Lecture 4

Zed

March 14, 2017

1 Stiff ODEs

1.1 Motivation

Consider the linear ode system:

$$\begin{cases} \boldsymbol{y}' = \boldsymbol{\Lambda}\boldsymbol{y}, \quad t \geq 0 \\ \boldsymbol{y}(0) = \boldsymbol{y}_0 \neq 0 \end{cases} \quad \text{where } \boldsymbol{\Lambda} = \begin{pmatrix} -100 & 1 \\ 0 & -\frac{1}{10} \end{pmatrix} \tag{1}$$

$\boldsymbol{\Lambda}$ is a diagonizable matrix, which we can write as $\boldsymbol{\Lambda} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^{-1}$, and

$$\boldsymbol{V} = \begin{pmatrix} 1 & 1 \\ 0 & \frac{999}{10} \end{pmatrix}, \quad \boldsymbol{D} = \begin{pmatrix} -100 & 0 \\ 0 & -\frac{1}{10} \end{pmatrix}$$

By theory, the entries of diagonal matrix $\boldsymbol{D}$ are the eigenvalues of $\boldsymbol{\Lambda}$. We can derive the exact solution of the system: $\boldsymbol{y} = e^{\boldsymbol{\Lambda}t}\boldsymbol{y}_0 = \boldsymbol{V}e^{\boldsymbol{D}t}\boldsymbol{V}^{-1}\boldsymbol{y}_0$. Where $e^{\boldsymbol{D}t}$, the exponential of a matrix is defined as a matrix of same dimension in taylor expansion of $e^{(\cdot)}$. In this case $\boldsymbol{e}^{\boldsymbol{D}t}$ is a diagonal matrix with entries e^{-100t} and $e^{-\frac{1}{10}t}$. So there exists constant $\boldsymbol{x}_1, \boldsymbol{x}_2$, such that $\boldsymbol{y}(t) = \boldsymbol{x}_1 e^{-100t} + \boldsymbol{x}_2 e^{-\frac{1}{10}t}$. Compared with the second term, e^{-100t} is small (for $t \geq 0$), so this is approximately $\boldsymbol{y} \sim \boldsymbol{1}e^{-\frac{1}{10}t}$.

On the other hand we try solve the system with euler's method: $\boldsymbol{y}_{n+1} = \boldsymbol{y}_n + h\boldsymbol{\Lambda}\boldsymbol{y}_n = (\boldsymbol{I} + h\boldsymbol{\Lambda})\boldsymbol{y}_n$. So we have

$$\begin{aligned} \boldsymbol{y}_n &= (\boldsymbol{I} + h\boldsymbol{\Lambda})^n \boldsymbol{y}_0 = \boldsymbol{V}(\boldsymbol{I} + h\boldsymbol{D})^n \boldsymbol{V}^{-1}\boldsymbol{y}_0 = \boldsymbol{V}\begin{pmatrix} 1 - 100h & 0 \\ 0 & 1 - \frac{1}{10}h \end{pmatrix}^n \boldsymbol{V}^{-1}\boldsymbol{y}_0 \\ &= \boldsymbol{c}_1(1 - 100h)^n + \boldsymbol{c}_2(1 - \tfrac{1}{10}h)^n \end{aligned} \tag{2}$$

The exact solution decays with t, we want the numerical solution to possess this property, i.e. to decay with n. This requires $|1 - 100h| < 1$ and $|1 - \frac{1}{10}h| < 1$, *depending on our choice of* h. We *should* choose $0 < h < \frac{1}{50}$ and $0 < h < 20$ to make the numerical solution decay with n. The problem is that we can not foresee this problem all the time, so we are possible to select an improper h, like $h = \frac{1}{10}$. Which will make the first term blow up with n, and clearly in this case the numerical solution does not match the decaying property of the exact solution.

1.2 Stiffness

Def. **Stiffness**: An ODE system is said to be *stiff* if the numerical solution requires a very *small* h, i.e. a significant depression of step size, to avoid blow up. We also define the *stiffness ratio* for the linear system $\boldsymbol{y}' = \boldsymbol{A}\boldsymbol{y}$ as the largest eigenvalue of $\boldsymbol{A}$ / the smallest eigenvalue of $\boldsymbol{A}$. We look at the (eigenvalues of) Jacobian $\nabla_{\boldsymbol{y}}\boldsymbol{f}$ as an approximation for nonlinear systems $\boldsymbol{y}' = \boldsymbol{f}(t, \boldsymbol{y})$.

Ex. *A Chemical Reaction ODE System*:

$$\begin{cases} y_1' = -0.04y_1 + 10^4 y_2 y_3 \\ y_2' = 0.04y_1 - 10^4 y_2 y_3 - 3\cdot 10^7 y_2^2 \\ y_3' = 3\cdot 10^7 y_2^2 \end{cases} \tag{3}$$

An observation is the conservation of mass: $(y_1 + y_2 + y_3)' = 0$, and we may want to preserve this property in numerical computing. In this system, y_3 is a *fast* variable, since it has the largest change rate; y_2 is *intermediate* and y_1 is *slow.* In general, fast variable requires a small h, and determines the appropriate step size.

2 Absolute Stability

Def. **Absolute Stability**: Apply a numerical method for $y' = f(t, y)$ to a linear ODE:

$$\begin{cases} y' = \lambda y, \quad t \geq 0 \\ y(0) = y_0 \neq 0 \end{cases} \tag{4}$$

for certain fixed $\bar{h} = \lambda h, \lambda \in \mathbb{C}$ (complex plane). $\{y_n\}$ is the path of numerical solution, if $\{y_n\}$ strictly decays to 0 as $n \to \infty$, i.e. $\lim\limits_{n\to\infty} y_n = 0$, we call the method as absolute stable (A-stable). Moreover, the region $\{\bar{h} : \bar{h} \text{ is A-stable}\} \subseteq \mathbb{C}$ is called the region of absolute stability of the method.

Ex. (*Explicit Euler*) check the A-stability of euler's method: $y_{n+1} = y_n + h\lambda y_n = (1 + \bar{h})y_n = (1 + \bar{h})^{n+1} y_0$. $\Rightarrow \left\|1 + \bar{h}\right\| < 1$ gives $\bar{h}$ A-stable. On the complex plane this the *interior* of a disk with radius 1, centered at $(-1, 0)$.

Ex. (*Backward Euler*): $y_{n+1} = y_n + h\lambda y_{n+1} = (1 - \bar{h})^{-1} y_n = (1 - \bar{h})^{-n-1} y_0$. $\Rightarrow \left\|1 - \bar{h}\right\| > 1$ gives $\bar{h}$ A-stable. On the complex plane this is the *outside* of a disk with radius 1, centered at $(1, 0)$.

For the explicit euler the A-stable region is a bounded region, when λ big, and we require the solution to decay with n, then $\bar{h} = \lambda h$ should be in that bounded region, hence we should pick a small h. However for the backward euler, the A-stable region include half of the complex plane ($\mathrm{Re}(\bar{h}) < 0$), which implies that we can even find an h to let the numerical solution decay for an exact solution that grows with t. (which is not desirable, hence we would not like to use this method for an ode that has growing solutions).

Ex. (*Trapezoid Method*)

$$y_{n+1} = y_n + \frac{h\lambda}{2}(y_n + y_{n+1}) = \left(\frac{1 + \frac{\bar{h}}{2}}{1 - \frac{\bar{h}}{2}}\right)^{n+1} y_0 \Rightarrow \left\|\frac{1 + \frac{\bar{h}}{2}}{1 - \frac{\bar{h}}{2}}\right\| < 1$$

gives the A-stable region, i.e. $\|1+\frac{\bar{h}}{2}\| < \|1-\frac{\bar{h}}{2}\| \Rightarrow \{z \in \mathbb{C} : \text{distance from } z \text{ to } (-2, 0) < \text{that to } (2, 0)\}$. Hence the A-stable region is the left-half plane: $\{z \in \mathbb{C} : \mathrm{Re}(z) < 0\}$. This matches the growing/decaying behavior of exactly solution in both cases: $\mathrm{Re}(\lambda) < 0 \Rightarrow$ the exact solution $\searrow$, and in this case for all $h \in \mathbb{R}$, $\lambda h \in$ A-stable region.

Ex. (*2-stage Runge-Kutta*)

$$\begin{cases} y_{n+1} = y_n + h(b_1 k_1 + b_2 k_2) \\ k_1 = f(t_n, y_n) \\ k_2 = f(t_n + c_2 h, y_n + a_{21} h k_1) \end{cases} \tag{5}$$

The order-2 condition requires $b_1+b_2=1, b_2c_2=b_2a_{21}=\frac{1}{2}$. Apply the method to $y'=\lambda y$ $\Rightarrow y_{n+1}=y_n(1+b_1\bar{h}+b_2\bar{h}+b_2a_{21}\bar{h}^2)=(1+\bar{h}+\frac{1}{2}\bar{h}^2)^{n+1}y_0$. Notice that the resulting formula is not dependent upon the exact value of parameters, only the order-2 constraints can determine it. We denote $R(z):=1+z+\frac{1}{2}z^2$, then the A-stable region is $\|R_2(z)\|<1$.

Ex. (*3-stage Runge-Kutta*)

$$\begin{cases} y_{n+1}=y_n+h(b_1k_1+b_2k_2+b_3k_3) \\ k_1=f(t_n,y_n) \\ k_2=f(t_n+c_2h,y_n+a_{21}hk_1) \\ k_3=f(t_n+c_3h,y_n+a_{31}hk_1+a_{32}hk_2) \end{cases} \tag{6}$$

It turns out that $y_{n+1}=(1+\bar{h}+\frac{1}{2}\bar{h}^2+\frac{1}{6}\bar{h}^3)^{n+1}y_0$. A-stable region: $\|R_3(z)\|<1$.

In fact, A-stable region does not depend on the specific choice of parameters until 4-stages. $R_4(z)=1+z+\frac{1}{2}z^2+\frac{1}{6}z^3+\frac{1}{24}z^4$. For $N\geq 5$ this is not true anymore.

Ex. Consider ode $y'=iwy$. The solution is $y(t)=e^{iwt}y_0$, which oscillates. And we want to keep the property $\|y(t)\|=\|y_0\|$ for numerical solutions. In fact, in this case we are seeking for a method whose A-stable region is A, and ∂A is the imaginary axis.

Ex. (*A-Stablity of Multistep Methods*) consider general formulation of multistep method:

$$\sum_{j=0}^{s} a_j y_{n+j}=h\sum_{j=0}^{s} b_j f(t_{n+j},y_{n+j}) \tag{7}$$

Apply to $y'=\lambda y$:

$$\sum_{j=0}^{s} a_j y_{n+j}=h\sum_{j=0}^{s} b_j\lambda y_{n+j}=\bar{h}\sum_{j=0}^{s} b_j y_{n+j} \tag{8}$$

$$\Rightarrow \sum_{j=0}^{s}(a_j-\bar{h}b_j)y_{n+j}=0$$

Recall fisrt and seond characteristic polynomials: $\rho(z)=\sum_0^s a_jz^j$, $\sigma(z)=\sum_0^s b_jz^j$, zero stability depends on the roots of $\rho(z)=0$. Now we define the *Stablity Polynomial*: $\pi(z,\bar{h}):=\rho(z)-\bar{h}\sigma(z)$. A-stablity $\iff$ all roots of $\pi(z,\bar{h})=0$ have norm <1 (within the unit circle on complex plane).

3 L Stability

Def. **L Stability**: We define the growth factor as $R(\bar{h})=\frac{y_{n+1}}{y_n}$. And we say that a method is L-stable if it is A-stable and $\lim\limits_{\bar{h}\to\infty}|R(\bar{h})|=0$.

Ex.
- The explicit Euler: $R(z)=1+z$, not L-stable.
- The implicit Euler: $R(z)=\frac{1}{1-z}$, L-stable.
- The trapezoid: $R(z)=\frac{1+\frac{z}{2}}{1-\frac{z}{2}}$, not L-stable.

Usually, we can see that an implicit method has better stability.