Absolute stability of numerical methods for ODEs

We have seen in previous lecture notes that if a method is zero-stable then for all Δt smaller than some Δt^*

$$\left\|\boldsymbol{y}(t_k) - \boldsymbol{u}_k\right\|_1 \leq MTe^{MLT} \left\|\boldsymbol{\tau}(\Delta t)\right\|_1 \quad \text{for all} \quad k = 0, 1, \ldots, N \tag{1}$$

where $\left\|\boldsymbol{\tau}(\Delta t)\right\|_1$ is the global truncation error of the scheme[1]. Equation (1) bounds the error between the analytical solution of the initial value problem

$$\begin{cases} \dfrac{d\boldsymbol{y}}{dt} = \boldsymbol{f}(\boldsymbol{y}, t) \\ \boldsymbol{y}(0) = \boldsymbol{y}_0 \end{cases} \tag{4}$$

evaluated at t_k and the numerical solution of (4) computed with the scheme

$$\begin{cases} \displaystyle\sum_{j=0}^{q} \alpha_j \boldsymbol{u}_{k+j} = \Delta t \boldsymbol{\Phi}_{\boldsymbol{f}}(\boldsymbol{u}_{k+q}, \ldots, \boldsymbol{u}_k, t_k, \Delta t), \\ \text{given} \quad \{\boldsymbol{u}_0, \ldots, \boldsymbol{u}_{q-1}\} \end{cases} \tag{5}$$

If we send Δt to zero we have that $\left\|\boldsymbol{\tau}(\Delta t)\right\|_1$ in (1) goes to zero (by consistency) and therefore we can make the error between the analytical solution $\boldsymbol{y}(t_k)$ and the numerical solution $\boldsymbol{u}_k$ as small as we like.

However, for *finite* Δt it is possible that the errors due to truncation and finite machine precision propagate form one iteration to then next, and eventually build up in a way that drives the numerical solution away from the exact solution.

A prototype problem for absolute stability analysis

To study the way local errors accumulate in time and eventually yield instabilities it is convenient to consider a prototype ODE system that has a well-defined time-asymptotic state. Of course, the simplest system we can think of is a linear system[2] of the form

$$\begin{cases} \dfrac{d\boldsymbol{y}}{dt} = \boldsymbol{B}\boldsymbol{y} \\ \boldsymbol{y}(0) = \boldsymbol{y}_0 \end{cases} \tag{6}$$

where $\boldsymbol{B}$ is a matrix with eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$ having strictly negative real part, i.e.,

$$\text{Re}(\lambda_i) < 0 \quad \text{for all } i = 1, \ldots, n. \tag{7}$$

[1] Recall that all norms in a finite-dimensional vector space are equivalent. Hence, we can replace the 1-norm in (1) with any other (equivalent) norm. Also, note that the bound at the right hand side of (1) has an amplification factor

$$C = MTe^{MLT} \tag{2}$$

that can be very big. For instance, if $T = 10$ (integration period), $L = 2$ (Lipschitz constant of $\boldsymbol{f}$ in (4)), and $M = 1$ (norm of the matrix $\boldsymbol{A}$ defined in the course note 4, Lemma 1) then we obtain

$$C = 10e^{20} \simeq 4.851 \times 10^9. \tag{3}$$

[2] A numerical method which cannot handle satisfactorily the linear system (6) shall not be considered a good method. Moreover, there is ample computational evidence that methods with ample absolute stability regions (see, e.g., Figure 1) outperform those with small regions.

Hereafter, we also assume that the matrix $\boldsymbol{B}$ is diagonalizable. This simplifies the mathematical derivations and it does not change the conclusions of the analysis, meaning that the same results can obtained for non-diagonalizable matrices using a slightly more involved analysis[3]. As is well known, if the matrix $\boldsymbol{B}$ is diagonalizable then there exists an invertible matrix $\boldsymbol{P}$ such that

$$\boldsymbol{B} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1}, \tag{8}$$

where

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} \qquad \text{(diagonal matrix of eigenvalues)}, \tag{9}$$

and

$$\boldsymbol{P} = \begin{bmatrix} \begin{bmatrix} v_{11} \\ \vdots \\ v_{n1} \end{bmatrix} \cdots \begin{bmatrix} v_{1n} \\ \vdots \\ v_{nn} \end{bmatrix} \end{bmatrix} \qquad \text{(matrix of eigenvectors)}. \tag{10}$$

With the representation (8) available, we can write the analytical solution to (6) as

$$\boldsymbol{y}(t) = \boldsymbol{P}e^{t\boldsymbol{\Lambda}}\boldsymbol{P}^{-1}\boldsymbol{y}_0, \tag{11}$$

where

$$e^{t\boldsymbol{\Lambda}} = \begin{bmatrix} e^{t\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{t\lambda_n} \end{bmatrix}. \tag{12}$$

The assumption $\mathrm{Re}(\lambda_i) < 0$ implies that

$$\lim_{t\to\infty} \|\boldsymbol{y}(t)\| = 0. \tag{13}$$

Note that the matrix $\boldsymbol{P}$ allows us to fully decouple the system of ODEs (6). In fact, a substitution of (8) into (6) yields

$$\begin{cases} \dfrac{d\boldsymbol{q}}{dt} = \boldsymbol{\Lambda}\boldsymbol{q} \\ \boldsymbol{q}(0) = \boldsymbol{q}_0 \end{cases} \tag{14}$$

where

$$\boldsymbol{q}(t) = \boldsymbol{P}^{-1}\boldsymbol{y}(t), \qquad \boldsymbol{q}_0 = \boldsymbol{P}^{-1}\boldsymbol{y}_0.$$

The matrix $\boldsymbol{\Lambda}$ is diagonal, and therefore the system of ODEs (14) is fully decoupled (meaning that we can solve each ODE independently of the others). On the other hand, if the matrix $\boldsymbol{B}$ is not diagonalizable then the system (14) can be written as $\dot{\boldsymbol{q}} = \boldsymbol{J}\boldsymbol{q}$, where $\boldsymbol{J}$ is the Jordan form of $\boldsymbol{B}$. In this case the system is not decoupled since $\boldsymbol{J}$ is not fully diagonal. Note also that, in general, the matrix of eigenvectors $\boldsymbol{P}$ is complex, i.e., $\boldsymbol{q}(t)$ can be a complex vector.

Next, we study under which conditions the numerical solution $\{\boldsymbol{u}_k\}$ produced by the scheme (5) applied the linear ODE (6) decays to zero as t_k goes to infinity.

Definition 1 (Absolute stability)**.** The numerical method (5) is said to be absolutely stable if when applied to the linear system (6) generates a numerical solution $\{\boldsymbol{u}_k\}$ that decays to zero as t_k goes to infinity, i.e.,

$$\lim_{k\to\infty} \|\boldsymbol{u}_k\| = 0 \tag{15}$$

[3] If we drop the assumption that $\boldsymbol{B}$ is diagonalizable, then we have that $\boldsymbol{B}$ is similar to a block-diagonal Jordan matrix $\boldsymbol{J}$.

Absolute stability analysis of elementary one-step methods

For a given matrix $\boldsymbol{B}$ with eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$ the absolute stability condition may be satisfied for some Δt but not for others. Let us provide a few simple examples of absolute stability analysis for elementary one-step methods.

- **Euler forward:** Let us approximate the numerical solution of (6) using the Euler forward scheme

$$\boldsymbol{u}_{k+1} = \boldsymbol{u}_k + \Delta t \boldsymbol{B} \boldsymbol{u}_k. \tag{16}$$

By using the similarity transformation $\boldsymbol{P}$ we can decouple this scheme exactly as we did for the system (6). To this end, note that

$$\boldsymbol{u}_{k+1} = \boldsymbol{u}_k + \Delta t \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1}\boldsymbol{u}_k \quad \Leftrightarrow \quad \underbrace{\boldsymbol{P}^{-1}\boldsymbol{u}_{k+1}}_{\boldsymbol{w}_{k+1}} = \underbrace{\boldsymbol{P}^{-1}\boldsymbol{u}_k}_{\boldsymbol{w}_k} + \Delta t \boldsymbol{\Lambda} \underbrace{\boldsymbol{P}^{-1}\boldsymbol{u}_k}_{\boldsymbol{w}_k} \tag{17}$$

which, upon definition of[4]

$$\boldsymbol{w}_k = \boldsymbol{P}^{-1}\boldsymbol{u}_k \tag{18}$$

can be written component by component as

$$w^j_{k+1} = w^j_k + \Delta t \lambda_j w^j_k = (1 + \Delta t \lambda_j)\, w^j_k = (1 + \Delta t \lambda_j)^{k+1} w^j_0 \qquad j = 1, \ldots, n. \tag{19}$$

By taking the modulus we obtain

$$\left| w^j_{k+1} \right| = |1 + \Delta t \lambda_j|^{k+1} \left| w^j_0 \right|. \tag{20}$$

Hence, a necessary and sufficient condition for absolute stability of the Euler forward method is

$$|1 + \Delta t \lambda_j| < 1. \tag{21}$$

This condition defines a region of the complex plane, called the *region of absolute stability* in which the Euler forward scheme is absolutely stable (see Figure 1). The region of absolute stability imposes conditions on Δt for a given set of eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$. Such conditions are sketched in Figure 1 and derived analytically hereafter. To this end, note that

$$\begin{aligned} |1 + \Delta t \lambda_j|^2 &= \left[\mathrm{Re}\,(1 + \Delta t \lambda_j)\right]^2 + \left[\mathrm{Im}\,(1 + \Delta t \lambda_j)\right]^2 \\ &= \left[1 + \Delta t\, \mathrm{Re}(\lambda_j)\right]^2 + \Delta t^2\, \mathrm{Im}(\lambda_j)^2 \\ &= 1 + \Delta t^2 \left[\mathrm{Re}(\lambda_j)^2 + \mathrm{Im}(\lambda_j)^2\right] + 2\Delta t\, \mathrm{Re}(\lambda_j) \\ &= 1 + \Delta t^2 \left|\lambda_j\right|^2 + 2\Delta t\, \mathrm{Re}(\lambda_j). \end{aligned} \tag{22}$$

Clearly,

$$|1 + \Delta t \lambda_j|^2 \leq 1 \quad \Leftrightarrow \quad \Delta t \left|\lambda_j\right|^2 + 2\,\mathrm{Re}(\lambda_j) < 0, \tag{23}$$

i.e.,

$$0 < \Delta t < \max_{j=1,\ldots,n} \left(-\frac{2\,\mathrm{Re}(\lambda_j)}{\left|\lambda_j\right|^2} \right). \tag{24}$$

Hence the Euler forward method is *conditionally absolutely stable*, the condition being Δt smaller than the maximum of $-2\,\mathrm{Re}(\lambda_j)/\left|\lambda_j\right|^2$.

[4] Note that the vector $\boldsymbol{w}_k$ defined in equation (18) has, in general, complex entries. In fact the matrix of eigenvectors $\boldsymbol{P}$ is complex if the eigenvalues are complex.

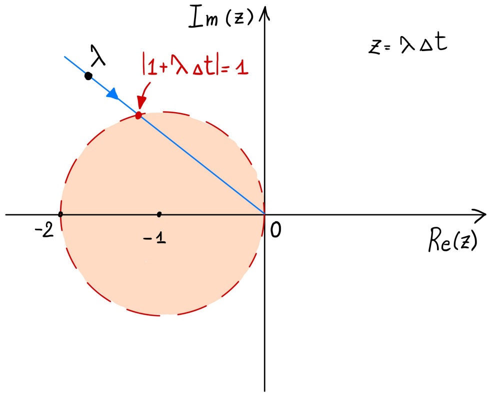

Figure 1: Region of absolute stability of the Euler forward method (shaded disk excluding the boundary). The largest Δt that guarantees absolute stability of the Euler Forward method is the one that re-scales the eigenvalues of the matrix $\boldsymbol{B}$ and brings them all within the disk (excluding the boundary). In the figure we sketch the re-scaling of one eigenvalue λ by a factor Δt that brings it exactly at the boundary of the circle.

- **Euler backward:** Let us approximate the numerical solution of (6) using the Euler backward scheme

  $$\boldsymbol{u}_{k+1} = \boldsymbol{u}_k + \Delta t \boldsymbol{B} \boldsymbol{u}_{k+1}. \tag{25}$$

  By using the similarity transformation defined by $\boldsymbol{P}$ we decouple this scheme exactly as we did for the ODE system (6) and for the Euler forward method. To this end, substitute (18) into (25) to obtain

  $$\boldsymbol{w}_{k+1} - \Delta t \boldsymbol{\Lambda} \boldsymbol{w}_{k+1} = \boldsymbol{w}_k. \tag{26}$$

  By writing (26) component by component we obtain

  $$(1 - \Delta t \lambda_j)\, w^j_{k+1} = w^j_k \quad \Rightarrow \quad w^j_{k+1} = \frac{1}{(1 - \Delta t \lambda_j)^{k+1}} w^j_0. \tag{27}$$

  Therefore, the Euler backward method is absolutely stable if and only if for all $j = 1, \ldots, n$ we have

  $$\frac{1}{|1 - \Delta t \lambda_j|} < 1 \quad \text{i.e.} \quad |1 - \Delta t \lambda_j| > 1. \tag{28}$$

  The inequality $|1 - z| > 1$ with $z \in \mathbb{C}$ defines the region outside a unit circle centered at 1 (see Figure 2). In terms of restrictions on Δt, a substitution of (22) into (28) yields

  $$\Delta t \underbrace{\left(\Delta t\, |\lambda_j|^2 - 2\,\mathrm{Re}(\lambda_j)\right)}_{>0} > 0 \quad \Leftrightarrow \quad \Delta t > 0 \tag{29}$$

  Since this condition is satisfied by any $\Delta t > 0$ we say that Euler Backward is *unconditionally absolutely stable.*

- **Crank-Nicolson:** Let us approximate the numerical solution of (6) using the Crank-Nicolson scheme

  $$\boldsymbol{u}_{k+1} = \boldsymbol{u}_k + \frac{\Delta t}{2} \left[\boldsymbol{B} \boldsymbol{u}_{k+1} + \boldsymbol{B} \boldsymbol{u}_k\right]. \tag{30}$$

  As before, we decouple this scheme by using the similarity transformation defined by $\boldsymbol{P}$. This yields,

  $$\boldsymbol{w}_{k+1} - \frac{\Delta t}{2} \boldsymbol{\Lambda} \boldsymbol{w}_{k+1} = \boldsymbol{w}_k + \frac{\Delta t}{2} \boldsymbol{\Lambda} \boldsymbol{w}_k, \tag{31}$$

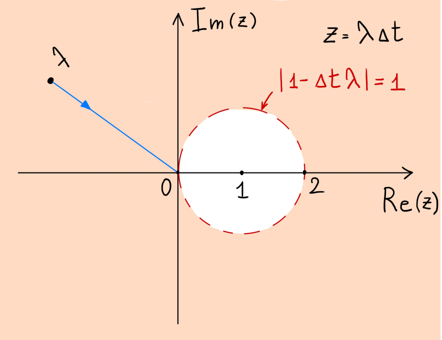

Figure 2: Region of absolute stability of the Euler backward method (shaded area outside the unit disk centered at $z = 1$ excluding the boundary of the disk). The Euler backward method is unconditionally absolutely stable (A-stable) since any eigenvalue with negative real part is in the region of absolute stability.

which can be written component by component as

$$\left(1-\frac{\Delta t\lambda_j}{2}\right)w^j_{k+1} = \left(1+\frac{\Delta t\lambda_j}{2}\right)w^j_k \quad\Rightarrow\quad w^j_{k+1} = \left|\frac{1+\frac{\Delta t\lambda_j}{2}}{1-\frac{\Delta t\lambda_j}{2}}\right|^{k+1} w^j_0. \tag{32}$$

Hence, the Crank-Nicolson method is absolutely stable if and only if

$$\left|1+\frac{\Delta t\lambda_j}{2}\right| < \left|1-\frac{\Delta t\lambda_j}{2}\right| \quad\Leftrightarrow\quad \mathrm{Re}(\lambda_j\Delta t) < 0. \tag{33}$$

The last condition follows from the following simple calculation. Set $z = \Delta t\lambda_j/2$. Then we have[5]

$$|1+z|^2 < |1-z|^2 \quad\Leftrightarrow\quad 1+2\,\mathrm{Re}(z)+|z|^2 < 1-2\,\mathrm{Re}(z)+|z|^2 \quad\Leftrightarrow\quad \mathrm{Re}(z) < 0. \tag{34}$$

Since $\mathrm{Re}(\lambda_j) < 0$ we conclude from (33) that the Crank-Nicolson method is absolutely stable for all $\Delta t > 0$. In other words it is *unconditionally absolutely stable*. The region of absolute stability of the Crank-Nicolson method is sketched in Figure 3

- **Heun method:** Let us approximate the numerical solution of (6) using the Heun method

$$\boldsymbol{u}_{k+1} = \boldsymbol{u}_k + \frac{\Delta t}{2}\left[\boldsymbol{B}\left(\boldsymbol{u}_k + \Delta t\boldsymbol{B}\boldsymbol{u}_k\right) + \boldsymbol{B}\boldsymbol{u}_k\right] = \boldsymbol{u}_k + \Delta t\boldsymbol{B}\boldsymbol{u}_k + \frac{\Delta t^2}{2}\boldsymbol{B}^2\boldsymbol{u}_k. \tag{35}$$

As before, we decouple the scheme by using the similarity transformation defined by $\boldsymbol{P}$ to obtain

$$\boldsymbol{w}_{k+1} = \boldsymbol{w}_k + \Delta t\boldsymbol{\Lambda}\boldsymbol{w}_k + \frac{\Delta t^2}{2}\boldsymbol{\Lambda}^2\boldsymbol{w}_k. \tag{36}$$

This can be written component by component as

$$w^j_{k+1} = \left(1 + \Delta t\lambda_j + \frac{\Delta t^2\lambda_j^2}{2}\right)^{k+1} w^j_0, \tag{37}$$

[5] Recall that for any $z \in \mathbb{C}$ we have:

$$\begin{aligned} |1+z|^2 =& (1+z)(1+z^*) = 1 + (z+z^*) + zz^* = 1 + 2\,\mathrm{Re}(z) + |z|^2\,, \\ |1-z|^2 =& (1-z)(1-z^*) = 1 - (z+z^*) + zz^* = 1 - 2\,\mathrm{Re}(z) + |z|^2\,. \end{aligned}$$

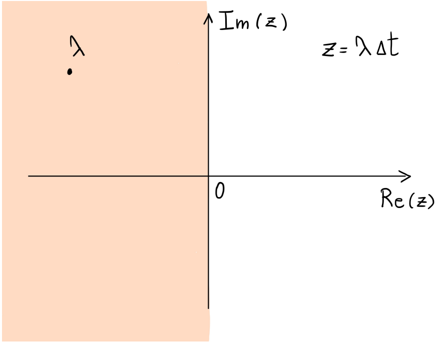

Figure 3: Region of absolute stability of the Crank-Nicolson method (shaded area representing half of the complex plane). The Crank-Nicolson method is unconditionally absolutely stable (A-stable) since any eigenvalue with negative real part is in the region of absolute stability.

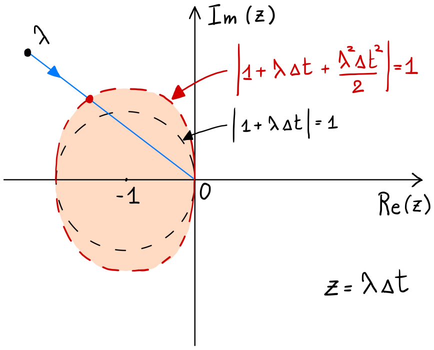

Figure 4: Region of absolute stability of the Heun method (shaded area). The largest Δt that guarantees absolute stability of the Heun method is the one that re-scales the eigenvalues of the matrix $\boldsymbol{B}$ and brings them all within the shaded area sketched in the Figure (excluding the boundary). In the figure we sketch the re-scaling of one eigenvalue λ by a factor Δt that brings it exactly at the boundary of the area. Note that the region of absolute stability of the Heun method is larger than the one of Euler forward, and therefore allows for slightly larger Δt (if the eigenvalues of the matrix $\boldsymbol{B}$ are complex).

which yields the absolute stability condition

$$\left|1 + \Delta t \lambda_j + \frac{\Delta t^2 \lambda_j^2}{2}\right| < 1 \quad \text{for all } j = 1, \ldots, n. \tag{38}$$

The region of absolute stability of the Heun method is sketched in Figure 4. The boundary of stability region is the *one level set* of the real-valued function

$$b(z) = \left|1 + z + \frac{z^2}{2}\right| \qquad z \in \mathbb{C}. \tag{39}$$

Similarly to the Euler forward method, the Heun method is conditionally absolutely stable.

At this point we provide a more rigorous definition of unconditional absolute stability. To this end, let

$$\mathbb{C}^- = \{z \in \mathbb{C} : \mathrm{Re}(z) < 0\}. \tag{40}$$

Definition 2 (A-stability)**.** Let R be the region of absolute stability of the numerical method (5). We say that the method is A-stable if

$$R \cap \mathbb{C}^- = \mathbb{C}^- \tag{41}$$

In other words, if the R includes $\mathbb{C}^-$ then the method is A-stable (or unconditionally absolutely stable).

Clearly, Euler backward and Crank-Nicolson methods are both A-stable, while Euler forward and Heun methods are all conditionally stable. More generally, one can prove that

Theorem 1. There is no explicit A-stable numerical method.

Hence, all explicit methods are *conditionally absolutely stable.* On the other hand, implicit methods can be *A-stable* (e.g., Crank-Nicolson, implicit midpoint, and all Gauss-Legendre implicit RK methods) or, *conditionally stable* (e.g., BDF methods with three or more steps, or Adams-Moulton methods with two or more steps). As we shall see hereafter, there is no A-stable implicit linear multistep method of order greater than 2 (second Dahlquist barrier).

Absolute stability analysis of linear multistep methods

Consider a general linear q-step method applied to the linear ODE system (6)

$$\sum_{j=0}^{q} \alpha_j \boldsymbol{u}_{k+j} = \Delta t \sum_{j=0}^{q} \beta_j \boldsymbol{B} \boldsymbol{u}_{k+j}. \tag{42}$$

We decouple the system by using the similarity transformation $\boldsymbol{P}$ defined in (8). To this end, define

$$\boldsymbol{w}_k = \boldsymbol{P}^{-1} \boldsymbol{u}_k, \tag{43}$$

and substitute it into (42) to obtain

$$\sum_{j=0}^{q} \alpha_j \boldsymbol{w}_{k+j} = \Delta t \sum_{j=0}^{q} \beta_j \boldsymbol{\Lambda} \boldsymbol{w}_{k+j}, \tag{44}$$

where $\boldsymbol{\Lambda}$ is the diagonal matrix (9). It is convenient to write (44) component by component as

$$\sum_{j=0}^{q} \underbrace{(\alpha_j - \Delta t \lambda_m \beta_j)}_{c_j} w^m_{k+j} = 0 \qquad m = 1, \ldots, n. \tag{45}$$

At this point we follow the same mathematical technique we used in the proof of Theorem 2 in the course note 4 (i.e., root condition implies zero-stability). To this end, we define[6]

$$\boldsymbol{z}^m_k = \begin{bmatrix} w^m_k \\ w^m_{k+1} \\ \vdots \\ w^m_{k+q-1} \end{bmatrix}, \qquad \boldsymbol{C} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -c_0/c_q & -c_1/c_q & -c_2/c_q & \cdots & -c_{q-1}/c_q \end{bmatrix}. \tag{46}$$

and write (45) as a recurrence (for a complex vector)

$$\boldsymbol{z}^m_{k+1} = \boldsymbol{C} \boldsymbol{z}^m_k. \tag{47}$$

The matrix $\boldsymbol{C}$ is the companion matrix of the characteristic polynomial

$$\pi(z) = \rho(z) - \lambda_j \Delta t \sigma(z) \qquad \text{(stability polynomial)}, \tag{48}$$

[6] Note that the vectors $\boldsymbol{z}^m_k$ and the matrix $\boldsymbol{A}$ have (in general) complex entries.

where

$$\rho(z)=\sum_{j=0}^{q}\alpha_j z^j \quad \text{(first characteristic polynomial)}, \tag{49}$$

$$\sigma(z)=\sum_{j=0}^{q}\beta_j z^j \quad \text{(second characteristic polynomial)}. \tag{50}$$

The recurrence (47) can be easily solved to obtain

$$\boldsymbol{z}^m_{k+1}=\boldsymbol{C}^{k+1}\boldsymbol{z}^m_0. \tag{51}$$

Clearly, a necessary and sufficient condition for $\|\boldsymbol{z}^m_k\|\to 0$ as $k\to\infty$ is that the matrix $\boldsymbol{C}$ is a contraction. This happens if and only if the eigenvalues of $\boldsymbol{C}$, i.e., the roots of the polynomial (48), are within the unit disk (excluding the boundary). We summarize these results as follows:

Theorem 2. The linear multistep method (42) is absolutely stable if and only if the roots of the stability polynomial (48) are within the unit disk (excluding the boundary of the disk).

Note that for $\Delta t\to 0$ the polynomial (48) tends to the first characteristic polynomial (49). Hence, in the limit of small Δt the condition for absolute stability tends to be the same as the root condition. This means that there exists a simple root of $\pi(z)$, say z^*, that approaches 1 for $\Delta t\to 0$. This is necessary for the consistency of the method. However, it should be kept in mind that zero-stability and absolute stability are different concepts. Indeed there exist convergent methods that are not absolutely stable. Let us provide an example

- **Leapfrog method:** Let us study absolute stability of the Leapfrog method

  $$\boldsymbol{u}_{k+2}=\boldsymbol{u}_k+2\Delta t\boldsymbol{f}(\boldsymbol{u}_{k+1},t_k). \tag{52}$$

  The first and second characteristic polynomials associated with the scheme are

  $$\rho(z)=z^2-1, \qquad \sigma(z)=2z. \tag{53}$$

  This gives us the following stability polynomial (see (48))

  $$\pi(z)=z^2-2\lambda_j\Delta t z-1. \tag{54}$$

  This is a polynomial with (in general) complex coefficients. To find the boundary of the region of absolute stability we look for all roots of $\pi(z)$ with modulus one, that is set[7]

  $$z=e^{i\vartheta}, \tag{55}$$

  substitute it into (54) and set the equation to zero

  $$e^{2i\vartheta}-2\lambda_j\Delta t e^{i\vartheta}-1=0 \quad \Leftrightarrow \quad \lambda_j\Delta t=\frac{e^{2i\vartheta}-1}{2e^{i\vartheta}}=\frac{e^{i\vartheta}-e^{-i\vartheta}}{2}=i\sin(\vartheta) \tag{56}$$

  As shown in Figure 5 the region of absolute stability in this case collapses to the interval $[-i,i]$ on the imaginary axis. Hence, the leapfrog method is unconditional absolutely unstable. This means that there is no hope for the method (52) to simulate accurately a linear system that has an attractor at the origin. The method is convergent through. Therefore as $\Delta t\to 0$ the global error becomes smaller and smaller (see Eq. (1)).

[7] Recall that the set of complex numbers with modulus one sits on the unit circle in the complex plane and can be represented in therm of the complex exponential function $e^{i\theta}=\cos(\theta)+i\sin(\theta)$.

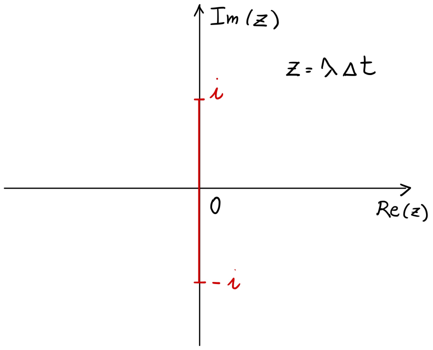

Figure 5: Region of absolute stability of the leapfrog method (52). The region collapses to the interval $[-i, i]$ on the imaginary axis. Hence, the leapfrog method is convergent but unconditionally absolutely unstable. In other words, there is no hope for the method (52) to simulate accurately a linear dynamical system that has an attractor at the origin.

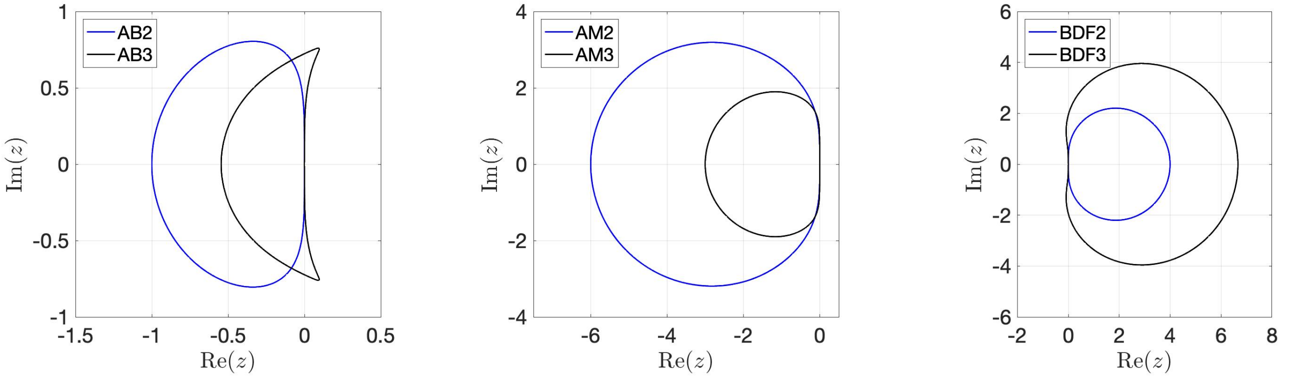

Figure 6: Boundary of the absolute stability region for various linear multistep methods. For Adams-Bashforth (AM) and Adams-Moulton (AM) methods, the region of absolute stability is the area inside the closed curve, while for BDF method is the area outside the curve. Note that the region of absolute stability of AM methods is larger than that of AB methods.

Plotting the absolute stability region of LMMs. The technique we used to compute the boundary of the absolute stability region of the leapfrog method can be generalized to arbitrary linear multistep methods. To this end, we just need to look for all roots of modulus one of the stability polynomial (48), i.e. plot the set of complex numbers[8]

$$\lambda_j \Delta t = \frac{\rho\left(e^{i\vartheta}\right)}{\sigma\left(e^{i\vartheta}\right)} = \frac{\displaystyle\sum_{j=0}^{q} \alpha_j e^{ij\vartheta}}{\displaystyle\sum_{j=0}^{q} \beta_j e^{ij\vartheta}}, \qquad \vartheta \in [0, 2\pi]. \tag{57}$$

In Figure 6 we provide a few plots of the boundary of the absolute stability region for a Adams-Bashforth, Adams-Moulton and BDF methods.

Remark: It is important to emphasize that the curves plotted in Figure 6 represent the set of values $\lambda\Delta t$

[8] Equation (57) follows immediately from the condition $\pi(e^{i\theta}) = 0$, which allows us to identify the set of $\lambda\Delta t$ such that the stability polynomial $\rho(z) - \lambda_j \Delta t \sigma(z)$ has roots with modulus one.

for which the stability polynomial (48) has at least one root with modulus one. As is well known, the roots of a polynomial are continuous functions of the coefficients of the polynomial. In the case of (48) we have one parameter, i.e., $\lambda\Delta t$, which multiplies all coefficients of $\sigma(z)$, hence affecting simultaneously multiple coefficients. To figure out which region of the complex plane is absolutely stable, e.g., the inner or the outer part of the curve defined in (57), it is sufficient to compute the roots of (48) for $\lambda\Delta t$ inside or outside the region defined by the curve. If such roots are within the unit disk, then the method is absolutely stable.

Zero-unstable LMMs are necessarily absolutely unstable. To show that zero-unstable LMMs are unconditionally absolutely unstable, we notice that in the limit $\Delta t \to 0$ we have

$$\pi(z) = \rho(z) - \lambda\Delta t\sigma(z) \to \rho(z). \tag{58}$$

If the method is zero-unstable then $\rho(z)$ has roots outside the unit disk. By continuity of polynomial roots as a function of $\Delta t\lambda$, we have that for all $\Delta t\lambda$ in a small neighborhood of 0 the polynomial (48) has roots outside the unit disk. If a method is consistent then the curve (57) passes through the origin (since $\rho(1) = 0$). Recalling that such curve represent the set of points $\lambda\Delta t$ for which at least one root of (48) has modulus one, we conclude by the continuity of the roots if $\pi(z)$ as a function of $\lambda\Delta t$ at $\lambda\Delta t = 0$ that both inner and outer regions of the curve are absolutely unstable. This proves the following lemma:

Lemma 1. A zero-unstable consistent linear multistep method is unconditionally absolutely unstable.

At this point we recall that no explict scheme can be A-stable. This implies, in particular, that there is no A-stable explicit linear multistep method. What can we say about implicit LMM methods?

Theorem 3 (Second Dahlquist barrier)**.** There is no A-stable LMM method with order greater than 2.

Recall that AM2 and BDF3 are both methods of order 3. It is seen in Figure 6 that these methods are in fact not A-stable.

Absolute stability analysis of Runge-Kutta methods

The absolute stability analysis we performed for one-step and LMM methods clearly shows that in order to compute the region of absolute stability of a numerical method it is sufficient to consider only one complex ODE of the form

$$\begin{cases} \dfrac{dq}{dt} = \lambda q \\ q(0) = q_0 \end{cases} \tag{59}$$

This ODE can be any in the decoupled system (14) corresponding to an arbitrary eigenvalue λ. Let us discretize (59) with the s-stage RK method

$$w_{k+1} = w_k + \Delta t \sum_{i=1}^{s} b_i K_i, \tag{60}$$

where

$$K_i = \lambda w_k + \lambda\Delta t \sum_{j=1}^{s} a_{ij} K_j \qquad i = 1, \dots, s. \tag{61}$$

At this point it is convenient to define

$$\boldsymbol{K} = \begin{bmatrix} K_1 \\ K_2 \\ \vdots \\ K_s \end{bmatrix}, \qquad \boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1s} \\ a_{21} & a_{22} & \cdots & a_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ a_{s1} & a_{s2} & \cdots & a_{ss} \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_s \end{bmatrix}, \qquad \boldsymbol{h} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \tag{62}$$

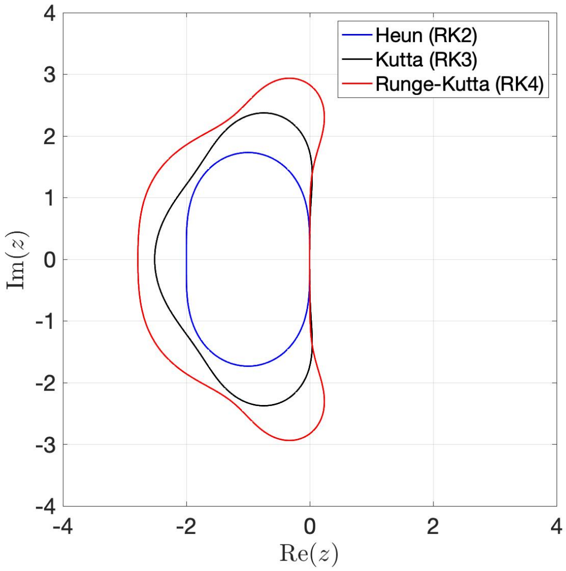

Figure 7: Boundary of the absolute stability region for various explicit RK methods. The region of stability is the interior of each closed curve.

and write (61) in a matrix-vector form as

$$(\boldsymbol{I} - \lambda\Delta t\boldsymbol{A})\boldsymbol{K} = \lambda w_k \boldsymbol{h} \quad \Leftrightarrow \quad \boldsymbol{K} = (\boldsymbol{I} - \lambda\Delta t\boldsymbol{A})^{-1}\boldsymbol{h}\lambda w_k. \tag{63}$$

Next, substitute expression we derived for $\boldsymbol{K}$ into (60) to obtain

$$w_{k+1} = w_k + \Delta t\boldsymbol{b}^T\boldsymbol{K} = \left[1 + \lambda\Delta t\boldsymbol{b}^T\left(\boldsymbol{I} - \lambda\Delta t\boldsymbol{A}\right)^{-1}\boldsymbol{h}\right] w_k. \tag{64}$$

At this point we define the *stability function*

$$S(z) = 1 + z\boldsymbol{b}^T\left(\boldsymbol{I} - z\boldsymbol{A}\right)^{-1}\boldsymbol{h}, \tag{65}$$

and iterate (64) to obtain

$$w_{k+1} = S(\lambda\Delta t)^{k+1}w_0. \tag{66}$$

Hence a necessary and sufficient condition for absolute stability of RK methods is that

$$|S(\lambda\Delta t)| < 1. \tag{67}$$

As shown in [1, p. 200], by using the Cramer's rule we can write the stability function (65) as

$$S(z) = \frac{\det\left(\boldsymbol{I} - z\boldsymbol{A} + z\boldsymbol{h}\boldsymbol{b}^T\right)}{\det(\boldsymbol{I} - z\boldsymbol{A})}. \tag{68}$$

Note that, in general $S(z)$ is a rational function, i.e., the ratio between two polynomials in z. In the particular case of explicit RK methods we have that the matrix $\boldsymbol{A}$ is strictly lower triangular. This yields $\det(\boldsymbol{I} - z\boldsymbol{A}) = 1$, which results in

$$S(z) = \det\left(\boldsymbol{I} - z\boldsymbol{A} + z\boldsymbol{h}\boldsymbol{b}^T\right) \qquad \text{(stability function for explicit RK methods).} \tag{69}$$

In Figure 7 we plot the boundary of the absolute stability region for the explicit RK methods corresponding to the following Butcher arrays[9]:

[9] The boundary of the stability regions are computed as *zero-level set* of $|S(z)| - 1$.

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1 & 0 \\ \hline & 1/2 & 1/2 \end{array}$$

Heun's method (RK2)

$$\begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1 & -1 & 2 & 0 \\ \hline & 1/6 & 2/3 & 1/6 \end{array}$$

Kutta's method (RK3)

$$\begin{array}{c|cccc} 0 & 0 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ \hline & 1/6 & 1/3 & 1/3 & 1/6 \end{array}$$

Runge-Kutta's method (RK4)

References

[1] J. D. Lambert. *Numerical methods for ordinary differential systems: the initial value problem.* Wiley, 1991.