Matrix Stability Analysis

Consider the initial boundary value problem (IBVP)

$$u_t = \sigma u_{xx}, \qquad 0 < x < 1, t > 0 \tag{1}$$
$$u(0,t) = g(t), \qquad u(1,t) = h(t) \tag{2}$$
$$u(x,0) = f(x) \tag{3}$$

Equation (1) can be written as

$$u_t = Lu, \tag{4}$$

where L is a linear differential operator.

We have seen three different numerical schemes to approximate the solution of IBVP (1)-(3). They are

1. **Forward in time–Centered in space**

$$U_i^{n+1} = rU_{i-1}^n + (1-2r)U_i^n + rU_{i+1}^n, \quad i = 1, \dots m, \tag{5}$$

where $r = \sigma\Delta t/\Delta x^2$. This scheme is $\mathcal{O}(\Delta t) + \mathcal{O}(\Delta x^2)$. The linear system that results from (5) can be represented by

$$\mathbf{U}^{n+1} = L_\Delta^F \mathbf{U}^n + \begin{bmatrix} rg^n \\ 0 \\ \cdot \\ \cdot \\ rh^n \end{bmatrix}. \tag{6}$$

2. **Backward in time–Centered in space**

$$-rU_{i-1}^{n+1} + (1+2r)U_i^{n+1} - rU_{i+1}^{n+1} = U_i^n, \quad i = 1, \dots m \tag{7}$$

This scheme is $\mathcal{O}(\Delta t) + \mathcal{O}(\Delta x^2)$ The linear system that results from (7) can be represented by

$$L_\Delta^B \mathbf{U}^{n+1} = \mathbf{U}^n + \begin{bmatrix} rg^{n+1} \\ 0 \\ \cdot \\ \cdot \\ rh^{n+1} \end{bmatrix}. \tag{8}$$

3. **Crank–Nicholson**

$$\frac{-r}{2}U_{i-1}^{n+1} + (1+r)U_i^{n+1} - \frac{r}{2}U_{i+1}^{n+1} = \frac{r}{2}U_{i-1}^{n} + (1-r)U_i^{n} + \frac{r}{2}U_{i+1}^{n}, \quad i = 1, \dots m \quad (9)$$

This scheme is $\mathcal{O}(\Delta t^2) + \mathcal{O}(\Delta x^2)$ The linear system that results from (9) can be represented by

$$L_\Delta^S \mathbf{U}^{n+1} = L_\Delta^G \mathbf{U}^n + \begin{bmatrix} r/2\, g^n + r/2\, g^{n+1} \\ 0 \\ \cdot \\ \cdot \\ r/2\, h^n + r/2\, h^{n+1} \end{bmatrix}. \quad (10)$$

0.1 Definition 1: Stability of Linear Finite Difference Methods

A linear finite difference method (FDM) of the form

$$\mathbf{U}^{n+1} = L_\Delta \mathbf{U}^n \quad (11)$$

corresponding to an IBVP of (4) (such as (1)-(3)) is stable if there exists $C > 0$, *independent of the mesh spacing and the initial data,* such that

$$||\mathbf{U}^n|| \leq C||\mathbf{U}^0||, \qquad n \to \infty, \quad \Delta t \to 0, \quad \Delta x \to 0, \quad n\Delta t \leq T \quad (12)$$

0.2 Theorem 1: Equivalent Condition

The FDM (11) is stable if and only if there exists a constant $C > 0$ independent of Δx and Δt such that

$$||(L_\Delta)^n|| \leq C, \qquad n \to \infty, \quad \Delta t \to 0, \quad \Delta x \to 0, \quad n\Delta t \leq T \quad (13)$$

Remark: Notice that C may be greater than 1.

Proof.
Notice that

$$\mathbf{U}^n = L_\Delta \mathbf{U}^{n-1} = L_\Delta \left(L_\Delta \mathbf{U}^{n-2}\right) = L_\Delta^2 \mathbf{U}^{n-2} = \cdots = L_\Delta^n \mathbf{U}^0$$

Therefore, for an arbitrary $\mathbf{U}^0 \neq \mathbf{0}$

$$||\mathbf{U}^n|| \leq C||\mathbf{U}^0|| \iff ||L_\Delta^n \mathbf{U}^0|| \leq C||\mathbf{U}^0|| \iff \frac{||L_\Delta^n \mathbf{U}^0||}{||\mathbf{U}^0||} \leq C \iff ||\left(L_\Delta\right)^n|| \leq C \quad (14)$$

0.3 Corollary 1: Practical Condition

If the discrete operator L_Δ of the FDM (11) satisfies

$$||L_\Delta|| \leq 1,$$

then the FDM (11) is stable.

Proof.
Notice that $||L_\Delta^n|| \leq ||L_\Delta||^n$. Therefore, if

$$||L_\Delta|| \leq 1 \Rightarrow ||\left(L_\Delta\right)^n|| \leq ||L_\Delta||^n \leq 1$$

The stability follows from Theorem 1.

Remark: Apply this condition to the explicit FDM FT-CS using the infinity norm.

In fact, if $r \leq 1/2$

$$||L_\Delta||_\infty = r + |1-2r| + r = r + 1 - 2r + r = 1$$

.

0.4 Corollary 2: More General Sufficient Condition

If there is is a $c > 0$ independent of Δx and Δt such that the discrete operator L_Δ of the FDM (11) satisfies

$$||L_\Delta|| \leq 1 + c\Delta t,$$

for $\Delta t < \Delta t^*$, then the FDM (11) is stable.

Proof.
Notice that $n\Delta t \leq T$ and $1 + c\Delta t \leq e^{c\Delta t}$, then $1 + c\Delta t \leq e^{cT/n}$. Therefore,

$$||\left(L_\Delta\right)^n|| \leq ||L_\Delta||^n \leq (1 + c\Delta t)^n \leq e^{cT} = e^{\tilde{c}} = C$$

0.5 Definition 2: Spectral Radius

The spectral radius $\rho(L_\Delta)$ of the FDM matrix L_Δ is the maximum of the absolute value of its eigenvalues. Assuming that λ_i, $i = 1, \ldots N$ are the eigenvalues of L_Δ, then

$$\rho(L_\Delta) = \max_{1 \leq i \leq N} |\lambda_i|$$

0.6 Theorem 2: Relationship Between Spectral Radius and Norm of L_Δ

If $\rho(L_\Delta)$ and $||L_\Delta||$ are the spectral radius and the vector-induced norm of L_Δ then,

$$\rho(L_\Delta) \leq ||L_\Delta||$$

Proof.
For any eigenvector $\mathbf{x}_i$, it holds $||L_\Delta \mathbf{x}_i|| = ||\lambda_i \mathbf{x}_i||$, for $i = 1, 2, \ldots, N$. Therefore,

$$|\lambda_i| = \frac{||L_\Delta \mathbf{x}_i||}{||\mathbf{x}_i||} \leq \max_{\mathbf{x} \neq \mathbf{0}} \frac{||L_\Delta \mathbf{x}||}{||\mathbf{x}||} = ||L_\Delta|| \Rightarrow \rho(L_\Delta) \leq ||L_\Delta||$$

0.7 Corollary 3: Necessary Condition

The condition

$$\rho^n(L_\Delta) \leq C,$$

for a constant $C > 0$ independent of Δx and Δt is a necessary condition for the stability of the FDM (11).

Proof.
Notice that $\rho^n(L_\Delta) = \rho\left((L_\Delta)^n\right) \leq ||\left(L_\Delta\right)^n)||$. Therefore, if $\rho^n(L_\Delta)$ is not bounded then $||\left(L_\Delta\right)^n)||$ is also not bounded and the FDM is not stable.

0.8 Corollary 4: A More Practical Sufficient Condition (special matrices)

If L_Δ of the FDM (11) is symmetric or similar to a symmetric matrix, then

$$\rho(L_\Delta) \leq 1,$$

for any Δx and Δt, is also a sufficient condition for stability in the Euclidean norm.

Proof.
If L_Δ is a symmetric matrix then the Eucledian norm $||L_\Delta||_2 = \sqrt{\rho(L_\Delta L_\Delta^T)} = \rho(L_\Delta)$. Therefore,

$$\rho(L_\Delta) \leq 1 \Rightarrow ||L_\Delta||_2 \leq 1$$

and the stability follows from Corollary 1.

Remark: Apply this condition to show stability of FT-CS and BT-CS FDM for IBVP (1)-(3) with homogeneous boundary conditions.

0.9 Definition 4: Convergence

A finite difference approximation $\mathbf{U}^n$ converges to the solution $\mathbf{u}^n$ (the restriction of the exact solution $u(x, t_n)$ to the mesh) on $0 < t \leq T$ in a particular vector norm if

$$||\mathbf{u^n} - \mathbf{U}^n|| \to 0, \qquad n \to \infty, \qquad \Delta x \to 0, \qquad \Delta t \to 0, \qquad n\Delta t \leq T \tag{15}$$

Why do we want to prove stability for FDM such as (11) approximating certain PDE problems modelled by (4)? The answer to this question is found in the next theorem

0.10 Theorem 3: Lax-Equivalence Theorem

A consistent linear FDM such as (11) is convergent if and only if it is stable.

In many problems of practical interest, we would like to study stability when $t \to \infty$. To analyze stability for these problems, we need an alternative stability definition.

0.11 Definition 3: Absolute Stability

A FDM such as (11) is absolutely stable for a given mesh (of size Δx and Δt) if

$$||\mathbf{U}^n|| \leq ||\mathbf{U}^0||, \qquad n > 0 \tag{16}$$

0.12 Definition 4: Unconditional Stability

A FDM such as (11) is unconditionally stable if it is absolutely stable for all choices of mesh spacing Δx and Δt.

3.4

LAX Equivalence Theorem.

Definition.– The IVP for the first-order (in time) PDE $u_t = L\,u$. (L differential operator) is well-posed if for any time $T \geq 0$, there is a constant C_T such that any solution $u(x,t)$ satisfies

$$\int_{-\infty}^{\infty} |u(x,t)|^2\,dx \leq C_T \int_{-\infty}^{\infty} |u(x,0)|^2\,dx.$$

for $0 \leq t \leq T$.

Theorem.– A Consistent finite difference scheme ($\vec{v}^{n+1} = L_\Delta \vec{v}^n$) for a PDE ($u_t = Lu$) for which the IVP is well-posed is convergent if and only if it is stable.

Proof.– ($\Leftarrow$) Stability $\Rightarrow$ convergence.

Consider the numerical scheme

$$\boxed{\vec{v}^{n+1} = L_\Delta \vec{v}^n} \qquad (1)$$

for example, FT-CS heat conduction

$$\underbrace{\begin{bmatrix} v_1^{n+1} \\ v_2^{n+1} \\ \vdots \\ v_{J-1}^{n+1} \end{bmatrix}}_{=\vec{v}^{n+1}} = \underbrace{\begin{bmatrix} 1-2r & r & 0 & \cdots & 0 \\ r & 1-2r & r & 0 \cdots & 0 \\ & \ddots & \ddots & \ddots & \\ & & r & 1-2r & r \\ & & & r & 1-2r \end{bmatrix}}_{=L_\Delta} * \underbrace{\begin{bmatrix} v_1^n \\ \vdots \\ v_{J-1}^n \end{bmatrix}}_{=\vec{v}^n}$$

Proof.– (Lax Equivalence theorem).

If u_j^n is a solution of the partial diff. equ.

then

$$\vec{u}^{n+1} = L_\Delta \vec{u}^n + \Delta t\, \vec{\tau}^n$$

For example, Heat cond.

$$u_j^{n+1} = r\, u_{j-1}^n + (1-2r)\, u_j^n + r\, u_{j+1}^n + \Delta t\, \tau_j^n$$

where

$$\tau_j^n = -\frac{\Delta t}{2} (u_{tt})_j^{n+\theta} + \frac{\sigma\, \Delta x^2}{12} (u_{xxxx})_{j+\xi}^n$$

or

$$\boxed{\vec{u}^{n+1} = L_\Delta \vec{u}^n + \Delta t\, \vec{\tau}^n.} \qquad (2)$$

The difference of the vector solution $\vec{u}^n$ of the PDE. and the vector solution $\vec{v}^n$ of the discrete approx. is called

$$\vec{e}^n = \vec{u}^n - \vec{v}^n \quad \text{(global discretization error)}$$

Subtracting (1) from (2)

$$\vec{e}^{n+1} = \vec{u}^{n+1} - \vec{v}^{n+1} = L_\Delta (\vec{u}^n - \vec{v}^n) + \Delta t\, \vec{\tau}^n = L_\Delta \vec{e}^n + \Delta t\, \vec{\tau}^n.$$

$$\Rightarrow \boxed{\vec{e}^{n+1} = L_\Delta \vec{e}^n + \Delta t\, \vec{\tau}^n} \qquad (3)$$

If L_Δ is independent of n by iterating on (3)

$$\vec{e}^{\,n} = L_\Delta \vec{e}^{\,n-1} + \Delta t\, \vec{\tau}^{\,n-1} = L_\Delta \left(L_\Delta \vec{e}^{\,n-2} + \Delta t\, \vec{\tau}^{\,n-2} \right) + \Delta t\, \vec{\tau}^{\,n-1}$$

$$= L_\Delta^2 \vec{e}^{\,n-2} + \Delta t \left[L_\Delta \vec{\tau}^{\,n-2} + \vec{\tau}^{\,n-1} \right] =$$

$$= L_\Delta^3 \vec{e}^{\,n-3} + \Delta t \left[L_\Delta^2 \vec{\tau}^{\,n-3} + L_\Delta \vec{\tau}^{\,n-2} + \vec{\tau}^{\,n-1} \right] =$$

$$\cdots\cdots = L_\Delta^n \vec{e}^{\,0} + \Delta t \left[L_\Delta^{n-1} \vec{\tau}^{\,0} + L_\Delta^{n-2} \vec{\tau}^{\,1} + L_\Delta^{n-3} \vec{\tau}^{\,2} + \cdots + \cdots + L_\Delta^{n-n} \vec{\tau}^{\,n-1} \right]$$

or

$$\boxed{\vec{e}^{\,n} = L_\Delta^n \underset{\substack{\| \\ 0,\ \text{if not rounded errors.}}}{\vec{e}^{\,0}} + \Delta t \left[L_\Delta^{n-1} \vec{\tau}^{\,0} + L_\Delta^{n-2} \vec{\tau}^{\,1} + \cdots + L_\Delta \vec{\tau}^{\,n-2} + \vec{\tau}^{\,n-1} \right]}$$

$$\Rightarrow \|\vec{e}^{\,n}\| \le \Delta t \left[\|L_\Delta^{n-1}\| \, \|\vec{\tau}^{\,0}\| + \|L_\Delta^{n-2}\| \, \|\vec{\tau}^{\,1}\| + \cdots + \|\vec{\tau}^{\,n-1}\| \right] \quad (4)$$

Let's choose a time of interest T and an arbitrary $\varepsilon > 0$. Since numerical scheme is consistent there exists $\delta_1 > 0$ such that $\|\vec{\tau}^{\,k}\| \le \varepsilon$ if $\Delta t, \Delta x < \delta_1$ for all k such that $k\Delta t \le T$!

Also, using the hypothesis that the scheme is stable and the previous theorem about stability we conclude that there exist C and $\delta_2>0$ such that

$$\text{if } \Delta x, \Delta t < \delta_2 \Rightarrow \|(L_\Delta)^k\| \leq C \quad \text{for all } k \text{ such that } k\Delta t \leq T.$$

Therefore, choosing $\delta = \min(\delta_1, \delta_2)$, for $\Delta x, \Delta t < \delta$

(4) reduces to

$$\|\hat{e}^n\| \leq \Delta t \left((n-1)C\varepsilon + \varepsilon \right) \begin{cases} \leq \Delta t\, n C\varepsilon, & \text{if } C \geq 1 \\ < \Delta t\, n\varepsilon, & \text{if } C < 1 \end{cases}$$

$\rightarrow \leq (n-1)C\varepsilon + C\varepsilon$

$\underset{C<1}{\rightarrow} < (n-1)C\varepsilon + C\varepsilon = nC\varepsilon < n\varepsilon$

Since $n\Delta t \leq T$

$$\Rightarrow \|\hat{e}^n\| \leq \begin{cases} TC\varepsilon, & C \geq 1 \\ \text{or} & \\ T\varepsilon, & C < 1 \end{cases}$$

In both cases, the scheme converges.