# Advanced computational methods
# X071521-Lecture 3

## 1 Analysis of schemes for ODEs-Continued

### 1.1 Convergence for one-step method

**Claim:**

*For one step solvers, $u^{n+1} = u^n + k\Psi(u^n, t_n, k)$, if $\Psi$ is continuous and Lipschitz continuous in $u$ with the Lipschitz constant being uniform in $t_n, k$, then the solver is stable. Here, 'stable' means that the global error introduced by the m-th step error will not be amplified too much. If further it is consistent, then it is convergent.*

Let $\tau_j$ be the local truncation error. Then, the one step error is

$$u(t_{j+1}) - u(t_j) - k\Psi(u(t_n), t_n, k) = k\tau_j.$$

Let $E^j = |u(t_j) - u_j|$. Then, we have

$$E^{j+1} \le E^j + k|\Psi(u(t_n), t_n, k) - \Psi(u^n, t_n, k)| + k|\tau_j| \le E^j + kLE^j + k|\tau_j|.$$

Then,

$$E^n \le E^0 e^{nkL} + \sum_{j=0}^{n-1} k|\tau_{n-1-j}|(1+kL)^j \le Ce^{TL}\|\tau\|_\infty.$$

In this proof, we have implicitly use the stability. The error term $k\tau_j$ is amplified by $(1+kL)^{n-j}$ which is uniformly bounded by $e^{TL}$, so we have stability.

This verifies the claim.

## 2 Zero stability and convergence for LMMs

(Chap. 6 in Leveque.)

Previously, we have seen that the stability for a one-step consistent method implies convergence. In this section, we look at a theory, which is particularly useful for LMMs.

**Zero stability**

Consider the LMM

$$\sum_{j=0}^{r} \alpha_j u^{n+j} = k\sum_{j=0}^{r} \beta_j f(u^{n+j}, t_{n+j}).$$

The zero stability considers the stability for $k \to 0$. Hence, we have

$$\sum_{j=0}^{r} \alpha_j u^{n+j} = 0,$$

which is a linear difference equation. This equation is the same as the equation when we apply the method to $u' = 0$.

The characteristic equation is

$$\rho = \sum_{j=0}^{r} \alpha_j \zeta^j,$$

and the general solution is determined by the roots.

Since it is equivalent to $f = 0$, then $u$ should not grow. If we find $|u^n| \to \infty$ as $n \to \infty$, then the method must be unstable.

The **root condition** is

**Condition 1.** *Suppose $\zeta_j$ are the roots of the characteristic equation. We require*

$$|\zeta_j| \leq 1$$

*and $|\zeta_j| < 1$ if it is repeated.*

The LMM is said to be **zero stable** if the root condition is satisfied.

Dahlquist proved that

**Theorem 1.** *When we apply LMMs for $u' = f(t, u)$ with $f$ being Lipschitz, then:*

$$\textit{Zero stability} + \textit{consistency} \leftrightarrow \textit{convergence}.$$

Though $FE$, $BE$ and trapezoidal methods are one-step methods, they can be regarded as special cases of LMM. For them, there is only one root $\zeta_j = 1$. They are zero-stable.

## 3 Stability region

(Chap. 7 in Leveque.)

We have seen that zero stability can ensure the convergence of LMMs. However, the convergence is in the $k \to 0$ limit. For some problems, we must choose very small $k$ to get convergence though it is zero stable. To figure out the behavior for finite $k$ and the restriction on the step size, we need the notion of stability region.

Now, we look at another concept that is useful for ODE solvers. Apply the method on the test equation $u' = \lambda u$ and define $z = k\lambda$. Usually, the method yields

$$u^{n+1} = R(z)u^n$$

for one step method and

$$\sum_{j=0}^{r} (\alpha_j - z\beta_j)u^{n+j} = 0.$$

for LMMs.

The **stability region** is the set of complex $z$-values for which the solutions $u^n$ is guaranteed to be bounded:

$$|u^n| < C.$$

For one step method, we require $|R(z)| \leq 1$. For LMM, we require,

$$\pi = \rho(\zeta) - z\sigma(\zeta) = \sum_{j=0}^{r} (\alpha_j - z\beta_j)\zeta^j$$

to satisfy the root-condition.

**Corollary 1.** *An LMM is zero stable if and only if $z = 0$ is in the stability region.*

It is clear that now we should choose $k$ so that $k\lambda$ falls into the stability region for any eigenvalue with $Re(\lambda) < 0$. The method is then stable whenever $z$ falls into the region of stability.

**Remark 1.** *The above analysis is reasonable since $\lambda$ can be understood as the Jacobian at $t = t_n$, and it is general enough.*

**Examples:**

- Note that the **midpoint method (leapfrog method)** $u^{n+1} = u^{n-1} + 2kf(t_n, u^n)$ is unstable for any finite $k$ but it is zero-stable.

  $$u^{n+1} = u^{n-1} + 2zu^n \Rightarrow \zeta^2 - 2z\zeta - 1 = 0.$$

  For $|\zeta| \leq 1$, both roots must have magnitude 1 since their product is $-1$. $\zeta_1 = e^{i\theta}$ and $\zeta_2 = -e^{-i\theta}$. $z = \frac{1}{2}(\zeta - \frac{1}{\zeta}) = i\sin\theta$ if $\zeta = e^{i\theta}$. $\theta \neq \pm\pi/2$ since $\zeta_1 \neq \zeta_2$. Hence, the stability region is the open interval from $-i$ to $i$.

- The stability region of the backward Euler is $\{z : |z-1| \geq 1\}$.

  $$u^{n+1} = u^n + zu^{n+1} \Rightarrow u^{n+1} = \frac{1}{1-z}u^n.$$

  Hence, $\zeta = \frac{1}{1-z}$ and $|\frac{1}{1-z}| \leq 1$.

- For the fourth order Runge-Kutta:

  $$\begin{aligned} Y_1 &= u^n \\ Y_2 &= u^n + \frac{1}{2}zY_1 = (1+\frac{1}{2}z)u^n \\ Y_3 &= u^n + \frac{1}{2}zY_2 = (1+\frac{1}{2}z(1+\frac{1}{2}z))u^n \\ Y_4 &= u^n + zY_3 = u^n(1+z+\frac{1}{2}z^2+\frac{1}{4}z^3) \\ u^{n+1} &= u^n + \frac{z}{6}(Y_1+2Y_2+2Y_3+Y_4) = (1+z+\frac{1}{2}z^2+\frac{1}{3!}z^3+\frac{1}{4!}z^4)u^n \end{aligned}$$

  Hence, the stability region is determined by

  $$|1+z+\frac{1}{2}z^2+\frac{1}{3!}z^3+\frac{1}{4!}z^4| \leq 1$$

  Note that the RK method always has Taylor polynomials of $e^z$ as the characteristic polynomial.

# 4 Stiff problems

In the so-called stiff problems, we care about a slowly varying solution while solutions nearby are rapidly varying with much smaller time scales. In some typical physical applications, the transition to the equilibrium solution is fast but the equilibrium solution itself changes slowly. We care the equilibrium solution instead of the fast transition.

For stiff problems, designing numerical schemes is challenging since the fast transition corresponds to negative eigenvalues with large absolute value. For the method to be stable, we need $k\lambda$ to fall into the stability region. However, for explicit methods, the intersection of the stability region and negative real axis usually has a finite length. The explicit schemes requires that $k$ to be very small for stiff problems.

The issue is that we don't care the fast transition, i.e. we only care the smaller eigenvalues but the eigenvalues for fast transition put restrictions. We hope $k \sim 1/|\lambda_{slow}|$ instead of $1/|\lambda_{fast}|$.

A scheme is said to be **A-stable** if its stability region contains the whole left half plane. A scheme is said to be $A(\alpha)$**-stable** if the region $\pi - \alpha \le arg(z) \le \pi + \alpha$ lies in the stability region.

Clearly, if we use $A$-stable schemes, we won't face instability even if our $k$ is large.

Sometimes, this is not enough since we hope the modes for the fast transitions to damp instead of just being stable. Then, we require a scheme to be $L$-stable.

Consider that a method applied to $u' = \lambda u$ and we have $u^{n+1} = R(z)u^n$. The method is said to be **L-stable** if it's $A$-stable and $\lim_{|z|\to\infty} R(z) = 0$.

**Example:** The trapezoidal method is $A$-stable but not $L$-stable. The backward Euler is $L$-stable. Actually, for trapezodial, we have

$$\frac{u^{n+1} - u^n}{k} = \frac{1}{2}\lambda(u^n + u^{n+1}) \Rightarrow R(z) = \frac{1 + z/2}{1 - z/2}.$$

Similarly, for backward Euler, we have

$$R(z) = \frac{1}{1 - z}.$$

# 5 FDM for PDE, Method of lines

(Sec. 9.2 and 10.2)

Idea: Approximate the spatial differential operators with finite difference and then we get a system of ODEs. Applying suitable ODE solvers, we then get the discretization of the PDEs.

- For the heat equation $u_t = u_{xx}$, we can approximate $u_{xx}$ by the centered difference and have

  $$u_j'(t) = \frac{1}{h^2}(u_{j+1}(t) - 2u_j(t) + u_{j-1}(t)).$$

  If we then apply the forward Euler method, we obtain the scheme:

  $$\frac{u_j^{n+1} - u_j^n}{k} = \frac{1}{h^2}(u_{j+1}^n - 2u_j^n + u_{j-1}^n),$$

  where $u_j^n$ means the the numerical value at $x_j = jh, t^n = nk$.

- For the advection equation $u_t + au_x = 0$ on $[0, 1]$ with periodic boundary condition $u(0, t) = u(1, t)$ (if it's not periodic, then, the boundary condition must be imposed at the boundary where the characteristics come out), we may again use centered difference:

$$u_j'(t) = -\frac{a}{2h}(u_{j+1} - u_{j-1}).$$

With the forward Euler, we have

$$\frac{u_j^{n+1} - u_j^n}{k} = -\frac{a}{2h}(u_{j+1}^n - u_{j-1}^n).$$

In the methods generated by MOL, the same ODE solver is used for all spatial discretization, which is sometimes not efficient and not appropriate. MOL, however, provides a useful tool and it is helpful for understanding the stability, as we'll see below.

## Stability for MOL

We say a scheme is stable if

$$\sup_{n:nk\le T} \|u^n\| \le C(T)\|u^0\|,$$

for some norm $\|\cdot\|$.

Analyzing stability in the viewpoint of MOL is often for $l^2$ stability analysis (the domain for $x$ is usually bounded and we have boundary conditions).

As in the ODE theory, we require $k\lambda$ to be in the stability region of the ODE method for any eigenvalue $\lambda$ of the spatial discretization.

- Consider the scheme

  $$\frac{u_j^{n+1} - u_j^n}{k} = \frac{1}{h^2}(u_{j+1}^n - 2u_j^n + u_{j-1}^n)$$

  for $u_t = u_{xx}$ with Dirichlet boundary conditions, the matrix $A$ is tridiagonal and the eigenvalues of the matrix are given by

  $$\lambda_p = \frac{2}{h^2}(\cos(p\pi h) - 1), \quad p = 1, 2, \dots, m.$$

  Note that $\cos(\xi) - 1$ is decreasing on $[0, \pi]$, so the eigenvalues are roughly in the interval $(\frac{-4}{h^2}, \lambda_1) \approx (\frac{-4}{h^2}, -\pi^2)$. Hence, we require $-\frac{4k}{h^2}$ to be in the stability region of the ODE method. The stability region of the forward Euler is $|1+z| \le 1$. Hence, the condition for the scheme to be stable is

  $$\frac{4k}{h^2} \le 2.$$

Hence, the time step is roughly the square of the spatial step. This is a severe restriction. Explicit schemes for parabolic equations usually have such restrictions.

- For the scheme

  $$\frac{u_j^{n+1} - u_j^n}{k} = -\frac{a}{2h}(u_{j+1}^n - u_{j-1}^n)$$

  with periodic boundary conditions, the matrix $A$ has eigenvalues

  $$\lambda_p = -\frac{ia}{h}\sin(2\pi ph), \qquad p = 1, 2, \dots, m+1.$$

  We therefor need the interval $(-i\frac{ak}{h}, i\frac{ak}{h})$ which is on the imaginary axis to be in the stability region. The stability region of the forward Euler is however $|1+z| \leq 1$. This means this scheme is **unstable** for any fixed ration $k/h > 0$.

  However, we consider the convergence in the limit $k, h \to 0$. If the ratio $k/h \to 0$, then $k\lambda$ will tend to the origin. We know forward Euler is zero stable, so we may expect the convergence if $k/h \to 0$. Actually, if $k = O(h^2)$, the convergence can be shown rigorously (read P205).

  If we instead use Leapfrog for time discretization:

  $$\frac{u_j^{n+1} - u_j^{n-1}}{2k} = -\frac{a}{2h}(u_{j+1}^n - u_{j-1}^n)$$

  we will however get stability for $ak/h < 1$.

# 6 Convergence: Lax Equivalence Theorem

Roughly speaking, this theorem says: for linear equations, a method is convergent if it is consistent and stable.

## Consistency: Local truncation error

(Sec. 9.1)

Given a scheme for a PDE, we use the local truncation error to measure the consistency.

We insert the exact solution $u(x,t)$ and determine how good it satisfies the PDE.

- For scheme
$$\frac{u_j^{n+1} - u_j^n}{k} = \frac{1}{h^2}(u_{j+1}^n - 2u_j^n + u_{j-1}^n),$$
the local truncation error is given by
$$\tau(x,t) = \frac{u(x,t+k) - u(x,t)}{k} - \frac{1}{h^2}(u(x+h,t) - 2u(x,t) + u(x-h,t))$$
$$= (u_t - u_{xx}) + \frac{1}{2}u_{tt}k + \frac{1}{12}u_{xxxx}h^2 + \dots$$
The error is $O(k+h^2)$, which is first order in time and second order in space.

- Similarly, the scheme for the advection equation we just proposed is also $O(k+h^2)$.

## The Lax equivalence theorem

Suppose a method can be written as
$$u^{n+1} = B(k)u^n + b^n(k),$$
where $u^n = (u_j^n)$ is a vector.

The method is called Lax-Richtmyer stable if for any fixed $T > 0$, there exists a constant $C_T$ such that
$$\|B(k)^n\| \le C_T,$$
whenever $nk \le T$.

**Theorem 2.** *Any* **consistent** *method of the above form is convergent if and only if it is Lax-Richtmyer stable.*

The 'if' part is straightforward. Suppose $u(x,t)$ is the exact solution, and $\bar{u}$ consists of the values of the exact solutions, then we have
$$\bar{u}^{n+1} = B(k)\bar{u}^n + b^n(k) + k\tau^n,$$
where $\tau^n$ is the local truncation error and goes to zero as $k \to 0$ since the method is consistent. Then, $E^n = u^n - \bar{u}^n$ and we have
$$E^{n+1} = BE^n - k\tau^n.$$
This relation implies that
$$\|E^N\| \le C_T\|E^0\| + TC_T \max_n \|\tau^n\|.$$
For the 'only if' part, it involves some uniform boundedness principle. We ignore it. Those who are interested can read the paper by Richtmyer.

# 7 Von Neumann-analysis for linear PDEs (Fourier analysis)

This is convenient for $l^2$ stability analysis for a unbounded domain or domain with periodic boundary condition. Suppose we have sample points $x_j = jh : j : -\infty \to \infty$. We have $u_j$ defined at $x_j$. Let's define the semi-discrete Fourier transform:

$$\hat{u}(\xi) = \frac{h}{\sqrt{2\pi}} \sum_{j=-\infty}^{\infty} u_j e^{-ix_j\xi} = \frac{h}{\sqrt{2\pi}} \sum_{j=-\infty}^{\infty} u_j e^{-ijh\xi}, \xi \in [-\pi/h, \pi/h].$$

Then, we can recover $u_j$ by

$$u_j = \frac{1}{\sqrt{2\pi}} \int_{-\pi/h}^{\pi/h} \hat{u}(\xi) e^{ijh\xi} d\xi.$$

We have the Parseval's equality:

$$\|u\|_2 = \sqrt{h \sum_j u_j^2} = (\int_{-\pi/h}^{\pi/h} |\hat{u}(\xi)|^2 d\xi)^{1/2} = \|\hat{u}\|_2.$$

The $l^2$ stability requires that $\|u^n\|_2$ is bounded. Hence, it is enough to check the $L^2$ norm of $\hat{u}$.

Usually, for the FDM of a linear PDE, the Fourier modes $e^{ix_j\xi}$ are decoupled. Just like the dispersion relation for linear PDE, for the discrete case, we'll have

$$\hat{u}^{n+1}(\xi) = g(\xi)\hat{u}^n(\xi).$$

**Theorem 3.** *If there exists $\alpha \geq 0$ such that the amplification factor $g$ satisfies*

$$|g(\xi)| \leq 1 + \alpha k,$$

*then the method is $l^2$ stable.*

*Proof.*

$$\|\hat{u}^{n+1}\|_2 \leq \sup_\xi |g(\xi)| \|\hat{u}_n\|_2 \leq (1 + \alpha k)\|\hat{u}_n\|_2.$$

Hence,

$$\|\hat{u}^N\|_2 \leq (1 + \alpha k)^N \|\hat{u}_0\|_2 \leq \exp(\alpha N k)\|\hat{u}_0\|_2 = \exp(\alpha T)\|\hat{u}_0\|_2.$$

□

To figure out $g(\xi)$, we can simply assume that $u_j^n = e^{ijh\xi}$ and then compute $u_j^{n+1} = g(\xi)e^{ijh\xi}$. (This works since $u_j^n$ is just the superposition of these modes. We can check what happens for each mode.)

- Consider again the scheme

  $$\frac{u_j^{n+1} - u_j^n}{k} = \frac{1}{h^2}(u_{j+1}^n - 2u_j^n + u_{j-1}^n).$$

  If we plug in $u_j^n = \exp(ijh\xi)$, we have

  $$u_j^{n+1} = e^{ijh\xi} + \frac{k}{h^2}(e^{ih\xi} - 2 + e^{-ih\xi})e^{ijh\xi} = g(\xi)e^{ijh\xi}.$$

  Hence, we find $g(\xi) = 1 + \frac{k}{h^2}2(\cos(\xi h) - 1)$. We find $2(\cos(\xi h) - 1) \in [-4, 0]$. Then, if $-4k/h^2 \geq -2$, $|g| \leq 1$, the method is stable. (Due to $1/h^2$, it's not possible to expect positive $\alpha$.) We obtain the same requirement.

- For the method,

  $$\frac{u_j^{n+1} - u_j^n}{k} = -\frac{a}{2h}(u_{j+1}^n - u_{j-1}^n),$$

  we find $g(\xi) = 1 + \frac{ak}{2h}2i\sin(\xi h)$. Then,

  $$|g| = \sqrt{1 + a^2\frac{k^2\sin^2(\xi h)}{h^2}} \lesssim 1 + a^2\frac{k^2}{2h^2}.$$

  Then, if $k/h^2$ is fixed. The method is stable. For any fixed $k/h$ ratio, the method is unstable.

**Remark 2.** *Besides the MOL and Fourier analysis, other method for stability include energy method, discrete maximal principle, direct estimation of the growth matrix et al.*