mathbf{U}^n \quad \text{with} \quad \mathbf{U}^n = [U_0^n, U_1^n, \ldots, U_{N-1}^n]^T$$

we get

$$\kappa_m \begin{bmatrix} 1 \\ e^{i2\pi m/N} \\ e^{i2\pi m2/N} \\ \vdots \\ e^{i2\pi m(N-1)/N} \end{bmatrix} = Q \begin{bmatrix} 1 \\ e^{i2\pi m/N} \\ e^{i2\pi m2/N} \\ \vdots \\ e^{i2\pi m(N-1)/N} \end{bmatrix} \tag{19}$$

We have then the same eigenvectors that are associated with the circulant matrix and therefore $\kappa_m = \lambda_m$. ■

Theorem 2. *The finite difference scheme (13)-(14) is von Neumann stable if and only if is Lax stable.*

Proof: A circulant matrix is a normal matrix (see for instance [1]). If the matrix is normal then $||Q|| = \rho(Q)$, where $||\cdot||$ is the 2-norm. Therefore if the finite difference scheme is von Neumann stable $||Q|| \leq 1$ since from the previous theorem $\rho(Q) \leq 1$. Also $||Q^n|| \leq ||Q||^n \leq 1$ and we conclude the scheme is Lax stable.

Let us now assume the interior scheme is not von Neumann stable.

We define the **Discrete Fourier Transform**:

$$\hat{U}(\theta) = \frac{1}{N}\sum_{j=0}^{N-1} U_j e^{ij\theta},$$

where $0 \leq \theta < 2\pi$, and $\theta = 2k\pi/N,\ k = 0, \ldots, N-1$.

The **Rayleigh Energy Theorem** (Discrete Parseval's relation) says [7]

$$\sum_{j=0}^{N-1} |U_j|^2 = N \sum_{k=0}^{N-1} |\hat{U}(\theta_k)|^2.$$

Assuming that the scheme (13) is not von Neumann stable, therefore the von Neumann condition is not satisfied. Suppose that for each $K > 0$, exists a θ_K, $0 \leq \theta_K < 2\pi$ such that

$$|\kappa(\theta_K)| > 1 + K\Delta t,$$

where $\kappa(\theta) = \sum_{k=-r}^{p} a_k e^{ik\theta}$. This proof follows an idea presented in Sod [11].

We have that $\kappa(\theta)$ is a continuous function of θ. Therefore there exists an interval I_K such that $\theta \in I_K$ and $|\kappa(\theta)| > 1 + K\Delta t,\ \forall \theta \in I_K$.

Consider the initial data u_j^0 so that the discrete Fourier transform of U_j^0, $\hat{U}^0(\theta)$ is 0 outside I_K. Thus, for $0 \leq \theta < 2\pi$ with $\hat{U}^0(\theta) \neq 0$, $|\kappa(\theta)| > 1 + K\Delta t$. We have that

$$\hat{U}^n(\theta) = \hat{QU}^{n-1}(\theta) = \kappa(\theta)\hat{U}^{n-1}(\theta).$$

Therefore, it follows that

$$\hat{U}^n(\theta) = \kappa(\theta)\hat{U}^{n-1}(\theta) = \ldots = \kappa^n(\theta)\hat{U}^0(\theta).$$

Then

$$\hat{U}^n(\theta) > (1 + K\Delta t)^n \hat{U}^0(\theta).$$

Now using the Discrete Parseval's relation we have

$$\begin{aligned} ||U^n||^2 &= \sum_{j=0}^{N-1} |U_j^n|^2 \\ &= N \sum_{k=0}^{N-1} |\hat{U}^n(\theta_k)|^2 \end{aligned}$$

$$
\begin{aligned}
&= N\sum_{k=0}^{N-1}|\kappa(\theta_k)|^{2n}|\hat{U}^0(\theta_k)|^2 \\
&> (1+K\Delta t)^{2n}N\sum_{k=0}^{N-1}|\hat{U}^0(\theta_k)|^2 \\
&= (1+K\Delta t)^{2n}||U^0||^2.
\end{aligned}
$$

We have $U^n = Q^n U^0$ and then

$$||Q^nU^0||^2 > (1+K\Delta t)^{2n}||U^0||^2$$

or

$$\frac{||Q^nU^0||^2}{||U^0||^2} > (1+K\Delta t)^{2n}.$$

But

$$
\begin{aligned}
||Q^n||^2 &= \max_{||U||\neq 0}\frac{||Q^nU||^2}{||U||^2} \\
&> \frac{||Q^nU^0||^2}{||U^0||^2} \\
&> (1+K\Delta t)^{2n}.
\end{aligned}
$$

Finally we have

$$||Q^n||^2 > (1+K\Delta t)^{2n}, \text{ for all } n.$$

This suggests that when the scheme is not von Neumann stable the operator Q^n follows the rise of the amplification factor κ^n. ■

4.2. Numerical boundary conditions. Suppose we have approximations U_j^n as described in section 2, given by the difference scheme

$$U_j^{n+1} = QU_j^n, \qquad j = r, r+1, \ldots \tag{20}$$

$$Q = \sum_{j=-r}^{p} a_j E^j, \quad E^a U_j^n = U_{j+a}^n \tag{21}$$

and the numerical boundary conditions

$$U_\beta^{n+1} = \sum_{j=0}^{q} b_{\beta j} U_j^n \quad \beta = 0, 1, \ldots, r-1. \tag{22}$$

We consider $p \leq q$. Usually $q = p + r$. Also in what follows $N > p + r$. Assume that the matricial form of the scheme is

$$\mathbf{U}^{n+1} = A\mathbf{U}^n, \tag{23}$$

and the iterative matrix A is now given by

$$A = \begin{bmatrix} B_{r\times N} \\ Q_{(N-(r+p))\times N} \\ F_{p\times N} \end{bmatrix} \tag{24}$$

where B contains the part of the boundary conditions and Q the part of the interior scheme, that is

$$U_\beta^{n+1} = B_\beta U^n, \quad \beta = 0, \ldots, r-1 \qquad\qquad U_j^{n+1} = Q_j U^n, \quad j = r, \ldots, N-(r+p).$$

Note that B_β represents the β-row and Q_j represents the j-row. The last block matrix F contains the lines that represent the discrete points computed using (20) and takes in consideration we are assuming the solution $u(x,t)$ is equal to zero as x goes to infinity, that is, we assume that

$$U_N^n = U_{N+1}^n = \ldots = U_{N+p-1}^n = 0.$$

Note that in most applications r and p takes the values 1,2 or 3.

Explicitly the matrix $B_{r\times N}$ is given by

$$B_{r\times N} = \begin{bmatrix} b_{00} & b_{01} & \ldots & b_{0q} & 0 & \ldots & 0 \\ b_{10} & b_{11} & \ldots & b_{1q} & 0 & \ldots & 0 \\ & & & & & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & & & & \\ b_{r-10} & b_{r-11} & \ldots & b_{r-1q} & 0 & \ldots & 0 \end{bmatrix}$$

$Q_{(N-(r+p))\times N}$ is of the form

$$\begin{bmatrix} a_{-r} & a_{-r+1} & \ldots & \ldots & a_0 & \ldots & \ldots & a_{p-1} & a_p & 0 & \ldots & \ldots & 0 \\ 0 & a_{-r} & a_{-r+1} & \ldots & \ldots & a_0 & \ldots & \ldots & a_{p-1} & a_p & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & & \ddots & & & \ddots & \ddots & \ddots & & \vdots \\ 0 & \ldots & 0 & a_{-r} & a_{-r+1} & \ldots & \ldots & a_0 & \ldots & a_{p-1} & a_p & & 0 \\ 0 & \ldots & \ldots & 0 & a_{-r} & a_{-r+1} & \ldots & \ldots & a_0 & \ldots & & a_{p-1} & a_p \end{bmatrix}$$

and the matrix $F_{p\times N}$ is

$$\begin{bmatrix}
0 & \dots & \dots & 0 & a_{-r} & a_{-r+1} & \dots & a_{-1} & a_0 & \dots & \dots & \dots & a_{p-1} \\
0 & \dots & \dots & \dots & 0 & a_{-r} & a_{-r+1} & \dots & a_{-1} & a_0 & \dots & \dots & a_{p-2} \\
\vdots & & & & & \ddots & \ddots & \ddots & & \ddots & \ddots & & \vdots \\
0 & \dots & \dots & \dots & \dots & \dots & 0 & a_{-r} & a_{-r+1} & \dots & a_{-1} & a_0 & a_1 \\
0 & \dots & \dots & \dots & \dots & \dots & \dots & 0 & a_{-r} & a_{-r+1} & \dots & a_{-1} & a_0
\end{bmatrix}$$

In what follows we prove properties that relates Lax stability with spectral and von Neumann stability.

Theorem 3. *If the numerical method (20)–(22) is Lax stable then is spectral stable.*

Proof: If the scheme is Lax stable then A is power-bounded, that is, $||A^n|| \leq K$ for $n = 1, 2, \ldots$. We know that $\rho(A) \leq ||A||$. Then

$$\rho(A^n) \leq ||A^n|| \leq K, \quad n = 1, 2, \ldots$$

But $\rho(A^n) = \rho^n(A)$, hence $\rho^n(A) \leq K, \quad n = 1, \ldots$ and therefore $\rho(A) \leq K^{1/n}$. Then $\lim_{n\to\infty} K^{1/n} = 1$. ■

Theorem 4. *If the interior scheme is von Neumann stable then*

$$||A^n|| \leq e^{n||L||}, \quad \textit{where} \quad L = A - Q,$$

for A and Q $N \times N$ matrices given by (24) and (16) respectively.

Proof: We prove by induction that

$$||(Q + L)^n|| \leq e^{n||L||}.$$

For $n = 1$ is true because

$$||Q + L|| \leq ||Q|| + ||L|| \leq 1 + ||L|| \leq e^{||L||}.$$

It is easy to prove (by induction) that

$$(Q + L)^n = Q^n + \sum_{k=0}^{n-1} Q^{n-k-1} L (Q + L)^k, \qquad n = 1, 2, \ldots$$

Therefore

$$||(Q+L)^n|| \leq ||Q^n|| + \sum_{k=0}^{n-1} ||Q^{n-k-1}||\ ||L||\ ||(Q+L)^k||.$$

By assumption, the interior scheme is von Neumann stable. Therefore, $||Q^n|| \leq 1$ and $||Q^{n-k-1}|| \leq 1$, for $k = 0, \ldots, n$. By the inductive hypothesis, $||(Q+L)^k||$ is bounded, that is, $||(Q+L)^k|| \leq e^{k||L||}$.

We have

$$\begin{aligned} ||(Q+L)^n|| &\leq 1 + \sum_{k=0}^{n-1} ||L|| e^{k||L||} \\ &\leq 1 + ||L|| \sum_{k=0}^{n-1} e^{k||L||}. \end{aligned}$$

It is easy to check that

$$e^{n||L||} \geq 1 + ||L|| \sum_{k=0}^{n-1} e^{k||L||}.$$

Therefore

$$||(Q+L)^n|| \leq e^{n||L||}.$$

■

Theorem 5. *Assume the interior scheme is von Neumann stable:*

(1) *If $A = Q + L$ such that $||L|| \leq \Delta t C$ then A is power-bounded.*
(2) *If $\rho(A) \leq 1$ then A is power-bounded.*

Proof: (1) If $A = Q + L$ such that $||L|| \leq \Delta t C$ then $||(Q+L)^n|| \leq e^{TC}$ for $T = n\Delta t$ and we can conclude that A is power-bounded, that is, $||A^n|| \leq K$.

(2) If $\rho(A) \leq 1$, since [14]

$$\lim_{k\to\infty} ||A^k||^{1/k} = \rho(A),$$

there is a k_0 such that $||A^k|| \leq 1$, for all $k \geq k_0$. Therefore, from Theorem 4.4, follows

$$||A^n|| \leq e^{k_0||L||}, \quad \text{for all} \quad n.$$

■

We have that von Neumann and spectral stability are sufficient conditions for $||A^n|| \leq K$ for all n.

How can we prove that $||A_N^n|| \leq C$ for all N, that is, $\lim_{N\to\infty} ||A_N^n|| \leq C$? Note that N is the size of the matrix A and is directly related with the space step Δx, such that, $\Delta x \to 0$ as $N \to \infty$.

If we have $\rho(A_N) \leq 1,$ for all N, it is guaranteed that $||A_N^n|| \leq K$, for all N.

If $A_N = Q_N + \Delta t L_N$ then $||A_N^n|| \leq e^{n\Delta t ||L_N||} = e^{T||L_N||}$. Let $||L_N|| = M$ for all $N > p + r$, where M is a constant. Then we have Lax stability and in this case $||A_N^n|| \leq e^{TM}$, for all n and $N > p + r$. The value of $||L_N||$ is in fact the same as $N \to \infty$ as we shall prove in the next theorem.

Theorem 6. *For L_N such that $L_N = A_N - Q_N$ we have*

$$||L_N|| = ||L_{N+1}||, \quad \textit{for all} \quad N > p + r.$$

Proof: Note that

$$L_N = A_N - Q_N = \begin{bmatrix} B_{r\times N} - Q_{r\times N} \\ O_{(N-(r+p))\times N} \\ F_{p\times N} - Q_{p\times N} \end{bmatrix}$$

where $O_{(N-(r+p))\times N}$ is the null matrix or zero matrix. The block matrix $Q_{r\times N}$ represent the first r lines of the matrix Q given by (16) and $Q_{p\times N}$ the last p lines.

Also,

$$L_{N+1} = A_{N+1} - Q_{N+1} = \begin{bmatrix} B_{r\times N+1} - Q_{r\times N+1} \\ O_{(N+1-(r+p))\times N+1} \\ F_{p\times N+1} - Q_{p\times N+1} \end{bmatrix}$$

The difference between $B_{r\times N} - Q_{r\times N}$ and $B_{r\times N+1} - Q_{r\times N+1}$ is that the latter has an additional column of zeros. The same difference occur between the matrices $F_{p\times N} - Q_{p\times N}$ and $F_{p\times N+1} - Q_{p\times N+1}$. Therefore it is easy to conclude that $||L_N|| = ||L_{N+1}||$. Note that usually in practical applications the block matrices B and F does not have more than three lines. ■

5. Numerical examples

In this section two examples are given: a classical example with physical boundary conditions and a more complex example with numerical boundary conditions.

5.1. A classical example. This example is a classical example showing that the eigenvalues of the matrix iteration does not give a sufficient and necessary condition for stability.

Suppose we consider the problem (1)-(3), where

$$\mathcal{L}u(x,t) = \frac{\partial u}{\partial x}, \quad x \in (0,1)$$

with an initial condition and boundary conditions given, where $u(1,t) = 0$.

We discretize the equation by

$$U_j^{n+1} = U_j^n + \sigma(U_{j+1}^n - U_j^n), \qquad \text{where} \quad \sigma = \frac{\Delta t}{\Delta x}.$$

Then $\mathbf{U}^{n+1} = A\mathbf{U}^n$, where $\mathbf{U}^n = [U_0^n, \ldots, U_{N-1}^n]^T$ and

$$A = \begin{bmatrix} 1-\sigma & \sigma & & \\ & 1-\sigma & \sigma & \\ \vdots & \vdots & \vdots & \vdots \\ & & & \sigma \\ & & & 1-\sigma \end{bmatrix}$$

The eigenvalues are $\lambda_i = 1 - \sigma,\ i = 0, \ldots, N-1$, and $\rho(A) < 1$ if and only if $\sigma < 2$.

On the other hand the von Neumann stability analysis for the interior scheme says that the amplification factor is given by

$$\kappa = 1 + \sigma(e^{ik\Delta x} - 1).$$

Therefore, for $\theta = k\Delta x$,

$$\kappa = 1 - \sigma + \sigma\cos\theta + i\sigma\sin\theta.$$

We have

$$|\kappa|^2 = 1 - 2\sigma(1-\sigma)(1-\cos\theta).$$

For that reason we have that $|\kappa|^2 \leq 1$ if and only if $\sigma(1-\sigma)(1-\cos\theta) \leq 1$ that is $\sigma \leq 1$.

In conclusion we have

$$\textit{von Neumann} : 0 < \sigma \leq 1 \qquad \textit{Spectral} : 0 < \sigma < 2$$

Therefore, the spectral radius suggests stability for $1 < \sigma < 2$, although the scheme is unstable for these values.

In this case it is easy to check that as N becomes larger the eigenvalues of A are the same. We can also conclude the stability condition, $0 < \sigma \leq 1$, is the intersection of both conditions and is the same as the von Neumann condition.

5.2. An example with numerical boundary conditions. Consider the one-dimensional problem with constant velocity V in the positive x direction and constant diffusion with coefficient $D > 0$:

$$\frac{\partial u}{\partial t} + V\frac{\partial u}{\partial x} = D\frac{\partial^2 u}{\partial x^2}, \quad t > 0, \quad x > 0 \tag{25}$$

with the initial condition

$$u(x,0) = f(x), \quad x \geq 0 \tag{26}$$

and the boundary conditions

$$u(0,t) = g(t), \quad t > 0 \quad u(x,t) \to 0, \quad x \to \infty. \tag{27}$$

We discretize at the interior points $j = 2, \dots, N-1$, using the scheme introduced by Leonard [6],

$$U_j^{n+1} = \{1 - \nu\Delta_0 + (\frac{1}{2}\nu^2 + \mu)\delta^2 + \frac{1}{6}\nu(1 - \nu^2 - 6\mu)\delta^2\Delta_-\}U_j^n \tag{28}$$

that interpolates the mesh points U_{j-2}^n, U_{j-1}^n, U_j^n, U_{j+1}^n and where

$$\nu = V\frac{\Delta t}{\Delta x} \quad \text{and} \quad \mu = D\frac{\Delta t}{\Delta x^2}.$$

The operators are the usual central, backward and second difference operators

$$\Delta_0 U_j := \frac{1}{2}(U_{j+1} - U_{j-1}), \quad \Delta_- U_j := U_j - U_{j-1}, \quad \delta^2 U_j := U_{j+1} - 2U_j + U_{j-1}.$$

At $j = 1$, we use a numerical boundary condition suggested in [12], that interpolates the mesh points U_{j-1}^n, U_j^n, U_{j+1}^n and U_{j+2}^n,

$$U_j^{n+1} = \{1 - \nu\Delta_0 + (\frac{1}{2}\nu^2 + \mu)\delta^2 + \frac{1}{6}\nu(1 - \nu^2 - 6\mu)\delta^2\Delta_+\}U_j^n, \tag{29}$$

where $\Delta_+ U_j := U_{j+1} - U_j$.

In Figure 1 we show the von Neumann condition for the scheme (28), considering we have a Cauchy problem, that is, periodic boundary condition. We also plot the spectral condition for the matrix iteration A that takes in consideration the interior scheme (28) and also the numerical boundary condition (29). The matrix iteration A is given by

$$A = \begin{bmatrix} a^* & b^* & c^* & d^* & 0 & 0 & \dots & \dots & 0 & 0 & 0 & 0 \\ a & b & c & d & 0 & 0 & \dots & \dots & 0 & 0 & 0 & 0 \\ 0 & a & b & c & d & 0 & \dots & \dots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & \dots & a & b & c & d \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & \dots & 0 & a & b & c \end{bmatrix}$$

where

$$\begin{array}{ll} a^* = c_1 + c_2 - c_3 & a = -c_3 \\ b^* = 1 - 2c_2 + 3c_3 & b = c_1 + c_2 + 3c_3 \\ c^* = -c_1 + c_2 - 3c_3 & c = 1 - 2c_2 - 3c_3 \\ d^* = c_3 & d = -c_1 + c_2 + c_3 \end{array}$$

for

$$c_1 = \frac{\nu}{2} \qquad c_2 = \frac{\nu^2}{2} + \mu \qquad c_3 = \frac{\nu}{6}(1 - \nu^2 - 6\mu).$$

Note that here $r = p = 1$.

It was proved in [13], for this example, using normal mode analysis that the stable region is given by the region in Fig. 1(b). This region is the same as the region shown in Fig. 1(a) that represents the intersection of the von Neumann condition and spectral condition.

In Figure 2 we plot $||A^n||$ as $n \to \infty$. For $\mu = 0.001$ and $\nu = 1.02$ we are in the spectral stable region but von Neumann unstable and we observe in Figure 2 (a) that the $\max_n ||A^n||$ is increasing as the size N of the matrix A increases. On the other hand, for $\mu = 0.001$ and $\nu = 0.1$ we are in the region where the scheme is spectral and von Neumann stable and it is shown in Figure 2 (b) that the maximum value is always the same as N increases.

The spectral condition guarantees eventual decay of the solution but does not control the intermediate growth of the solution, being this guaranteed by the von Neumann condition, that is, the von Neumann condition seems to assure that as N becomes larger there will not be a strong effect in the value $\max_n ||A^n||$.

In Figure 3 we plot the amplification factor for the scheme (28) given by

$$\begin{aligned} |\kappa(s)| &= 1 - 8\mu s + 4[(2\mu + \nu^2)^2 - \nu^2 + 2\alpha(1 - 2\nu)]s^2 \\ &\quad -16\alpha(2\mu + \nu^2 - \nu - \alpha)s^3, \end{aligned}$$

where $\alpha = 2\nu\mu - (\nu/3)(1-\nu^2)$ and $s = \sin^2(\xi/2)$. Also if we observe Figure 4 we see that $||A^n||$ has a behavior very close to $|\kappa(\pi)|^n$ where $\xi = \pi$ is the value for which the amplification factor takes the biggest value.

It was pointed out in the end of section 4.1 that the norm of the powers of the matrix iteration, namely $||Q^n||$, follows the rise of the amplification factor. Although, in this example, A is not a normal matrix and the first line represents the numerical boundary condition, the norm $||A^n||$ seems to follow also the rise of the amplification factor associated to the interior scheme.

6. Conclusion

The aim of this paper is to present some properties that relates the von Neumann analysis, the spectral analysis and the Lax theory in order to give a new way to verify when a finite difference scheme is stable even if the matrix iteration is not normal. To control the growing of the norm of the power matrices is usually more difficult than to verify the von Neumann condition for the interior scheme and to calculate the maximum eigenvalue of the matrix iteration.

References

[1] P. J. Davis, *Circulant matrices*, Wiley-Interscience: New York, 1979.

[2] Godunov, S.K. and V.S. Ryabenkii, *Spectral criteria for the stability of boundary problems for non-self-adjoint difference equations* Uspekhi Mat. Nauk. **18** (1963), 1-12.

[3] Godunov, S.K. and V.S. Ryabenkii, *The theory of difference schemes*, Amsterdam: North-Holland, 1964

[4] B. Gustafsson, H-T. Kreiss and J. Oliger, *The time dependent problems and difference methods*, Wiley-Interscience: New York, 1995.

[5] C. Hirsch, *Numerical computation of internal and external flows*, Wiley–Interscience: New York, 1988.

[6] B. P. Leonard, *A stable and accurate convective modeling procedure based on quadratic upstream interpolation*, Computer Methods in Applied Mechanics and Engineering **19** (1979) 59-98.

[7] D. W. Kammler, *A first course in Fourier Analysis* Prentice-Hall: New Jersey, 2000.

[8] K. W. Morton and D. F. Mayers, *Numerical solution of Partial Differential Equations*, Cambridge University Press: Cambridge, 1996.

[9] R.D. Richtmyer and K.W. Morton, *Difference methods for initial-value problems.* Wiley-Interscience: New York, 1967.

[10] G. D. Smith, *Numerical Solution of Partial Differential Equations: Finite Difference Methods*, Oxford University Press: Oxford, 1985.

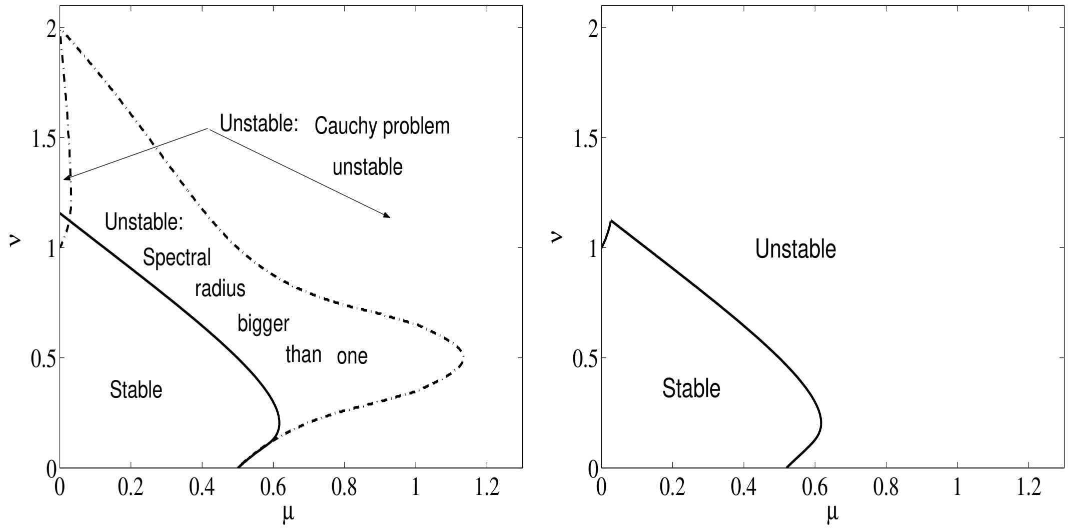
[11] G. A. Sod, *Numerical methods in fluid dynamics: initial and initial boundary-value problems*, Cambridge University Press: Cambridge, 1988.

[12] E. Sousa and I. J. Sobey, *On the influence of numerical boundary conditions* Applied Numerical Mathematics, **41**, (2002) 325-344.

[13] E. Sousa, *Stability analysis of difference methods for parabolic initial value problems*, Journal of Scientific Computing **26** (2006), 45-66.

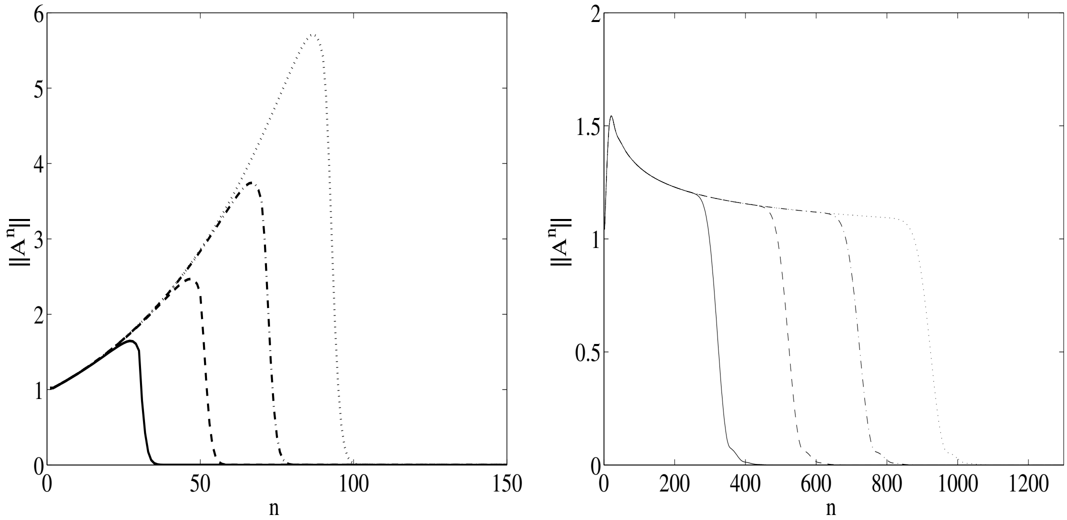
[14] L. N. Trefethen and M. Embree, *Spectra and Pseudospectra*, Princeton University Press: New Jersey, 2005.

Ercília Sousa
Departamento de Matemática, Universidade de Coimbra, Portugal



(a) (b)

FIGURE 1. Stability region



(a) (b)

FIGURE 2. $N = 30$, $(-), N = 50$, $(--)$, $N = 70, (- \cdot -)$, $N = 90(\cdot \cdot \cdot)$ (a)$\mu = 0.001$, $\nu = 1.02$ $\rho_{30}(A) = 0.3031$, $\rho_{50}(A) = 0.3053$, $\rho_{70}(A) = 0.5448$, $\rho_{90}(A) = 0.5671$ (b) $\mu = 0.001$, $\nu = 0.1$ $\rho_{30}(A) = 0.9608$, $\rho_{50}(A) = 0.9608$, $\rho_{70}(A) = 0.9608$, $\rho_{90}(A) = 0.9642$

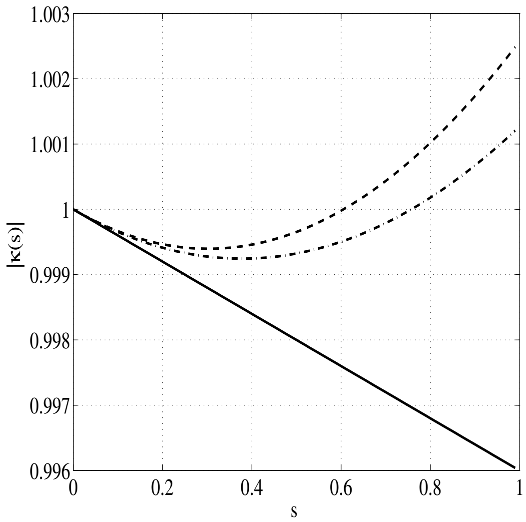


FIGURE 3. Amplification factor $\kappa(s)$, $0 \leq s \leq 1$ for $\mu = 0.001$ and $\nu = 1$ (-−), $\nu = 1.005(- \cdot -)\nu = 1.02(--)$

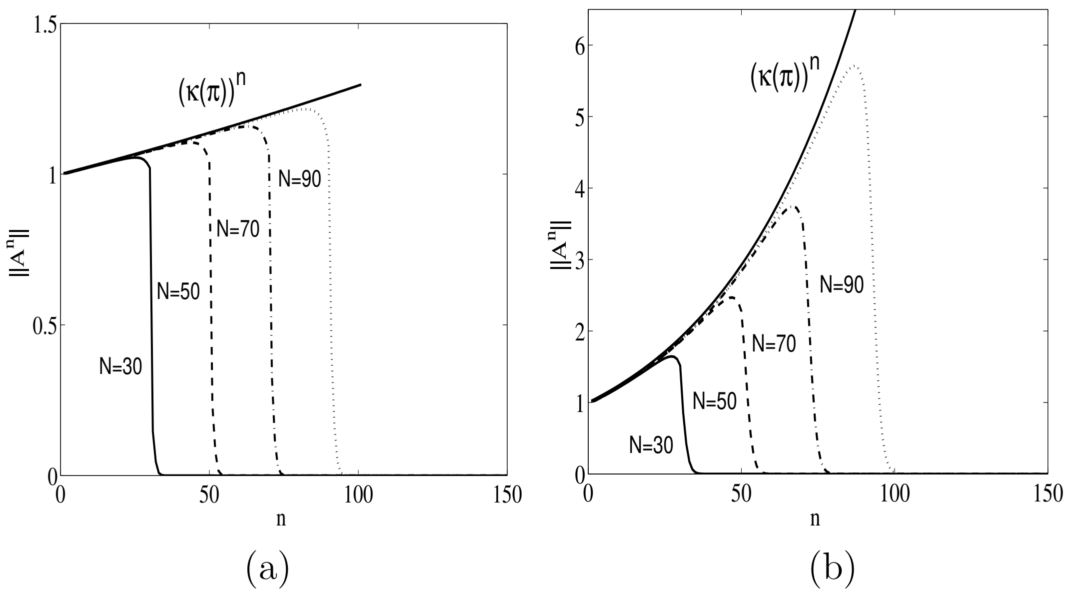


FIGURE 4. $N = 30,\ (-), N = 50,\ (--),\ N = 70, (- \cdot -),\ N = 90(\cdots)$ $\mu = 0.001$, (a) $\nu = 1.005$ (b) $\nu = 1.02$