Von Neumann stability analysis

In numerical algorithms for differential equations the concern is the growth of round-off errors and/or initially small fluctuations in initial data which might cause a large deviation of final answers from the exact solution. The method of stability analysis shown next was developed by the mid-twentieth century Hungarian mathematician — and father of the electronic computer — John von Neumann. The von Neuman stability analysis is based on the decomposition of numerical errors of numerical approximations into Fourier series. Fourier series decomposes any periodic function or periodic signal into the sum of a (possibly infinite) set of simple oscillating functions, namely sines and cosines (or, equivalently, complex exponentials). We have chosen complex exponentials to represent errors as they are much easier to work with than real trigonometric functions.

Consider the one-dimensional heat equation

$$\frac{\partial U}{\partial t} = \alpha \frac{\partial^2 U}{\partial x^2} \quad [1]$$

Stability Analysis of FTCS scheme.
Approximate the numerical error. Using the FTCS method the discretized form of equation [1] is:

$$U_j^{n+1} = U_j^n + \lambda\left(U_{j+1}^n - 2U_j^n + U_{j-1}^n\right) \; [2]$$

where $\lambda = \alpha\Delta t/\Delta x^2$ (called the mesh ratio), j is the space index, and n is the time index.

Define the error in the numerical approximation as

$$\epsilon_j^n = U_j^n - u_j^n \quad [3]$$

where u_j^n is the exact solution at grid point (j, n). Both the numerical solution U_j^n and the exact solution u_j^n satisfy equation [2], therefore, the error ϵ_j^n also follows the discretized ODE equation [2]

$$\epsilon_j^{n+1} = \epsilon_j^n + \lambda\left(\epsilon_{j+1}^n - 2\epsilon_j^n + \epsilon_{j-1}^n\right) \quad [4]$$

The equations [2] and [4] show that both the error and the numerical solution have the same growth or decay behavior with respect to time. For linear differential equations with periodic boundary condition, the spatial variation of error may be expanded in a finite Fourier series, in the interval L, as

$$\epsilon(x) = \sum_{m=1}^{M} A_m e^{ik_m x} \quad [5]$$

where the wavenumber $k_m = \frac{\pi m}{L}$ where $m = 1,2, \ldots, M$ and $M = L/\Delta x$, $i = \sqrt{-1}$, and x is the independent space variable. $e^{ik_m x}$ is the complex exponential (recall the Euler's formula: $e^{i\varphi} = \cos\varphi + i\sin\varphi$, where e is the base of natural logarithm and φ is a real number).

The time dependence of the error is included by assuming the amplitude of error A_m is a function of time. Since the error tends to grow or decay exponentially with time, it is reasonable to assume that the amplitude varies exponentially with time; hence

$$\epsilon(x,t) = \epsilon_j^n = \sum_{m=1}^{M} e^{at} e^{ik_m x} \quad [6]$$

where a is a constant.

Since the difference equation for error is linear (the behavior of each term of the series is the same as the series itself), it is enough to consider the growth of error of a typical term:

$$\epsilon_m(x,t) = \epsilon_j^n = e^{at} e^{ik_m x} \quad [7]$$

The stability characteristics can be studied using just this form for the error with no loss in generality.

Define a Stability Criterion. The goal is to show that the error incurred by a particular numerical scheme doesn't grow in the evolving steps of time. Define the amplification factor

$$G \equiv \frac{\epsilon_j^{n+1}}{\epsilon_j^n} \quad [8]$$

Then, the necessary and sufficient condition for the error to remain bounded, maintaining numerical stability is

$$|G| \leq 1 \quad or \quad [-1 \leq G \leq 1\,] \quad [9]$$

How the FTCS scheme complies with the Stability Criterion. The goal now is to show that equation [9] holds for the particular FTCS numerical scheme. To find out how the error varies in steps of time, substitute equation [7] into each term of equation [4], as shown below

$$\epsilon_j^n = e^{at} e^{ik_m x} \quad [10.a]$$

$$\epsilon_j^{n+1} = e^{a(t+\Delta t)} e^{ik_m x} \quad [10.b]$$
$$\epsilon_{j+1}^n = e^{at} e^{ik_m (x+\Delta x)} \quad [10.c]$$
$$\epsilon_{j-1}^n = e^{at} e^{ik_m (x-\Delta x)} \quad [10.d]$$

By eqs. [10] in equation [4], the last yields

$$e^{a(t+\Delta t)} e^{ik_m x} = e^{at} e^{ik_m x} + \lambda\left(e^{at} e^{ik_m (x+\Delta x)} - 2e^{at} e^{ik_m x} + e^{at} e^{ik_m (x-\Delta x)}\right) \quad [11]$$

Divide by $e^{at} e^{ik_m x}$ to yield after simplification

$$e^{a\Delta t} = 1 + \lambda\left(e^{ik_m \Delta x} + e^{-ik_m \Delta x} - 2\right) \quad [12]$$

Using the identities (see Appendix)

$$sin\left(\frac{k_m\Delta x}{2}\right) = \frac{e^{\frac{ik_{m\Delta x}}{2}} - e^{-\frac{ik_m\Delta x}{2}}}{2i} \quad [13], \text{ and}$$

$$\sin^2\left(\frac{k_m\Delta x}{2}\right) = -\frac{\left[e^{ik_m\Delta x} + e^{-ik_m\Delta x} - 2\right]}{4} \quad [14]$$

By eq. [14] in eq. [12], the last may be written as

$$e^{a\Delta t} = 1 - 4\lambda \sin^2\left(\frac{k_m\Delta x}{2}\right) \quad [15]$$

However, by eqs [10] in eq [8], the last yields

$$G = \frac{\epsilon_j^{n+1}}{\epsilon_j^n} = \frac{e^{a(t+\Delta t)}e^{ik_m x}}{e^{at}e^{ik_m x}} = e^{a\Delta t} \; [16]$$

Thus, by equations [15], [16] plugged into [9], the condition for stability is given by

$$[17] \quad |G| = |1 - 4\lambda \sin^2(k_m\Delta x/2)| \leq 1 \quad \begin{cases} (1 - 4\lambda \sin^2\left(\frac{k_m\Delta x}{2}\right)) \geq -1 \; [17.a] \\ (1 - 4\lambda \sin^2\left(\frac{k_m\Delta x}{2}\right)) \leq 1 \; [17.b] \end{cases}$$

For the above conditions, equations [17.a] and [17.b], to hold for all m (and therefore all $\sin^2(k_{m\Delta x}/2)$), we need to consider [17.a] and [17.b], separately.

Equation [17.a] yields

$$[18] \quad 4\lambda \sin^2(k_m\Delta x/2) \leq 2$$

Note that the term $4\lambda \sin^2(k_m\Delta x/2)$ is always positive, i.e., always in the range [0,1]. The worst case is when $\sin^2(k_m\Delta x/2) = 1$, therefore, equation [18] yields

$$[19] \quad \lambda \leq \frac{1}{2} \qquad \text{or} \qquad \lambda = \frac{\alpha\Delta t}{\Delta x^2} \leq \frac{1}{2}$$

Equation [17.b] yields

$$[20] \quad 4\lambda \sin^2(k_m\Delta x/2) \geq 0$$

Since $\sin^2(k_m\Delta x/2)$ range is [0,1] equation [20] yields

$$[21] \quad \lambda \geq 0$$

which always holds. Hence, combining [19] and [21]: $0 \leq \lambda \leq \frac{1}{2}$

Equation [19] gives the stability requirement for the FTCS scheme as applied to the one-dimensional heat equation. We say that the method is *conditionally stable*; for a given Δx, the allowed value of Δt must be small enough to satisfy equation [19] or

$$[22]\ \ \Delta t \leq \frac{\Delta x^2}{2\alpha}$$

Usually we establish the space discretization by assuming Δx, then we compute the allowed Δt with equation [22] so to be sure our method won't exceed the stability requirement. For instance, if we have $\Delta x = 0.01$, and $\alpha = 1$, then we can only use a time step size $\Delta t \leq 0.00005$, which is minuscule. It would take an inordinately large number of time steps to compute the value of the solution at even moderate final times, e.g., t = 1. Moreover, owing to the limited accuracy of computers, the propagation of round-off errors might then cause a significant reduction in the overall accuracy of the final solution values. Since not all choices of space and time steps lead to a convergent scheme, that's the reason the explicit scheme FTCS is called *conditionally stable*, capisce my friend?

Stability Criterion for the Crank-Nicolson Scheme

By the CN scheme, the discretized form of equation [1]

$$U_i^{k+1} - U_i^k = \frac{\lambda}{2}\left[U_{i-1}^{k+1} - 2U_i^{k+1} + U_{i+1}^{k+1}\right] + \frac{\lambda}{2}\left[U_{i-1}^{k} - 2U_i^{k} + U_{i+1}^{k}\right] \quad [23]$$

Using similar arguments as in the FTCS method, the error ϵ_j^malso follows the discretized ODE equation of the CN scheme

$$\epsilon_i^{k+1} - \epsilon_i^k = \frac{\lambda}{2}\left[\epsilon_{i-1}^{k+1} - 2\epsilon_i^{k+1} + \epsilon_{i+1}^{k+1}\right] + \frac{\lambda}{2}\left[\epsilon_{i-1}^{k} - 2\epsilon_i^{k} + \epsilon_{i+1}^{k}\right] \quad [24]$$

Divide [24] by ϵ_i^k

$$\frac{\epsilon_i^{k+1}}{\epsilon_i^k} - \frac{\epsilon_i^k}{\epsilon_i^k} = \frac{\lambda}{2}\left[\frac{\epsilon_{i-1}^{k+1}}{\epsilon_i^k} - \frac{2\epsilon_i^{k+1}}{\epsilon_i^k} + \frac{\epsilon_{i+1}^{k+1}}{\epsilon_i^k}\right] + \frac{\lambda}{2}\left[\frac{\epsilon_{i-1}^{k}}{\epsilon_i^k} - \frac{2\epsilon_i^{k}}{\epsilon_i^k} + \frac{\epsilon_{i+1}^{k}}{\epsilon_i^k}\right] \quad [25]$$

Simplify each term in equation [25]:

$$\frac{\epsilon_i^{k+1}}{\epsilon_i^k} = G = \frac{e^{a(t+\Delta t)}}{e^{at}}\ [26.a]$$

$$\frac{\epsilon_{i-1}^{k+1}}{\epsilon_i^k} = \frac{e^{a(t+\Delta t)}e^{ik_m(x-\Delta x)}}{e^{at}e^{ik_m x}} = G\frac{e^{ik_m x}e^{ik_m(-\Delta x)}}{e^{ik_m x}} = Ge^{-ik_m\Delta x}\ [26.b]$$

$$\frac{\epsilon_{i+1}^{k+1}}{\epsilon_i^k} = \frac{e^{a(t+\Delta t)}e^{ik_m(x+\Delta x)}}{e^{at}e^{ik_m x}} = G\frac{e^{ik_m x}e^{ik_m\Delta x}}{e^{ik_m x}} = Ge^{ik_m\Delta x}\ [26.c]$$

$$\frac{\epsilon_{i-1}^{k}}{\epsilon_i^k} = \frac{e^{at}e^{ik_m(x-\Delta x)}}{e^{at}e^{ik_m x}} = \frac{e^{ik_m x}e^{ik_m(-\Delta x)}}{e^{ik_m x}} = e^{-ik_m\Delta x}\ [26.d]$$

$$\frac{\epsilon_{i+1}^{k}}{\epsilon_{i}^{k}} = \frac{e^{at}e^{ik_m(x+\Delta x)}}{e^{at}e^{ik_m x}} = \frac{e^{ik_m x}e^{ik_m \Delta x}}{e^{ik_m x}} = e^{ik_m \Delta x} \quad [26.e]$$

Plugging [26.a,b,c,d,e] into [25] yields

$$G - 1 = \left(\frac{\lambda}{2}\right) G\left(e^{-ik_m \Delta x} - 2 + e^{ik_m \Delta x}\right) + \left(\frac{\lambda}{2}\right)\left(e^{-ik_m \Delta x} - 2 + e^{ik_m \Delta x}\right) \quad [27]$$

Using a trigonometric identity involving complex exponentials, such as

$$2\cos(\theta) = e^{i\theta} + e^{-i\theta} \quad [28]$$

Using [28] into [24]

$$G - 1 = \left(\frac{\lambda}{2}\right) G(\cos(k_m \Delta x) - 2) + \left(\frac{\lambda}{2}\right)(\cos(k_m \Delta x) - 2) \quad [29]$$

Few kilometer later

$$G = \frac{1 - \lambda(1 - \cos(k_m \Delta x))}{1 + \lambda(1 - \cos(k_m \Delta x))} \quad [30]$$

Remembering that the cosine range is [-1,+1], the value we should explore for cosine values are $\cos(k_m \Delta x) = -1, 0, +1$, hence

Case 1: $\cos(k_m \Delta x) = -1$

$G = \frac{1-\lambda(1-(-1))}{1+\lambda(1-(-1))} = \frac{1-2\lambda}{1+2\lambda} \Longrightarrow since\ \lambda > 0, then\ \ G \leq 1$, always [31.a]

Case 2: $\cos(k_m \Delta x) = 0$

$G = \frac{1-\lambda(1-0)}{1+\lambda(1-0)} = \frac{1-\lambda}{1+\lambda} \Longrightarrow since\ \lambda > 0, then\ \ G \leq 1$, always [31.b]

Case 3: $\cos(k_m \Delta x) = +1$:

$G = \frac{1-\lambda(1-1)}{1+\lambda(1-1)} = \frac{1}{1} \Longrightarrow\ then\ \ G \leq 1$, always [31.c]

Therefore, we can simple say,

$$|G| \leq 1 \quad [32]$$

Hence, unconditionally stable

Stability criterion for the BTCS scheme

For this method the recurrence formula is

$$-\lambda u_{i-1}^{k+1} + (1 + 2\lambda)u_i^{k+1} - \lambda u_{i+1}^{k+1} = u_i^k \quad [33]$$

For the error is

$$-\lambda \epsilon_{i-1}^{k+1} + (1 + 2\lambda)\epsilon_i^{k+1} - \lambda \epsilon_{i+1}^{k+1} = \epsilon_i^k \quad [34]$$

Ten and half miles later, equation [30] yields a magnification factor of

$$G = \frac{1}{1 + 2\lambda(1 - \cos(k_m \Delta x)} = \frac{1}{1 + 4\lambda \sin^2 \frac{1}{2} k_m \Delta x} \quad [35]$$

Since $\lambda > 0$, and $\sin^2 \frac{1}{2} k_m \Delta x$ worst case is zero for the current condition, the magnification factor always is less than or equal to 1 in absolute value (i.e., $|G| \leq 1$), and so the stability criterion of equation [35] is satisfied for any choice of step sizes. We conclude that the implicit BTCS scheme is *unconditionally stable*.

REFERENCES

(1) MIT18_336S109_lec14.pdf
(2) https://en.wikipedia.org/wiki/Von_Neumann_stability_analysis
(3) Richard Fitzpatrick 2006-03-29
http://farside.ph.utexas.edu/teaching/329/lectures/node79.html
(4) Exponential trigonometric identities:
https://wstein.org/edu/winter06/20b/notes/html/node30.html

APPENDIX

$$[10.a] \quad sin\left(\frac{k_m \Delta x}{2}\right) = \frac{e^{\frac{ik_{m\Delta x}}{2}} - e^{-\frac{ik_m \Delta x}{2}}}{2i}$$

$$\sin^2\left(\frac{k_m \Delta x}{2}\right) = \left[\frac{e^{\frac{ik_{m\Delta x}}{2}} - e^{-\frac{ik_m \Delta x}{2}}}{2i}\right]\left[\frac{e^{\frac{ik_{m\Delta x}}{2}} - e^{-\frac{ik_m \Delta x}{2}}}{2i}\right]$$

$$= \frac{\left\{\left(e^{\frac{ik_m \Delta x}{2}}\right)^2 - 2e^{\frac{ik_{m\Delta x}}{2}} e^{-\frac{ik_m \Delta x}{2}} + \left(e^{-\frac{ik_m \Delta x}{2}}\right)^2\right\}}{(2i)(2i)} =$$

$$= \frac{\left\{e^{ik_m \Delta x} - 2 + e^{-ik_{m\Delta x}}\right\}}{(4)(-1)} =$$

$$= -\frac{\left\{e^{ik_m \Delta x} + e^{-ik_{m\Delta x}} - 2\right\}}{(4)}$$