Math 6620: Analysis of Numerical Methods, II
Solvers for initial value problems, Part II

See Ascher and Petzold 1998, Chapters 1-5

Akil Narayan[1]

[1] **Department of Mathematics, and Scientific Computing and Imaging (SCI) Institute**
University of Utah

$$\boldsymbol{u}'(t) = \boldsymbol{f}(t;\boldsymbol{u}), \qquad \boldsymbol{u}(0) = \boldsymbol{u}_0.$$
$$\boldsymbol{u}_n \approx \boldsymbol{u}(t_n)$$
$$\boldsymbol{u}_{n+1} \approx \boldsymbol{u}_n + \int_{t_n}^{t_{n+1}} \boldsymbol{f}(t,\boldsymbol{u}(t))\mathrm{d}t$$

The forward Euler discretization is:

$$D^+\boldsymbol{u}_n = \boldsymbol{f}_n, \qquad \boldsymbol{u}_{n+1} = \boldsymbol{u}_n + k\boldsymbol{f}_n, \qquad k = \Delta t.$$

This is an explicit scheme.

$$\boldsymbol{u}'(t) = \boldsymbol{f}(t;\boldsymbol{u}), \qquad \boldsymbol{u}(0) = \boldsymbol{u}_0.$$

$$\boldsymbol{u}_n \approx \boldsymbol{u}(t_n)$$

$$\boldsymbol{u}_{n+1} \approx \boldsymbol{u}_n + \int_{t_n}^{t_{n+1}} \boldsymbol{f}(t, \boldsymbol{u}(t))\mathrm{d}t$$

The forward Euler discretization is:

$$D^+\boldsymbol{u}_n = \boldsymbol{f}_n, \qquad \boldsymbol{u}_{n+1} = \boldsymbol{u}_n + k\boldsymbol{f}_n, \qquad k = \Delta t.$$

This is an explicit scheme.

We've seen that this method

- Is *consistent*: The LTE is $\mathcal{O}(k)$
- Is 0-*stable*: There is some $C > 0$ such that for all sufficiently small k,

$$\max_{n\in[N]} \|\boldsymbol{e}_n\| \leqslant C\left(\|\boldsymbol{e}_0\| + \max_{n\in[N]} \|R_n\boldsymbol{u}(t_n)\|\right),$$

$$R_n\boldsymbol{u}(t_n) := D^+\boldsymbol{u}(t_n) - \boldsymbol{f}(t_n, \boldsymbol{u}(t_n))$$

$$\boldsymbol{u}'(t) = \boldsymbol{f}(t;\boldsymbol{u}), \qquad \boldsymbol{u}(0) = \boldsymbol{u}_0.$$

$$\boldsymbol{u}_n \approx \boldsymbol{u}(t_n)$$

$$\boldsymbol{u}_{n+1} \approx \boldsymbol{u}_n + \int_{t_n}^{t_{n+1}} \boldsymbol{f}(t,\boldsymbol{u}(t))\mathrm{d}t$$

The forward Euler discretization is:

$$D^+\boldsymbol{u}_n = \boldsymbol{f}_n, \qquad \boldsymbol{u}_{n+1} = \boldsymbol{u}_n + k\boldsymbol{f}_n, \qquad k = \Delta t.$$

This is an explicit scheme.

Pairing these facts with the result,

$$\text{Consistency} + \text{0-stability} \implies \text{Convergence}$$

we conclude that Forward Euler is first-order convergent.

There is a constant C such that for all sufficiently small k,

$$\max_{n\in[N]} \|\boldsymbol{u}_n - \boldsymbol{u}(t_n)\| \leqslant Ck.$$

One minor detail is that via analysis, $C \sim e^{LT}$.

$$\boldsymbol{u}'(t) = \boldsymbol{f}(t;\boldsymbol{u}), \qquad \boldsymbol{u}(0) = \boldsymbol{u}_0.$$
$$\boldsymbol{u}_n \approx \boldsymbol{u}(t_n)$$
$$\boldsymbol{u}_{n+1} \approx \boldsymbol{u}_n + \int_{t_n}^{t_{n+1}} \boldsymbol{f}(t,\boldsymbol{u}(t))\mathrm{d}t$$

The forward Euler discretization is:

$$D^+\boldsymbol{u}_n = \boldsymbol{f}_n, \qquad \boldsymbol{u}_{n+1} = \boldsymbol{u}_n + k\boldsymbol{f}_n, \qquad k = \Delta t.$$

This is an explicit scheme.

$$\max_{n\in[N]} \|\boldsymbol{u}_n - \boldsymbol{u}(t_n)\| \leqslant Ck, \qquad C \sim e^{LT}$$

This is fine in principle, but as a practical tool this bound can be somewhat useless.

Part of the technical reason why this bound is not sharper is that we ask for a certain notion of stability for all k sufficiently small.

Towards new notions of stability

To explore potential alternatives for stability, consider the (very simple!) IVP:

$$u'(t) = \lambda u(t), \qquad\qquad u(0) = u_0,$$

for some given constants u_0 and λ. We allow λ to be complex valued, $\lambda \in \mathbb{C}$.

Towards new notions of stability

To explore potential alternatives for stability, consider the (very simple!) IVP:

$$u'(t) = \lambda u(t), \qquad u(0) = u_0,$$

for some given constants u_0 and λ. We allow λ to be complex valued, $\lambda \in \mathbb{C}$.

The value of λ is indicative of what we expect a scheme should do.

$$u(t) = u_0 \exp(\lambda t) = u_0 e^{\mu t} \cos \omega t + i u_0 e^{\mu t} \sin \omega t, \qquad \lambda = \mu + i\omega.$$

- If $\mu > 0$, then $u(t) \sim e^{\mu t}$, growing to infinity
- If $\mu = 0$, then $u(t) \sim 1$, having oscillatory behavior
- If $\mu < 0$, then $u(t) \sim e^{-|\mu| t}$, decaying to zero.

Towards new notions of stability

To explore potential alternatives for stability, consider the (very simple!) IVP:

$$u'(t) = \lambda u(t), \qquad\qquad u(0) = u_0,$$

for some given constants u_0 and λ. We allow λ to be complex valued, $\lambda \in \mathbb{C}$.

The value of λ is indicative of what we expect a scheme should do.

$$u(t) = u_0 \exp(\lambda t) = u_0 e^{\mu t} \cos \omega t + i u_0 e^{\mu t} \sin \omega t, \qquad\qquad \lambda = \mu + i\omega.$$

- If $\mu > 0$, then $u(t) \sim e^{\mu t}$, growing to infinity
- If $\mu = 0$, then $u(t) \sim 1$, having oscillatory behavior
- If $\mu < 0$, then $u(t) \sim e^{-|\mu| t}$, decaying to zero.

This last situation is of particular interest since we would reasonably expect a stable scheme to satisfy the condition,

$$|u_{n+1}| \leqslant |u_n|.$$

Step sizes attaining stability

$$u'(t) = \lambda u(t), \qquad\qquad u(0) = u_0,$$

To impose this type of (informal) stability, let's consider forward Euler:

$$u_{n+1} = u_n + k\lambda u_n.$$

Note that conceptually both λ and k should be allowed to vary, so we'll combine them into a single (complex) constant $z = \lambda k$.

$$u_{n+1} = u_n(1+k\lambda)$$

$$|u_{n+1}| \leq |u_n| \iff |1+k\lambda| \leq 1$$

Step sizes attaining stability

$$u'(t) = \lambda u(t), \qquad\qquad u(0) = u_0,$$

To impose this type of (informal) stability, let's consider forward Euler:

$$u_{n+1} = u_n + k\lambda u_n.$$

Note that conceptually both λ and k should be allowed to vary, so we'll combine them into a single (complex) constant $z = \lambda k$.

Then the condition $|u_{n+1}| \leqslant |u_n|$ is attained if,

$$|\phi(z)| \leqslant 1, \qquad\qquad \phi(z) = 1 + z, \qquad\qquad z = \lambda k.$$

The function $\phi(z)$ is called the amplification factor for the scheme.

$$|1+z| \leq 1 \rightsquigarrow |1 + k\lambda| \leq 1, \quad \lambda = \mu + i\omega$$

$$|(1+k\mu) + i k\omega| \leq 1$$

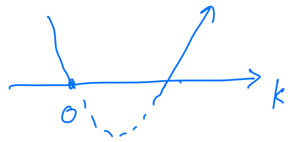
$$(1+k\mu)^2 + k^2\omega^2 \leq 1$$

$$k^2\mu^2 + 2k\mu + k^2\omega^2 \le 0$$

$$k(k\mu^2 + k\omega^2 + 2\mu) \le 0$$

$$\Downarrow$$

$$k = 0, \quad \frac{-2\mu}{|\lambda|^2}$$



$$|u_{n+1}| \le |u_n| \implies k \in \left[0, \frac{-2\,\mathrm{Re}\,\lambda}{|\lambda|^2}\right]$$

$$u'(t) = \lambda u(t), \qquad u(0) = u_0,$$

To impose this type of (informal) stability, let's consider forward Euler:

$$u_{n+1} = u_n + k\lambda u_n.$$

Note that conceptually both λ and k should be allowed to vary, so we'll combine them into a single (complex) constant $z = \lambda k$.

Then the condition $|u_{n+1}| \leqslant |u_n|$ is attained if,

$$|\phi(z)| \leqslant 1, \qquad \phi(z) = 1 + z, \qquad z = \lambda k.$$

The function $\phi(z)$ is called the amplification factor for the scheme.

We can write this in terms of k:

$$k \leqslant -\frac{2\Re\lambda}{|\lambda|^2}.$$

In particular, if λ is real (and negative), then this requires $k \leqslant 2/|\lambda|$, which is somewhat reasonable.

This particular example with forward Euler reveals a *qualitative* concept that is quite useful.

Suppose we try to solve,

$$u' = \lambda u, \qquad u(0) = 1, \qquad \Re\lambda < 0,$$

over the interval $t \in [0, 1]$ using Forward Euler.

Our *consistency* (accuracy) realizes error $\sim k$ (with a small constant). In particular, say, $k \sim 0.1$ seemingly suffices for accuracy.

This particular example with forward Euler reveals a *qualitative* concept that is quite useful.

Suppose we try to solve,

$$u' = \lambda u, \qquad u(0) = 1, \qquad \Re\lambda < 0,$$

over the interval $t \in [0, 1]$ using Forward Euler.

Our *consistency* (accuracy) realizes error $\sim k$ (with a small constant). In particular, say, $k \sim 0.1$ seemingly suffices for accuracy.

Our *absolute stability* requirement is that $k \lesssim 1/|\lambda|$, which is far smaller than what accuracy suggests is required.

Such problems (IVP's), where the stability criterion (say for forward Euler) is much stricter than the corresponding accuracy criterion, are called stiff problems.

This particular example with forward Euler reveals a *qualitative* concept that is quite useful.

Suppose we try to solve,

$$u' = \lambda u, \qquad u(0) = 1, \qquad \Re\lambda < 0,$$

over the interval $t \in [0, 1]$ using Forward Euler.

Our *consistency* (accuracy) realizes error $\sim k$ (with a small constant). In particular, say, $k \sim 0.1$ seemingly suffices for accuracy.

Our *absolute stability* requirement is that $k \lesssim 1/|\lambda|$, which is far smaller than what accuracy suggests is required.

Such problems (IVP's), where the stability criterion (say for forward Euler) is much stricter than the corresponding accuracy criterion, are called stiff problems.

Loosely speaking, over the interval $t \in [0, 1]$, the problem above is stiff if $\Re\lambda \ll -1$.

The previous analysis, however simple, is actually extraordinarily useful in more complicated scenarios, and so warrants its own name.

Definition

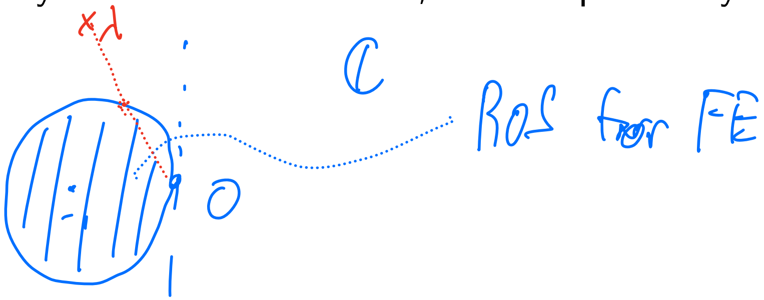
The notion of *absolute stability* is the requirement $|u_{n+1}| \leqslant |u_n|$ applied to ODE problem $u'(t) = \lambda u(t)$ using a time step k.

The set of values of $z = \lambda k$ in the complex plane attaining $|u_{n+1}| \leqslant |u_n|$ is called the *region of stability* (ROS) for the scheme.

The region of stability is a property of the overall scheme, *not* independently of the time step k or of the value of λ.

FE: $|1+z| \leq 1$

$z = k\lambda$



The previous analysis, however simple, is actually extraordinarily useful in more complicated scenarios, and so warrants its own name.

Definition

The notion of *absolute stability* is the requirement $|u_{n+1}| \leqslant |u_n|$ applied to ODE problem $u'(t) = \lambda u(t)$ using a time step k.

The set of values of $z = \lambda k$ in the complex plane attaining $|u_{n+1}| \leqslant |u_n|$ is called the *region of stability* (ROS) for the scheme.

The region of stability is a property of the overall scheme, *not* independently of the time step k or of the value of λ.

Using the concept of absolute stability, there is a stronger notion of stability:

Definition

If the ROS of a numerical scheme contains the entire closed left half-plane in $\mathbb{C}$, then the scheme is **A-stable**.

FE <u>not</u> A-stable.

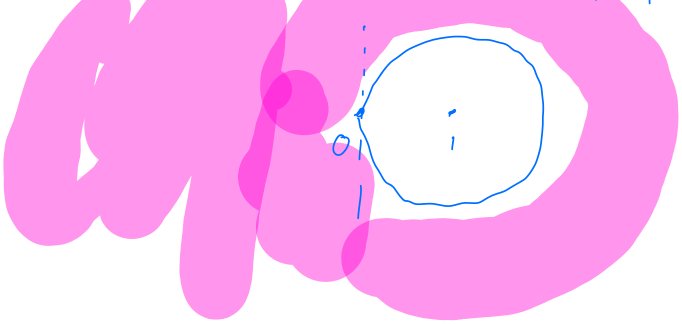
ROS for BE?

$$u_{n+1} = u_n + k\lambda u_{n+1}$$

$$(1-k\lambda)\, u_{n+1} = u_n$$

$|u_{n+1}| \le |u_n| \longrightarrow \Phi(z) = \frac{1}{1-z}$ satisfies $|\Phi(z)| \le 1$

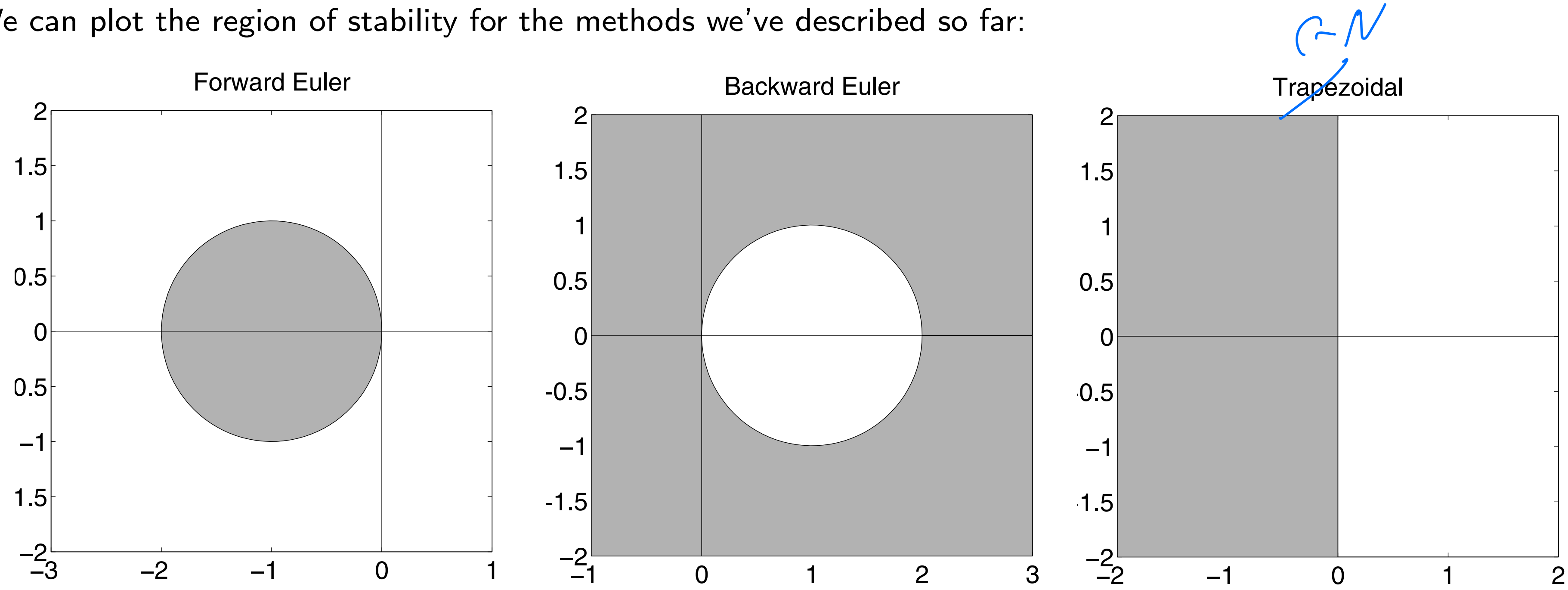
$$|1 - z| \ge 1$$



BE is A-stable (!!)

Regions of stability

We can plot the region of stability for the methods we've described so far:



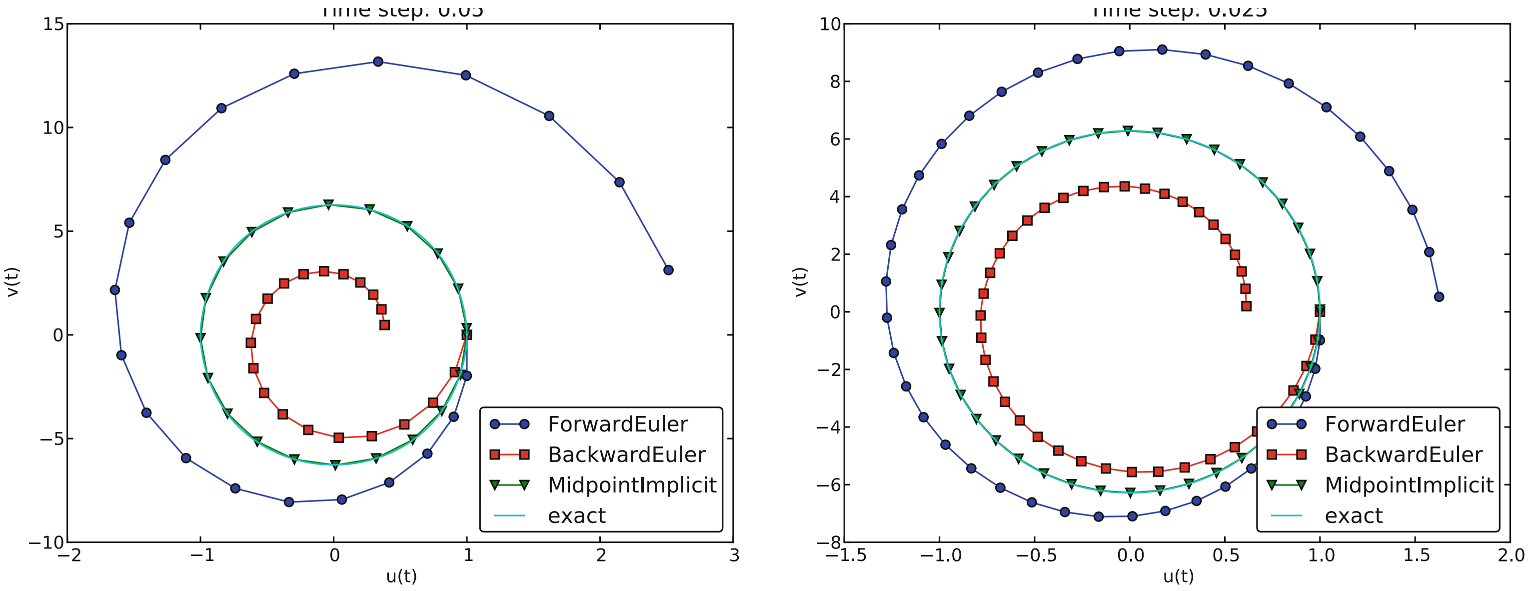
LeVeque 2007, Figure 7.1

Example

This absolute stability idea actually surfaces in practice. Consider a simple harmonic oscillator:

$$\begin{pmatrix} u'(t) \\ v'(t) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\omega^2 & 0 \end{pmatrix} \begin{pmatrix} u(t) \\ v(t) \end{pmatrix}$$

which has oscillating solutions, $u, v \sim \sin(\omega t), \cos(\omega t)$, and hence do not grow in time.



Langtangen and Linge 2017, Figure 1.7

This notion of stability motivates why implicit methods are useful:

Although both backward Euler and Crank-Nicolson involve the inversion of a (generally) nonlinear system, they are both A-stable, i.e., absolutely stable for *any* $k > 0$ for any λ with negative real part.

Implicit methods

This notion of stability motivates why implicit methods are useful:

Although both backward Euler and Crank-Nicolson involve the inversion of a (generally) nonlinear system, they are both A-stable, i.e., absolutely stable for *any* $k > 0$ for any λ with negative real part.

Generally, explicit methods are

+ Easy to implement, computationally efficient
− Can suffer from instability for large timesteps restrictions

Generally, implicit methods are

− More difficult to implement, more computationally expensive
+ Are typically (much) more stable than explicit counterparts

Linear systems, I

The utility of stability regions can be seen by considering an ODE system:

$$\boldsymbol{u}'(t) = \boldsymbol{A}\boldsymbol{u}, \qquad \boldsymbol{u}(0) = \boldsymbol{u}_0 \in \mathbb{R}^M.$$

What time step restriction should we impose to maintain A-stability?

The utility of stability regions can be seen by considering an ODE system:

$$\boldsymbol{u}'(t) = \boldsymbol{A}\boldsymbol{u}, \qquad \boldsymbol{u}(0) = \boldsymbol{u}_0 \in \mathbb{R}^M.$$

What time step restriction should we impose to maintain A-stability?

If we assume $\boldsymbol{A}$ is diagonalizable,

$$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1},$$

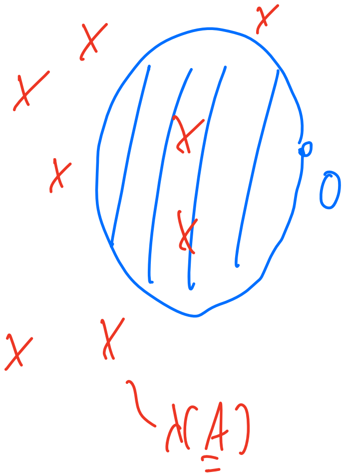
where $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues $\lambda_1, \ldots, \lambda_M$ of $\boldsymbol{A}$, then we have,

$$\boldsymbol{w} := \boldsymbol{V}^{-1}\boldsymbol{u} \implies \boldsymbol{w}'(t) = \boldsymbol{\Lambda}\boldsymbol{w}.$$

$$\boldsymbol{w} := \boldsymbol{V}^{-1}\boldsymbol{u} \quad \Longrightarrow \quad \boldsymbol{w}'(t) = \boldsymbol{\Lambda}\boldsymbol{w}.$$

Since the system for $\boldsymbol{w}$ is diagonal, it is an uncoupled set of M scalar IVP's, and hence a reasonable notion of absolute stability here is,

$$|(\boldsymbol{w}_{n+1})_m| \leqslant |(\boldsymbol{w}_n)_m| \quad \Longrightarrow \quad k\lambda_m \in \text{ROS}$$



$$\boldsymbol{w} := \boldsymbol{V}^{-1}\boldsymbol{u} \quad \Longrightarrow \quad \boldsymbol{w}'(t) = \boldsymbol{\Lambda}\boldsymbol{w}.$$

Since the system for $\boldsymbol{w}$ is diagonal, it is an uncoupled set of M scalar IVP's, and hence a reasonable notion of absolute stability here is,

$$|(\boldsymbol{w}_{n+1})_m| \leqslant |(\boldsymbol{w}_n)_m| \quad \Longrightarrow \quad k\lambda_m \in \mathrm{ROS}$$

Thus, we could say that that a particular scheme for solving $\boldsymbol{u}' = \boldsymbol{A}\boldsymbol{u}$ satisfies the notion of absolute stability if the step size k is small enough to satisfy,

$$k\lambda(\boldsymbol{A}) \subset \mathrm{ROS}$$

$$\boldsymbol{w} := \boldsymbol{V}^{-1}\boldsymbol{u} \quad \Longrightarrow \quad \boldsymbol{w}'(t) = \boldsymbol{\Lambda}\boldsymbol{w}.$$

Since the system for $\boldsymbol{w}$ is diagonal, it is an uncoupled set of M scalar IVP's, and hence a reasonable notion of absolute stability here is,

$$|(\boldsymbol{w}_{n+1})_m| \leqslant |(\boldsymbol{w}_n)_m| \quad \Longrightarrow \quad k\lambda_m \in \mathrm{ROS}$$

Thus, we could say that that a particular scheme for solving $\boldsymbol{u}' = \boldsymbol{A}\boldsymbol{u}$ satisfies the notion of absolute stability if the step size k is small enough to satisfy,

$$k\lambda(\boldsymbol{A}) \subset \mathrm{ROS}$$

This is particularly useful since it relates stability to the spectrum of $\boldsymbol{A}$.
E.g., linear IVP's whose spectrum for $\boldsymbol{A}$ extends very far away from the origin will likely require a rather small time step.

Absolute stability as we've defined it does not apply for nonlinear systems directly, but typically one can get a sense of stability via linearization.

For the general IVP,

$$\boldsymbol{u}'(t) = \boldsymbol{f}(t, \boldsymbol{u}), \qquad \boldsymbol{u}(0) = \boldsymbol{u}_0,$$

a version of the problem above linearized at $t = t_n$ is,

$$\boldsymbol{u}'(t) = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{u}}\left(t_n, \boldsymbol{u}(t_n)\right)\boldsymbol{u},$$

Absolute stability for nonlinear IVP's

Absolute stability as we've defined it does not apply for nonlinear systems directly, but typically one can get a sense of stability via linearization.

For the general IVP,

$$\boldsymbol{u}'(t) = \boldsymbol{f}(t, \boldsymbol{u}), \qquad \boldsymbol{u}(0) = \boldsymbol{u}_0,$$

a version of the problem above linearized at $t = t_n$ is,

$$\boldsymbol{u}'(t) = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{u}}(t_n, \boldsymbol{u}(t_n))\,\boldsymbol{u},$$

Therefore, a qualitative condition for stability at the next time step is that the step size k is small enough so that,

$$k\lambda\left(\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{u}}(t_n, \boldsymbol{u}(t_n))\right) \subset \mathrm{ROS}$$

(In practice, only $\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{u}}(t_n, \boldsymbol{u}_n)$ is computable.)

When $\phi(z)$ is "simple", plotting RoS is "easy".

If $\phi(z)$ is not simple?

Numerically: ϕ is generally a rational fcn.

Plot boundary of RoS.

$$|\phi(z)| = 1$$

pick $\theta \in [0, 2\pi)$

Solve $\phi(z) = e^{i\theta}$ for z

$$\frac{P(z)}{Q(z)} = e^{i\theta}$$

$\Rightarrow$ solve for z $\underbrace{P(z) - e^{i\theta} Q(z)}_{\text{poly. in } z.} = 0$

How useful is absolute stability?

$u_t = a u_{xx} \rightsquigarrow D^+ \underline{u}^n = -a \underline{\underline{A}}\, \underline{u}^n$

$\underline{\underline{A}}$: symmetric, p.d.

$\lambda(-a\underline{\underline{A}})$: purely real, negative. ~~x x x x x~~ o

$k\lambda(-a\underline{\underline{A}}) \in$ RoS (forward Euler) $\Longrightarrow k \le \frac{-2\,\mathrm{Re}\,\lambda}{|\lambda|^2} = \frac{2}{|\lambda|}$

References I

- Ascher, Uri M. and Linda R. Petzold (1998). *Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations*. SIAM. ISBN: 978-1-61197-139-2.
- Kreiss, Heinz-Otto, Joseph Oliger, and Bertil Gustafsson (2013). *Time-Dependent Problems and Difference Methods*. John Wiley & Sons. ISBN: 978-1-118-54852-3.
- Langtangen, Hans Petter and Svein Linge (2017). *Finite Difference Computing with PDEs: A Modern Software Approach*. Springer. ISBN: 978-3-319-55456-3.
- LeVeque, Randall J. (2007). *Finite Difference Methods for Ordinary and Partial Differential Equations: Steady-State and Time-Dependent Problems*. SIAM. ISBN: 978-0-89871-783-9.