数值分析

— 科学与工程计算基础

任课教师：黄忠亿

清华大学数学科学系

目录 I

常微方程差分方法基本理论

许多常见、重要的自然科学(物理、化学、生物)、人文科学及工程技术中许多问题都可以用常微方程（组）初值问题来描述. 通常来说这些问题不能写出其解的解析表达式, 因此需要用数值方法求解.

考虑一阶常微方程（组）初值问题:

$$\begin{cases} \dfrac{d\mathbf{y}}{dt} = \mathbf{f}(t, \mathbf{y}), \quad t \in (0, T], \\ \mathbf{y}(0) = \mathbf{y}_0. \end{cases} \tag{9.1}$$

这里 $\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}, \mathbf{f} = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}, \mathbf{y}_0 = \begin{pmatrix} y_{0,1} \\ \vdots \\ y_{0,m} \end{pmatrix} \in \mathbb{R}^m$.

常微方程差分方法基本理论

首先我们来看该问题 (9.1) 的适定性.

定义 9.1 (适定性)

如果初值问题 (9.1) 满足

1. 存在唯一解;
2. 对于任给的 $\varepsilon > 0$, 存在正数 δ^*, 使得当 $|\vec{\varepsilon}_0| < \delta^*$ 时, 在 $[0, T]$ 上有 $\vec{\delta}(x)$ 连续, 且 $\left\|\vec{\delta}\right\| < \delta^*$ 时, 初值问题

$$\begin{cases} \mathbf{z}_t = \mathbf{f}(t, \mathbf{z}) + \vec{\delta}, & t \in (0, T], \\ \mathbf{z}(0) = \mathbf{z}_0 + \vec{\varepsilon}_0 . \end{cases}$$

存在唯一解 $\mathbf{z}$, 且 $\forall t \in [0, T]$ 有 $\|\mathbf{z} - \mathbf{y}\| < \varepsilon$.

那么就称初值问题 (9.1) 是适定的.

常微方程差分方法基本理论

为了保证问题 (9.1) 的适定性 (即解的存在唯一性及对初值和源项的连续依赖性), 通常假设 $\mathbf{f}(t,\mathbf{y})$ 关于 $\mathbf{y}$ 满足 **Lipschitz** 连续条件, 即 $\exists L>0$, $\forall t\in[0,T]$, $\forall \mathbf{x},\mathbf{z}\in\mathbb{R}^m$, 有

(9.2) $$\|\mathbf{f}(t,\mathbf{x})-\mathbf{f}(t,\mathbf{z})\|\le L\,\|\mathbf{x}-\mathbf{z}\|\,.$$

这里 $\|\cdot\|$ 为 $\mathbb{R}^m$ 中的某种范数.

关于适定性我们有以下结论:

定理 9.1

设 $D=[0,T]\times\mathbb{R}^m$, $\mathbf{f}(t,\mathbf{y})$ 在D上连续并满足$Lipschitz$条件 (9.2), 那么初值问题 (9.1) 是适定的.

差分方法求解常微分方程初值问题

鉴于一般常微分方程初值问题的解

$$\mathbf{y}(t) = \mathbf{y}_0 + \int_0^t \mathbf{f}\big(s, \mathbf{y}(s)\big)\,ds,$$

不一定有简单解析表达式, 我们一般都用数值方法来求近似解.

为简单起见, 假设对区间 $[0, T]$ 采用等距剖分:

取自然数 N, 令步长 $h = \dfrac{T}{N}$, $t_n = nh, \quad n = 0, 1, \cdots, N.$

下面看如何用数值方法计算出 $\mathbf{y}(t_n)$ 的近似值 $\mathbf{y}_n$.

差分格式的一般形式

为简单起见, 以下取 $m=1$. 可以很容易推广到 $m>1$ 情形.

一般来说，如果知道了 $y_0, y_1, \cdots, y_{k-1}$ 的值, 可用以下算法来求 y_k:

$$\begin{cases} \sum_{j=0}^{k} \alpha_j y_{n+j} = h\phi_f(t_n; y_n, y_{n+1}, \cdots, y_{n+k}; h), \ n \geq 0, \\ y_j = S_j(h), \quad j = 0, 1, \cdots, k-1. \end{cases} \tag{9.3}$$

这里假设 $\alpha_k \neq 0$.

$k=1$ 称为单步法. $k>1$ 称为多步法.

ϕ_f 取不同形式的函数，可得到不同的方法。

最简单的单步法—Euler 法

我们先来看最简单的单步法如何构造 (我们将给出几种不同的构造思路).

已知道初始值 $y(t_0)=y_0$**, 假设** $y(t)$ **充分光滑, 由 Taylor 展开**

$$y(t_1)=y(t_0+h)=y(t_0)+hy'(t_0)+\frac{h^2}{2}y''(\xi),\quad \xi\in(t_0,t_1).$$

略去二阶项我们得到 $y(t_1)\approx y_1=y_0+hf(t_0,y_0)$**.**

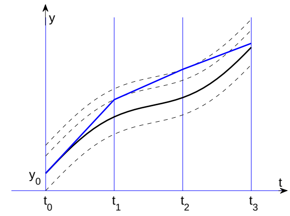
有了 y_0 **便可以由此计算出** $y(t_1)$ **的近似值, 类推下去便可求得** $y_2, y_3, \cdots$**:**

$$y_{n+1}=y_n+hf(t_n,y_n),\quad n=0,1,\cdots,N-1 \tag{9.4}$$

此即 Euler 法.

最简单的单步法—Euler 法

我们也可以从其几何意义来看:



由常微分方程理论我们知道 $\dfrac{dy}{dt} = f(t,y)$ 的通解为 (t,y) 平面上的一族积分曲线. 由适定性理论可知, 过 (t,y) 平面上任一点 (t^*, y^*) 恰好有一条解曲线.

Euler 法即过 (t_0, y_0) 做过该点的积分曲线的切线交直线 $t = t_1$ 于 (t_1, y_1), 然后过 (t_1, y_1) 做过该点的积分曲线的切线交直线 $t = t_2$ 于 (t_2, y_2), $\cdots\cdots$ 依次做下去便得到近似解 $y_1, y_2, y_3, \cdots$

显然, 步长 h 越小, 折线段会越靠近过 (t_0, y_0) 的那条解曲线.

由适定性的保证

最简单的单步法—Euler 法

还可以从积分形式来推导 **Euler 法:**

由初值问题 (9.1) 的等价积分形式

(9.5) $$y(t) = y_0 + \int_{t_0}^{t} f\big(s, y(s)\big)ds,$$

可得 $y(t_1) = y_0 + \int_{t_0}^{t_1} f\big(s, y(s)\big)ds$.

然后如果用'左矩积分公式'计算上面积分即得

$$y_1 = y_0 + hf(t_0, y_0),$$

此即 **Euler 法.**

隐式 Euler 法

如果在 (9.5) 中使用右矩公式, 我们可以得到

(9.6) $$y_{n+1} = y_n + hf(t_{n+1}, y_{n+1}).$$

这称为隐式 Euler 法或称为后退 Euler 法. 如果 $f(t, y)$ 为 y 的非线性函数, 那么上面隐式 Euler 法就需要迭代求解. 可用显式 Euler 法给出初值, 然后进行 Picard 迭代: 稳定性会更好一些

$$\begin{aligned} y_{n+1}^{(0)} &= y_n + hf(t_n, y_n), \\ y_{n+1}^{(1)} &= y_n + hf(t_n, y_{n+1}^{(0)}), \\ &\cdots\cdots \\ y_{n+1}^{(k+1)} &= y_n + hf(t_n, y_{n+1}^{(k)}) \end{aligned}$$

当 $\left|y_{n+1}^{(k+1)} - y_{n+1}^{(k)}\right| < \varepsilon$ (ε 事先给定) 时停止迭代, 令 $y_{n+1} = y_{n+1}^{(k+1)}$.

隐式 Euler 法

对于以上简单迭代方法

$$y_{n+1}^{(k+1)} = y_n + hf\left(t_n, y_{n+1}^{(k)}\right), \quad k = 0, 1, \cdots \tag{9.7}$$

由于我们一般都假设 $f(t,y)$ 关于 y 满足 **Lipschitz** 条件,

$$|f(t,y_1) - f(t,y_2)| \le L|y_1 - y_2|$$

将(9.7)与(9.6)式相减, 我们得到　　**利用**y(n+1)的隐式表达方式

$$\begin{aligned}
\left|y_{n+1} - y_{n+1}^{(k+1)}\right| &= h\left|f(t_{n+1}, y_{n+1}) - f\left(t_n, y_{n+1}^{(k)}\right)\right| \\
&\le Lh\left|y_{n+1} - y_{n+1}^{(k)}\right| \le \cdots \\
&\le (Lh)^{k+1}\left|y_{n+1} - y_{n+1}^{(0)}\right|
\end{aligned}$$

也就是说只要 $Lh < 1$, 上述迭代方法 (9.7) 会收敛.

梯形法与改进的 Euler 法

从上面积分形式推导的过程可以知道, 显然用高精度的数值积分格式应该可以得到更好的差分方法!

如果我们使用梯形公式计算 (9.5) 中的积分, 得到

$$y_{n+1} = y_n + \frac{h}{2}[f(t_n, y_n) + f(t_{n+1}, y_{n+1})], \quad n \geq 0,$$

此即梯形法. 显然这也是一种隐式格式, 一般需要迭代求解. 如果我们仍然使用显式 **Euler** 法提供初值进行 **Picard** 迭代, 有

迭代不是必须的，有些情况下可以隐式格式可以求解

$$\text{(9.8)} \quad \begin{cases} y_{n+1}^{(0)} = y_n + hf(t_n, y_n), \\ y_{n+1}^{(k+1)} = y_n + \frac{h}{2}\left[f(t_n, y_n) + f\left(t_{n+1}, y_{n+1}^{(k)}\right)\right], \quad k = 0, 1, \cdots \end{cases}$$

类似于上面的推导可知, 当 $\frac{Lh}{2} < 1$ 时以上迭代是收敛的.

梯形法与改进的 Euler 法

上述迭代过程中每一步都需要计算非线性函数 $f(t,y)$ 的值, 如果每一时间步都计算多次的话，通常来说这个计算量较大. 而我们知道通常时间步长 h 并不大, 所以我们用 **Euler** 法提供初值 (预估) 之后, 可以迭代(校正)一次即可, 这便是改进的 **Euler** 法:

$$\begin{cases} \bar{y}_{n+1} &= y_n + hf(t_n, y_n), \\ y_{n+1} &= y_n + \frac{h}{2}[f(t_n,y_n) + f(t_{n+1}, \bar{y}_{n+1})]. \end{cases}$$

把 $\bar{y}_{n+1}$ 代入第二式也可以写成一步的形式

$$y_{n+1} = y_n + \frac{h}{2}[f(t_n,y_n) + f(t_{n+1}, y_n + hf(t_n,y_n))]. \tag{9.9}$$

这个改进的 **Euler** 法也是一种显式单步法.

几种单步法的对比

我们来看一个例子:

例 9.1 (用以上几种单步法解初值问题)

$$\begin{cases} y'(t) = -y(t), & t > 0; \\ y(0) = 1. \end{cases}$$

初值问题的解为 $y(t) = e^{-t}$.

取步长 $h = 0.1$, 从 $t = 0$ 计算到 $t = 4$.

t_n	**Euler** y_n	改进 **Euler** y_n	梯形法 y_n	$y(t_n)$
1.0	**3.4867E-1**	**3.6854E-1**	**3.6757E-1**	**3.6788E-1**
2.0	**1.2158E-1**	**1.3582E-1**	**1.3511E-1**	**1.3534E-1**
3.0	**4.2391E-2**	**5.0056E-2**	**4.9663E-2**	**4.9787E-2**
4.0	**1.4781E-2**	**1.8447E-2**	**1.8255E-2**	**1.8316E-2**

可以看到改进Euler与梯形法差别不大, 但是都远好于Euler法.

高阶单步法

显然上面给出的几种方法都还不算是精度太高的方法 (上例中如果继续计算到 $t=5$ 有效位会继续减少). 如何系统化地构造更高精度的格式呢?

从 **Euler** 法的构造来类推的话, 最容易想到的当然是用 **Taylor** 级数展开截断办法:

$$y(t+h)=y(t)+hy'(t)+\frac{h^2}{2}y''(t)+\cdots+\frac{h^p}{p!}y^{(p)}(t)+\mathcal{O}(h^{p+1}),$$

如果我们舍去高阶项 $\mathcal{O}(h^{p+1})$, 便可以得到一种差分格式. 下面主要看如何计算 $y^{(p)}(t)$? 因为 $y(t)$ 为初值问题的解, 那么有

高阶单步法

$$
\begin{aligned}
y'(t) &= f\big(t, y(t)\big) \\
y''(t) &= \frac{df}{dt} = \frac{\partial f(t,y)}{\partial t} + \frac{\partial f(t,y)}{\partial y}\frac{dy}{dt} = \frac{\partial f}{\partial t}(t,y) + \frac{\partial f}{\partial y}(t,y)f(t,y) \\
\cdots & \quad \cdots
\end{aligned}
$$

如果我们截断到 $p=1$, 得到的就是显式 **Euler** 法:

$$
y_{n+1} = y_n + hf(t_n, y_n).
$$

但是如果 $p>1$, 我们就需要计算 $f(t,y)$ 的各种偏导数了.

计算高阶偏导数显然工作量太大, 而且我们有时候仅知道 f 的一些离散值列表, 那么计算这些偏导数的精度也很有限了.

因此我们实际计算中很少采用这个办法.

Runge-Kutta 方法—数值积分

我们下面主要从数值积分角度推导一下高阶单步法 — **Runge-Kutta 方法:**

将 (9.1) 中的方程改写成积分形式

$$y(t) = y(t_0) + \int_{t_0}^{t} f\big(s, y(s)\big)ds. \tag{9.10}$$

这样如果我们知道了 y_n, 可以如下计算 y_{n+1}:

$$y(t_{n+1}) = y(t_n) + \int_{t_n}^{t_{n+1}} f\big(s, y(s)\big)ds.$$

问题是如何计算上面的定积分. 要想获得高精度近似解, 一般在 $[t_n, t_{n+1}]$ 上引入几个新节点: $t_n \le s_1 \le s_2 \le \cdots \le s_m \le t_{n+1}$. 记 $s_j = t_n + a_j h$.

Runge-Kutta 方法—数值积分

如果知道了 f 在 $(s_j, y(s_j))$ 上的值 K_j, 可以用公式

$$y_{n+1} = y_n + h\sum_{j=1}^{m} c_j K_j$$

来计算 y_{n+1}, 其中

$$\int_{t_n}^{t_{n+1}} f\big(s, y(s)\big)ds \approx h\sum_{j=1}^{m} c_j K_j = h\sum_{j=1}^{m} c_j f\big(s_j, y(s_j)\big)$$

为数值积分公式. 即用 f 在 $s_j(j=1,\cdots,m)$ 上的值构造 $m-1$ 次插值多项式, 然后代替 f 积分便求得上面定积分的近似值.

如何计算 $y(s_j) = y(t_n) + \int_{t_n}^{t_n+a_jh} f(s, y(s))ds$ 又有多个选择:

如果使用显式格式计算，即用 f 在 $s_1,\cdots,s_{j-1}$ 上的值做插值来计算上述积分: $y(s_j) \approx y_n + h\sum_{l=1}^{j-1} b_{jl} f\big(s_l, y(s_l)\big)$.

Runge-Kutta 方法—数值积分

代入 K_j 的表达式即得: $K_j = f(t_n + a_j h, y_n + h\sum_{l=1}^{j-1} b_{jl}K_l)$, $j = 1, 2, \cdots, m$.

由显式计算的要求, 有 $a_1 = 0$, 即 $s_1 = t_n$, $K_1 = f(t_n, y_n)$.

最终得到显式 **Runge-Kutta** 格式为

$$
(9.11)\quad \begin{cases} y_{n+1} = y_n + h\sum_{j=1}^{m} c_j K_j, \\ K_j = f(t_n + a_j h, y_n + h\sum_{l=1}^{j-1} b_{jl}K_l), \quad j = 1, \cdots, m. \end{cases}
$$

这相当于在(9.3) 中取 $\phi_f = \sum_{j=1}^{m} c_j K_j$.

Runge-Kutta 方法—待定系数法

我们也可以基于待定系数法和 **Taylor** 展开来推导 **R-K** 格式.

我们假设 **R-K** 格式有 (9.11) 形式. 下面问题是如何确定系数 $\{a_j, b_{jl}, c_j\}$? 由 **Taylor** 展开 (假设 $0 < h \ll 1$):

$$y(t_{n+1}) = y(t_n) + hy'(t_n) + \cdots + \frac{h^p}{p!}y^{(p)}(t_n) + \mathcal{O}(h^{p+1}).$$

若令 $\phi(t, y(t); h) = \sum_{j=1}^{p} \frac{h^{j-1}}{j!} \frac{d^{j-1}f(t, y(t))}{dt^{j-1}}$, 则有

$$y(t_{n+1}) = y(t_n) + h\phi(t_n, y(t_n); h) + \mathcal{O}(h^{p+1}).$$

舍去高阶项便得到计算公式

(9.12) $$y_{n+1} = y_n + h\phi(t_n, y_n; h).$$

Runge-Kutta 方法－－待定系数法

当然，我们已经说了上面的 (9.12) 一般不使用. 但是我们可以通过比较 (9.11) 与 (9.12) 来确定系数 $\{a_j, b_{jl}, c_j\}$.

我们将 (9.11) 中的 K_j 也在 (t_n, y_n) 处做**Taylor**展开, 也能得到如下形式

$$y_{n+1} = y_n + h\widetilde{\phi}(t_n, y_n; h).$$

其中 $\widetilde{\phi}(t_n, y_n; h)$ 也按照 h 的升幂次序排列, 其中的系数自然包含了 $\{a_j, b_{jl}, c_j\}$. 将之与 (9.12) 中的 ϕ 关于 h 的相应幂次项的系数做比较, 令其相同便可求出 $\{a_j, b_{jl}, c_j\}$.

下面来看一个例子.

Runge-Kutta 方法——待定系数法

例 9.2

$m=1$ 即为最简单的 *Euler* 法: $y_{n+1}=y_n+hf(t_n,y_n)$.

$m=2$ 时, 比较系数有 $c_1+c_2=1$, $a_2c_2=\frac{1}{2}$.

这自然有无穷多组解. 常用的有以下两种 ($f_n=f(t_n,y_n)$):

1. 显式中点格式: $c_1=0, c_2=1, a_2=\frac{1}{2}$, 即
   $y_{n+1}=y_n+hf(t_n+\frac{h}{2}, y_n+\frac{h}{2}f_n)$.
2. 改进 *Euler* 法: $c_1=c_2=\frac{1}{2}, a_2=1$, 即
   $y_{n+1}=y_n+\frac{h}{2}[f_n+f(t_n+h, y_n+hf_n)]$.

$m=4$ 时, 首先由显式要求及相容性有

$a_1=0$, $a_2=b_{21}$, $a_3=b_{31}+b_{32}$, $a_4=b_{41}+b_{42}+b_{43}$.

Runge-Kutta 方法——待定系数法

再比较 ϕ 与 $\widetilde{\phi}$ 的系数(利用了前面的公式)有

$$\sum_{j=1}^{4} c_j a_j^{l-1} = \frac{1}{l}, l = 1, 2, 3, 4; \quad c_3 a_2 b_{32} + c_4(b_{42} a_2 + b_{43} a_3) = \frac{1}{6},$$

$$c_3 b_{32} a_2^2 + c_4 b_{43} a_3^2 = \frac{1}{12}, \quad c_4 b_{43} a_2 b_{32} = \frac{1}{24},$$

$$c_3 a_3 a_2 b_{32} + c_4 a_4 (b_{42} a_2 + b_{43} a_3) = \frac{1}{8}.$$

这显然也有无穷多组解. 取一组特殊值

$$\mathbf{a} = \left(0, \frac{1}{2}, \frac{1}{2}, 1\right), \quad \mathbf{c} = \left(\frac{1}{6}, \frac{1}{3}, \frac{1}{3}, \frac{1}{6}\right),$$

$$b_{21} = \frac{1}{2},\ b_{31} = 0,\ b_{32} = \frac{1}{2},\ b_{41} = 0 = b_{42},\ b_{43} = 1.$$

此即经典 4 级 Runge-Kutta 方法.

隐式 Runge-Kutta 方法

我们也可以采用隐式的**Runge-Kutta**方法, 即用 f 在所有 s_l 上的值做插值来计算 $y(s_j) = y(t_n) + \int_{t_n}^{s_j} f\big(t, u(t)\big)dt$ 中的积分:

$$y(s_j) = y(t_n + a_j h) \approx y_n + h\sum_{l=1}^{m} b_{jl} f\big(s_l, y(s_l)\big),$$

代入 K_j 的表达式即有以下隐式**R-K**算法:

$$(9.13) \quad \begin{cases} y_{n+1} &= y_n + \sum_{j=1}^{m} c_j K_j, \\ K_j &= f\Big(t_n + h a_j, y_n + h\sum_{l=1}^{m} b_{jl} K_l\Big), \quad j = 1, \cdots, m. \end{cases}$$

当然隐式**R-K**格式的缺点在于无法显式计算上面的 K_j, 通常需要求解一个非线性方程组才能得到.

隐式 Runge-Kutta 方法

虽然我们也可以用待定系数法来得到隐式 **R-K** 格式, 以下我们仅从数值积分角度来推导.

我们既然用隐式格式，自然希望格式的精度尽可能高, 从而我们希望数值积分公式的代数精度尽可能高. 这样一来我们应该采用 **Gauss** 型求积公式来计算积分:

(9.14) $$y_{t_n+1} = y(t_n)+\int_{t_n}^{t_{n+1}} f(s,y(s))ds \approx y_n + h\sum_{j=1}^{m} c_j K_j,$$

(9.15) $$y(s_j) = y(t_n)+\int_{t_n}^{s_j} f(s,y(s))ds \approx y_n+h\sum_{l=1}^{m} b_{jl}K_l,\ 1\le j\le m.$$

即$s_1,\cdots,s_m$是在区间$[t_n,t_{n+1}]$上的 m 次**Legendre**多项式零点.

隐式 Runge-Kutta 方法

有了高斯积分节点 s_j, 由 (9.14) 积分公式应该具有 $2m-1$ 阶代数精度, 可取 $f(t_n+\tau h)=(\tau h)^{l-1}, l=1,\cdots,m$ 代入 (9.14) 即得

(9.16)
$$\sum_{j=1}^{m} c_j a_j^{l-1} = \frac{1}{l}, \quad l=1,\cdots,m.$$

求解上述方程组便可得到 c_j. 类似由 (9.15) 具有 $2m-1$ 阶代数精度, 同样可得

(9.17)
$$\sum_{j=1}^{m} b_{ij} a_j^{l-1} = \frac{1}{l} a_i^l, \quad l=1,\cdots,m;\ i=1,\cdots,m.$$

同样求解上述方程组便可得到 b_{ij}.

隐式 Runge-Kutta 方法

即计算 a_j, b_{ij}, c_j 的步骤为:

1. **求出$[-1,1]$上m次Legendre多项式的零点 $\xi_j \Longrightarrow a_j = \frac{1+\xi_j}{2}$;**
2. **由 (9.16) 式计算出 c_j, $j = 1, \cdots, m$;**
3. **对 $i = 1, \cdots, m$ 由 (9.17) 式计算出 b_{ij}, $j = 1, \cdots, m$.**

例 9.3

$m = 1$, $P_1(t) = t \Longrightarrow a_1 = \frac{1}{2}$, $c_1 = 1$, $b_{11} = \frac{1}{2}$. 此即隐式中点格式:

$$y_{n+1} = y_n + hf\left(t_n + \frac{h}{2}, \frac{1}{2}(y_n + y_{n+1})\right).$$

$m = 2$, $P_2(t) = \frac{3t^2-1}{2} \Longrightarrow$

$$a_{1,2} = \frac{1}{2} \pm \frac{\sqrt{3}}{6}, \quad c_{1,2} = \frac{1}{2}, \quad B = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} - \frac{\sqrt{3}}{6} \\ \frac{1}{4} + \frac{\sqrt{3}}{6} & \frac{1}{4} \end{pmatrix}$$

单步法的相容性、收敛性与稳定性

下面我们来分析一下这些单步法的性质.

定义 9.2 (截断误差)

对于公式 $y_{n+1}=y_n+h\phi(t_n,y_n;h)$, 令

$$R_n=L[t_n,y(t_n);h]=y(t_{n+1})-y(t_n)-h\phi(t_n,y(t_n);h)$$

为单步法的 局部截断误差, 即假设之前函数值准确的前提下，用该单步法计算一步产生的误差 (或者说用数值积分公式计算积分 $\int_{t_n}^{t_{n+1}} f(s,y(s))ds$ 带来的误差). 进一步的，令 $\varepsilon_n=y(t_n)-y_n$ (为近似解与真解之间的误差) 称为整体截断误差.

一般来说直接估计整体截断误差不太容易, 我们可以通过局部截断误差来估计.

单步法的相容性、收敛性与稳定性

例 9.4 (Euler法的误差分析)

对于 *Euler* 法, 利用 *Taylor* 展开$R_n = \frac{h^2}{2}y''(\xi)$, 其中 $\xi \in (t_n, t_{n+1})$.
假设 $y(t)$ 充分光滑, 即有局部截断误差为 $R_n = \mathcal{O}(h^2)$ 二阶.
下面看如何得到整体截断误差？ 由

$$\begin{aligned} y(t_{n+1}) &= y(t_n) + hf(t_n, y(t_n)) + R_n, \\ y_{n+1} &= y_n + hf(t_n, y_n), \end{aligned}$$

相减得到 $\varepsilon_{n+1} = \varepsilon_n + h[f(t_n, y(t_n)) - f(t_n, y_n)] + R_n$.
由f关于y满足 *Lipschitz* 条件, 以及令 $R = \max_n |R_n| = \mathcal{O}(h^2)$

$$\begin{aligned} \Longrightarrow \quad & |\varepsilon_{n+1}| \le |\varepsilon_n| + Lh|\varepsilon_n| + |R_n| \le (1+Lh)|\varepsilon_n| + R \le \cdots \\ \le \quad & (1+Lh)^{n+1}|\varepsilon_0| + R\sum_{j=0}^{n}(1+Lh)^j \end{aligned}$$

单步法的相容性、收敛性与稳定性

$\Longrightarrow |\varepsilon_{n+1}| \le (1+Lh)^{n+1}|\varepsilon_0| + R\frac{(1+Lh)^{n+1}-1}{Lh}$. 再由 $(1+\frac{x}{n})^n \le e^x$

$$\Longrightarrow (1+Lh)^{n+1} = \left(1+L\frac{T}{N}\right)^{n+1} \le \left(1+L\frac{T}{N}\right)^{N} \le e^{LT}$$

$$\Longrightarrow |\varepsilon_{n+1}| \le |\varepsilon_0|e^{LT} + \frac{R}{Lh}\left(e^{LT}-1\right).$$

设 $\varepsilon_0=0$, 由 $R=\mathcal{O}(h^2) \Longrightarrow \varepsilon_n = \mathcal{O}(h)$. 即 **Euler** 法整体误差为一阶. □

类似可以得到一般常用的 $m(\le 4)$ 级显式 **Runge-Kutta** 方法整体截断误差为 m 阶,

m 级隐式 **Runge-Kutta** 方法整体截断误差为 $2m$ 阶.

单步法的相容性、收敛性与稳定性

另外衡量一种差分方法好坏的性质是，我们显然希望该差分格式是原微分方程的一个好的近似. 相容性即为描述此性质而引入的.

定义 9.3 (相容性)

如果 $\lim_{h\to 0}\frac{1}{h}L[t,y(t);h]=0$ (即差分格式收敛到常微方程), 则称该格式与原问题相容. 一般如果整体截断误差为 $\varepsilon_n=\mathcal{O}(h^p)$, 就称之为 p 阶相容格式.

不失一般性，我们以下假设 $f(t,y)$ 在其定义域上是一致连续和一致有界的. 我们有以下定理结论

单步法的相容性、收敛性与稳定性

定理 9.2

单步法 $y_{n+1}=y_n+h\phi(t_n,y_n;h)$ 相容的充分必要条件是

(9.18) $$\lim_{h\to 0}\phi(t,y;h)=f(t,y).$$

$\triangleleft$ 由方程有 $y(t+h)-y(t)=\displaystyle\int_0^h f\big(t+s,y(t+s)\big)ds$.

利用 f 的一致有界性, 显然有 $h\to 0$ 时, 上式右式趋于0.

这样有 $h\to 0$ 时, 再利用 f 的一致连续性, 可以得到

$$\begin{aligned}\frac{1}{h}L[t,y;h]+\phi(t,y;h)-f(t,y) &= \frac{1}{h}[y(t+h)-y(t)]-y'(t)\\ &= \frac{1}{h}\int_0^h [y'(t+s)-y'(t)]ds\to 0\end{aligned}$$

从而证明了定理结论. $\triangleright$

单步法的相容性、收敛性与稳定性

但是实际上初值的误差 ε_0 一般也不为零, 例如是测量数据的话可能包含测量误差, 如果是用其它方法计算得到的，也可能包含舍入误差等. 这样我们需要考虑此误差随计算过程的传递是否会带来恶性影响.

定义 9.4

如果初值误差带来的影响可以被初始误差的某个常数倍控制，则称该方法对初值是稳定的.

对于线性问题，如果使用一个相容的、稳定的格式求解，离散解会收敛到真解（在忽略每一步计算的舍入误差影响下）. 稍后我们再考虑每一步舍入误差带来的影响。

单步法的相容性、收敛性与稳定性

对于**Euler**法，假设有不同初值(或者相当于初值有扰动):

$$y_{n+1} = y_n + hf(t_n, y_n), \quad z_{n+1} = z_n + hf(t_n, z_n).$$

令 $e_n = y_n - z_n$, 两式相减有 $e_{n+1} = e_n + h[f(t_n, y_n) - f(t_n, z_n)]$, 利用 f 的**Lipschitz**连续性条件得到

$$|e_{n+1}| \le (1 + Lh)|e_n| \le (1 + Lh)^{n+1}|e_0| \le e^{LT}|e_0|.$$

即 **Euler** 法对初值是稳定的。

类似可证明其他 **R-K** 方法在此条件下也是对初值稳定的.

单步法的相容性、收敛性与稳定性

定义 9.5 (收敛性)

设 $T = Nh$, $t_n = nh$, 令 $E(h) = \max\limits_{0\le n\le N} |\varepsilon_n| \equiv \max\limits_{0\le n\le N} |y(t_n) - y_n|$.

如果 $\lim\limits_{h\to 0} E(h) = 0$, 就称该单步法是**收敛的**. 如果 $\forall h \in (0, h_0]$, 存在常数 $C > 0$ s.t. $E(h) \le Ch^p$, 就称该单步法是 ***p* 阶收敛的**.

关于收敛性、相容性的关系我们有以下定理结论:

定理 9.3

设单步法 $y_{n+1} = y_n + h\phi(t_n, y_n; h)$ 的增量函数 $\phi(t, y; h)$ 是连续函数, 且关于 y 满足 *Lipschitz* 条件, 即

$$\exists L > 0,\ \forall y_1, y_2 \in \mathbb{R}, \quad |\phi(t, y_1; h) - \phi(t, y_2; h)| \le L\,|y_1 - y_2|\,.$$

那么该单步法收敛的充要条件是该单步法是相容的.

单步法的相容性、收敛性与稳定性

◁ **先证明从相容性可以得到收敛性. 即假设单步法相容, 那么**

$$
\begin{aligned}
\varepsilon_{n+1}-\varepsilon_n &= y(t_{n+1})-y_{n+1}-[y(t_n)-y_n] \\
&= [y(t_{n+1})-y(t_n)]-(y_{n+1}-y_n) \\
&= h\phi(t_n,y(t_n);h)+L[t_n,y(t_n);h]-h\phi(t_n,y_n;h) \\
&= h\big[\phi(t_n,y(t_n);h)-\phi(t_n,y_n;h)\big]+L[t_n,y(t_n);h]
\end{aligned}
$$

利用相容性有 $L[t_n,y(t_n);h]=o(h)$. 再利用 $\phi(t,y;h)$ 关于 y 的 Lipschitz 连续性有

$$
|\varepsilon_{n+1}-\varepsilon_n| \le hL\,|y(t_n)-y_n|+o(h) \Longrightarrow |\varepsilon_{n+1}| \le (1+hL)\,|\varepsilon_n|+o(h)
$$

$$
\Longrightarrow |\varepsilon_n| \le (1+hL)^n|\varepsilon_0|+o(h)\sum_{k=0}^{n-1}(1+hL)^k = (1+hL)^n|\varepsilon_0|+o(h)\frac{(1+hL)^n-1}{hL}
$$

单步法的相容性、收敛性与稳定性

由 $\varepsilon_0 = 0$ (现在没有考虑舍入误差), 及 $(1+Lh)^n \le e^{LT}$ 即得 $h \to 0$ 时 $E(h) \to 0$.

再看由收敛性证相容性. 假设单步法收敛, 即 $E(h) \to 0$, 这说明任给 $t_0, y_0 \in \mathbb{R}$, 有 $h \to 0$ 时,

(9.19) $$y_{n+1} = y_n + h\phi(t_n, y_n; h), \quad y_0 = y(t_0)$$

的解收敛到常微分方程初值问题

$$y'(t) = f(t, y), \quad y(t_0) = y_0$$

的解. 这样令 $g(t,y) = \phi(t,y;0)$, 那么利用定理9.2的结论知, 单步法 (9.19) 与以下初值问题相容

(9.20) $$y'(t) = g(t, y), \quad y(t_0) = y_0$$

单步法的相容性、收敛性与稳定性

因而由上面推导知单步法(9.19)的解收敛到上面初值问题(9.20)的解.

自然这两个初值问题的解应该是相同的, 即 $\forall t_0, y_0 \in \mathbb{R}$, $f(t_0, y_0) = g(t_0, y_0)$.

这样由 $\phi(t, x; h)$ 的连续性知

$$\lim_{h \to 0} \phi(t, y; h) = f(t, y).$$

这样再利用定理 **9.2** 可以得到单步法(9.19)的相容性. $\triangleright$

目录 I

线性多步法

从单步法的构造中我们可以看到, 每一步我们需要通过插值来计算 K_j ($\approx f\big(s_j, y(s_j)\big)$). 当我们已经获得了多个节点的值 y_j ($j = 1, 2, \cdots, k-1$) 之后, 我们可以尝试利用这些现成的函数值来做插值, 而不需要再引入新的节点 s_j. 这就是多步法, 相当于之前的增量函数取成

$$\phi_f = \sum_{j=0}^{k} \beta_j f(t_{n+j}, y_{n+j}),$$

即线性多步法为

$$\sum_{j=0}^{k} \alpha_j y_{n+j} = h \sum_{j=0}^{k} \beta_j f(t_{n+j}, y_{n+j}). \tag{9.21}$$

这里一般设 $\alpha_k \neq 0$. 如果 $\beta_k = 0$, 为显式格式; 反之若 $\beta_k \neq 0$, 则为隐式格式.

线性多步法—待定系数法

类似于前面单步法，我们也可令

$$L[y(t_n);h]=\sum_{j=0}^{k}[\alpha_j y(t_{n+j})-h\beta_j f(t_{n+j},y(t_{n+j}))]$$

为多步法 (9.21) 的局部截断误差. 将之在t_n做 **Taylor** 展开, 写成

$$L[y(t_n);h]=c_0y(t_n)+c_1hy'(t_n)+\cdots+c_ph^py^{(p)}(t_n)+\cdots$$

易得

$$\text{(9.22)}\quad \begin{cases} c_0=\alpha_0+\alpha_1+\cdots+\alpha_k \\ c_1=\alpha_1+2\alpha_2+\cdots+k\alpha_k-(\beta_0+\beta_1+\cdots+\beta_k) \\ \cdots\cdots \\ c_p=\frac{\alpha_1+2^p\alpha_2+\cdots+k^p\alpha_k}{p!}-\frac{\beta_1+2^{p-1}\beta_2+\cdots+k^{p-1}\beta_k}{(p-1)!} \\ \cdots\cdots \end{cases}$$

线性多步法—待定系数法

若(9.22)中 $c_0 = \cdots = c_p = 0$, $c_{p+1} \neq 0$, 则称多步法 (9.21) 是 p 阶方法, 局部截断误差为 $L[y(t_n); h] = c_{p+1}h^{p+1}y^{(p+1)}(\xi_n)$.

一般我们取 $\alpha_k = 1$, 若在 (9.22) 中令 $c_0 = c_1 = \cdots = c_p = 0$, 而且恰好 $p + 1 = 2k + 1$, 则可解出 $\alpha_0, \cdots, \alpha_{k-1}$ 和 $\beta_0, \cdots, \beta_k$. 这样就是用待定系数法来确定 (9.21) 中的系数.

$k = 1$ 时, 若只要求 $p = 1$, 即 $\alpha_0 + \alpha_1 = 0$, $\alpha_1 - \beta_0 - \beta_1 = 0$, 我们有$\alpha_0 = -1$, $\beta_0 + \beta_1 = 1$, 即 β_0, β_1 有无穷多种选取方法.

如果取 $\beta_0 = 1,\ \beta_1 = 0$, 此即显式 Euler 法; 如果取 $\beta_0 = 0,\ \beta_1 = 1$, 此即隐式 Euler 法.

若要求 $p = 2$, 则得到 $\alpha_0 = -1$, $\beta_0 = \beta_1 = \frac{1}{2}$, 此即梯形法.

线性多步法—计算步骤

当然, 多步法无法自行启动, 需要由其他单步法给出初值 $y_1, y_2, \cdots, y_{k-1}$. 另外如果 $\beta_k \neq 0$, 即对于隐式格式,

$$y_{n+k} = -\sum_{j=0}^{k-1} \alpha_j y_{n+j} + h \sum_{j=0}^{k} \beta_j f(t_{n+j}, y_{n+j}),$$

一般还需要迭代才能求解. 将之改写成

$y_{n+k} = h\beta_k f(t_{n+k}, y_{n+k}) + g_{n,k}$, 我们可以用以下迭代格式

$$y_{n+k}^{(m+1)} = h\beta_k f(t_{n+k}, y_{n+k}^{(m)}) + g_{n,k},$$

$y_{n+k}^{(0)}$) 可以由其他显式方法给出.

类似于之前的推导可知, 该迭代方法的收敛性条件为 $h|\beta_k|L < 1$, 其中 L 为 $f(t, y)$ 关于 y 的 **Lipschitz** 常数.

线性多步法—数值积分法

当然也可以用数值积分办法来推导. 但是要注意, 数值积分推导方法只能得到部分形式的线性多步法.

我们把常微分方程 (9.1) 在 $[t_{n+k-l}, t_{n+k}]$ 上积分可以得到

$$y(t_{n+k}) = y(t_{n+k-l}) + \int_{t_{n+k-l}}^{t_{n+k}} f\big(t, y(t)\big)dt,$$

然后如果用等距节点的各种插值多项式来代替被积函数 $f\big(t, y(t)\big)$, 便可以得到各种各样的线性多步法.

如取 $k = l = 4$, 那么得到

$$y(t_{n+4}) = y(t_n) + \int_{t_n}^{t_{n+4}} f\big(t, y(t)\big)dt.$$

线性多步法—数值积分法

然后如果我们取 $F(t) \equiv f\big(t, y(t)\big)$ 在 $t_{n+1}, t_{n+2}, t_{n+3}$ 三点上的值进行二次插值, 得到

$$L_2(t) = \sum_{j=0}^{2} f\big(t_{n+1+j}, y(t_{n+1+j})\big) l_j(t),$$

代替 $F(t)$ 求积分我们得到

$$\int_{t_n}^{t_{n+4}} L_2(t)dt = \frac{4h}{3}[2f_{n+1} - f_{n+2} + 2f_{n+3}],$$ 这里 $f_m = f(t_m, y_m)$.

即得到格式 $y_{n+4} = y_n + \frac{4h}{3}[2f_{n+1} - f_{n+2} + 2f_{n+3}]$, 此格式称为 **Milne** 方法. 易见其局部截断误差为

$$T_{n+4} = \frac{14}{45}h^5 y^{(5)}(t_n) + \mathcal{O}(h^6),$$

即 **Milne** 方法为四阶方法.

线性多步法—Adams 多步法

如果取 $l=1$, 得到

$$y(t_{n+k}) = y(t_{n+k-1}) + \int_{t_{n+k-1}}^{t_{n+k}} f\big(t, y(t)\big)\,dt,$$

如果利用 f 在 $t_{n+k-1}, \cdots, t_n$ 上的值构造一个 $k-1$ 阶插值多项式 $L_{k-1}(t)$ 来代替 f 求积分，便得到 **Adams** 外插法格式:

显式格式

$$y_{n+k} = y_{n+k-1} + h\sum_{j=0}^{k-1} \beta_j f_{n+j}. \tag{9.23}$$

这称为显式 **Adams** 格式 (外插法).

Adams 多步法

如果利用 f 在 $t_{n+k}, \cdots, t_n$ 上的值构造一个 k 阶插值多项式 $\widetilde{L}_k(t)$ 来代替 f 求积分，便得到 **Adams** 内插法格式:

$$y_{n+k} = y_{n+k-1} + h\sum_{j=0}^{k}\widetilde{\beta}_j f_{n+j}. \tag{9.24}$$

这称为隐式 **Adams** 格式 **(**内插法**).**

Adams 线性多步法是使用较多的一类线性多步法**.**

比较 (9.23) 与 (9.23) 的构造过程易见**,** k 步隐式 **Adams** 方法比 k 步显式 **Adams** 方法用了更高一阶的多项式做插值**,** 因而其精度自然会高一阶**.** 但是隐式格式的缺点是需要迭代求解**.**

线性多步法的相容性与稳定性

由前面局部截断误差的定义，可以说，如果

$$L[y(t_n);h]=\sum_{j=0}^{k}[\alpha_j y(t_{n+j})-h\beta_j f(t_{n+j},y(t_{n+j}))]=c_{p+1}h^{p+1}y^{(p+1)}(\xi_n),$$

则称多步法 (9.21) 与原初值问题是 p 阶相容的.

如果引入多步法 (9.21) 的第一、第二特征多项式:

(9.25) $$\rho(\lambda)=\sum_{j=0}^{k}\alpha_j\lambda^j,\quad \sigma(\lambda)=\sum_{j=0}^{k}\beta_j\lambda^j.$$

当 $h\to 0$ 时, 若

$$\frac{1}{h\rho'(1)}\sum_{j=0}^{k}[\alpha_j y(t_{n+j})-h\beta_j f(t_{n+j},y(t_{n+j}))]-[y'(t_n)-f(t_n,y(t_n))]=o(1),$$

即差分方程收敛到原微分方程, 则称(9.21)与(9.1)相容.

线性多步法的相容性与稳定性

可以验证 (将 $y_{t_{n+j}}$, $f(t_{n+j}, y_{t_{n+j}})$ 在 t_n 处 **Taylor** 展开):

多步法 (9.21) 与 (9.1) 相容 $\Longleftrightarrow \rho(1)=0$, 且 $\rho'(1)=\sigma(1)$.

下面我们来看其对初值的稳定性.

定义 9.6

称多步法 (9.21) 是对初值稳定的, 是指若存在不依赖于 h 的常数 C 及 $h_0, \varepsilon > 0$, s.t. $\forall h \in (0, h_0)$ 及任取两组不同初值得到的解 $\{u_n\}_{n=0}^N$ 和 $\{u_n\}_{n=0}^N$, 当 $\max\limits_{0\le j\le k-1} |u_j - v_j| < \varepsilon$ 时, 就有

$$\max_{nh\le T} |u_n - v_n| \le C \max_{0\le j\le k-1} |u_j - v_j|. \tag{9.26}$$

换句话说，只要步长充分小，多步法的解可以连续依赖于初值.

利用以下Gronwall不等式, 我们可以得到多步法的稳定性定理.

Gronwall 不等式

引理 9.1 (连续情形 Gronwall 不等式)

设 $\eta(t)$ 对 $t \in [a,b]$ 连续且满足:

$$\exists \alpha, \beta \geq 0, \forall t \in [a,b], \quad |\eta(t)| \leq \beta + \alpha \int_a^t |\eta(\tau)| d\tau.$$

则有 $|\eta(t)| \leq \beta e^{\alpha(t-a)}, \quad \forall t \in [a,b]$.

$\lhd$ **先设** $\beta > 0$. **令** $\xi(t) = \beta + \alpha \int_a^t |\eta(\tau)| d\tau \Longrightarrow |\eta(t)| \leq \xi(t)$. **且** $\xi'(t) = \alpha|\eta(t)| \leq \alpha\xi(t)$. **结合** $\xi(t) \geq \beta > 0 \Longrightarrow \frac{\xi'}{\xi} \leq \alpha$ **积分并利用** $\xi(a) = \beta$ **即得** $|\eta(t)| \leq \xi(t) \leq \beta e^{\alpha(t-a)}$

对 $\beta = 0$ **情形加一个摄动** $\varepsilon > 0$: $\beta_\varepsilon = \beta + \varepsilon > 0$ **即可证明.** $\rhd$

离散 Gronwall 不等式

引理 9.2 (离散情形 Gronwall 不等式)

设 $\alpha, \beta \geq 0$, $T, h > 0$, 序列 $\{\eta_n\}_{n\geq 0}$满足:

$$|\eta_n| \leq \beta + \alpha h \sum_{j=0}^{n-1} |\eta_j|, \quad \text{对 } n = k, k+1, \cdots, \text{且 } nh \leq T \text{ 成立},$$

则当 $n \geq k$ 且 $nh \leq T$时，有$|\eta_n| \leq e^{\alpha nh}(\beta + \alpha k h M_0)$, 其中 $M_0 = \max(|\eta_0|, \cdots, |\eta_{k-1}|)$.

◁ **将前面积分换成求和, 令 $\xi_n = \beta + \alpha h \sum_{j=0}^{n-1} |\eta_j|$,**

有$|\eta_n| \leq \xi_n$, 且 $\xi_n - \xi_{n-1} = \alpha h |\eta_{n-1}| \leq \alpha h \xi_{n-1}$

$\Longrightarrow \xi_n \leq (1 + \alpha h)\xi_{n-1} \leq (1 + \alpha h)^{n-k} \xi_k$.

又 $\xi_k = \beta + \alpha h \sum_{j=0}^{k-1} |\eta_j| \leq \beta + \alpha h k M_0$,

且 $(1 + \alpha h)^{n-k} \leq (1 + \alpha h)^n \leq e^{\alpha nh}$. 代入上面即得. ▷

线性多步法的相容性与稳定性

多步法 (9.21) 什么时候具有对初值的稳定性呢？

定理 9.4 (多步法对初值稳定的充要条件)

多步法 (9.21) 对初值稳定 $\Longleftrightarrow$ 第一特征多项式 $\rho(\lambda)$ 满足所谓的根条件：$\rho(\lambda)$ 的根都在单位圆内，且在单位圆周上的根全为单根.

$\lhd$ 先证必要性“$\Longrightarrow$”. 即假设多步法对初值稳定. 可无妨设 $f \equiv 0$.

先证明 $\rho(\lambda)$ 的根都在单位圆内. 用反证法，假设 $\rho(\lambda) = 0$ 有一解 μ s.t. $|\mu| > 1$. 考虑如下两组初值:

$$\left.\begin{array}{rcl} u_j &=& \delta\mu^j, \\ v_j &=& 0, \end{array}\right. \quad 0 \le j \le k-1, \quad \text{这里 } \delta > 0 \text{ 为待定常数}$$

显然有如果用 (9.21) 计算，有其解为:

$$u_n = \delta\mu^n, \quad v_n \equiv 0, \quad n = 0, 1, \cdots$$

线性多步法的相容性与稳定性

这样尽管初始误差 $\max_{0\le j<k}|u_j - v_j| = \delta|\mu|^{k-1}$ 可以很小, 但是当 $h\to 0, nh\le T$ 时, 会有 $n\to+\infty$, 因而 $|u_n - v_n| = \delta|\mu|^n \to +\infty$. 显然此时不稳定，与前提矛盾, 因此 $\rho(\lambda)$ 不会有模大于1 的根.

再证单位圆周上只能有单根: 依然用反证法, 假设 $\rho(\lambda)$ 有重根 $\mu\in\mathbb{C}$ s.t. $|\mu| = 1$ 且 $\rho(\mu) = \rho'(\mu) = 0$.

此时考虑如下初值选取：

$$\begin{array}{rcl} u_j &=& \delta j\mu^j, \\ v_j &=& 0, \end{array} \quad 0\le j\le k-1, \quad \text{这里 } \delta>0 \text{ 为常数}$$

易见如果用 (9.21) 计算，有其解为：

$$u_n = \delta n\mu^n, \quad v_n \equiv 0, \quad n = 0, 1, \cdots$$

线性多步法的相容性与稳定性

这样仍然有初始误差 $\max\limits_{0\le j<k}|u_j-v_j|=\delta(k-1)|\mu|^{k-1}=\delta(k-1)$ 可以足够小, 但 $|u_n-v_n|=\delta n\to+\infty$, 当 $n\to+\infty$. 与稳定性矛盾.

因此上面我们证明了: 如果多步法 (9.21) 对初值稳定，必然有$\rho(\lambda)$满足根条件.

下面看充分性:"$\Longleftarrow$", 即假设$\rho(\lambda)$满足根条件, 欲证明对初值的稳定性. 要证明这个，我们先要把差分方程的解结构弄清楚.

先看齐次差分方程: $\alpha_k y_{n+k}+\cdots+\alpha_0 y_n=0$. 其通解可以写成

(9.27) $$Y_n=\sum_{j=1}^{J}\sum_{l=1}^{r_j}c_{jl}n^{l-1}\xi_j^n,\quad (c_{jl}\text{ 为一些常数})$$

其中 ξ_j 为 $\rho(\lambda)=0$ 的 r_j 重根, 且 $\sum_{j=1}^{J}r_j=k$.

线性多步法的相容性与稳定性

对于非齐次差分方程: $\alpha_k z_{n+k} + \cdots + \alpha_0 z_n = \gamma_{n+k}$, 给了初值 $\{z_0, \cdots, z_{k-1}\}$ 后的解为 (设 $\alpha_k \neq 0$)

$$(9.28) \qquad z_n = \sum_{j=0}^{k-1} z_j Y_n^{(j)} + \frac{1}{\alpha_k} \sum_{j=0}^{n-k} \gamma_{j+k} Y_{n-j-1}^{(k-1)}, \qquad n = k, k+1, \cdots$$

这里 $\{Y_n^{(j)}\}$ 是齐次差分方程 $\begin{cases} \sum_{l=0}^{k} \alpha_l Y_{n+l}^{(j)} = 0, \\ Y_i^{(j)} = \delta_{ij}, \quad 0 \leq i < k; \end{cases}$ 的解.

设 u_n, v_n 为用 (9.21) 式计算的任两组解, 令 $\varepsilon_n = u_n - v_n$, 有

$$\sum_{j=0}^{k} \alpha_j \varepsilon_{n+j} = \gamma_{n+k} \equiv h \sum_{j=0}^{k} \beta_j [f(t_{n+j}, u_{n+j}) - f(t_{n+j}, v_{n+j})].$$

由前面结果知其解为: $\varepsilon_n = \sum_{j=0}^{k} \varepsilon_j Y_n^{(j)} + \frac{1}{\alpha_k} \sum_{j=0}^{n-k} \gamma_{j+k} Y_{n-j-1}^{(k-1)}$.

线性多步法的相容性与稳定性

由于 $\rho(\lambda)$ 满足根条件，即 $|\xi_j| \le 1$, 且若 $|\xi_j| = 1$ 必有 $r_j = 1$.
$\Longrightarrow \exists C > 0$, s.t. $|Y_n| \le C$, (当 $nh \le T$) 即齐次方程的解总是有界. $\Longrightarrow |\varepsilon_n| \le C\left(\sum_{j=0}^{k-1} |\varepsilon_j| + \frac{1}{|\alpha_k|}\sum_{j=0}^{n-k} |\gamma_{j+k}|\right)$, 当 $n \ge k$.

令 $M_0 = \max\limits_{0\le j<k} |\varepsilon_j|$, 再利用 $f(t,y)$ 关于 y 满足 Lipschitz 连续性得:

$$|\varepsilon_n| \le C\left(kM_0 + \frac{hL}{|\alpha_k|}\sum_{j=0}^{n-k}\sum_{l=0}^{k} |\beta_l||\varepsilon_{l+j}|\right)$$

再令 $B_0 = \dfrac{kL}{|\alpha_k|}\max\limits_{0\le l\le k} |\beta_l|$, 有

$$|\varepsilon_n| \le CkM_0 + hB_0\sum_{j=0}^{n} |\varepsilon_j|.$$

线性多步法的相容性与稳定性

显然当 h 充分小时, 可以有

$$|\varepsilon_n| \le CkM_0 + h\widetilde{C}\sum_{j=0}^{n-1}|\varepsilon_j|.$$

从而由离散 **Gronwall** 不等式 (引理 **9.2**)

$$\Longrightarrow |\varepsilon_n| \le e^{\widetilde{C}nh}\left(CkM_0 + \widetilde{C}khM_0\right), \quad n \ge k,\ nh \le T.$$

即证明了格式对初值的稳定性。 $\triangleright$

由上面定理, 如果 (9.21) 满足相容性以及根条件, 且 $h \to 0$ 时, 假设初值误差 $\to 0$, 则 $t_0 + nh \to t$ 时, 有 $y_n = y(t_0 + nh) \to y(t)$, 即有离散解的收敛性.

线性多步法的相容性与稳定性

例 9.5

考虑初值问题 $\begin{cases} y'(t)=4t\sqrt{y},\ 0\le t\le 1, \\ y(0)=1; \end{cases}$ 其解为 $y(t)=(1+t^2)^2$.

用多步法: $y_{n+2}-(1+a)y_{n+1}+ay_n=\frac{h}{2}[(3-a)f_{n+1}-(1+a)f_n]$,

可以检验 $c_0=c_1=c_2=0$, $c_3=\frac{a+5}{12}$. 又由其第一特征多项式

$\rho(\lambda)=\lambda^2-(1+a)\lambda+a=(\lambda-1)(\lambda-a)$ 的根为$\lambda_1=1,\lambda_2=a$.

我们知道 $a=-5$ 时, 该方法为不稳定三阶格式; $a\neq-5$

(例如取 $a=0$) 为二阶稳定格式. 假如取 $a=-5$, $h=0.1$, 可得

t	0	0.1	0.2	$\cdots$	0.7	0.8	0.9	1.0
$y(t)$	1	1.0201	1.0816	$\cdots$	2.2201	2.6896	3.2761	4.0000
y_n	1	1.0201	1.0812	$\cdots$	2.9130	−0.6026	—	—

目录 I

舍入误差的影响

前面我们一直没有考虑每一步计算中舍入误差带来的影响, 但这事实上是无法避免的.

考虑如下一般形式的差分法:

$$\sum_{j=0}^{k} \alpha_j y_{n+j} = h\phi_f(t_n; y_n, \cdots, y_{n+k}; h). \tag{9.29}$$

实际计算时, 由于舍入误差的影响, 实际计算式为

$$\sum_{j=0}^{k} \alpha_j \bar{y}_{n+j} = h\phi_f(t_n; \bar{y}_n, \cdots, \bar{y}_{n+k}; h) + \delta_{n+k}. \tag{9.30}$$

其中 δ_{n+k} 表示每步计算带来的舍入误差影响.

通常可以认为 δ_l 是可以控制的, 可以足够小.

舍入误差的影响

虽然我们前面估计了 y_n 与 $y(t_n)$ 之间的截断误差, 但是还需 $|\bar{y}_n - y_n|$ 充分小, 才能有实际计算精度足够高.

令 $e_n = \bar{y}_n - y_n$, 有

$$\sum_{j=0}^{k} \alpha_j e_{n+j} = h[\phi_f(t_n; \bar{y}_n, \cdots, \bar{y}_{n+k}; h) - \phi_f(t_n; y_n, \cdots, y_{n+k}; h)] + \delta_{n+k}.$$

若 f 是一个比较复杂的函数, 显然不好估计舍入误差的增长情况. 因此一般我们都只在假设 $f(t, y)$ 仅是 y 的线性函数情况下讨论. 即以下假设

(9.31) $$f(t, y) = \mu y, \quad \mu \in \mathbb{C} \text{ 为一个常数}.$$

可以将 μ 看成 $\frac{\partial f}{\partial y}(t, y)$ 的近似.

舍入误差的影响

自然，若 $f=\mu y$, 问题 $\begin{cases} y'(t)=\mu y, \\ y(0)=y_0; \end{cases}$ 的解为 $y(t)=y_0e^{\mu t}$.

用多步法 (9.21) 来求解 (单步法(9.3)可以看成 $k=1$ 的特例)

(9.32) $$\sum_{j=0}^{k}(\alpha_j-\bar{h}\beta_j)y_{n+j}=0,$$

其中 $\bar{h}=\mu h$. 若令 $\gamma_j(\bar{h})=\alpha_j-\bar{h}\beta_j$, 有

(9.33) $$\sum_{j=0}^{k}\gamma_j e_{n+j}=\delta_{n+k}.$$

引入向量形式记号(假设 $\gamma_k=\alpha_k-\bar{h}\beta_k\neq 0$)

$$\vec{E}_n=(e_{n+k-1},e_{n+k-2},\cdots,e_n)^T,\quad \vec{\eta}_n=(\gamma_k^{-1}\delta_{n+k},0,\cdots,0)^T\in\mathbb{R}^k,$$

舍入误差的影响

再令矩阵 $A = \begin{pmatrix} -\gamma_k^{-1}\gamma_{k-1} & -\gamma_k^{-1}\gamma_{k-2} & \cdots & -\gamma_k^{-1}\gamma_0 \\ 1 & 0 & \cdots & 0 \\ & \ddots & \ddots & \vdots \\ 0 & & \ddots & 0 \end{pmatrix}$, 得到以下矩阵形式的误差传递式:

(9.34) $$\vec{E}_{n+1} = A\vec{E}_n + \vec{\eta}_n.$$

这样其解为

(9.35) $$\vec{E}_{n+1} = A^n\vec{E}_0 + \sum_{j=0}^{n-1} A^{n-j-1}\vec{\eta}_j, \quad n = 1, 2, \cdots$$

一般可假设 $\|\vec{E}_0\|$, $\|\vec{\eta}\|$ 可以充分小 (取决于计算机字长).

绝对稳定性

但要保证 $\|\vec{E}_n\|$ 有界 (注意 $h \to 0$ 时, $nh \le T$ 会有 $n \to +\infty$), 我们需要 $\|A^n\| \to 0$ 才行. 即需要 A 的谱半径< 1. 由矩阵 A 的表达式易得其特征方程为

(9.36) $$\rho(\lambda) - \bar{h}\sigma(\lambda) = 0.$$

这样，要保证 $\|\vec{E}\|$ 有界, 即需要上面方程的根的模均小于1.

定义 9.7 (绝对稳定性)

称线性多步法(9.21)关于 $\bar{h} = \mu h$ 是绝对稳定的，是指上述特征方程 (9.36) 的根均在单位圆周内, 即 $\lambda_j(\bar{h}) < 1, j = 1, \cdots, k$.

如果在复平面上存在区域 D s.t. $\forall \bar{h} \in D$, 多步法 (9.21) 都关于该 $\bar{h}$ 绝对稳定, 则称 D 为该方法的绝对稳定域.

由于 e_n 与 y_n 满足同样的差分方程, 故而有以下定理.

绝对稳定性

定理 9.5 (绝对稳定的必要条件)

若 (9.21) 是相容且对初值稳定的格式，且当 $h \to 0$ 时关于 $\bar{h} = \mu h$ 绝对稳定，则必有 $Re(\mu) < 0$ (即 $y(t)$ 本身是指数衰减的).

$\triangleleft$ **如果格式相容且对初值稳定，那么 $y_n \to y(t_n)$，当 $h \to 0$ 时.**

而 y_n 与 e_n 满足的差分方程相同，如果格式关于 $\bar{h} = \mu h$ 绝对稳定，那么特征值方程的所有根的模都小于1.

由前面差分方程解的表达式知，(参看定理 9.4 中的公式 (9.27)-(9.28)) 应该有，当 $n \to +\infty$ 时 $y_n, e_n \to 0$.

再由试验方程的解为 $y(t) = y(0)e^{\mu t}$，立刻应该有 $Re(\mu) < 0$.

$\triangleright$

绝对稳定性

对单步法可类似定义绝对稳定性.

由于问题的复杂性, 我们也仅对试验方程考虑, 即假设 $f(t,y)=\mu y$, 这样有

$$y_{n+1}=y_n+h\phi_f(t_n,y_n;h)\equiv E(\bar{h})y_n.$$

即特征方程是 $\lambda-E(\bar{h})=0$.

例如对于 **Euler** 方法, 有

$$y_{n+1}=y_n+hf(t_n,y_n)=(1+h\mu)y_n=(1+\bar{h})y_n.$$

由前面线性多步法的绝对稳定性条件, 需要特征方程的根的模小于1，也就是说单步法的绝对稳定性条件为 $|E(\bar{h})|<1$.

显式 R-K 方法的绝对稳定性

例 9.6 (考虑二阶显式 R-K 格式)

$$\begin{cases} y_{n+1} = y_n + \frac{h}{2}(K_1 + K_2), \\ K_1 = f_n, \quad K_2 = f(t_{n+1}, y_n + hK_1). \end{cases}$$

应用于试验方程 $f \equiv \mu y$, 得到 (令 $\bar{h} = \mu h$):

$$y_{n+1} = y_n + \bar{h}\left(1 + \frac{\bar{h}}{2}\right) y_n \equiv \left(1 + \bar{h} + \frac{\bar{h}^2}{2}\right) y_n.$$

特征方程: $\lambda - \left(1 + \bar{h} + \frac{\bar{h}^2}{2}\right) = 0$, 稳定性条件为: $|1 + \bar{h} + \frac{\bar{h}^2}{2}| < 1$.

可以用边界轨迹法来确定稳定区域 $D \subset \mathbb{C}$:

令 $\lambda = e^{i\theta}$ 代入上面特征方程 (D 的边界点对应模为1的根):

$$1 + \bar{h} + \frac{\bar{h}^2}{2} - e^{i\theta} = 0 \Longrightarrow \bar{h} = -1 \pm \sqrt{2e^{i\theta} - 1}.$$

隐式 R-K 方法的绝对稳定性

例 9.7 (类似地考虑二阶隐式 R-K 格式)

$$y_{n+1} = y_n + hf\left(t_n + \frac{h}{2}, \frac{y_n + y_{n+1}}{2}\right),$$

令 $f \equiv \mu y$, 得到 (令 $\bar{h} = \mu h$):

$$y_{n+1} = y_n + \bar{h}\frac{y_n + y_{n+1}}{2} \Longrightarrow y_{n+1} = \frac{1+\bar{h}/2}{1-\bar{h}/2} y_n.$$

特征方程为: $\lambda - \dfrac{1+\bar{h}/2}{1-\bar{h}/2} = 0$, 即要求: $\left|\dfrac{1+\bar{h}/2}{1-\bar{h}/2}\right| < 1.$

易见其等价于 $Re(\bar{h}) < 0$.

即隐式二阶 *R-K* 方法的绝对稳定域包含了整个左半复平面！

可类似地证明其他各阶隐式 *R-K* 方法的绝对稳定域都包含了左半复平面！我们称这样的格式为 *A*-稳定方法.

多步法的绝对稳定性

例 9.8 (考虑以下隐式格式–Simpson 方法)

$$y_{n+2} = y_n + \frac{h}{3}\left(f_{n+2} + 4f_{n+1} + f_n\right).$$

可以证明这是精度最高的二步法, 为二步四阶隐式格式. 同样考虑 $f = \mu y$, 其特征方程为

$$\begin{aligned} 0 &= \rho(\lambda) - \bar{h}\sigma(\lambda) = \lambda^2 - 1 - \frac{\bar{h}}{3}(\lambda^2 + 4\lambda + 1) \\ &= \left(1 - \frac{\bar{h}}{3}\right)\lambda^2 - \frac{4\bar{h}}{3}\lambda - \left(1 + \frac{\bar{h}}{3}\right). \end{aligned}$$

自然有两个根 $\lambda_\pm = \frac{2\bar{h} \pm \sqrt{3\bar{h}^2+9}}{3-\bar{h}}$. 当 $h \to 0$ 时, 总有一个根模大于1

$$\lambda_\pm(\bar{h}) = \frac{2\bar{h}}{3} + \frac{2\bar{h}^2}{9} + \mathcal{O}(\bar{h}^3) \pm \left(1 + \frac{\bar{h}}{3} + \frac{5\bar{h}^2}{18} + \mathcal{O}(\bar{h}^3)\right).$$

当 $Re(\mu) < 0$, $h \ll 1$ 时, 就有 $Re(\lambda_-) < -1$, 即 $|\lambda_-| > 1$. 这说明无论 h 多小, 总有 $|\lambda_-| > 1$. 即 *Simpson* 方法为“绝对不稳定”.

相对稳定性

前面我们已经说过 (*cf.* 定理 **9.5**), 只能在 $Re(\mu) < 0$ 时考虑格式的绝对稳定性. 那么如果 $Re(\mu) \geq 0$, 此时什么样的算法是可以接受的呢？ 仍考虑试验方程:

$$\begin{cases} y'(t) = \mu y, \\ y(0) = y_0; \end{cases} \implies y(t) = y_0 e^{\mu t}.$$

及多步法 $\sum_{j=0}^{k} \alpha_j y_{n+j} = h \sum_{j=0}^{k} \beta_j f_{n+j}$. 设其为相容、对初值稳定的. 也就是说, $\rho(1) = 0,\ \rho'(1) = \sigma(1),\ \rho(\lambda) = 0$ 的根均在单位圆内, 且单位圆周上为单根. 由此也说明 $\lambda_1 = 1$ 是 $\rho(\lambda) = 0$ 的单根.

设该方法为 p 阶格式, 即局部截断误差为 $c_{p+1} h^{p+1} y^{(p+1)}(\xi_n)$. 也就是有 $L[y(t_n); h] = \sum_{j=0}^{k} (\alpha_j - \bar{h}\beta_j) y(t_{n+j}) = \mathcal{O}(h^{p+1})$.

相对稳定性

由 $y(t_j)=y_0e^{\mu jh}=y_0(e^{\bar{h}})^j \Longrightarrow y_0e^{n\bar{h}}\sum_{j=0}^{k}(\alpha_j-\bar{h}\beta_j)\,(e^{\bar{h}})^j=\mathcal{O}(h^{p+1})$.

结合其特征方程: $\rho(\lambda)-\bar{h}\sigma(\lambda)=0$, 我们知道特征方程必有一根

$$\lambda_1(\bar{h})=e^{\bar{h}}+\mathcal{O}(h^{p+1}).$$

$h\to 0$ 时, 即有 $\lambda_1(\bar{h})\to\lambda_1=1$.

再考虑舍入误差的方程 $\sum_{j=0}^{k}(\alpha_j-\bar{h}\beta_j)e_{n+j}=\delta_{n+k}$.

结合差分方程解的表达式我们知道, e_n 的表达式中必有一项

$$C\cdot\lambda_1^n(\bar{h})=C\cdot e^{n\bar{h}}+\mathcal{O}(h^{p+1})=C\cdot e^{\mu t_n}+\mathcal{O}(h^{p+1}).$$

也就是说 $e_n=C\cdot e^{\mu t_n}+\mathcal{O}(h^{p+1})+\cdots$

对比 $y(t)$ 的表达式我们知道, e_n 的增长性至少与 $y(t_n)$ 相同.

相对稳定性

假设上面解 e_n 的表达式中 $\lambda_1^n(\bar{h})$ 为主项, 其他项为该项的高阶小量 (当 $h \to 0$), 即 $|\lambda_1(\bar{h})| > |\lambda_j(\bar{h})|, j = 2, \cdots, k$. 这时可以认为舍入误差带来的影响的增长率至多与真实解同阶. 此时可以认为计算结果是可接受的. 这样引入了另一种稳定性概念:

定义 9.8 (相对稳定性)

称多步法(9.21)对给定的 $\bar{h}$ 是相对稳定的, 是指若特征方程 (9.36) 的根满足 $\lambda_1(\bar{h}) = e^{\bar{h}} + \mathcal{O}(h^{p+1})$ (自然有 $\lambda_1(0) = 1$), 且

$$|\lambda_1(\bar{h})| > |\lambda_j(\bar{h})|, j = 2, \cdots, k.$$

否则就称该方法对于 $\bar{h}$ 不是相对稳定的.

如果在复平面上存在区域 D s.t. $\forall \bar{h} \in D$, 多步法 (9.21) 都关于该 $\bar{h}$ 相对稳定, 则称 D 为该方法的相对稳定域.

相对稳定性

例 9.9 (考虑以下两步法)

$$y_{n+2}-(1+a)y_{n+1}+ay_n=\frac{h}{12}[(5+a)f_{n+2}+8(1-a)f_{n+1}-(1+5a)f_n]$$

假设 $-1\le a\le 1$. 显然，当 $a\neq -1$ 时，上式为三阶方法; $a=-1$ 时为*Simpson*方法. 其特征多项式为

$$\rho(\lambda)-\bar{h}\sigma(\lambda)=[1-\frac{\bar{h}(5+a)}{12}]\lambda^2-[(1+a)+\frac{2\bar{h}}{3}(1-a)]\lambda+a+\frac{\bar{h}}{12}(1+5a)$$

简单起见，仅考虑 $\mu\in\mathbb{R}$ 情形. 可得

其绝对稳定域为 $\left(6\dfrac{a+1}{a-1},0\right)$. 其相对稳定域为 $\left(\dfrac{3(a+1)}{2(a-1)},+\infty\right)$.

例如取 $a=0$，来求解

$$\begin{cases} y'=-100y,\ 0\le t\le 1, \\ y(0)=1; \end{cases} \Longrightarrow y(t)=e^{-100t}.$$

相对稳定性

要使方法绝对稳定, 需要 $-6 < -100h < 0$, 即 $h < 0.06$.

要使方法相对稳定, 需要 $-\frac{3}{2} < -100h < +\infty$, 即 $h < 0.015$.

取 $h = 0.01, 0.02, 0.1$ 分别计算:

$h=0.01$	$y_2=0.1454$	$y_3=0.0558$	$y_4=0.0217$	$y_5=8.39E-3$	$\cdots$
$y(t_n)$	0.1353	0.0498	0.0183	$6.74E-3$	$\cdots$

$h=0.02$	$y_2=0.0663$	$y_3=2.48E-4$	$y_4=5.98E-3$	$y_5=-1.07E-3$	$\cdots$
$y(t_n)$	0.0183	$2.48E-3$	$3.35E-4$	$4.54E-5$	$\cdots$

$h = 0.1$	$y_2=0.1612$	$y_3=-0.1768$	$y_4=0.2200$	$y_5=-0.2698$	$\cdots$
$y(t_n)$	$2.06E-9$	0	0	0	$\cdots$